CON-INFO: A Context-based Methodology for Designing and Assessing the Quality of Adaptable MUIs in Healthcare Applications

Reem Abdulaziz Alnanih

A Thesis
In the Department
of
Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Computer Science and Software Engineering) at
Concordia University
Montreal, Quebec, Canada

April 2015

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Reem Alnanih

Entitled: CON-INFO: A Context-based Methodology for Designing and Assessing the Quality of Adaptable MUIs in Healthcare Applications

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________ Chair
Dr. R. Dssouli

______________________________ External Examiner
Dr. J.-M. Desharnais

______________________________ External to Program
Dr. A. Agarwal

______________________________ Examiner
Dr. S. Mudur

______________________________ Examiner
Dr. P. Grogono

______________________________ Thesis Supervisor
Dr. O. Ormandjieva

Approved by: ______________________________
Dr. V. Haarslev , Graduate Program Director

____
October 29, 2014 Dr. A. Asif, Dean
Faculty of Engineering and Computer Science

ABSTRACT

CON-INFO: A Context-based Methodology for Designing and Assessing the Quality of Adaptable MUIs in Healthcare Applications

Reem Alnanih, Ph.D.

Concordia University, 2015

Mobile technology is an integral part of the modern healthcare environment. The mobile user interface (MUI) serves as the bridge between the application and healthcare professionals. It is important that the physician be able to easily express his needs on the MUI and correctly interpret the information displayed. However, there are many challenges that face the designer in designing and developing context-sensitive MUIs in this environment. The adaptability of the MUI is considered to be one of the most important issues to address. According to the World Health Organization (WHO), MUI adaptability is a major problem in the healthcare context. For the designer, the hope is that new technologies will be developed, such as mobile devices adaptable to different environments, to enable customization of the application to the user's context.

In this thesis, we propose a new methodology for designing a context-based adaptable MUI for healthcare applications. This methodology offers a new approach to automated MUI context adaptation, and provides a solution for both the provider (designer) of the healthcare application and the consumer (physician). New techniques for adapting MUIs offer new opportunities for the MUI designer to maximize the benefits of mobile health technology by providing the best possible way for healthcare professionals to perform their tasks efficiently and effectively.

The proposed methodology is based on research contributions in four areas: (1) a new quality-in-use measurement model for validation purposes; (2) user stereotype modeling

with a set of context *descriptors*, which formalize the domain expertise of the users; (3) context information modeling; and (4) use of the decision table technique to adapt the MUI features based on the context and the user stereotypes. The proposed quality-in-use model is inspired by the ISO/IEC 25010 and ISO/IEC 25022 international standards and adapted to healthcare applications. The first contribution is used in validating the quality-in-use of a software product developed according to the CON-INFO methodology, and the last three contributions are linked to form a methodology for development. The MUI features adapted to the needs of healthcare professionals have been implemented on the iPhone™ for validation purposes.

An example of software for medical application is the Phoenix Health Information System (PHIS), which is in use at King Abdulaziz University Hospital (KAUH). PHIS2 is an updated desktop version developed based on Human-Computer Interaction (HCI) principles. A new mobile-based version of PHIS2 (PHIS2-M) has since been introduced, to make PHIS accessible from a mobile-based platform. The proposed context-based and rule-based approach for MUI feature adaptability resulted in a new version of PHIS2-M – PHIS2-MA (MA stands for *mobile adaptation*).

This thesis validates the proposed methodology and clearly demonstrates its usefulness, providing details of the four empirical studies conducted with the end-users (physicians) in a real environment at the KAUH.

The results of the formal studies reveal that our CON-INFO methodology for designing an adaptable MUI led to improvements to the current application and allowed researchers to test successive versions of the ‘final’ application.

ACKNOWLEDGMENTS

I am grateful to God for his guidance and protection throughout all my life;

First and foremost, I would like to express my thanks and gratitude to my thesis advisor Dr. Olga Ormandjieva for her confidence, guidance, moral support, inspiration and constant encouragement during this journey of the research;

I am greatly indebted to the late Dr. Thiruvengadam Radhakrishnan, my co-advisor, for his rich advices, discussions and motivation through every step of my Ph.D. work.
My advisors, I have learned a lot from you!

I would also like to express my utmost appreciation to Prof. Jean-Marc Desharnais, Canadian representative at ISO SC/7 WG-6 (Quality measures) and the editor of the Quality Measure Element standard (ISO/IEC 25021), for providing especially valuable feedback on my dissertation;

My deepest thanks and gratitude to my dear parents for giving and bringing so much happiness into my life and for their special encouragement, prayers and invaluable support;

A special feeling of gratitude to my dear husband for being a role model in academic interest and for his emotional support, patience and understanding throughout this long process;

I dedicate this work to my lovely kids who added a great pleasure to my life during the challenging times of this work;

Acknowledgement is to King Abdulaziz University Hospital for giving me the chance to apply the case study of my research to the hospital's system; I am thankful to all house officers and residents for their cooperation and time;

Lastly, I would like to thank everyone who has contributed to the success of this research; I pray that God reward you all abundantly.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

Mobile computing has become one of the most dominant computer use paradigms, and there are a number of various visions of how best to realize it. A great deal of interest has been shown recently in the use of mobile applications to support healthcare (Alnanih, 2008); (Orwat, Graefe, & Faulwasser, 2008). As these applications become more sophisticated, a trend will inevitably develop towards providing comprehensive support for healthcare practitioners. It is expected that mobile healthcare applications will improve patients' quality of life, while at the same time reducing system costs, paper records, delays and errors (Archer, 2009).

Today's smartphones are powerful computers combined with the features of a telephone. Users have come to expect to have all the benefits of a desktop in their pocket. Mobile applications can help them manage their health and wellness, promote healthy living, and provide them with access to useful information when and where they need it. These tools are being adopted almost as quickly as they can be developed. According to industry estimates, 500 million smartphone users worldwide will be using a healthcare application by 2015, and, by 2018, 50 percent of the more than 3.4 billion smartphone and tablet users will have downloaded mobile health applications (Jahns, 10 November 2010). These users include healthcare professionals, consumers and patients. In the United States, the Food and Drug Administration (FDA) is encouraging the development of mobile medical applications that improve healthcare, and provide consumers and healthcare professionals with valuable health information. The FDA also has a public health responsibility, which is to oversee the safety and effectiveness of medical devices, including mobile medical applications (Food & Administration, 2013).

The latest smartphones have implemented new input mechanisms through the user interface, such as touch and voice. As these increasingly sophisticated mobile devices are now mass-market commodities, careful consideration must be given to smartphones and their applications, in terms of their aesthetics, usability and utility, as well as the emotional aspects of the user experience (Jones & Marsden, 2006).

The user interface (UI) of the smartphone incorporates the data based functions of both desktop and mobile computer products, and determines this phone's usability and commercial success.

In this chapter, we introduce the motivation for this research work, define the research objectives, and clarify some key terms. We also give an overview of the organization of the thesis.

1.1 Research Motivation and Challenges

"Going mobile" is the next logical step in an era in which computers are all-pervasive. Currently, mobile computing is a supplemental service in hospitals, complementing existing clinical information systems by providing an alternative means to access medical information (Bardram, 2004) and supporting interpersonal communication (Bardram & Hansen, 2004). Given the increasing role of technology, the risk of errors caused by poor design, and the complexity of healthcare itself, the mobile user interface (MUI) design for healthcare applications is subject to growing scrutiny, precisely because the stakes are so high and the potential gains from technology development in this area are so significant (Kinkade & Verclas, 2008).

Interface designers face a challenge in the healthcare context, in terms of meeting future technology needs, and the ever growing complexity of technology. In the medical field, the reality is that medical errors are seldom a result of carelessness or negligence. More commonly, they are caused by low quality interface design. Both Human-Computer Interaction (HCI), which is the study of the interaction between users and computers that occurs through the user interface (Dix, Finlay, Abowd and Beale, 2004), and human factors have a significant role to play in increasing the quality of healthcare, and consequently in reducing the number of errors, especially those associated with the use of medical devices (Alvarado et al., 2004). The traditional HCI approaches to interface design are essential to the design of MUI, but they are unable to cope with the complexity of the typical interactive devices available today in the safety-critical context of medical devices (Thimbleby, 2007; Thimbleby & Thimbleby, 2007). This has led to the development of a broad range of user-centered design (UCD) methods (Holzinger, 2004, 2005; Holzinger & Errath, 2007); however, traditional UCD practices are insufficiently powerful to solve problems at this level of complexity.

The dynamic work habits of healthcare practitioners, along with the variety and complexity of the healthcare environments in which they operate, represent other challenges for software designers, in terms of designing mobile medical applications that meet the needs of these users (Alnanih, Ormandjieva, & Radhakrishnan, 2011). Healthcare practitioners need interfaces tailored to their specific requirements to facilitate their work and to help them avoid the misunderstandings that can result in medical errors (Alnanih, Alnuaim, & Ormandjieva, 2009) .

1.2 Research Objectives

The research described in this thesis constitutes an attempt to address these and related challenges by proposing a new context-based MUI design methodology, which we call CON-INFO, that bridges HCI and Pervasive Healthcare (PH). Together, HCI and PH provide the potential to assist the designer in solving the most important issues in the realm of healthcare.

Objective 1: Quality-in-use measurement model for MUIs

The quality of MUIs is crucial in the healthcare domain, as the attention of healthcare professionals is usually on the patient and not on the system, and so low-quality UIs may lead to critical medical errors. Based on the existing standards referenced in our literature review, no method currently exists to measure the quality of MUI design for healthcare applications (see chapter 3, Table 3-1). New guidelines for designing MUIs for quality-in-use models and quality evaluation/decision-making models have to be derived, as the current standards may not meet the specific needs of healthcare application (app) users.

Therefore, the first objective of this thesis is to propose a new MUI quality-in-use measurement model validated both theoretically and empirically for healthcare (see chapter 3) (Alnanih, Ormandjieva, & Radhakrishnan, 2013c).

Objective 2: User modeling

MUIs are being used by more and more individuals in the healthcare domain in all walks of life, from different backgrounds and with varying levels of experience. Consequently, MUI designers are certain to face an increasing number of challenges in designing customized MUIs for the healthcare context. They will have to optimize user interactions with the devices in that operating environment, identifying user needs and understanding the activities that are being supported by the devices.

The second objective of this thesis is aimed at eliciting user stereotype categories in order to make the MUI strategically adaptable to the 'type' of user (see chapter 5) (Alnanih, Ormandjieva, & Radhakrishnan, 2013b).

Objective 3: Context model

Mobility provides additional opportunities for leveraging context, but a potentially rapidly changing context and the need to synthesize it, and act on it, place an extra burden on the mobile computing platform.

In this thesis, the third objective is to address this challenge by introducing a two-level characterization of the term *healthcare context* (a basic level and a domain level) as a requirement and guide for designing and developing a context-based MUI (see chapter 6) (Alnanih, Radhakrishnan, & Ormandjieva, 2012).

Objective 4: MUI design adaptation

MUI design should benefit from the idea of adaptation of the UI according to the context. Therefore, there is a need for a formal definition of context in the healthcare domain and for corresponding context-based adaptation techniques.

The fourth objective of this thesis is to capture domain expertise as a foundation for designing and developing adaptable, context-based MUIs.

Two categories of context model adaptation are tackled here: (a) the designer adapts the design to the mobile device, and (b) the device recognizes the context at runtime and adapts its interface widgets based on pre-programmed requirements. Our proposed adaptation technique extracts values from sensors in smartphones, maps them to the user stereotype model, and then maps them to the MUI using decision rules designed specifically for the context (see chapter 7) (Alnanih, Ormandjieva, & Radhakrishnan, 2013a).

1.3 Scope and Significance of the Research

The main research significance of this thesis lies in the novelty of the approach, in which the adaptation of MUIs is systematized, allowing for objective and subjective analysis in an incremental fashion. The proposed CON-INFO approach is structured in five primary phases: 1) measurement model for quality assessment, 2) healthcare user stereotype model, 3) context descriptors which formalize the domain expertise of the users, 4) healthcare context model, and 5) adaptation design approach using decision tables.

This five-phase structured framework is important for designing adaptable context-dependent MUIs from a number of perspectives. The quality-in-use of the MUI is monitored using a new measurement model inspired by the ISO 25010 international standard (ISO/IEC-25010:2011(E), 2011) and ISO 25022 international standard (ISO/IEC CD 25022:2014, 2014) and adapted to healthcare. This model is validated both theoretically and empirically (Alnanih et al., 2013c). The context model allows for flexible characterization of user needs, which is a key issue in user modeling for healthcare apps. The context descriptors, which formalize the domain expertise of the users captured in the context model, are expressed in a language that designers who are not experts in the domain can understand, and they can be used to adapt the MUI to a specified context.

Designing adaptable MUIs for the healthcare domain provides a solution that adapts essential patient information for physicians in an easily accessible, clear and accurate way, and makes it available to them at any time. The benefits of the proposed CON-INFO methodology for healthcare professionals include improved productivity, performance and level of satisfaction, as well as increased patient safety, as physicians can access patient information whenever and wherever it is needed. It also reduces the physician's cognitive load by specifying a limited number of actions per view.

Our CON-INFO methodology is a top-down modeling approach applicable in the beginning stages of analysis and development of an adaptable MUI. It is based on ISO/IEC 12207 that provide a comprehensive set of life cycle processes, activities and tasks for software that is part of a larger system, and for stand alone software products and services**.** The major phases of this methodology are illustrated in Figure 1-1. Each phase represents the thesis objectives in some way:

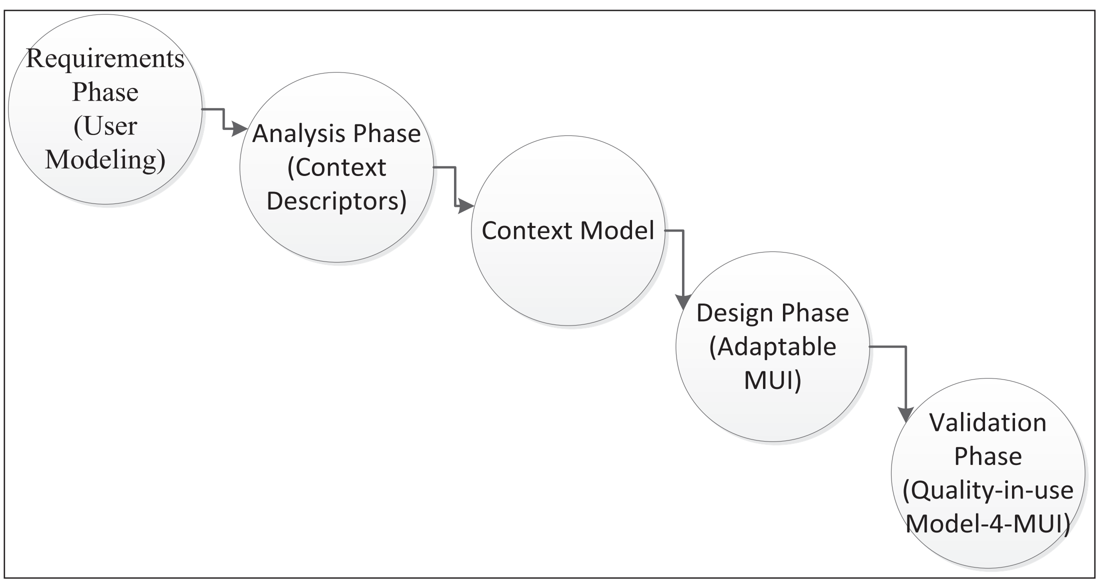


Figure 1-1 The CON-INFO methodology process

1.4 Overview of the Contributions

The specific contributions of this research are listed as follow:

(1) Definition and modeling of the Quality-in-Use-4-MUI.

(2) Definition and modeling of the user, to incorporate user needs, behaviors and experiences as key inputs to the design process, and to extract a set of context descriptors from those users.

(3) Definition and modeling of contextual information, to define a systematic process to guide the designers in translating the inputs of the user analysis phase into adaptable design solutions.

(4) Capturing the data from the user modeling and the context model, and representation of these data in an MUI using a suitable representation rules and techniques.

The significance of this research lies in the interdisciplinary nature of the CON-INFO methodology of adaptable MUI operation, specifically the interaction between the user and the app and the perspectives to be taken into account: 1) the user's perspective, based on user stereotype modeling and the set of statements elicited from the user, referred to as context descriptors; 2) the domain perspective, based on the characterization of the healthcare context model; and 3) the designer's perspective, based on the adaptation rules and the real, empirical evaluation of the design of the MUI's adaptation features. The CON-INFO methodology is schematically depicted in Figure 1-2.

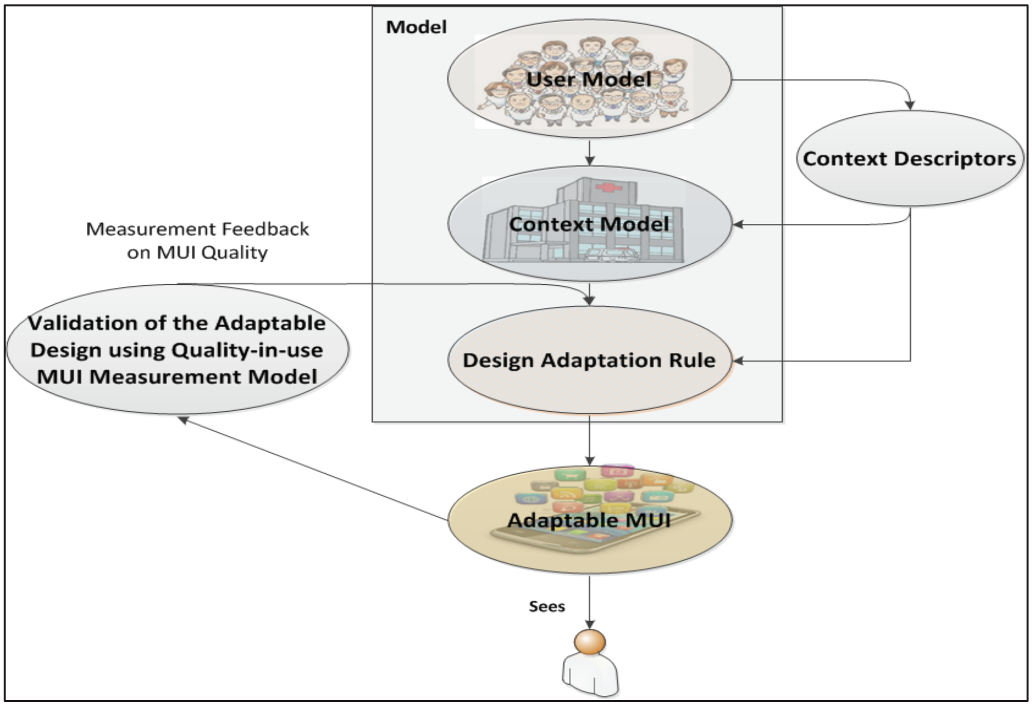


Figure 1-2 The CON-INFO workflow

1.5 Organization of the Thesis

The organization of this thesis is as follows:

In **chapter 2**, we discuss background and summarize the main case study on which we base our proof of concept for the thesis. We then introduce current techniques relevant to empirical studies carried out in the MUI field. Finally, we highlight the main topics that are important in understanding the quality models and measurements that we are proposing.

In **chapter 3**, we present our quality-in-use measurement model for MUIs as an assessment method to validate our CON-INFO methodology. In this chapter, our proposed quality-in-use measurements for evaluating MUI are theoretically and empirically validated.

In **chapter 4**, we present our MUI design for medical application based on HCI principles, and validate our design by applying the proposed quality-in-use measurement model in the healthcare context.

In **chapter 5**, we present our user stereotype approach to classifying healthcare physicians to enable us to design an adaptable MUI for different types of users of healthcare apps. A set of context descriptors is provided in this chapter to help the designer understand the specific context in which healthcare professionals operate. The results in this chapter are based on interviews with 30 physicians and a questionnaire to which they all responded.

In **chapter 6**, we examine the way in which context can be characterized for developing context-sensitive user interfaces for smartphone application in a hospital environment.

In **chapter 7**, we propose a context-based and rule-based approaches as knowledge requirements for designing adaptable MUIs for healthcare application, and apply our quality-in-use model to evaluate the adaptable design as a pilot test with 12 end-users belong to junior stereotype.

In **chapter 8**, we present an empirical evaluation of the final version of the adaptable MUI in a real environment, among 45 end-users representing all the user stereotypes, including junior, intermediate, and experienced physicians.

Finally, in **chapter 9**, we summarize our work and its major contributions, and present avenues for future research in this area.

In the following chapter, we present a summary of the main case study to which we applied our CON-INFO methodology, as well as the various techniques we used in our controlled experiments and for our quality measurements.

Chapter 2 Background

In this chapter, we provide a brief synopsis of the main case study on which we based our proof of concept for the main thesis and the empirical investigation of the research hypotheses. We then introduce key concepts underlying the empirical studies, which helped us to determine whether the study was appropriate for our evaluation, or whether we needed to design our own study and derive meaningful results from that. Finally, we highlight the main topics that are important to an understanding of the quality models and quality measurements.

2.1 Case Study

In this section, a brief description of the medical software we used to empirically investigate the research objectives of this thesis is presented.

2.1.1 The Phoenix Health Information System (PHIS) Application

An example of medical software is the Phoenix Health Information System (PHIS), which was created in 1998 after a team from New Zealand and other countries met in Dubai to evaluate the health information systems currently available at that time. After reviewing those systems, the team concluded that a new system was needed. The objective was to use the latest development tools to build the most advanced and technologically innovative health information system, rich in functionality and reasonable in price. Development was begun in late 1998 by Phoenix Health Systems in Jeddah, Saudi Arabia. The team in Jeddah was led by Ian Bailey and Liza Bacolod. The financial modules were developed in Dubai, under the leadership of Mohan Attavar.

King Abdulaziz University Hospital (KAUH) replaced the Oasis system they had been using with the PHIS in 2004. That new system was supposed to enable medical personnel to order, trace and store patient information, and facilitate communication among services involved in patient care. Like many information systems, each patient has an identification number that opens what is known as the ‘patient file’, which contains all the options available for the patient, including ordering tests and drugs, viewing results and scheduling an operation. It also contains patient demographic information and administrative information specific to that patient.

The purpose of my Master’s thesis in 2008 was to attempt to alleviate the problems suffered by PHIS users at the KAUH by performing an iterative evaluation of the system, and by assessing user needs, involving them at every stage of the design process, with the aim of developing a complex, user-centered health information system (Alnanih, 2008). Phoenix Health Information System 2 (PHIS2) is a redesigned version of the desktop-based PHIS currently in use at the KAUH. The PHIS was formally evaluated by novice and expert users at the hospital to identify problems they were having with the system. This researcher concluded from the evaluation that all the problems found, based on the commonly used tasks physicians performed on a daily basis, constitute violations of the HCI principles documented in the literature: Mental Model, Metaphor, Visibility, Affordance and Feedback (defined in chapter 4).

Results of the formal evaluation of PHIS2-D show that the system is accurate in terms of performing tasks very quickly, and in an acceptable amount of time; efficient and effective in terms of the limited number of wrongly chosen menu items and wrongly selected icons; and is easy to use, even with no training whatsoever. The implementation of an iterative, user-centered approach not only resulted in major improvements to the design of the current system, but allowed the researchers to test successive approximations of the "final" system as well; and, by studying user-centered design models, they were able to better understanding the users, their work, and their needs and preferences for PHIS 2-D, and fine-tune it accordingly.

2.1.2 The Phoenix Health Information System Mobile (PHIS-M) Application

In order to transition the PHIS2-D to mobile devices, we redesigned its interface for the iOS platform, referring to the mobile app as PHIS2-M, and evaluated the applicability of medical mobile versions in the healthcare context. We then applied the general guidelines for an iOS app to the design (Manickam, 2011). Details are provided in chapter 4.

2.1.3 The Phoenix Health Information System Mobile Adaptable (PHIS2-MA) Application

In order to validate our methodology on an adaptable mobile app interface, the proposed approach for MUI feature adaptability was implemented on an iPhone 4.1™ loaded with the PHIS2-M using Xcode 4 and SQLite. It also uses the special framework for speech recognition, iSpeech iOS SDK. The result was a new version of PHIS2-M, which we call PHIS2-MA (mobile adaptation).

After apple introduced the iPhone 5™ in December 2013, we shifted the design of PHIS2-MA from the iPhone 4.1™ to the iPhone 5™, using the updated version of iOS, Xcode, and the iSpeech framework.

The list of MUI features implemented in PHIS2-MA and adapted to the context of healthcare professionals is explained in detail in chapter 7.

2.2 Empirical Studies in Software Engineering

An experimental design is a complete plan for applying differing experimental conditions to the subjects in an experiment, so that researchers can determine how the conditions affect behavior or the results of some activity. In our case, we needed to plan how the application of these conditions would help us to test our hypotheses (Fenton & Bieman, 2014).

In this section, we explain the assessment techniques available, and provide guidelines for applying the scientific method to empirically assess whether or not the research objectives were achieved.

2.2.1 Types of Empirical Investigation in Software Engineering

There are different types of empirical investigation, such as: controlled experiments; case studies; survey research; ethnographies; and action research. Controlled experiments, case studies, and survey research are the methods that we believe are most relevant to software engineering, as each makes a valuable contribution to the body of software engineering knowledge, and all of them involve careful measurement.

Below, we describe in detail the methods most likely to be applied in software engineering contexts; these methods are adapted from a number of different fields (Easterbrook, Singer, Storey, & Damian, 2008).

- **Controlled Experiment:**

A controlled experiment is an investigation of a testable hypothesis where one or more independent variables are manipulated to measure their effect on one or more dependent variables. Controlled experiments, which must be planned in advance, allow us to determine in precise terms how the variables are related, and, specifically, whether a cause-effect relationship exists between them. Each combination of values of the independent variables is a treatment (Easterbrook et al., 2008). The simplest experiments involve just two treatments, representing two levels of a single independent variable (e.g. using a tool vs. not using a tool). More complex experimental designs involve more than two levels, or more than one independent variable. In most software engineering experiments, human subjects are required to perform some task and the effect of the treatments on the subjects is measured. Empirical studies involving observations where potential confounding variables cannot be controlled and/or subjects cannot be assigned to treatment or control groups are called observational studies, natural experiments, or quasi experiments (Fenton & Bieman, 2014).

- **Case Study:**

A case study is a quasi experiment where key factors are identified that may affect the outcome of an activity, and then the inputs, constraints, resources, and outputs of the activity are documented (Fenton & Bieman, 2014). The term *case study* is often used to mean a working example. However, as an empirical method, a case study is something very different. Yin (2009) introduces the case study as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident.” Case studies can be planned or retrospective.

- **Survey:**

A survey is a retrospective study of a situation to try to document relationships and outcomes. As such, a survey is carried out after an event has occurred. In performing a survey, researchers have no control over the activity that is under study. That is, we can record a situation and compare it with similar ones that have occurred because it is a retrospective study, but we cannot manipulate variables as we do with controlled experiments and case studies.

- **Our Approach:**

We followed the guidelines proposed by (Fenton & Bieman, 2014) to determine which method would be the best for our research. Controlled experiments require a great deal of monitoring, and tend to involve small numbers of people or events. We can describe them as ‘**research in the small**.’ Case studies usually look at a typical project, rather than trying to capture information about all possible cases. They can be thought of as ‘**research in the typical**.’ Surveys can be referred to as ‘**research in the large’** because they are used to poll what is happening broadly, over large groups of projects**.**

In this thesis, we focus primarily on controlled experiments. Based on the criteria for defining these methods, we have chosen to apply this method here for the following reasons:

1. Our investigation is planned, and not retrospective.
2. The treatments that we propose have not been applied previously.
3. The level of replication in our study is high, since we conducted the same test many times, with different apps, different types of UI, and different types of users.
4. We investigate alternative UIs for performing a particular standalone task.
5. A limited number of participants took part, and they were carefully controlled.

6. We had a high level of control over the variables that could affect the outcome.

In this research, we applied usability testing techniques to design our experiments. Usability testing is explained below.

2.2.2 Usability Testing

In order to complete the controlled experiment with a very high degree of MUI monitoring, we undertake usability testing. Usability testing is an "empirical method", focusing on the observations of actual behaviors. Observation of the test team and the comments of test participants are often given more weight in diagnosing problems than they are in a research study.

All usability testing methods share the following five characteristics (Dumas & Redish, 1999):

1. Their primary goal is to improve the usability of the product, either its software or its hardware, along with their secondary, specific goals and concerns related to this goal.
2. The participants are real users.
3. The participants perform real tasks.
4. The experimenters observe and record what the participants do and say.
5. The experimenters analyze the data, diagnose the real problems, and recommend changes to fix those problems.

There are materials we need, in order to prepare to perform every usability test, as suggested by Dumas and Redish, (1999). These include:

1) List of Tasks

This list contains the set of tasks given to each participant before the test. A brief description of the scenario in which each task is to be performed is provided, in order to clarify the task and simplify the experience for the user.

2) Performance Measures (Paper Login Form)

Performance measures are quantitative, in that they can track how much time people take to perform a task, how many errors they make, and how many times they repeat the same error. Most performance measures depend on careful observation by the tester, who must ensure that the task has been carried out; however, some observable behaviors depend on the tester's judgment, for example frustration and confusion. Since judgment can vary from tester to tester, it is strongly recommended that the same person be responsible for collecting the performance measures throughout a usability test, which requires experts in usability testing.

To evaluate the performance of the participants during our testing, we used a paper login form to collect the performance data, which included all the required assessment criteria. The tasks on the paper login form matched the tasks on the list given to the participants during the test.

3) Subjective Measure (Questionnaire)

A satisfaction questionnaire was handed out at the end of the test to gather and evaluate the participants' impressions of the system. The questionnaire contained two types of questions: general questions that applied to any system, and specific questions that only apply to the system tested. All the questions were structured into formats, such as ratings and questions requiring a yes or no answer. The participants were given the opportunity to fill out the questionnaire in private. When they were ready to hand in the completed questionnaire, the answers were reviewed by a tester to clarify any ambiguous answers.

4) The Test Environment

Before conducting the pilot test, which is the last step before conducting the formal test, it was necessary to prepare the physical environment and the people who would be administering the test. The physical environment included the hardware used: a laptop or mobile device, and a stopwatch to measure the time taken to complete each task; the software used, such as the operating system, the app software, and the database.

In the following section, we describe the various quality models and measurements required to perform the validation.

2.3 Quality Model Framework

Software products and software-intensive computer systems need to maintain considerable level of personal satisfaction, improve business success, and human safety by incorporating high-quality software and systems (ISO/IEC 25010:2011(E)).

It is important that the quality characteristics be specified, measured, and evaluated whenever possible using validated measures and measurement methods.
The new SQuaRE (Software product Quality Requirements and Evaluation) series of standards consists of the following divisions (ISO/IEC 25000: 2014):

- ISO/IEC 2500n - Quality Management Division,
- ISO/IEC 2501n - Quality Model Division,
- ISO/IEC 2502n - Quality Measurement Division,
- ISO/IEC 2503n - Quality Requirements Division,
- ISO/IEC 2504n - Quality Evaluation Division, and
- ISO /IEC 25050 – 25099 SQuaRE Extension Division.

Figure 2-1 illustrates the organization of the SQuaRE series of international standards with its divisions.

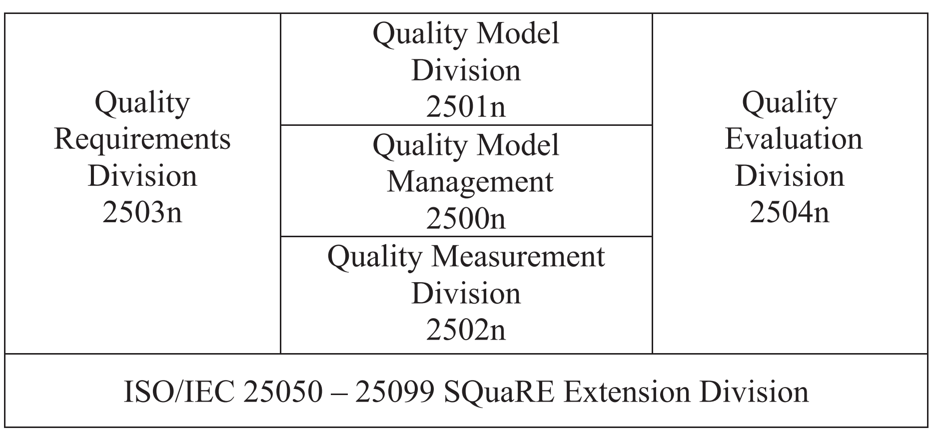


Figure 2-1 Organizations of SQuaRE series (ISO/IEC 25000: 2014)

A brief description of ISO/IEC 25010, 25021 and 25022 are presented in the following section.

2.3.1 Quality Models

The quality of a system is the degree to which the system satisfies the stated and implied needs of its various users. ISO/IEC 2501n International Standards present detailed quality models for computer systems and software products, quality-in-use, and data. This International Standard defines (ISO/IEC 25010: 2011(E)):

a) A quality-in-use model composed of five characteristics that relate to the outcome of interaction when a product is used in a particular context of use. This system model is applicable to the complete human-computer system, including both computer systems and software products.

b) A product quality model composed of eight characteristics that relate to static properties of software and dynamic properties of the computer system. The model is applicable to both computer systems and software products.

The above listed quality models together serve as a framework to ensure that all characteristics of quality are considered from the point of view of each of each stakeholder.

According to the SQuaRE series, "the measurable quality-related properties of a system are called quality properties, with associated quality measures. To arrive at measures of the quality characteristic or subcharacteristic, unless the characteristic or subcharacteristic can be directly measured, it will be necessary to identify a collection of properties that together cover the characteristic or subcharacteristic, obtain quality measures for each, and combine them computationally to arrive at a derived quality measure corresponding to the quality characteristic or subcharacteristic" (ISO/IEC 25010:2011 (E)). Figure 2-2 shows the relationship between quality characteristics and subcharacteristics, and quality properties.

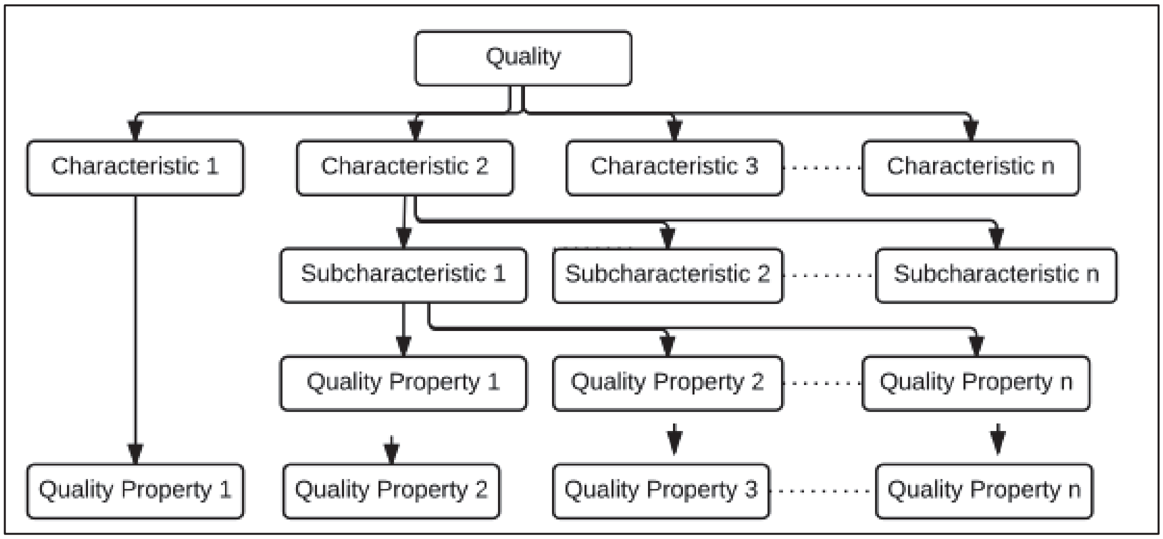


Figure 2-2 Structure used for the quality models (ISO/IEC 25010:2011 (E))

2.3.2 Quality-in-Use Models

Quality-in-use based on ISO/IEC 25010:2011(E) and ISO/IEC 25022:2014 is the degree to which a product or system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency, freedom from risk, and satisfaction in specific contexts of use.

The ISO/IEC 25010, and ISO/IEC 25022 standard definitions of the quality-in-use characteristics are as follows:
Effectiveness measures the accuracy and completeness with which goals can be achieved. Efficiency measures the level of effectiveness achieved to the expenditure of resources. Relevant resources can include mental or physical effort, time, materials or financial cost. Freedom from risk measures the risk of operating the software or computer system over time, conditions of use and the context of use. Satisfaction measures the extent to which users are free from discomfort and their attitudes towards the use of the product. Satisfaction can be specified and measured by subjective rating on scales such as: liking for the product, satisfaction with product use, acceptability of the workload when carrying out different tasks, or the extent to which particular quality-in-use objectives

(such as efficiency or learnability) have been met. Context of use include users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a system, product or service is used.

2.4 Quality Measurement

According to many studies on the application of measurements and models in industrial environments, measurement, in order to be effective must be:

1. Focused on specific goals;
2. Applied to all life cycle products, processes, and resources;
3. Interpreted based on characterization and understanding of the organizational context, environment, and goals.

This means that measurement must be defined in top-down fashion, that is, it must be based on goals and models. A bottom-up approach will not work, because the measurements used to evaluate the many observable characteristics in software and how they are interpreted (e.g. duration, number of defects, complexity, lines of code, severity of failures, effort, productivity, defect density) require clarification, which is impossible without the appropriate models and goals to define the context.

The goal question metric (GQM) approach is the most commonly used mechanism for defining measureable goals (Fenton & Bieman, 2014). A brief summary of GQM is presented in the following section.

2.4.1 Goal Question Metric Approach

The Goal Question Metric (GQM) approach is based on the fact that, for an organization to measure in a purposeful way, it must first specify goals for itself and its projects, then it must trace those goals to the data that are intended to define them operationally, and, finally, it must provide a framework for interpreting the data with respect to those goals (Basili, Caldiera, & Rombach, 1994 ; Fenton & Bieman, 2014). The result of applying the GQM approach is a description of a measurement model targeting a particular set of issues and a set of rules for the interpretation of the measurement data. A GQM model is a hierarchical structure (Figure 2-3) that starts with a conceptual goal. That goal is refined into several sub goals using appropriate questions which usually break the issue down into its quantitative or qualitative indicators. Each indicator is then refined into a measurement procedure and rules for data interpretation. As a result, GQM is useful for ensuring that the proper metrics are allocated for the purpose of achieving the conceptual goal.

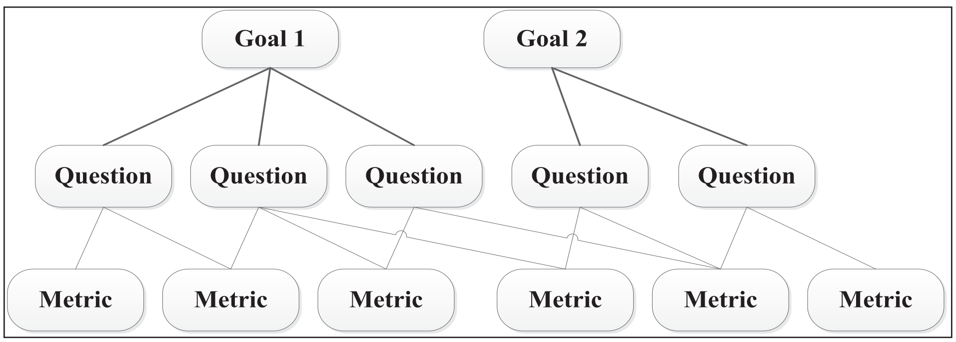


Figure 2-3 Goal question metric approach

Fenton and Bieman claim that GQM is useful for identifying measurement objectives, but it does not address the actual problems of measurement, such as ensuring that it properly characterizes the attribute in question numerically (Fenton & Bieman, 2014). In the following section we describe the measurement process information that address the activities of the measurement process that are required to specify what measurement information is required.

2.4.2 Measurement Process Information

ISO/IEC 15939: 2007 (ISO/IEC 15939: 2007) defines a measurement process applicable to software engineering and management disciplines. The process is described through a model that defines the measurement process activities that are required to specify what measurement information is required, how the measures and analysis results are to be applied, and how to determine the validity of the analysis results. The measurement information model is a structure linking information needs to the relevant entities and attributes of concern by describing how the relevant attributes are quantified and converted to indicators that provide a basis for decision-making.

2.4.3 Quality Measurement Division

In the next section, we define and explain ISO/IEC 25021 and ISO/IEC 25022 since they are related to our approach in objective 1 in (Chapter 3).

2.4.3.1 Quality Measurement Elements

This standard (ISO/IEC FDIS 25021: 2012) defines an initial set of Quality Measure Elements (QME) to be used through the product life cycle for the purpose of SQuaRE. A number of QMEs for quality measures that quantify some of the characteristics and subcharacteristic constitute an initial list to be used during the construction of the quality measures as referenced in (ISO/IEC TR 9126-2: 2003), (ISO/IEC TR 9126-3:2003) and (ISO/IEC TR 9126-4: 2004).

The benefits of defining and using the QMEs include (ISO/IEC FDIS 25021: 2012):

- To provide guidance for organisations developing and implementing their own QMEs;
- To promote the consistent use of specific QME for measuring and using the product properties that are relevant to different product quality characteristics and subcharacteristics;
- To help identify a set of QMEs that are uniquely required in developing all the quality measures for a given set of product characteristics or a set of subcharacteristics.

The QMEs are the common components of a number of quality measures. It will allow users to select and define relevant valid QMEs to define internal, external, data or quality-in-use quality measures. These can then be used to define quality requirements, evaluate products and assess quality, among other things (ISO/IEC FDIS 25021: 2012).

In order to understand and indicate quality (sub) characteristics, Quality Measure (QM) is defined and then QMEs are defined. A measurement function is applied to a QME to generate QM. A measurement method must be applied to a property to define and identify a way to quantify a QME (ISO/IEC FDIS 25021: 2012).

The user of the measurement method should identify and collect data related to quantifying the property (Figure 2-4). Depending on the context of usage and objectives of the QME, a number of properties and sub-properties can be identified. These are the input of the measurement method. They are extracted and defined from the artifacts, components, content or behavior of the target entity.

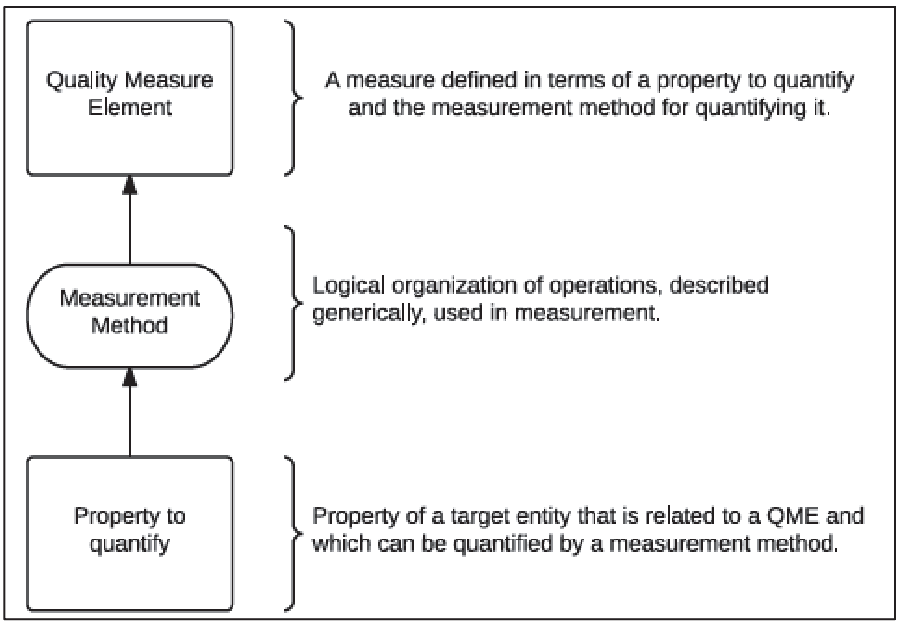


Figure 2-4 Relationship between property to quantify, measurement method and QME (ISO/IEC 25021: 2012)

2.5 Measurement of Quality-in-Use

Measurement of quality-in-use (ISO/IEC CD 25022: 2014) provides measures, including associated measurement methods and QMEs for the quality characteristics in the quality-in-use model. This International Standard contains a basic set of measures for each quality-in-use characteristic and an explanation of how to how quality-in-use is measured. It provides a suggested set of quality-in-use measures to be used with the

ISO/IEC 25010 quality model. The quality measures are applicable to the use of any operational human-computer system. The standard does not assign ranges of values of the quality measure to rated levels or to grades of compliance, because these values are defined for each system, product or a part of the product by its nature and depending on the context of use and users' needs. Some attributes could have a desirable range of values, which does not depend on specific user needs but on generic factors; for example, human cognitive factors (ISO/IEC CD 25022: 2014). The proposed quality-in-use measures are primarily intended to be used for quality assurance of software products.

Quality-in-use measures the outcome of interaction between a user and a system. The final quality-in-use can be measured when an implemented system is used in the user's environment for its intended purpose. Quality-in-use measures can be also used at the conceptual design and development stages (ISO/IEC CD 25022: 2014).

Effectiveness, efficiency and satisfaction can be assessed by observing representative users carrying out representative tasks in a realistic context of use. The measures can be obtained by simulating a realistic usage environment in a usability laboratory or by observing operational use of the product. In order to specify or measure quality-in-use it is necessary to identify each component of the intended context of use: the users, their goals, and the environment of use (ISO/IEC CD 25022: 2014). The evaluation needs to be designed to match this context of use as closely as possible. It is also important that users be given only the type of help and assistance that would be available to them in the operational environment (ISO/IEC CD 25022: 2014).

In the following section we present the different approaches for validating the quality models.

2.6 Software Measurement Validation

Two approaches to validation have been prescribed and practiced in software engineering: (a) theoretical validation e.g. (Fenton & Kitchenham, 1991), and (b) empirical validation e.g. (Schneidewind, 1992). These two types of validation are respectively used to demonstrate that: (a) a measure is really measuring the attribute it is purporting to measure, and (b) the measure is useful, in the sense that it is related to other variables in expected ways (as defined in the following section) (Briand, El Emam, & Morasca, 1995).

1) Theoretical Validation

A measure is valid if it reflects the real meaning of the concept under consideration and is based on the representational theory of measurement.

Measurement validation is "the act or process of ensuring that (a measurement) reliably predicts or assesses a quality factor" (IEEEStd1061, 1998). A measurement can be theoretically validated with respect to ***Tracking*** and ***Consistency*** criteria:

a) The Tracking Criterion

This criterion assesses whether or not a measurement is capable of tracking changes in product or process quality over the life cycle of that product or process (IEEEStd1061). A change in the attributes at different times should be accompanied by a corresponding change in the measurement data. It can be expressed as follows:

If a measure *M* is directly related to a quality characteristic *F*, for a given product or process, then a change in a quality characteristic value from F_{T1} to F_{T2}, at times *T1* and *T2*, shall be accompanied by a change in the measurement value from M_{T1} to M_{T2}.This change shall be in the same direction (e.g. if *F* increases, *M* increases). If *M* is inversely related to *F*, then a change in *F* shall be accompanied by a change in *M* in the opposite direction (e.g. if *F* increases, *M* decreases).

b) The Consistency Criterion

This criterion assesses whether or not there is consistency between the ranks of the quality characteristic of a set of software components and the ranks of the measurement values of the same set of software components. It is used to determine whether or not a measurement can accurately rank, by quality, a set of products or processes (IEEEStd1061, 1998). The change of ranks should be in the same direction in both quality characteristics and measurement values, that is, the order of preference of the attributes will be preserved in the measurement data and can be expressed as follows:

If quality characteristic values $F_1, F_2, \ldots, F_n$, corresponding to products or processes 1, 2, n, have the relationship $F_1 \succ F_2 \succ \ldots \succ F_n$, then the corresponding measurement values shall have the relationship $M_1 > M_2 > \ldots > M_n$:

$$a \succ b \Longleftrightarrow \mathbf{M}(a) > \mathbf{M}(b)$$

Tracking and consistency are a way to validate the representational condition without collecting and analyzing large amounts of measurement data, thus can be done manually.

2) Empirical Validation:

Empirical validation is a process for establishing software measurement accuracy by empirical means. In a given context, we believe the measure of a particular internal attribute is useful if it is related to a measure of some external attribute(s) of the object of study, which is defined in the GQM context as the *quality focus* (Fenton & Bieman, 2014).

Validating base measures and quality measures is fraught with problems. Often, validation is performed by correlating one measure with another. But surrogate measures used in this correlation can mislead the evaluator. It is very important to validate using a second measure that is a direct and valid measure of the factor it reflects. Such measures are not always available or easy to measure. Moreover, the measures used must conform to human notions of the factor being measured. For example, if system A is perceived to be more reliable than system B, then the measure of reliability of A should be higher than that for system B; that is, the perception of ‘more’ should be preserved in the mathematics of the measure. This preservation of the relationship means that the measure must be objective and subjective at the same time: objective in that it does not vary with the measurer, but subjective in that it reflects the intuition of the measurer (Fenton & Bieman, 2014).

Ultimately, it is clear that both theoretical and empirical validation, are necessary and complementary.

2.7 Conclusion

In this chapter we highlighted the main case study that we used to proof our concept for the main thesis. The main topics that are important to understand the quality measurements and empirical study were introduced.

In the following chapter, we introduce a new quality-in-use model for mobile user interfaces and desktop user interfaces, in order to compare various user interfaces in medical apps in the healthcare domain. The proposed quality-in-use measurements for evaluating the MUIs of mobile applications are validated theoretically and empirically in chapter 3.

Chapter 3 Quality-in-Use Measurement Model for the MUI

3.1 Introduction

Quality-in-use is “the degree to which a product or system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency and satisfaction and reduced risk in specific contexts of use” (ISO 25022: 2014). Thus the aim in this chapter is to propose a new quality-in-use model, inspired by the ISO 25010 international standard (ISO/IEC-25010:2011(E)) and adapted specifically to the MUI in the healthcare context.

Our objective here is twofold: (1) to theoretically validate the proposed measurement model, and (2) to compare the quality-in-use of an MUI and a desktop user interface (DUI) for the same application to be used by the same user in the same context.
Two validation approaches are applied: (a) theoretical validation, and (b) empirical validation. These types of validation are used to demonstrate, respectively, that: (a) a measure is really measuring the attribute it is purporting to measure, and (b) the measure is useful in the sense that it is related to other variables in expected ways (as defined in the hypotheses) (Fenton & Bieman, 2014).
We compare the quality-in-use of MUIs with that of DUIs in terms of improving the effectiveness, productivity, efficiency, error safety, and cognitive load success of users in their daily tasks in different contexts using a social networking app as an example.

The chapter is organized as follows: In section 3.2, we state the motivation for this research and describe measurement models available in the literature and their applicability to our work here. In section 3.3, we present our quality-in-use model for MUIs. In section 3.4, we describe the theoretical methods we use to validate our quality assessment model, and, in section 3.5, we describe the empirical method that we use to validate our proposed model. In section 3.6, we empirically validate our model in a case study involving a social networking app. Results of the hypotheses are given in section 3.7. In section 3.8, we present a discussion, and we summarize the chapter in section 3.9.

3.2 Motivation and Related Work

For the purpose of our research is to study the effect of using the software in a specified context instead of the software itself. Quality-in-use is based on the user's view of the quality of the system containing the software, and is measured in terms of the results of using the software, rather than the properties of the software itself. To specify and measure quality-in-use requires not only measures of effectiveness, efficiency, satisfaction, and reduced risk but also details of the characteristics of the users and their goals, as well as the relevant context of use (Bevan, 1999) (ISO/IEC 25022, 2014). In the healthcare environment, quality-in-use is related to a well-defined set of tasks that enables a physician to achieve his or her goals, while usability is more technical, and concentrates on the quality of the specific task and how easy it will be to perform. In short, the upper limit of usability is the lower limit of quality-in-use.

The existing standards in the literature do not satisfy the requirements for measuring the quality-in-use of MUIs in healthcare. In Table 3-1, we compare various standards, such as ISO/IEC 25010 (ISO/IEC 25010:2011(E)), ISO/IEC 25022 (ISO/IEC (CD) 25022:2014), IEC 62366 (IEC 62366: 2007), ISO 9241-210 (ISO:9241-210: 2010(E)), and ISO 9241-11(ISO 9241-11: 1998), with respect to quality-in-use in terms of designing UIs for software (SW) and hardware (HW) in the healthcare domain, specifically for mobile devices in different contexts of use.

Table 3-1 Comparison of selected ISO standards

Standard/Criterion	SW/HW	Healthcare	UI design	UCD method	Mobile use	Context of use
ISO/IEC 25010	SW	√	√	×	×	√
ISO/IEC 25022	SW	√	√	×	×	√
IEC 62366	HW	√	√	×	×	×
ISO 9241-210	SW, HW	×	√	√	×	√
ISO 9241-11	SW	×	√	×	×	√
New Quality-in-Use Model	SW	√	√	√	√	√

The ISO/IEC 25010 quality-in-use characteristics are: effectiveness, efficiency, satisfaction, freedom from risk, and context coverage (ISO/IEC-25010:2011(E)).

The quality model proposed in this standard is highly dependent on the domain, and its definition of quality is too abstract for our purposes. So, we can't apply this standard as is, to the evaluation of the usability of MUIs in healthcare apps.

ISO/IEC 25022 measurement of quality-in-use provides measure for the quality characteristics in the quality-in-use model. However the quality measures included in this international standard were selected based on their practical value. They are not comprehensive, and users are encouraged to refine them if necessary (ISO/IEC CD 25022, 2014).

IEC 62366:2007 focuses on the identification of use-related hazards and risks through usability engineering methods and risk control. It aims to "design in" usability and "design out" user error for medical devices, but it does not apply to clinical decision making that may be facilitated by the use of MUIs in healthcare (Hodgson, 2010).

ISO 9241-210:2010 outlines the basic phases of the human-centered design methodology. It defines the general process throughout the life cycle of computer-based interactive systems, but does not specify the precise methods that are applicable to an MUI in a healthcare environment.

Designing for practical use is a central aim in UCD. This, of course, makes the concept of usability highly relevant (Dahl, 2007), because, according to ISO 9241-11 (ISO 9241-11: 1998), 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. To improve the quality-in-use of mobile applications in healthcare, there is a need to focus on who the users of a product are, what they want to use it for, and where and in what context it will be used. So, deterring the user context of use is a critical step in systematic quality-in-use engineering.

We conclude that a new methodology should be developed especially for determining MUI user context in healthcare, and more generally for managing the process for designing context-based MUI.

A new quality-in-use model and quality-in-use evaluation guidelines have to be derived, as the current standards may not meet the specific needs of MUI users in healthcare. In this chapter, we identify the main phases of our methodology, and propose specific methods for designing an MUI and evaluating its quality in the healthcare context.

Our new quality-in-use model and quality-in-use evaluation/decision making criteria are introduced below.

3.3 A Quality-in-Use Model for MUIs

The objective in this section is to present a new quality-in-use model tailored specifically to the MUI design process. In our approach, the quality-in-use characteristics are decomposed through several layers in a hierarchical structure (see Figure 3-1). The model we propose is based on the ISO 25010 international standard definition of quality-in-use (ISO/IEC-25010:2011(E)). To meet the user's needs in an MUI for use in a healthcare environment, we propose a new characteristic for this model, which we call *cognitive load*. The model is aimed at helping the designer evaluate the mobile app in terms of its ability to execute work-related tasks with: i) increased effectiveness, productivity, efficiency, and satisfaction, because it can be used anywhere and at any time; and ii) increased safety, through error reduction by restricting the use of the UI to the options available, as well as making the UI more pleasant to use and easier to manipulate; and iii) reduced cognitive load on the user, through a reduction in the number of navigation-related actions.

The root of the proposed new quality-in-use model is divided into objective and subjective characteristics (see Figure 3-1).

Objective characteristics measurements are derived from quality measure elements that are applied to the measurement function in order to collect quantitative data based on the results of the user during conducting the test. By doing so, we make sure that different users produce the same measures, regardless of whether they are measuring product or resources because the quantification of the objective is based on numerical rules (ISO/IEC 15939: 2007). This consistency of measurement is very important (Fenton & Bieman, 2014).

Subjective characteristics measurements can vary with the person measuring, and they reflect the judgment of the measure. Subjective characteristics are indicated by subjective ratings on an ordinal scale in the post questionnaires that users fill out in order to collect qualitative data based on their experience using the application after conducting the test. So, the quantification of the subjective characteristics involves human judgment (ISO/IEC 15939: 2007), (ISO/IEC CD 25022:2014).

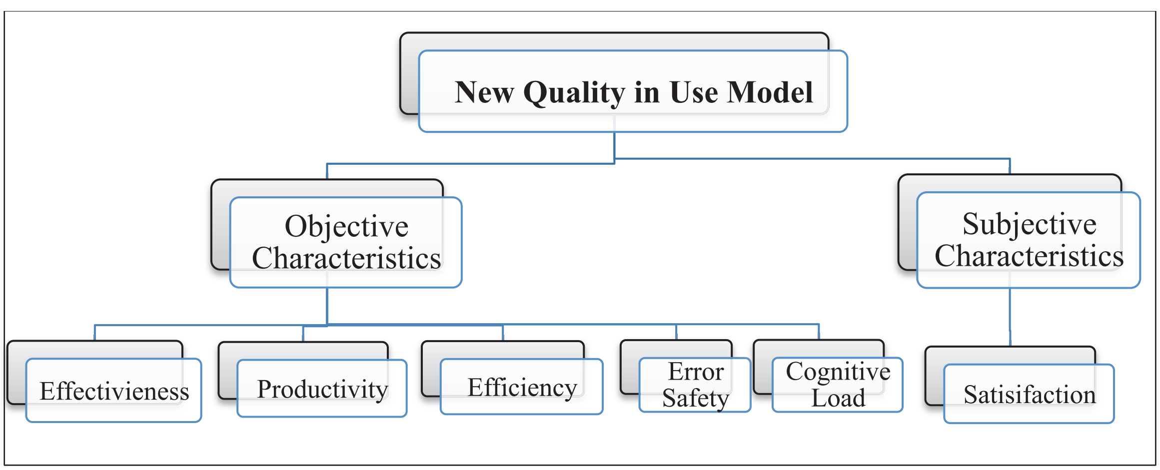


Figure 3-1 New quality-in-use characteristics.

The characteristics of the new quality-in-use model for MUIs proposed in this dissertation are defined as follows:

- ***Effectiveness:*** Effectiveness is defined as the completion with which tasks can be performed with minimum number of actions. It is quantified using the quality measure task completion ratio defined in terms of the minimum number of correct actions required to complete each task in a specified context of use. The unit of measurement is actions per task;
- ***Productivity:*** Productivity is defined as the amount of completed tasks in relation to the time spent on the tasks. It is quantified using the quality measure task productivity defined in terms of the number of actions performed in a specified context of use relative to the time taken by the user to complete the task. The unit of measurement is actions per second;
- ***Efficiency:*** Efficiency is defined as the level of effectiveness with which users achieve tasks in relation to the time required to complete the tasks. It is quantified using the quality measure cost effectiveness defined in terms of the efficiency of the user in completing the task in a specified context of use. The unit of measurement is actions per second;
- ***Error Safety (error prevention and recovery from error):*** Error safety is concerned with minimizing the number of errors during the completion of the tasks by the users. It is quantified using the quality measure error free task completion in terms of the number of errors committed in each action of each task performed in a specified context of use. The unit of measurement is errors per action;
- ***Cognitive load (task navigation):*** It is defined as the level of the cognitive load of the users with which tasks can be achieved in a specified context of use. Movement required on the part of the user from one screen to another, perhaps even back and forth, to perform a 'user task'. This will depend on three things: task complexity, screen size or shape factor, and the way the designer has packed information onto the various screens. Moving from screen to screen adds to the cognitive load on the user. For a given user task, each screen view is weighted by the number of actions

performed on that screen, which must be minimized to keep the user focused on the task at hand, but sufficient in number to increase the user's confidence in using the app and reduce the possibility of the user losing interest during a task. It is quantified using the quality measure task navigation defined in terms of the number of views required to complete each task relative to the number of actions in a specified context of use. The unit of measurement is views per actions;

- ***Satisfaction:*** the user's level of enjoyment as a result of interacting with the app in a specified context of use, in terms of learning the app, using the app, performing a particular task, finding the features, understanding the navigation process, recovering from error, and performing a task anywhere and at any time. It is measured on a Likert scale. Satisfaction is measured by subjective rating on scales that quantify the strength of a user's subjectively expressed attitudes or opinions. The process of quantification can be done in a number of ways, for example, by using an attitude scale based on a questionnaire (ISO/IEC CD 25022:2014).

Table 3-2 presents the definition of our quality measurement elements including the measurement unit and the measurement method (Desharnais et al., 2011).

Table 3-2 Definition of quality measurement elements

QME	Definition
A	The minimum number of actions required to complete a task, measured by the number of clicks recorded.
C	The number of correct actions the user is required to take to complete a task, measured by the number of clicks recorded.
X	The number of incorrect actions that a user performs when completing a task involving risk, recorded to help assess the level of safety of the MUI, measured by the number of incorrect clicks recorded
T	Duration of task, time between start point to and finish point of specified task of software, recorded to help assess whether or not the MUI is sufficiently usable and simple. Measured in seconds.
V	The number of screen views involved in performing a task. Measured by the number of screen views recorded.

Table 3-3 presents the proposed quality-in-use objective characteristics' quality measures and measurement functions.

Table 3-3 Quality-in-use characteristics' quality measure and measurement functions

Quality Characteristics	Quality Measure	Measurement Function
Effectiveness	Task completion ratio	= A / (C + X)
Productivity	Task productivity	= C / T
Efficiency	Cost efficiency	= ((A / (C+X)) / T
Error Safety	Error free task completion	= 1- (X / (C + X))
Cognitive Load	Task navigation	= V/ (C + X)

Figure 3-2 presents the quality measure that used to indicate the quality-in-use objective characteristics, along with their measurement data's interpretation, and the relation between the property to quantify, the measurement method and QME.

Our quality-in-use measurement model is motivated by the ISO 25010 and classified into five objective characteristics. Derived from the ISO 25021 guidelines are the corresponding five properties to quantify that represent the source of the quality-in-use measurement input. Each property to quantify is measured by a measurement method - a logical sequence of operations required to quantify the property. The result of applying the measurement method is a value assigned to the corresponding quality measurement element (QME), which is the most important component of a quality measure. Based on the ISO 25022, a measurement function is defined to combine two or more QMEs into a formula that is used to assign a value to the corresponding quality measure. The quality measures' values with their interpretations represent the quantification of the five quality characteristics.

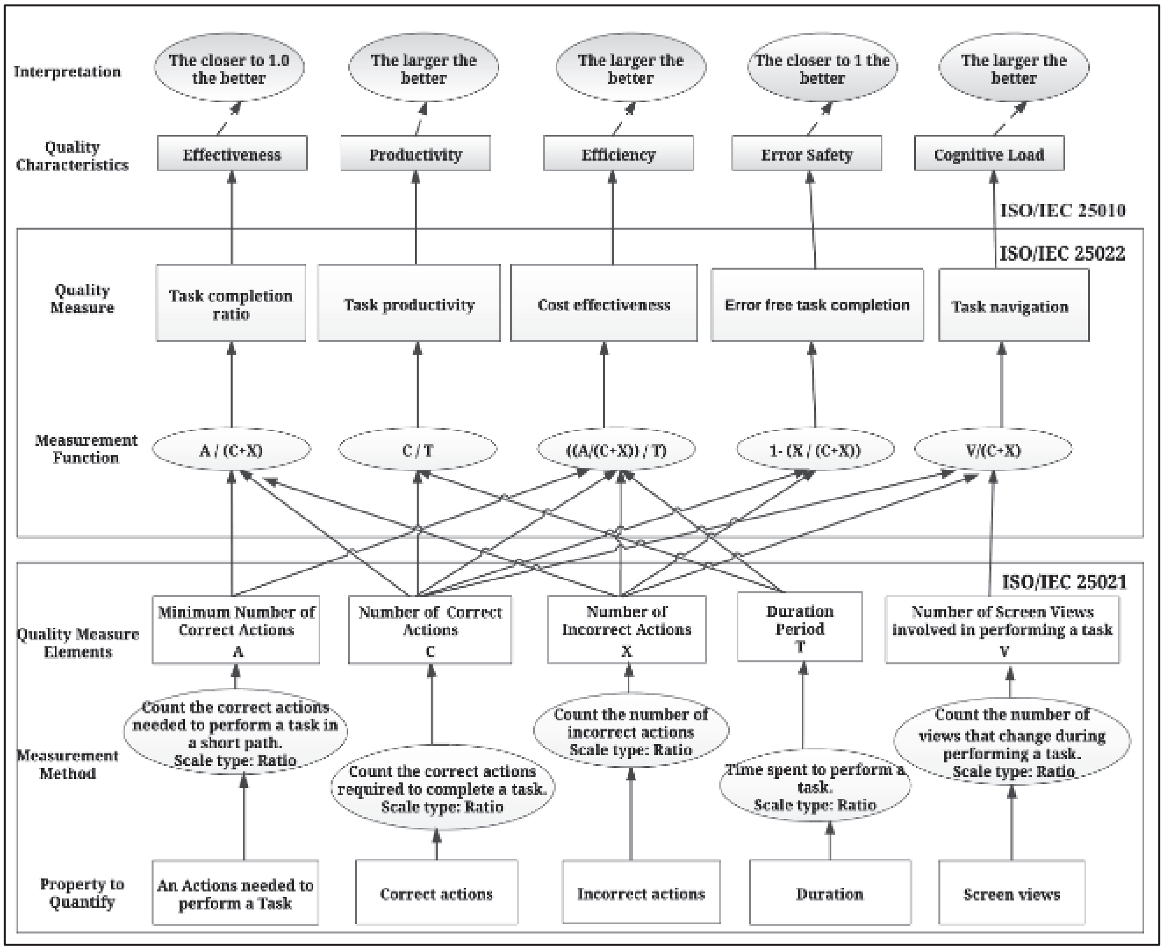


Figure 3-2 New quality-in-use model for MUIs

In the following section, we describe the theoretical validation process of the proposed measurements.

3.4 Theoretical Validation of the New Quality-in-Use Measurements

In this section, we look at measurement validation: both the process of ensuring that we are measuring what we say we are measuring, so that we satisfy the representation condition, and the process of demonstrating the usefulness of a measure, by applying the characteristics and the quality measures that will permit to explain what you are measuring is what we say you are measuring. The validity of a measurement is defined intuitively as the extent to which the measurement reflects the real meaning of the concept under consideration. Consequently, our intuition is the starting point for all measurement.

Validating measurements is enormously important, as this step guarantees the correctness not just of the data, but of decisions taken based on their analysis in a specific environment. The theoretical validity of the measurement methods proposed in this thesis is guaranteed by the representational theory of measurement (Fenton & Bieman, 2014), (Roberts, 1979).

This theory provides a theoretical basis for treating quality-in-use characteristics as mathematical objects and investigates them by means of formal analysis. This means that domain-specific knowledge is used to prove that quality-in-use measurements satisfy certain conditions about an individual's judgments, such as preferences, which make the measurement possible. An example of such an empirical relation would be the following preference: "The user performing a task on UI1 is more effective than the user performing the same task on UI2." In this case, the empirical relation "more effective than" is mapped to a numerical relation between the effectiveness values corresponding to UI1 and UI2 expressed by the symbol >. The ability to incorporate intuition and observation as representational conditions of measurement is the notion underlying our proposed theoretical validation of the measurement model.

The representational theory of measurement seeks to formalize our intuition about the way the world works. In this section, we demonstrate theoretically that our measurements reflect the empirical understanding of the attributes of the entities we observe (such as effectiveness of the MUI), and that our manipulation of the data preserves the relationships (such as "more effective than") that we see among empirical entities. That is, the data we obtain as measures should represent attributes of the entities we observe, and manipulation of the data should preserve the relationships that we observe among the entities. There are two important criteria that a representation condition must meet: tracking and consistency. Tracking validates whether or not a change in the attribute at different times is accompanied by a corresponding change in the measurement data. Consistency complements the tracking criterion by ensuring that the direction of change is the same for both the empirical attribute and its measurements. The theoretical validation of the quality measures of the quality-in-use objective characteristics is described below:

1) Task completion ratio: Based on the meaning of effectiveness, the closer the task completion ratio's value is to 1.0, the better the user's level of effectiveness in a performing a task. For instance, let the minimum number of actions required to perform a task be 3 (A = 3). If user 1 completes the task in three actions (C = 3), then task completion ratio (User 1) = 3/3 = 1, which is excellent, and what we expect. In this case, the level of effectiveness is high, because the user has performed the task with a minimum number of required actions without increasing the total number of actions. If User 2 performs the same task in four actions (C = 4), the effectiveness of User 2 will be lower than the effectiveness of user 1 (for instance, compare the task completion ratio of User 2 which is 3/4 and User 1, which is 1: 3/4 < 1). This is again what we expect, and it makes sense, because if the user performs the task in a minimum number of required actions (short path), he will be more effective than if he performs the task in a larger number of actions (long path).

2) Task productivity: Based on the meaning of productivity, the larger the task productivity value, the better productivity of the user performing the task. For instance, if C = 3 and T = 10 seconds for User 1, then task productivity (User 1) = 3/10 = 0.3. In this case, this user's productivity is greater than that of User 2, who performs the same task with C = 3, but with T = 20 seconds, and so his task productivity (User 2) = 3/20 = 0.15. If we increase the time elapsed, the productivity will decrease, as expected.

3) Cost effectiveness: Based of the meaning of efficiency, the larger the cost effectiveness value, the better efficiency. For instance, if effectiveness = 1 (the highest value) for User 1 and T = 10 seconds, then cost effectiveness (User 1) = 1/10 = 0.1. In this case, User 1's efficiency is greater than that of User 2, who performs the same task in the same time, T = 10 seconds, but this user is less effective (0.75): cost effectiveness (User 2) = 0.75/10 = 0.075), which is a decrease in efficiency, as expected. Now, if the user performs the task with the highest possible effectiveness value (1), but takes more time (20 seconds), then efficiency = 1/20 = 0.5, which is also a decrease in efficiency, as expected.

4) Error free task completion: The ability to successfully complete a task with no errors is essential to ensure the safe use of a medical software in the hospitals. The larger the error free task completion value, the better error safety level; the highest level of error safety corresponds to an error free task completion. For instance, if X = 0 for User 1, and C = 3, error free task completion (User 1) = 1- (0/3) = 1. In this case, the error safety of User 1 is greater than the error safety of User 2, who performs the same task with X = 1 and C = 3: error free task completion = 1- (1/(3+1)) = 0.75. If we increase the number of incorrect actions to 1, the level of error safety decreases, as expected.

5) Task navigation: Based on the real meaning of cognitive load, a decrease in its level is better than an increase in its amount. So, the higher the task navigation value, the better the reduction in the cognitive load. For instance, if the user completes a task with no errors (X = 0), in three correct actions (C = 3) and on three views (V = 3), task navigation = 3/3 = 1. This is interpreted as a perfect score, because the user performed one action in each view, which minimizes the cognitive load. If we increase the number of incorrect actions to 1 and decrease the number of views to 2, the level of the cognitive load will increase, because the user has to perform two or three actions in one screen view.

These calculations establish the theoretical validity of the quality-in-use measurements, as required by the representational theory of measurement. We have seen how the measurement of an attribute assigns a representation or mapping from an observed (empirical) relation to some numerical relation. The purpose of performing the mapping is to be able to manipulate measurement data and use the results to draw conclusions about the empirical attribute. The quality-in-use QME are on the absolute scale type, because the measurement is made simply by counting (actions, views, etc.), and there is only one possible measurement mapping, namely the actual count (Fenton & Bieman, 2014). All arithmetic analysis of the resulting counts is meaningful on the absolute scale type; so, all mathematical and statistical operations are meaningfully applied on the quality-in-use measurement data.

In the section below, quality-in-use measurements are validated empirically through a carefully controlled experiment.

3.5 Controlled Experiment I: Usability Testing of the Quality-in-Use Model for MUIs

In this chapter, we provide empirical evidence that the quality-in-use measurement model is effective on a well-understood problem (use of a mobile device vs. a desktop) and in a familiar environment, such as social networking, of the same nature as the ones where we plan to apply it (healthcare). An example of a social networking app used by graduate students is Twitter. Although Facebook is reportedly the most popular social networking site used by American college students to date, educators are increasingly willing to try to integrate Twitter into the learning process (Junco, Heiberger, & Loken, 2011). Twitter is more amenable to an ongoing, public dialog than Facebook, because it is primarily a microblogging platform (Ebner, Lienhardt, Rohs, & Meyer, 2010). In fact, Twitter has been described by some as a blog that is restricted to 140 characters per post, but at the same time providing the functionality of a social networking site (Ebner et al., 2010).

The empirical study presented in this section involves a controlled experiment, Controlled Experiment I, in which the usability of the Twitter social app is tested on a DUI and an MUI used by the same participant in the same environment, in order to measure the effects of applying the proposed quality-in-use characteristics to DUI and MUI quality assessment. We need to investigate whether or not the theoretical interpretation of each of the characteristics in the proposed new quality-in-use model matches the values of the empirical study's variables and their directions, in the sense that each characteristic is related to other variables in expected ways.

The goal of Controlled Experiment I can be expressed as a set of hypotheses that we want to test. Two sets of hypotheses, one relating to the objective characteristics of the quality-in-use model (effectiveness, efficiency, productivity, Error safety, cognitive load) and the other to its subjective characteristic (satisfaction), were formulated and investigated empirically. Sample hypotheses for the above objective and subjective characteristics are listed below:

HYP^0: There is no significant difference between the effectiveness of the Twitter app using an MUI and the effectiveness of the same Twitter app using a DUI.

HYP^1: There is a significant difference between the effectiveness of the Twitter app using an MUI and the effectiveness of the Twitter app using a DUI.

HYP$^{0'}$: There is no significant difference between the user's satisfaction of the Twitter app using an MUI and the user's satisfaction of the Twitter app using a DUI.

HYP$^{1'}$: There is a significant difference between the user's satisfaction of the Twitter app using an MUI and the user's satisfaction of the Twitter app using a DUI.

3.5.1 Experimental Design

Our participant samples were made up of students registered in a Software Engineering Development Process course, SOEN 6611, during the summer of 2013 at Concordia University. Twenty graduate students were selected at random for the study. Usability testing for the Twitter app was conducted twice for each participant, once using the DUI and once using the MUI, in random order. The independent variables in the study were the base measurements (A, C, X, T, and V) of the two UIs for the Twitter app. The dependent variables were the quality-in-use characteristics (Figure 3-1).

Prior to conducting the formal evaluation of the Twitter app with the participants, a list of materials to be used during testing was prepared, as suggested by Dumas and Redish (1999) (Dumas & Redish, 1999), and includes the following:

1) **Task list*:*** A list of tasks (Table 3-4) was prepared for the test participants. Users normally perform the same tasks on a daily basis, such as search for a friend, follow a friend, write a comment (a tweet), or send a message. The tasks included in Table I are those most frequently performed.

2) **Objective measure (paper login Form):** The base measurements (A, C, X, T, and V) for each participant for each UI type were recorded in a paper login form (Appendix A).

3) **Subjective measure (user satisfaction questionnaire):** The participants were asked to complete a questionnaire containing two types of questions: general questions related to the user profile, and specific questions designed to capture their level of satisfaction using the DUI and the MUI (Appendix A).

Table 3-4 List of participant tasks

Task	
1	Search for a friend
2	Follow a friend
3	Write a tweet, e.g. 'Have a nice day'
4	Send a message to a friend, e.g. 'Good luck!'

3.5.2 Test Environment

The test was conducted one participant at a time in a closed lab in the university building. Each participant was asked to perform a set of tasks (Table 3-4), once using the desktop version of the app and once using the mobile version.

3.6 Results of the Empirical Study

Interviews and observation were the two methods used to gather information from the participants while they carried out the listed tasks on the two platforms. The researchers evaluated the participants' ability to use the DUI-based and MUI-based Twitter app.

3.6.1 Objective Characteristics Results

The objective characteristics of the proposed quality-in-use model provided measures relating to the effectiveness, productivity, efficiency, Error safety, and cognitive load success of 20 graduate students performing 4 tasks on a DUI and an MUI for the Twitter app.

The raw data for the empirical study were tabulated in MS Excel for each of the 20 participants. We automatically calculated the dependent variables for each task separately. Then, we used the mean of all 4 tasks, for all 20 participants, and for each characteristic to compare the quality-in-use of the DUI and the MUI. Our data are presented in two separate tables. The first table, for the DUI, contains the results for the 5 objective characteristics for all the tasks and all the participants. The second table, for the MUI, contains the results for the 5 objective characteristics for all the tasks and the same participants. Since the mean and median of the data for each characteristic in the two tables are the same, our data are normal, and we have two conditions (using the DUI, and using the MUI) for the same participants, and so the data are paired. Consequently, we used the paired Student t-Test for the data analysis. We relied on the mean, because we are 95% confident that the mean of each characteristic lies within the confidence interval.

We first compared the means of all the quality-in-use characteristics collected for the DUI-based and MUI-based Twitter app for all the tasks and all the participants in the test case. Based on the mean (Figure 3-3), there is no difference between the DUI and the MUI for the effectiveness and error safety characteristics. However, there is a difference in the mean between the DUI and the MUI for productivity, efficiency, and cognitive load.

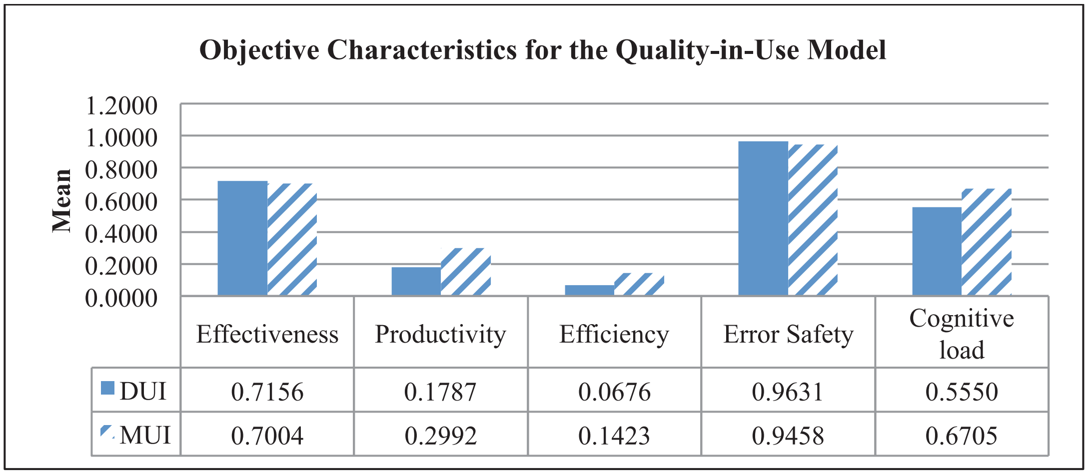


Figure 3-3 Objective characteristics for all tasks for the DUI-based and MUI-based app.

For statistical validation Table 3-5 shows the t-Test values and the P-value for all the characteristics. The hypotheses are verified for each characteristic, based on the t-Test value and the P-values, the critical value approach of the t-Test at 19 degrees of freedom, α =0.025 for the two-tailed test, and a critical level of t = ± 2.09. Our decision rule is to reject the null hypothesis if the computed t statistic is less than -2.09 or more than 2.09.

Table 3-5 Paired t-Test and P-value for all the objective characteristics

	Effectiveness	Productivity	Efficiency	Error Safety	Cognitive Load
t-value	0.3550	-2.2738	-2.0652	0.9088	-2.7069
P-value	0.3632	0.0173	0.0264	0.1873	0.0069

1) **Effectiveness**: The mean for the task completion ration for the DUI is 0.7156, and for the MUI it is 0.7004. So, the effectiveness of the DUI is virtually the same as that of the MUI. Since the t-Test values obtained for effectiveness do not fall into the critical region, we cannot reject the null hypothesis for the effectiveness characteristic. Based on the P-value approach of the t-Test, the P-value for effectiveness is more than alpha. Therefore, we cannot reject the null hypothesis for this characteristic. So, we accept HYP^0, which states that there is no difference in effectiveness between the DUI and the MUI.

Our conclusion is that there is no significant difference between the DUI and MUI for the effectiveness characteristic.

2) **Productivity**: The mean for the task productivity for the DUI is 0.1787, and for the MUI it is 0.2992. So, the productivity of the MUI is greater than that of the DUI. Since the t-Test values obtained for productivity fall into the critical region, we reject the null hypothesis for the productivity characteristic. Based on the P-value approach of the t-Test, the P-value for productivity is less than alpha. So, we reject the null hypothesis and accept HYP^1. Therefore, there is a significant difference between the two UIs, and, based on the mean, the productivity of the MUI is greater than that of the DUI.

3) **Efficiency**: The mean for the cost effectiveness for the DUI is 0.0676, and for the MUI it is 0.1423. So, the efficiency of the MUI is virtually better than the DUI. Since the t-Test values obtained for efficiency do not fall into the critical region, we cannot reject the null hypothesis for the efficiency characteristic. Based on the P-value approach of the t-Test, the P-value for efficiency is greater than alpha. Therefore, we cannot reject the null hypothesis for this characteristic. So, we accept HYP^0. Our conclusion is that there is no significant difference between the DUI and the MUI for the efficiency characteristic.

4) **Error Safety**: The mean for the error free task completion for the DUI is 0.9631, and for the MUI it is 0.9458. So, the error safety of the DUI is virtually the same as that of the MUI. However, the t-Test values obtained for error safety do not fall into the critical region. So, we cannot reject the null hypothesis for error safety characteristic. Based on the P-value approach of the t-Test, the P-value for safety is greater than alpha. Therefore, we cannot reject the null hypothesis for this characteristic. So, we accept HYP^0. Our conclusion is that there is no significant difference between the DUI and the MUI for the error safety characteristic.

5) **Cognitive Load**: The mean for the task navigation for the DUI is 0.5550, and for the MUI it is 0.6705. So, the cognitive load of the MUI is virtually better than that of the DUI (the cognitive load is decreased using the MUI when task navigation is better). Since, the

t-Test values obtained for cognitive load fall into the critical region, we reject the null hypothesis for cognitive load characteristic. Based on the P-value approach of the t-Test, the P-value for cognitive load is less than alpha. Therefore, we reject the null hypothesis for this characteristic and accept HYP^1. Our conclusion is that there is a significant difference between the two UIs, and that, based on the mean, the cognitive load of the MUI is better than that of the DUI.

3.6.2 Subjective Characteristic Results

The subjective characteristics of the quality-in-use model measure the level of satisfaction of the 20 participants using the DUI and the MUI for the Twitter app. They were asked to complete a user satisfaction questionnaire to determine their impressions and opinions of the DUI-based app and the MUI-based app, and to rate their response to interacting with the two UIs.

The results of the questionnaire reveal that 85% of the participants were male and 15% were female. Of the total number of participants, 40% were not familiar with Twitter prior to the study, and 100% use a computer or a mobile device, or both, daily for various tasks. With regard to the satisfaction characteristic, the questionnaire results reveal that all the users rated the following features of both the DUI and the MUI as 'easy' or 'very easy': learning the app, using the app, performing a particular task, finding features, understanding navigation, recovering from error(s), and performing a task anywhere and at any time. In addition, they reported being "very satisfied" with the MUI, which is a higher rating than they gave for the DUI for task performance and for the features it provides. These results agree with $HYP^{0'}$, which means that the level of satisfaction experienced by the participants when using the MUI was no different from that using the DUI. So, overall, they rated the quality-in-use of the MUI for the Twitter app as equivalent to that of the DUI. In terms of the subjective characteristic, the difference between the UI types was not significant, based on the t-Test values.

3.7 Discussion of the Empirical Validation of the Quality-in-Use Model for MUIs

Table 3-6 displays the Pearson correlation values for all the objective characteristics of the new quality-in-use model from the DUI and MUI datasets for the Twitter app. Note that the upper triangle-shaped part of the table (gray background) contains the results of correlating the objective characteristics for the MUI. The lower triangle-shaped part of the table contains the results of correlating the objective characteristics for the DUI.

Table 3-6 Correlation between objective characteristics for DUIs and MUIs

			MUI				
			Effectiveness	Productivity	Efficiency	Error Safety	Cognitive Load
DUI	Effectiveness	Pearson Correlation	1	.088	.163	.216	.357
		Sig. (2-tailed)		.705	.480	.347	.113
	Productivity	Pearson Correlation	.306	1	.828**	.074	-.111
		Sig. (2-tailed)	.177		.000	.750	.631
	Efficiency	Pearson Correlation	.504*	.805**	1	-.236	-.027
		Sig. (2-tailed)	.020	.000		.302	.908
	Error Safety	Pearson Correlation	.458*	.180	.280	1	.263
		Sig. (2-tailed)	.037	.435	.219		.250
	Cognitive Load	Pearson Correlation	.433*	.084	.233	.385	1
		Sig. (2-tailed)	.050	.718	.310	.085	

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

The paired characteristics (effectiveness, productivity, efficiency, and error safety) for the DUI Twitter app users were correlated. The correlation between effectiveness and error safety is significant, and the direction is positive (R=0.458, P-value=0.037), which means that when safety increased, effectiveness increased. Also, there is a positive significant correlation between effectiveness and efficiency (R=0.504, P-value=0.02), which means that greater effectiveness leads to greater efficiency, and vice versa. For the final significant correlation, between productivity and efficiency, the results are (R=0.805, P-value=0). These two variables are highly and positively correlated, which means that increased productivity leads to greater efficiency.

Interestingly, the only significant correlation between variables for the MUI Twitter app users is between efficiency and productivity (R=0.828, P-value=0). The relationship is positive, which means that greater efficiency leads to increased productivity.

Figure 3-2 in section 3.3 provides a theoretical interpretation of each of the characteristics in the proposed new quality-in-use model. These interpretations match the values of the correlated variables and their directions.

Increased effectiveness leads to greater efficiency, and vice versa. Increased productivity leads to greater efficiency, and vice versa. This coincides with the interpretation of both variables. Also, increasing effectiveness increases error safety, which means the higher the better for effectiveness, and the lower the number of incorrect actions the better for erorr safety. In the real-life experience of using the Twitter app, users prefer to access the app and follow their friends on a mobile device, rather than on a desktop. We draw this conclusion from the results of the empirical study, which show that productivity using the MUI is better than that using the DUI. The other characteristics were rated equally in the two UIs. Based on the new quality measures as we defined them, and the interpretation that matches each quality characteristic, the results of this study support our new quality-in-use model. In terms of cognitive load, which is the new characteristic that we introduced into this model for measuring the task navigation on the user when he or she is using the MUI to access the social networking app, the results in Table 3-6 show that there is no correlation between this characteristic and the rest of the quality-in-use characteristics for the DUI or the MUI. Because any new characteristic proposed for a quality-in-use model must be independent on the rest of the characteristics and not highly correlated to them, our decision to include cognitve load in the model to measure task navigation is justified.

Threats to validity: There are many ways in which a study can provide misleading results. Potential problems with empirical studies are classified in terms of categories of threats to validity. Wohlin et al. (Wohlin et al., 2012) describe the following four categories of threats to validity:

1) Conclusion validity. In this controlled experiment, we had two primary purposes: to confirm a theory by applying paired Student t-test to our data (see section 3.6), and to explore the relationships among datasets using correlation analysis to confirm whether or not there is a relationship between two attributes. With this technique, we are generating measures of association that indicate the closeness of the behavior associated with the two variables.

2) Construct validity. We have theoretically validated all five measurements in a narrow sense, which means a measure is valid if it reflects the real meaning of the concept under consideration (see section 3.4).

3) Internal validity. A study has internal validity if the treatment actually caused the effect shown in the dependent variables (Fenton & Bieman, 2014). The reason why the independent variable would affect the dependent variable is that an independent variable can be manipulated to affect the outcome. The outcome, i.e. the result, is, in turn, supported by the values of the dependent variable, which means that the value of the dependent variable is affected by altering the value of one or more of the independent variables.

4) External validity. The results of the controlled experiment can be generalized to anyone accustomed to working with a smartphone.

From the above discussion, it is clear that both the theoretical and empirical validations, as they are defined in this chapter, are necessary and complementary.

3.8 Conclusion

The purpose of this chapter has been to introduce a new quality-in-use model for DUIs and MUIs. The research reported here is built on our previous work (Alnanih et al., 2009), which focused on the design of a DUI for the healthcare domain, and identified specific quality-in-use goals based on HCI principles. The research reported, carried out in the social networking domain, moves into the area of MUI design.

Our main contribution in this chapter is to achieve the central objective of the methodology, which is to design theoretically valid measurements as a foundation for collecting and analyzing data on the new quality-in-use model of MUIs by applying it to a social networking app as an initial example.

We have identified two types of validation for this model: theoretical and empirical. The former addresses the question of whether or not the measurement does, in fact, measure the attribute it is purporting to measure, and the latter addresses the question of whether or not the measurement is useful, in the sense that it is related to other variables in expected ways. We tested the model among 20 graduate students with both objective and subjective characteristics.

In the next chapter, we present our design for MUIs for the medical app based on HCI principles, and validate our design by applying the proposed quality-in-use measurement model in the healthcare context

Chapter 4 HCI Principles and MUIs in Healthcare

In this chapter, now that we have proposed our new quality-in-use measurement model for MUIs, we investigate the applicability of the objective characteristics of the quality-in-use model to a mobile user interface (QiU-4-MUI) as quality subcharacteristics of the impact of HCI principles on user interface design. A controlled experiment is carried out among 23 physicians in a hospital environment to empirically investigate the impact of HCI principles on the quality-in-use of a mobile user interface.

4.1 Introduction

The focus in this chapter is the interaction between a human and a mobile device, in other words, the use of a mobile device by a human to access an app. However, the quality of an MUI is limited by its structure, in terms of the way it presents data and forces navigation from screen to screen. One of the major challenges for UI designers is how to help the novice user to become proficient quickly, and then to become an expert user without the benefit of the training aids that were useful at the start (Alnanih et al., 2009). Novices need an interface that is easy to master and easy to use (Hwang & Yu, 2011), and so the system should keep them informed about what is going on through appropriate feedback provided within a reasonable time. It should also communicate in the user's language, using words, phrases, and concepts that are familiar to them, rather than system-oriented technical terms; follow real-world conventions; and present information in a natural and logical order (J. Nielsen, 2005).

Interface designers face two challenges in the healthcare context: meeting future technology needs, and the ever-growing complexity of technology. Both HCI principles and human factors have a significant role to play in enhancing the quality of healthcare apps, and consequently in reducing the number of errors, especially those associated with the use of medical software. The traditional approaches of HCI to interface design are essential to the design of MUIs, but they are unable to cope with the complexity of the typical interactive devices available today in the safety critical context of medical devices (Thimbleby, 2007).

Simplicity is the most important principle of interface design, which means that the design should make common tasks simple to perform, communicate clearly in the user's language, and provide useful shortcuts that are meaningfully related to longer procedures (J. Nielsen, 2005). To achieve simplicity, the design has to be consistent, and avoid cognitive overload and disorientation on the part of the user. Consistency in design is about presenting elements in a uniform way throughout the interface. This reduces the cost of mistakes and misuse by allowing the user to undo and redo actions, preventing errors wherever possible by tolerating variations in inputs and sequences, and interpreting and implementing all reasonable actions (J. Nielsen, 2005).

User disorientation (also called the "lost in hyperspace syndrome") is experienced by users who browse an information space which has a complex hypermedia structure, and then become lost in it (Dix et al., 2004). Also, users can experience cognitive overload when they face a massive amount of information that is difficult to understand. Exposing users to either of these conditions is a serious shortcoming of a UI, and, for this reason, pertinent information must be organized, managed, and displayed in an efficient and intuitive way on the interface (Dix et al., 2004).

Our aim in this chapter is to explore the impact of HCI principles on the quality-in-use of the MUI design (QiU-4-MUI) for medical mobile apps. We consider HCI principles mental model, metaphor, feedback, affordance, and visibility measuring as quality characteristics of the MUI design, and propose to use the QiU-4-MUI's effectiveness, productivity, efficiency, error safety, and cognitive load, as quality subcharacteristics that rank the MUIs in terms of the impact of the HCI principles on their design quality. Note that the QiU-4-MUI model has already been validated, both theoretically and empirically, on several case studies (see chapter 3) (Alnanih et al., 2013c). In this research, the QiU-4-MUI model is used to empirically investigate the impact of the HCI principles on the design quality of an MUI as compared to their impact on the UI of a desktop app, in terms of the user's effectiveness, productivity, efficiency, error safety, and of improving his or her cognitive load by reducing it. A controlled experiment II is carried out among 23 physicians using the Phoenix Health Information System (PHIS2) app currently installed at King Abdulaziz University Hospital (KAUH).

It is important to point out that the research reported in this chapter builds on our previous work (Alnanih et al., 2009), which focused on the design of a desktop UI in the healthcare domain based on HCI principles. The research for thesis, also carried out in the healthcare domain, moved into the area of MUI design, focusing on human-mobile interaction in the healthcare context.

The rest of the chapter is organized as follows: Related work on designing medical apps based on HCI principles is reviewed in section 4.2. In section 4.3, the HCI principles required to understand this chapter are introduced, and the connection between them and the QiU-4-MUI measurement model is clarified. The PHIS2-M user interface design on the iOS platform is described in section 4.4. Guidelines for designing MUIs based on HCI principles are discussed in section 4.5. Controlled Experiment II is described and the results of the empirical validation on a case study are summarized in section 4.6. Our findings are discussed in section 4.7, and our conclusions and directions for future work are presented in section 4.8.

4.2 Research Motivation and Related Work

Mobile devices can be used to perform tasks that were previously possible only on a PC. Moreover, people are increasingly using applications on the go, as their mobile devices are always with them. However, as the technological sophistication of these devices has grown, the UIs of mobile apps are becoming increasingly complex. Unfortunately, because of the limited set of resources in a mobile device compared to those in a PC, in terms of input and output capabilities, processing power, connectivity, and memory, UI design cannot be directly transferred to a mobile device from a PC. In addition, the mobile context can be totally different to that of a PC. For example, a user on a mobile device might be on the move and have only a limited, and possibly fragmented, time to spend on a task (Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005). Consequently, MUI design is an important characteristic in determining success or failure in the highly competitive mobile market, especially for use in the healthcare domain. The MUI should allow the user to complete tasks. In other words, the design must be both usable and useful, and it must be user-centered. Mobile devices that do not live up to the high expectations of their users will not be able to compete, and hence the needs for new context-based MUI design methodology.

HCI design is aimed at creating UIs which can be operated with ease and efficiency. Most HCI researchers are interested in developing new design methodologies, experimenting with new hardware devices, prototyping new software systems, and exploring new interaction paradigms (Huang, 2009). Another consideration in studying HCI is rapid technological change, which offers new interaction possibilities to which previous research findings may not apply. Also, users' preferences change as they gradually master new interfaces (Dix et al., 2004). The application of HCI principles can enable designers to enhance error detection by capturing clear and timely feedback on the state of the device or the effect of a user's action, for example (Alnanih et al., 2009).
Because of the multidisciplinary nature of HCI, designing UIs for mobile devices poses several interaction challenges (Huang, 2009). One is the fact that the mobile environment is fragmented, and this needs to be considered in the design of interactive mobile apps. Others are designing effective hierarchical menu structures, creating usable navigation and browsing techniques, designing images and icons that are optimized for the small screen. Nevertheless, a study by Localytics estimates that 22 percent of downloaded apps are only opened once. As a result, MUI designers should avoid common errors, such as forcing users to register before demonstrating value; providing overly detailed tutorials, including unusual interface elements or gesture controls; and making customers fill out lengthy forms on mobile devices. These mistakes are all related to their failure to base their designs on HCI principles (Yao, 2013).

Human errors in medical device use account for a large proportion of medical mistakes. Most of these errors are caused by inappropriate designs for user interactions, rather than mechanical failures (Zhang, Johnson, Patel, Paige, & Kubose, 2003). The designers of an MUI for the healthcare environment match its properties to the characteristics of healthcare professionals. This may lead us to consider such devices used in some situations in a hospital setting as medical devices. This is especially true in demanding situations in healthcare environments, such as when the physician is moving between patients' rooms and wants to access patient files through the medical mobile app, or is in a meeting room and wants to review a patient's X-rays.

The process of incorporating HCI principles into MUI design and applying them in the healthcare domain has been explored in a limited number of projects. Albinsson and Zhai (Albinsson & Zhai, 2003) refer to touch screen interaction as literally the most "direct" form of HCI, as information display and control take place on a single surface. This demonstrates that visualization and interaction can work together simultaneously, particularly in mobile situations. The goal of these authors was to design interactive tools that enable rapid and efficient pixel level pointing. Based on several cycles of iterative design and testing, they note that the touch screen has two major drawbacks: the display can be obscured by the user, and interaction with the fingers can be imprecise. In their study, they review earlier selection techniques, such as zoom-pointing and take-off. After investigating a number of techniques, they decided to compare two of the best, cross-keys and precision-handle selection, to the zoom-pointing and take-off techniques. They conclude that some techniques work better than others, depending on usage and situation: "In practice, it is probably more desirable to switch tools according to different needs, just as in the physical world we use pliers, wrenches, screw drivers, and other tools selectively." The focus of the research we report in this thesis is more specific, in that it investigates the impact of HCI principles on MUI design in the healthcare domain.

In (Acharya, Thimbleby, & Oladimeji, 2010), the authors show that some healthcare devices fall far short of expectations, indicating that their designers may not have fully adhered to HCI principles. For example, their study of a basic, interactive hospital bed reveals that the bed's interactive control panel design violates standard HCI principles. It was also badly programmed by the manufacturer. Clearly, something went wrong between design and procurement, and we argue that most of the problems could have been managed, or avoided altogether, by conventional HCI processes. The authors of (Acharya et al., 2010) explore the problems and make recommendations for redesigning the control panel of the hospital bed based on HCI principles. We have to point out here that their work differs from ours, in that they focused on hardware, the hospital bed, while we focus on software, and designing a mobile medical app based on HCI principles.

Zhang et al. (Zhang et al., 2003) are among the few regular contributors to HCI development in healthcare. In 2003, the authors focused on medical devices rather than medical software. They described two human characteristics involved in engineering methods designed for the prediction and evaluation of medical errors in medical device use. Predicting and evaluating patient safety in medical device use are critical for developing interventions to reduce such errors, either by redesigning the devices or by training users on the trouble spots identified in the devices. They demonstrated the initial success of these two methods for infusion pumps, which can be applied to health information systems, as well as to other medical devices, to enable quick and inexpensive evaluation.

In (Tavasoli & Archer, 2009), the authors focus on the eHealth environment with the goal of making the system interface adaptable and easy to learn, since the various users of a mobile eHealth system may have quite different requirements. However, their approach differs from ours in some important respects. Their aim was to drive the interface by means of intelligent policies instead of HCI principles, which can be gradually refined using inputs from users through intelligent agents.

CAMob (Martinez-Garcia & Favela, 2003) is a mobile handheld system designed for communication and collaboration among healthcare professionals in a hospital environment, in which user roles facilitate the communication and collaboration process. However, contextual elements, such as User Category and MUI elements, are not considered.

Buranatrived et al. (Buranatrived & Vickers, 2002) conduct an empirical study of the effect of device type by comparing the performance of the same J2ME app running on a PDA and a smartphone. However, in their study, there is no comparison of app types, as described below for mobile and desktop UIs, and which we empirically test in this chapter.

In this research, the QiU-4-MUI measurement model is used to empirically investigate the impact of HCI principles on MUI quality-in-use. The rationales for mapping the QiU-4-MUI quality characteristics to the HCI principles, such as mental model, metaphor, feedback, affordance, and visibility, are described below.

4.3 Mapping HCI Principles to the Quality-in-Use Characteristics

In this section, we evaluate the effects of applying HCI principles to MUI design in healthcare applications, in terms of the objective quality characteristics of the QiU-4-MUI model. Our approach is to match these characteristics to HCI principles specifically adapted to the type of UI used in the healthcare domain. The quality characteristics of the QiU-4-MUI model correspond to the quality subcharacteristics of the quality characteristics of HCI principles, namely: mental model, affordance, visibility, feedback, and metaphor (see Figure 4-1).

Below we discuss the mapping between the HCI quality characteristics (mental model, affordance, etc.) and the QiU-4-MUI subcharacteristics (effectiveness, productivity, etc.).

1) **Mental Model to Effectiveness:** MUI designers must keep in mind the knowledge that a user may have from his or her experience in the real world, such as using a desktop UI, and how this may be applied to an MUI. HCI practitioners define a mental model as a set of beliefs about how a system works. A mental model is powerful because it is the lens through which we interpret the world, and we interact with any system based on these beliefs.

Applying the principle of the mental model should lead to an increase in the effectiveness of the user when using a mobile application in the context of healthcare, in terms of completing the tasks with minimum number of actions.

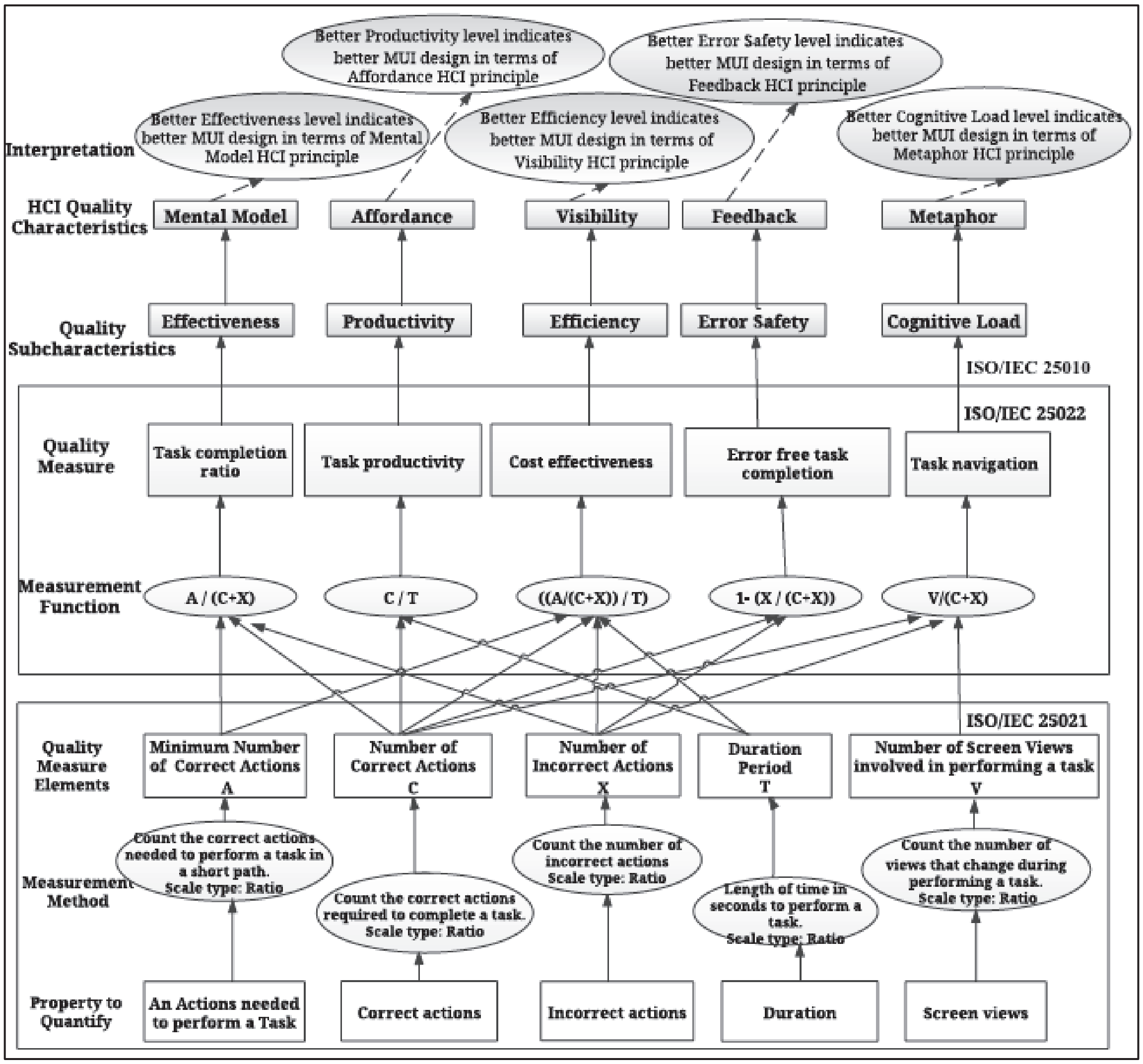


Figure 4-1 Mapping HCI principles (quality characteristics) to the QiU-4-MUI objective quality subcharacteristics

2) Affordance to Productivity: In HCI terms, perceived affordance is the quality that makes it easy for a user to spot and identify the functionalities that a UI offers (McCracken & Wolfe, 2004).

Applying the affordance principle to the design leads to an increase in productivity by decreasing the time taken by the user to complete the task successfully when using the MUI in the context of healthcare domain.

3) Visibility to Efficiency: With visibility, the MUI should help the user understand the current state of the app and the operations that can be performed (Dix et al., 2004).
Applying the visibility principle leads to an increase in the efficiency of the user in completing the task using the MUI in the context of healthcare.

4) Feedback to Error Safety: Feedback is the information that the user should receive concerning the response of the UI to any action performed. Better feedback can eliminate errors in actions performed on the MUI; for example, when an action appropriate for one type of UI is mistakenly performed on another type.
Applying the feedback principle leads to an increase in the safety of the user when using MUI in the context of healthcare, in the sense that the user performs the task correctly with a minimum number of incorrect actions.

5) Metaphor to Cognitive Load: The metaphor principle provides the user with short-cuts to understanding difficult concepts in using the MUI, which can be used to shape user behavior in circumstances that are unfamiliar and that they might otherwise find confusing (Norman, 2002).
Applying the metaphor principle leads to a decrease in the cognitive load on the user when using the MUI in the context of healthcare by reducing the number of actions (clicks) in each view.
In the following section, we present the PHIS2-M's MUI development methodology, and describe the user interface of the iPhone™.

4.4 PHIS2-M User Interface Design on the iOS Platform

We redesigned the PHIS2-D user interface for the iOS platform, which we refer to as PHIS2-M, and evaluated the applicability of the QiU-4-MUI model to the desktop and mobile versions of the UI in the healthcare context. We applied the general guidelines for an iOS application to the design (Manickam, 2011).

To help explain the various MUI views, we have separated the interface into levels, so that each time the user makes a selection, they enter a new level. A level may contain more than one view, and many choices can stem from a single view.

Figure 4-2 shows the flow of the PHIS2-M app, starting with the Login view. A solid arrow represents navigation forward. It also means that going backward will take the user back to the previous view. For example, navigating back from Patient Summary will bring the user back to the Main Menu view. The dashed arrow pointing to a previous view skips the normal backwards flow behavior and takes the user straight to that view. For example, going back from the Pharmacy Order view will take the user back to the "Main Menu" view, instead of the New Order view, and normal backwards flow will continue from there.

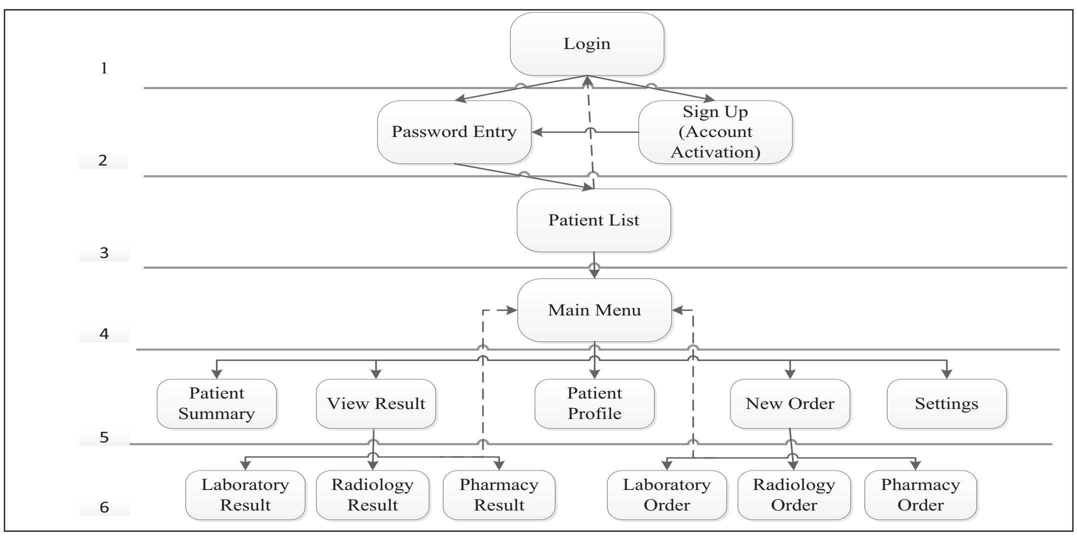


Figure 4-2 PHIS2-M application flow diagram.

The notion is that going to a subsequent level represents going forward, while going to a previous level represents going backward in the app.

The Messaging view is set apart, since it works differently from the other views. Much like text messaging, the smartphone gives a notification when a new message is received, which the user can view by dragging down the Status Bar and pressing on the notification to access the Conversation view.

Our methodology for developing the PHIS2-M user interface on the iOS platform consists of the following steps:

Step 1: Set the goal, which is to design an MUI for physicians' tasks in a healthcare environment, to enable them to enter data and access patient information.

Step 2: Build a low fidelity MUI prototype considering the HCI principles and using paper-based materials, and evaluate it with a pilot test and then with real users (physicians).

Step 3: Build a medium fidelity MUI prototype taking into account the feedback from the low fidelity prototype testing. This prototype was designed to include some visual design and a level of detail somewhere between high and low fidelity, and play a major role in identifying potential layout problems. Our medium fidelity MUI prototype was designed using MobiOne, version 2.0.3, which is the emerging standard for mobile Web development, including iPhone™ application development.

Step 4: Test the medium fidelity MUI prototypes with pilots and real users using interviews and observation testing methods to evaluate the effect of HCI principles on the quality of MUIs.

Step 5: Build a high fidelity prototype, with the emphasis on coding and implementing the images to be displayed; choosing button shapes, font sizes, and colors which enhance understandability and visibility; and adding to the information in the database and retrieving data from it.

Since our aim was to build the PHIS2-M as an iPhone™ app, and Apple mobile devices run exclusively on their own mobile operating system, iOS, we had to use the following in our implementation:

- Xcode: an implementation tool that allows a productive and easy-to-use development environment to be created.
- SQLite: a software library that implements a self-contained, server less, zero configuration, transactional SQL database engine.

In the following section, we provide a list of guidelines for MUIs design inspired by HCI principles.

4.5 MUI Design Guidelines based on HCI Principles

We designed the MUI for the PHIS2-M application on an iPhone™ smartphone, based on HCI principles (Alnanih et al., 2009). The proposed design is expected to reduce user error and enhance quality in terms of the QiU-4-MUI model characteristics. The proposed design follows guidelines that match these HCI Principles, and helps prevent user disorientation:

4.5.1 Mental Model

- Create three tab views, one each for Laboratory, Radiology, and Pharmacy, with a different symbol for each view to speed up information retrieval (Figure 4-3 A).
- Display the words that start with the same first letter in the search of a patient name, drug name, laboratory test result, or radiology test result (Figure 4-3 B).
- Add a Settings option to the main menu to allow the user to adjust font size and brightness level, and to select the text-to-speech option or the speech recognition option (Figure 4-4 and 4-5).

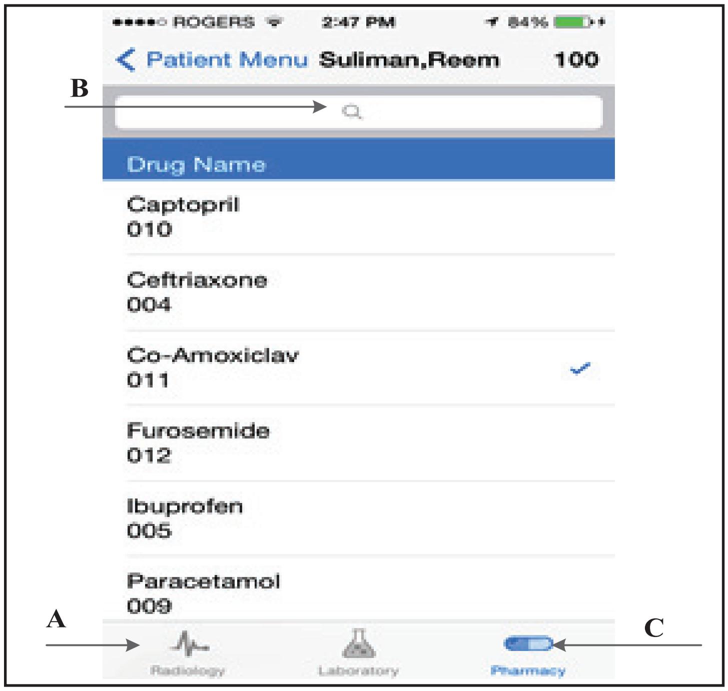


Figure 4-3 Pharmacy: order view.

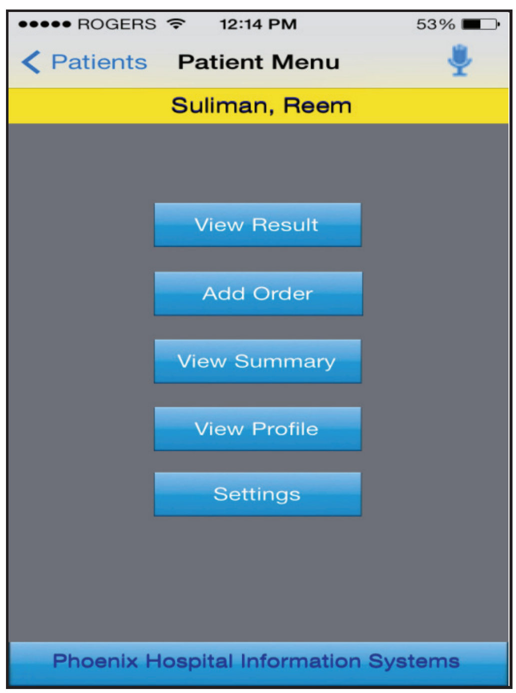


Figure 4-4 Patient menu view.

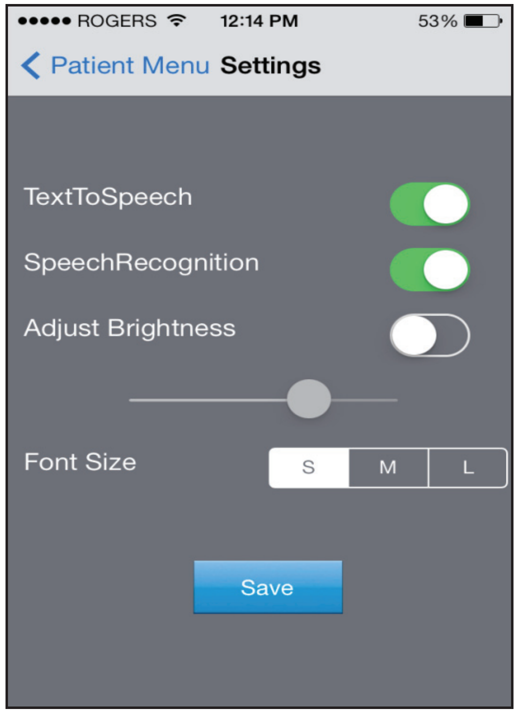


Figure 4-5 Settings view.

4.5.2 Affordance

- Avoid false or misleading affordance. For example, the medication symbol appears on a tab (Figure 4-3 C).
- Reduce the need for typing on the views, by providing a pop-up menu on most of the entry views to help prevent errors (Figure 4-6).
- Facilitate entry on the order views (Laboratory, Radiology, Pharmacy) by providing menu options instead of requiring typed entries (Figure 4-7)
- Set an alert message to be triggered if a selection error occurs, so that the user can recover from the error immediately (Figure 4-8).
- Add a Compare Item function to allow the user to compare lab results by tapping on the row that includes the test to give immediate information (Figure 4-9).

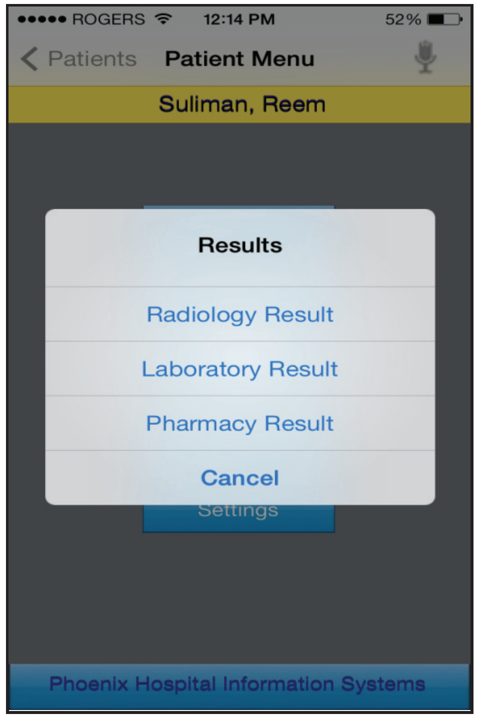


Figure 4-6 A pop-up menu.

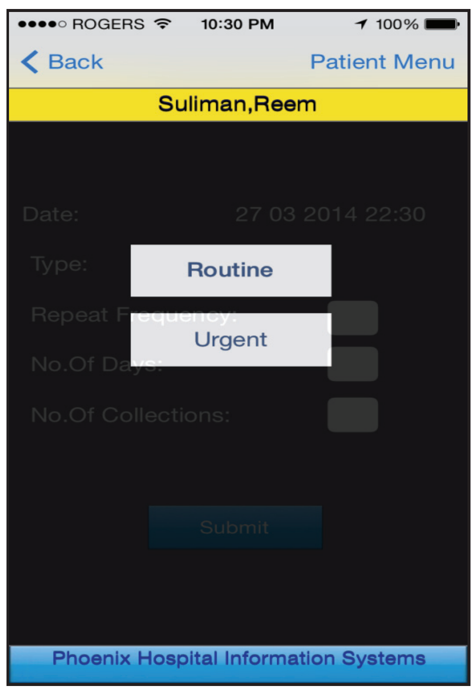


Figure 4-7 Type option in laboratory order view.

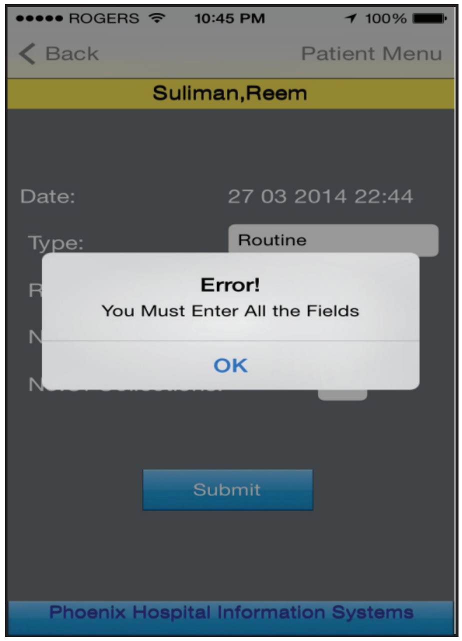


Figure 4-8 Laboratory: order view.

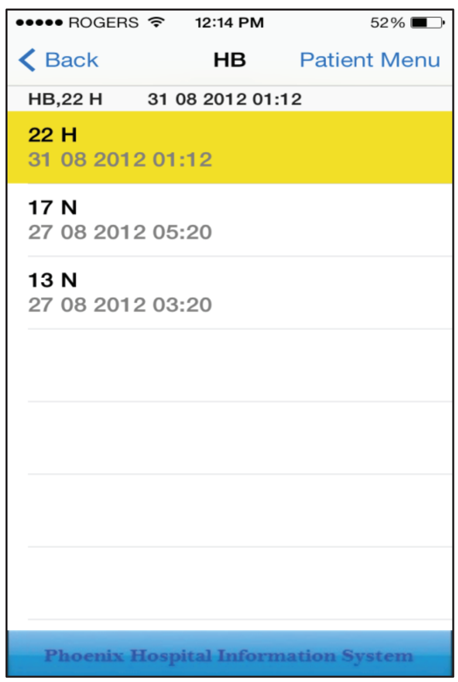


Figure 4-9 Compare Items view.

4.5.3 Metaphor

- On the Radiology, Laboratory, and Pharmacy views, provide options to help the user perform order and result tasks quickly, using familiar medical symbols for navigation purposes (Figure 4-3 C).
- Provide only frequently used navigation symbols, to simplify the app.

4.5.4 Visibility

- Activate a colorful tab symbol for the current view, and deactivate the other tab symbols (Figure 4-3 A).
- Allow the user to select the font size for the views to match his or her preferences (Figure 4-5).
- Provide Order tabs (Figure 4-3) and Results tabs (Figure 4-10) to make the transition from one view to another easy and clear.
- Provide a Patient Menu to give the user access to all the views related to patient care, in order to reduce the cognitive load on the users and to facilitate their access to patient data and make changes if required (Figure 4-4).
- Standardize the design of the views, in terms of button size and color, as well as terminology. Only provide the buttons that users need. Use colors like blue, gray, yellow, and white to help the user feel comfortable using the system (Figure 4-4, Figure 4-10).

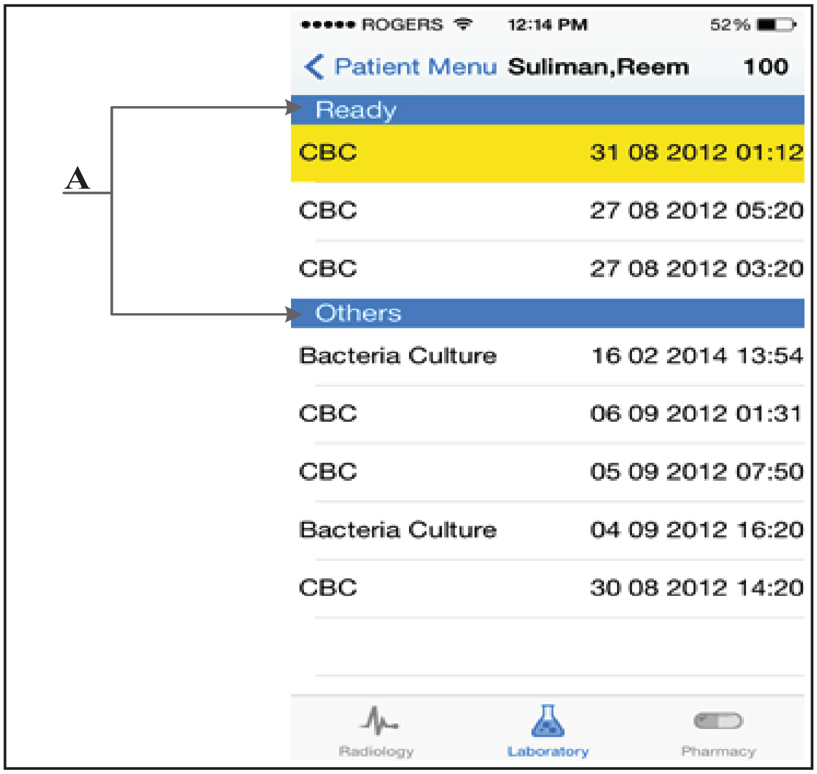


Figure 4-10 Laboratory: results view.

4.5.5 Feedback

- Provide good feedback on all the views. For example, highlight the most recent laboratory result in yellow and display it at the top of the list (Figure 4-10).
- Display the patient name and the patient's medical record number (MRN) on all the main views, such as the Laboratory order and results views (Figure 4-3, Figure 4-10).
- Show the name of the main view in the middle of the screen at the top, such as Patient Menu (Figure 4-4), Settings (Figure 4-5), and the Patients list (Figure 4-11).

Figure 4-11 Patients view.

- Display an alert message if the user attempts to skip a step in an Order view or forgets to select an item from the list of options.
- Provide a banner at the top of the Main Menu to prominently display demographic patient information (Figure 4-4).
- Highlight the view headings in a different color, to help the user select the appropriate task (Figure 4-10 A).
- Provide buttons in the Main Menu view to display a new action, or a pop-up menu offering choices, to help the user manage his or her work effectively (Figure 4-4, Figure 4-6).
- Add letters, for example H for high and L for low, to the lab result column which users can easily understand (Figure 4-12).
- Add a Range column to the lab result table to compare the patient's test results with what is considered normal (Figure 4-12).

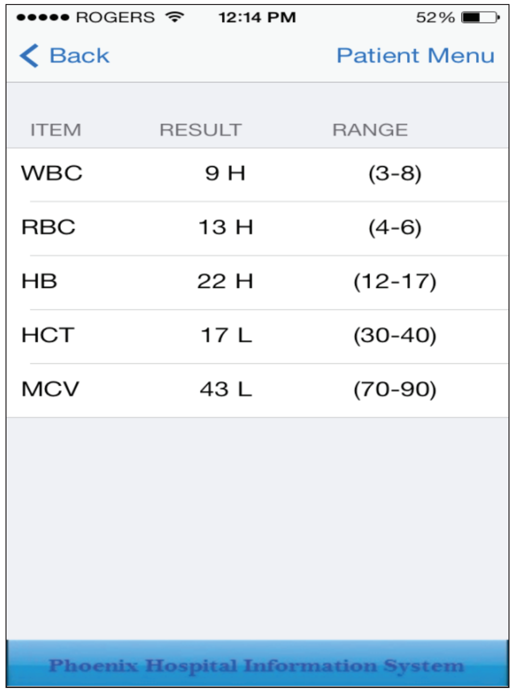


Figure 4-12 Patient's CBC laboratory results.

In the following section, the effect of applying HCI principles to the design of mobile UIs vs. desktop UIs is investigated empirically through a carefully controlled experiment using the QiU-4-MUI model.

4.6 Controlled Experiment II: Usability Testing of PHIS2

The empirical study presented in this section concerns the usability testing of two scenarios of use of the PHIS2, deployed in two different configurations and used by a group of 23 male and female physicians, in order to validate the effect of UI type on the objective quality characteristics of the QiU-4-MUI model. The study was conducted on an iPhone™ 4.1 iOS version 5.1 smartphone.

The QiU-4-MUI characteristics were formally evaluated using PHIS2-D and PHIS2-M by expert physicians at the KAUH to identify problems they were having with the system. The test was conducted one participant at a time in a closed environment with minimum background noise, in order to determine the effect of UI type. Most of the time, the physicians met in conference rooms located on the first and second floors of the hospital, during the normal working hours of each user and for a full week. The participants were asked to perform a set of tasks in a random way, once using the desktop version of the app and once using the mobile version. The test environment, including the participant, the desktop device, the mobile device, the app, and the experimenter, is shown in Figure 4-13.

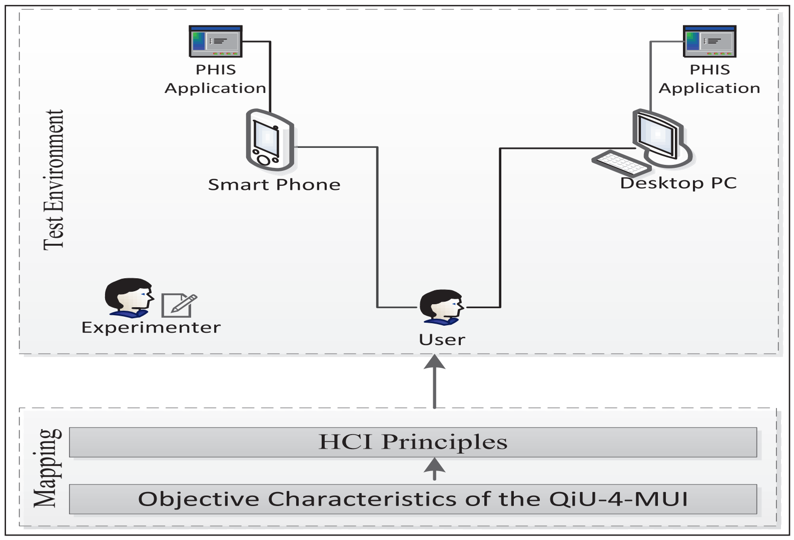


Figure 4-13 Our approach to mapping HCI principles to the QiU-4-MUI characteristics of DUIs and MUIs.

4.6.1 Hypotheses

In order to empirically investigate the effect of the five HCI principles on the quality of desktop UIs and mobile UIs in terms of the five objective characteristics of the QiU-4-MUI measurement model (see section 4.3), pairs of hypotheses related to PHIS-D2 and PHIS2-M were formulated for each characteristic and HCI principle, and are listed below:

HYP1: There is no significant difference in the *task completion ratio* between the desktop UI and mobile UI developed based on the HCI mental model principle.

HYPa: There is a significant difference in the *task completion ratio* between the desktop UI and mobile UI developed based on the HCI mental model principle.

HYP2: There is no significant difference between the *task productivity* of the desktop UI and mobile UI developed based on the HCI affordance principle.

HYPb: There is a significant difference between the *task productivity* of the desktop UI and mobile UI developed based on the HCI affordance principle.

HYP3: There is no significant difference in the *cost effectiveness* between the desktop UI and mobile UI developed based on the HCI visibility principle.

HYPc: There is a significant difference in the *cost effectiveness* between the desktop UI and mobile UI developed based on the HCI visibility principle.

HYP4: There is no significant difference in the *error free task completion* between the desktop UI and mobile UI developed based on the HCI feedback principle.

HYPd: There is a significant difference in the *error free task completion* between the desktop UI and mobile UI developed based on the HCI feedback principle.

HYP5: There is no significant difference in the *task navigation* between the desktop UI and mobile UI developed based on the HCI metaphor principle.

HYPe: There is a significant difference in the *task navigation* between the desktop UI and mobile UI developed based on the HCI metaphor principle.

4.6.2 Usability Testing Materials

Prior to testing the usability of the PHIS2-D and PHIS2-M designs, a list of materials to be used during the test, as suggested by Dumas and Redish (1999) (Dumas & Redish, 1999), was prepared. The materials include the following:

- A list of tasks (Table 4-1) to be given to all the participants in the test. The physicians normally perform the same tasks on a daily basis. These tasks can be summarized, based on those most frequently performed.

Table 4-1 List of participant tasks

Task	
1	Check the most recent Complete Blood Count (CBC)
2	Compare the most recent HB level with the previous level
3	Check the most recent medication prescribed
4	Order a chest X-ray
5	Order a Captopril prescription

- A paper login form was used to collect quantitative measurements on each participant, which includes criteria for objective measures. The time taken to complete each task was recorded. The number of actions performed, the number of incorrect actions, and the screen views were counted. In addition, the participants' comments were noted (Appendix B).
- A satisfaction questionnaire was given to the participants at the end of the test to gather their impressions about using the desktop- and mobile-based PHIS2 (Appendix B).

4.6.3 Data Collection

Two data collection methods were used to gather information from the participants: interviews, and observation while they performed a list of tasks on PHIS2-D and PHIS2-M. With these methods, the researchers evaluated the participants' ability to use the desktop- and mobile-based healthcare applications.

4.7 Results of Measuring the QiU-4-MUI Objective Quality Characteristics

The objective characteristics of our QiU-4-MUI model were used as quantitative indicators of the effectiveness, productivity, efficiency, error safety, and cognitive load of the UIs of desktop and mobile apps designed based on HCI principles. The QiU-4-MUI measurement data were collected during Controlled Experiment II, described in section 6, with 23 physicians in a healthcare environment.

In order to investigate the statistical significance of the observed differences in the objective characteristics of the QiU-4-MUI, the raw data for the empirical study were tabulated in MS Excel for each of the 23 participants. We calculated the objective characteristics automatically for each task. Then we used the mean of all 5 tasks, for all 23 participants, and for each characteristic to compare the MUI and the DUI using the QiU-4-MUI model. Our data are presented in two separate tables. The first table, for the DUI, contains the results for the five objective characteristics for all the tasks and all the participants. The second table, for the MUI, contains the results for the 5 objective characteristics for all the tasks and the same participants. We relied on the mean, because we are 95% confident that the mean of each characteristic lies within the confidence interval.

Since we have two conditions (using the DUI, and using the MUI) for the same participants, the data are paired. To use the paired Student t-Test, we have to check the normality of each pair first. Table 4-2 shows that all the pairs, except for effectiveness and efficiency, follow the normal distribution (the distribution is normal if the P-Value >0.05). Accordingly, we can use the paired Student t-Test to check the difference in the productivity, error safety, and cognitive load characteristics between the desktop and mobile UIs. For the remaining characteristics, we used the Wilcoxon Test, which does not need the normality assumption.

Table 4-2 Normality tests

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	P-value	Statistic	df	P-value
Effectiveness	.234	23	.002	.900	23	.025
Productivity	.101	23	.200*	.979	23	.884
Efficiency	.244	23	.001	.734	23	.000
Error Safety	.112	23	.200*	.949	23	.280
Cognitive Load	.141	23	.200*	.929	23	.106

a. Lilliefors Significance Correction

* This is a lower bound of the true significance.

Table 4-3 shows the t-Test values and the P-value for the three normal characteristics for statistical validation. The hypotheses are verified for each characteristic, based on the t-Test value and the P-values, the critical value approach of the t-Test at 22 degrees of freedom, α =0.025 for the two-tailed test, and a critical level of $t = \pm 2.07$. Our decision rule is to reject the null hypothesis if the computed t statistic is less than -2.07 or more than 2.07.

Table 4-3 Paired t-Test and P-value for three objective characteristics

Objective Characteristics	Productivity	Error Safety	Cognitive Load
P-value	1.29E-05	0.00308	2.60E-08
t-value	-5.58543	-3.32414	-8.3996516

Our t-Test values for productivity, error safety, and cognitive load fall into the critical region, and so the null hypothesis is rejected. The P-values of the t-Test for these characteristics are less than alpha. So the null hypothesis is rejected and accepts HYPb, HYPd, and HYPe. Since the t values for the above characteristics fall into the negative region, we can conclude the following:

- The impact of the HCI affordance principle on the MUI user's productivity is greater than on the DUI user's productivity, as indicated by the task *productivity*.
- The impact of the HCI feedback principle on the MUI user's error safety is greater than on the DUI user's error safety, as indicated by the *error free task completions*.
- The impact of the HCI metaphor principle on the MUI user's cognitive load is greater than on the DUI user's cognitive load, as indicated by the *task navigation*.

Effectiveness and efficiency did not follow the normal distribution, so we used the Wilcoxon Test to check the difference between each pair. In both pairs (Table 4-4), positive ranks are higher than negative ranks. So, in both pairs, the mobile UI version is better than the desktop UI, and we can conclude the following:

- The impact of the HCI mental model principle on the MUI user's effectiveness is greater than on the DUI user's effectiveness, as indicated by the *task completion ratio*.
- The impact of the HCI visibility principle on the MUI user's efficiency is greater than on the DUI user's efficiency, as indicated by the *cost effectiveness*.

Table 4-4 Ranks: Wilcoxon test

		N	Mean Rank	Sum of Ranks
MUI_effectiveness - DUI_effectiveness	Negative Ranks	1[a]	7.00	7.00
	Positive Ranks	22[b]	12.23	269.00
	Ties	0[c]		
	Total	23		
MUI_task_efficiency - DUI_task_efficiency	Negative Ranks	3[d]	6.33	19.00
	Positive Ranks	16[e]	10.69	171.00
	Ties	4[f]		
	Total	23		

a. MUI_effectiveness < DUI_effectiveness
b. MUI_effectiveness > DUI_effectiveness
c. MUI_effectiveness = DUI_effectiveness
d. MUI_task_efficiency < DUI_task_efficiency
e. MUI_task_efficiency > DUI_task_efficiency
f. MUI_task_efficiency = DUI_task_efficiency

4.8 Discussion

The results of Controlled Experiment II prove that applying HCI principles to the MUI design leads to an increase in the quality-in-use of the MUI:

- ***Effectiveness***: Considering the mental model in designing the MUI leads to better control over the actions that are required to complete a task successfully by helping the user make the right decisions and take the right action. This leads to an improvement in user's effectiveness.

- ***Productivity***: Considering affordance in the design of the MUI leads to a more efficient design by suggesting the right functionality and choosing suitable icons and symbols to guide the user in selecting the correct action in less time. This leads to an increase in user's productivity.
- ***Error safety***: Considering the feedback principle in designing MUI provides the user with evidence of closure in the form of direct manipulation interactions, which satisfies the communication expectations users have when engaging in a dialogue via the MUI, and provides confirmation that the operation is being carried out correctly. This leads to an increase in safety by reducing the number of incorrect actions.
- ***Efficiency***: Considering the visibility principle in designing the MUI improves the way information is presented and requested from the user in a clear way, and lets the user know what is happening and what is about to happen. This leads to increased user efficiency in performing a task.
- ***Cognitive load***: The metaphor principle constitutes the overall concept that can help the user to organize all the objects and actions in the MUI into a coherent whole. Considering the right metaphor supports fast and efficient interaction for the user by reducing their cognitive load and enhancing their performance.

The empirical study described in this chapter is analyzed for potential problems classified as the following four categories of threats to validity (C. Wohlin et al., 2012):

1) *Conclusion validity:* This refers to the statistical relationship between independent and dependent variables. Our experimental study has conclusion validity because we applied the paired Student t-Test to our data (see section 4.7), and explored the relationship among the datasets of good user interfaces by evaluating the effects of applying HCI principles on the quality of mobile UIs and desktop UIs in healthcare, in terms of the QiU-4-MUI model. By implementing this technique, we are proving statisticallythe stregth of the association or dependence between the HCI principles and the MUI quality improvement capacity as indicated by the QiU-4-MUI characteristics.

2) *Construct validity:* A study with construct validity uses measures that are relevant to the study and meaningful. We have theoretically validated the fact that all five HCI principles are meaningful in terms of the QiU-4-MUI model in a narrow sense, which means that a measure is valid if it reflects the real meaning of the concept under consideration (see section 4.3).

3) *Internal validity:* A study has internal validity if the treatment actually caused the effect shown in the dependent variables (Fenton & Bieman, 2014). The reason why the independent variables would affect the dependent variables is that the independent variables can be manipulated to affect the outcome. The outcome, i.e. the result, is, in turn, supported by the values of the dependent variables. This means that the dependent variables (objective characteristics of the QiU-4-MUI model) are those whose values are affected by altering the value of one or more independent variables (HCI principles).

4) *External validity:* The results of the controlled experiment II can be generalized to anyone accustomed to working with a smartphone. In this experiment, the controlled environment consisted of a group of healthcare professionals who have had a great deal of exposure to modern technology. For this reason, the results can be generalized to any similar group.

From the above discussion, it is clear that there are no threats to the validity of this empirical study as defined in this chapter.

4.9 Conclusion

In this research, we have attempted to assess the impact of HCI principles on the quality of MUIs in the healthcare context, in terms of: i) increasing user error safety by minimizing human error, ii) increasing the effectiveness and efficiency of the execution of the tasks carried out by physicians, iii) increasing the productivity of the physicians, and iv) reducing the cognitive load on the physicians.

The Phoenix Health Information System (PHIS) that is currently used at King Abdulaziz University Hospital was designed on an iPhone™ smartphone (PHIS2-M) based on HCI principles. PHIS2-M was formally evaluated by users at the hospital. The researchers conclude from the evaluation that considering the HCI principles mental model, metaphor, visibility, affordance, and feedback in the MUI design leads to increased quality-in-use of the MUI.

Our analysis of the effect of applying HCI principles to MUI design evaluated using the QiU-4-MUI model resulted in major improvements to the design of the mobile version of the UI. Results of the formal evaluation of the PHIS-based mobile app revealed that the app is a success in terms of helping physicians perform tasks very rapidly and effectively, in terms of selecting the minimum number of correct actions required, safely, in terms of reducing the number of incorrect actions, and productively, in terms of completing the task successfully in less time. In addition, the physician's cognitive load was reduced by decreasing the number of actions required per view. Therefore, the QiU-4-MUI objective characteristics can serve not only to evaluate the quality of an existing app, but also as a guide in the design of the UI for new mobile medical apps.

In the following chapter, we introduce the user stereotype approach to classifying healthcare physicians, in order to design an adaptable MUI for different users of healthcare applications.

Chapter 5 User Classification in the Healthcare Domain

5.1 Introduction

There are several dynamic changes and challenges that are happening constantly in mobile apps. We cite some of them here: The form factors are relatively small when compared to laptop or desktop apps, the sizes and capabilities of the mobile devices are constantly being improved, competing OS emerge, wireless connections pose new limitations and security issues, users' attention and needs during mobility are very different, and so on. As a result, the characterization and clear understanding of user needs are key issues in the development of MUI, particularly in the context of healthcare apps. To deal with the complex and subjective needs involved in a product's use, new design methodologies may be required to improve our understanding of those needs (Tang & Patel, 1994). It is also important to acknowledge the difference in experience between a junior physician (a physician in training) and a senior (experienced) physician in the healthcare system. The former, who is normally younger, will likely have more experience with mobile technology and less experience in the medical domain, while the latter will have more experience in the medical domain, but probably less experience with mobile technology (see section 5.3).

In this research, the main stakeholders in healthcare delivery are the healthcare providers (physicians), who are directly involved in the process of caring for patients.

The goal in this chapter is to propose a knowledge-based MUI that is adaptable to the needs of the various categories of physicians (user classification), based on age, domain experience, and smartphone experience, specifically using the iPhone™ in our case. Adaptation should consider both the context of use and the user stereotype, unlike traditional approaches, which generate static UIs considering only general user characteristics.

In this chapter, we define a user stereotype model for healthcare apps considering two general characteristics: domain experience, and experience using the iPhone™ technology. Physicians from various hospitals in Saudi Arabia participated in the empirical study we conducted in this research, which we refer to in this thesis as Controlled Experiment III. Note that all the participants share the same background and belong to the same Arabic culture. However, some of them have experience working in Canadian hospitals. It is important to point out that this proposed user model for healthcare apps builds on our proposed context model (Alnanih et al., 2012) introduced in chapter 6, and works in tandem with it.

The chapter is organized as follows. Section 5.2 summarizes related work on user modeling techniques and justifies our choice of the user stereotype in this work. In section 5.3, we present our model, which provides a clear understanding of the user stereotype in the context of our healthcare app. Section 5.4 concludes the chapter.

5.2 Modeling User-related Work

Design and development involves several systematic stages. UCD is, by definition, committed to involving users in every phase of development, and the task of designers is to deal with simplified accounts of users that are suitable for every stage. Users might be skilled or not very experienced, have different levels of background knowledge, and more or less familiarity with technology.

Several techniques have been proposed by the HCI community to help researchers understand users and model their needs. These techniques each have their strengths and weaknesses, and include, for example, the concept of the *persona* (Floyd, Cameron Jones, & Twidale, 2008; L. Nielsen, 2012), the concept of the *archetype* (Dantin, 2005), and the concept of the *user profile* (Versloot, 2005), among others. We determined the concepts that met our requirements based on these models, which are detailed below. Turner and Turner (Turner & Turner, 2011) argue, for example, that the concept of the *user stereotype* might have some validity, in our case the fact that a young iPhone™ user reflects the reality that most iPhone™ users are young, as well as the reality that these phone are very effective tools, as they are easy to understand and as society is cognitively ready to embrace the smartphone technology. We apply the user stereotype technique in this thesis precisely because of the assumptions made regarding the age and experience of users (related to their domain and experience using mobile technology), which is verified through empirical research (in the form of the interviews and the questionnaire).

5.2.1 Persona

In essence, a persona is a fictional, composite description of an individual, complete with name, gender, age, occupation, friends, and potentially all the attributes of real people, including membership in an ethnic group, likes and dislikes, particular educational achievements, and the trappings of their socioeconomic status. Alan Cooper (Cooper, 1999) brought the concept of the persona from marketing to UCD, and so it is geared more to marketing professionals than it is to healthcare workers. A persona is primarily an abstract representation (i.e. a model), in this case of a particular user group.

Pruitt and Adlin (Pruitt & Adlin, 2010) state in "The Persona Lifecycle" that personas are successful in promoting user-centeredness over self-centeredness, that they work well as proxies, since the users themselves aren't always available, and they help process and manage complex needs and preferences. However, the derivation and use of personas have not been without controversy. Floyd et al. (Floyd et al., 2008) note a tendency among some members of design teams to construct personas that reflect their own favorite design ideas, which, contrary to Pruitt and Adlin's research, suggests that personas are, in fact, self-centered constructs.

It is commonly understood that personas imply agreement between the designer and the user. However, Pruit and Grudin (Pruitt & Grudin, 2003) observe that using a persona does not, in fact, mean to imply agreement, because the social and political aspects of the relationship are not integrated. On the contrary, the highly cognitive organization associated with the use of personas implies a lack of agreement with the characteristics of the people for whom they are designing (Turner & Turner, 2011). Personas, they argue, provide a foundation on which to build scenarios, and they are a technique that is powerful only when combined with other methods in an effort to restore these social and political dimensions.

Ronkko et al. (Rönkkö, Hellman, Kilander, & Dittrich, 2004) chose the persona as a means of addressing the lack of agreement about users' requirements. They found that, in practice, the primary design influence turned out to be "new technology, market- and competition-related issues". Competing clients all had their own private technological priorities that effectively overpowered the personas. More recently, Massanari (Massanari, 2010) suggests in a similar vein that persona creation is dominated by internal political realities rather than user needs. Categorization and the use of flat characters is not the preferred approach of the persona method (L. Nielsen, 2012). The research method for our proposed model demands categorization (based on the users' age groups, and their experience in their domain and their experience with mobile technology), which is why we chose the user stereotype (see section 5.4 for a discussion), and not the persona.

5.2.2 Archetype

The term *archetype* comes from the Greek, and means form or pattern. The archetype is a representation of a combination of characteristics that form a basic human pattern, and individuals can be classified based on one or a number of combinations of these characteristics or patterns (Jung, 1981). Some designers, e.g. Dantin (Dantin, 2005), prefer to conceptualize users as somewhat more generic archetypes.

The terms *archetype* and *stereotype* are sometimes confused. But for designers, the differences between them are significant. Both draw from a type of person to create a character, the difference being that the archetype will take the template as the starting point, and the stereotype uses it as the end point. For example, the archetype is the model from which the character is created. In design prototyping terms, the archetype is the medium: paper, pen, stick notes, or MS PowerPoint. But from there, the designer creates the UI design. From the archetype, the writer builds an individual character. Stereotypes are slightly different. A stereotyped character takes a general type of person and oversimplifies their qualities into a predictable type. This is why we chose the stereotype instead of the archetype for modeling the users in our healthcare app.

5.2.3 User Profile

The user profile models individual users, which is not conducive to the achievement of the aims of this research, which are group-based. Factors which are taken into account are user characteristics, motivation and attitude, and context of use information. A user profile adapts to a single user's needs as quickly as possible, while the user stereotype represents the interests of a group of people; in the case of this thesis, three groups organized by experience – junior, intermediate, and seasoned healthcare workers. "The hardest category for the profile to learn is clearly the neutral category with documents that are interesting for some people and not for others" (Versloot, 2005), and, because this research design is, in fact, based on neutral categories (the age group and experience of healthcare workers related to their domain and using mobile technology) with documents that pertain to their experience in the field, the user profile approach is clearly not indicated here. While the user profile contains a great deal of information, there is no widely accepted standard representation of that information. Statistical analyses are used to find the average user profile, which is very similar to market segmentation in the field of marketing, where the aim is to overcome consumer heterogeneity. By trying to design for average users, crucial user information may be lost.

5.2.4 User Stereotype

Stereotypes and stereotyping appear frequently in user representations (Turner & Turner, 2011). As a concept, the stereotype can be understood as a cognitive, social, or narrative approach. From the cognitive perspective, creating stereotypes is an innate activity of the human being, and we use it internally (mentally) to generate knowledge (Hinton, 2013). For example, when we meet new people, we stereotype them immediately in terms of how similar they are to us, and how different they are from us or from others we know. Stereotyping is a natural mechanism of categorization (L. Nielsen, 2012). In a detailed review, Hamilton and Sherman (Hamilton & Sherman, 1996) conclude that a stereotype is a cognitive structure containing our knowledge and beliefs about, in this instance, a social group, and our expectations of it.

From a social perspective, stereotypes typically comprise abstract knowledge about a group. Stereotypes as a concept was, in fact, first introduced to social psychology by Lippmann, who described them rather vaguely as “the little pictures we carry around inside our heads” (Abrams & Hogg, 2006). Allport (Allport, 1979) developed this notion further, describing them as “an exaggerated belief associated with a category” and this remains the dominant and defining characteristic of the term today. Stereotypes help humans navigate in a socially ordered world by assigning people to groups based on their characteristics; for example, if we see a middle-aged person drinking coffee, wearing glasses, reading a book, and acting in a rather absent-minded fashion, we might stereotype this person as a professor. Any such stereotype is, of course, a simplification which puts the individual ‘in a box’, yet is useful for this research to be able to assign a single user to a group of similar users.

One of the goals of this research is to adapt the MUI to individual users, while at the same time assigning each user to one of a number of groups. The user stereotype helps us achieve this goal the most successfully. For this research, we needed to consider two aspects of stereotyping: categorization and accentuation. Through categorization, the differences between people are simplified, and the individuals are assigned to membership groups. Through accentuation, the differences between groups are exaggerated, and the differences between individual members of the same group are minimized (Oakes, Haslam, & Turner, 1994). The symbiosis between categorization and accentuation is an important characteristic of our proposed model. Another is that it is based on the acquisition and use of long-term models of individual users.

A user stereotype model designed for the healthcare domain is more neutral than the word *stereotype* suggests, since this is not an anthropological study of group hierarchies, politics, or dynamics, but more a method of developing accurate categories in order to make the MUI strategically adaptable to types of user.

In the next section, we explain the model of the user stereotype that we derived for our healthcare app.

5.3 User Modeling in Healthcare

In this section, we investigate the idea of creating a user stereotype model through an empirical study based on a questionnaire circulated among physicians and interviews conducted with them, with a view to implementing an adaptable MUI from the perspective of the designer.

Empirical research has been conducted to better understand the nature of the stereotypes involved in healthcare, with a view to developing a practical MUI for these users. Fifty Saudi physicians were recruited from various departments in Saudi Arabian hospitals, some of whom are currently working at the Montreal Children's Hospital in Montreal, Canada.

In order to understand the nature of the role that smartphone technology can play in a healthcare environment, and the relationship among the various user stereotype models proposed, a questionnaire was sent to these physicians, who perform different medical functions. The reason why we decided on a questionnaire at this stage was because of its value as a research tool to help us gather both general and statistical information about our users. The questionnaire was designed with variables in order to collect (Appendix C):

- user characteristics, such as age group and gender;
- user experience in a domain or career;
- user behaviors in an environment or with respect to an application, such as mobile technology adoption and use at work; and
- user preferences (likes and dislikes).

The interview was designed, based on our context variables, to enable us to derive context descriptors (chapter 6, see section 6.5) related to the various physical environments in which the users and the mobile technology will be operating.

To generate the user model, we expressed the responses collected as percentages combining the individual responses given and the individual feedback provided by the physicians. In order to maintain a complete dataset for the users in the various age groups, 30 interviews were selected from the 50 conducted: 10 interviews per stereotype, and 5 male respondents and 5 female respondents per stereotype were considered. We are not looking at a collective sum, but rather at individual percentages related to each user group created among the 30 physicians selected. Table 5-1 illustrates how drilling down through the percentages yielded the three physician stereotypes.

Table 5-1 Experimental data

User Characteristics	Value	User Stereotype 1	User Stereotype 2	User Stereotype 3
Stereotype Name		Junior	Intermediate	Experienced
Age group		25-34	35-44	45+
Gender	Male	50%	50%	50%
	Female	50%	50%	50%
Domain experience *	Novice	100%	10%	-
	Intermediate	-	90%	40%
	Experienced	-	-	60%
Mobile experience **	Basic	-	-	-
	Intermediate	-	-	-
	Advanced	100%	100%	100%
Mobile use at work	Yes	100%	100%	40%
	No	-	-	60%
Mobile use for daily tasks	Strongly agree	100%	80%	20%
	Agree	-	-	10%
	Disagree	-	20%	70%
Mobile use for search location (maps, people, etc.)	Strongly agree	20%	10%	-
	Agree	70%	80%	20%
	Disagree	10%	10%	80%
Mobile use for social networking tools to communicate with others	Strongly agree	60%	30%	-
	Agree	40%	70%	10%
	Disagree	-	-	90%
Common daily mobile apps: medical, Facebook & Twitter, SMS & MMS, Email, WhatsApp ***	Expert	80%	30%	-
	Intermediate	20%	60%	30%
	Novice	-	10%	70%

* Domain experience is defined as **Novice** (less than 10 years), **Intermediate** (more than 10 years and less than 15 years), and **Experienced** (more than 15 years).

** Mobile experience is defined as **Basic** (less than 1 year), **Intermediate** (more than 1 year and less than 2 years), and **Advanced** (more than 2 years).

*** Common daily mobile app experience is defined as **Expert** (uses all the listed applications), **Intermediate** (uses all the listed applications, except medical), and **Novice** (sends messages and makes calls only).

Advanced experience using a mobile phone was set at only two years, because we are interested in designing an MUI that is adaptable to the iPhone 4™, which was released on June 27, 2010.

Our findings show that all the physicians in all the age groups have been using a smartphone for more than 2 years. While this fact does not help us categorize them in a stereotype, it does allow us to assume that they have all been exposed to the basic smartphone platform. Our findings also show that most of the physicians use their smartphone during working hours. Drilling down into the age group data, we found that young and middle-aged physicians are more likely to use their smartphone in their work than older physicians.

Specifically, we found that most junior and intermediate physicians strongly favor using their smartphone in their daily work, as opposed to not doing so. They also favor using it as a means to connect with others in their daily activities, e.g. social networking. Most of the physicians in the experienced category don't use their smartphone for such daily activities. We believe this is due to a lack of awareness of apps that could be helpful to them.

During working hours, young and middle-aged physicians are those who most frequently consult medical websites, such as Medscape; drug dosage and medical calculation apps; and networking apps, like Facebook and Twitter, WhatsApp, Email, SMS, and MMS, while experts in the experienced category use their smartphones mostly for sending messages and making calls.

Table 5-2 demonstrates that the pattern of domain expertise emerging from this study is the following: 100% of the junior physicians are novices in their domain, 90% of the intermediate physicians are intermediates in their domain, and 60% of the experienced physicians are experts in their domain. However, in terms of iPhone™ app experience, this study reveals the following: 80% of the junior physicians are experts, 60% of the intermediate physicians are intermediates, and 70% of the experienced physicians are novices. Based on the age group characteristic in Table 5-2, and following the two categorization rules, which are domain experience and smartphone experience with different apps, our findings support three significant groups out of nine, taking into account their relative expertise based on the categorization rules. These are: Junior (novice: expert); Intermediate (intermediate: intermediate); and Experienced (expert: novice), as illustrated in Table 5-2.

Table 5-2 User stereotype model for healthcare applications

User Stereotype	Junior	Intermediate	Experienced
Age group	25- 34	34- 44	45+
Domain experience	Novice 100%	Intermediate 90%	Experienced 60%
Smartphone app experience	Expert 80%	Intermediate 60%	Novice 70%

5.3.1 User Stereotype 1: Junior Physicians

Junior physicians (medical students and residents) belong to the 25- to 34-year-old age group. These users have the best understanding of the smartphone platform. They spend more time working with their phones, trying out new apps, Web browsing, and using social media apps than their older counterparts. They have very high tolerance towards using new apps, especially if they can help them with their daily work tasks. We note that the users in this group are usually the younger physicians with limited experience in their domain.

From the feedback we collected in the interviews, it was clear that users belonging to this stereotype like the idea of using a smartphone at work. For example, they said they would like to see their smartphone include information about all their patients and their current medications, along with the expiry dates, so they can alert their patients when needed. They also discussed current uses of smartphones in healthcare environments; for example, they can be sitting in one room, watching a live operation taking place in another room, and be able to ask questions and learn, all at the same time. Their only concern is the potential risk associated with smartphones being used near patients and the wireless communication perhaps interfering with medical devices. In the opinion of many of the junior physicians, mobile technologies can provide information on the newest evidence-based medical practices. Also, apps could be developed that track treatment outcomes, to help physicians understand where the results of treatment plans diverge from predicted outcomes and why, so that professionals can work to repair them in the future.

5.3.2 User Stereotype 2: Intermediate Physicians

Intermediate physicians (fellows and newly hired staff) are in the 35- to 45-year-old age group. These users have an understanding of the smartphone platform. They don't spend a lot of time with their smartphones, and so have not had as much exposure to them as the junior physicians. However, they show average to little resistance to using new apps on their smartphones. They are usually experienced physicians who have worked in their field for at least 10 years.

From the feedback collected in the interviews, we note that these users have as many likes as dislikes about using smartphones in their work. They do not seem interested in using a smartphone to record information about their patients, for example. Some said that the procedure they currently use gives them all the efficiency they need. They, like the junior physicians, are worried about using smartphones near patients. This issue probably requires more research. With respect to using new apps on their smartphones, they seem to have little tolerance for making the effort. They don't even seem to have carefully weighed the benefits and disadvantages of doing so.

5.3.3 User Stereotype 3: Experienced Physicians

Experienced physicians belong to the 45-year-old-plus age group. These users have little or no understanding of the smartphone platform. The physicians who spend little or no time with their smartphones and show marked resistance to using them in their work are usually the older ones, who have been in the field for at least 15 years.

From the feedback collected in the interviews, we note that users belonging to this stereotype have more dislikes than likes when using smartphones in their work. This is understandable, given that they are older and in the habit of using more traditional methods. They have almost zero tolerance for being introduced to anything new. They prefer holding papers in their hand when visiting a patient, exchanging information one-on-one with other physicians, instead of using a smartphone or iPad™. They discourage the use of smartphones in the workplace, as they believe that it distracts young physicians from paying attention to their work.

5.4 Conclusion

The purpose of this chapter is to introduce and classify a user model of healthcare physicians based on the user stereotype approach in order to design an adaptable MUI. MUI adaptation is based on the user stereotype model that we develop in this work. This research led us to the conclusion that smartphones, which are currently being used in this environment, can benefit healthcare organizations. In a rapidly changing and increasingly technological world, these organizations are finding that they need better, more efficient, and more effective methods of internal communication and information sharing. The findings from the interviews and the questionnaire have prepared the way for designing and implementing an MUI with features and functions to satisfy every user's tastes. We are now in a position to create friendly, responsive, and scalable MUIs for any smartphone app intended for the healthcare environment.

Among the challenges we encountered in this research was the reluctance of experienced physicians to use smartphones, and their tendency to prefer desktop solutions instead. There are many possible reasons for this resistance, for example: (a) age; (b) fear of the unknown; and (c) physical constraints, such as compromised vision, slower reactions, unsteady fingers, less time available for learning new things, their preference for the larger format of the desktop, etc. This challenge reflects the general resistance we encountered among the senior physicians to trying new things – not only smartphones – which supports our stereotyping method. As revealed in some of the interviews, older physicians may prefer their traditional methods over smartphone use. However, some admitted that presentations explaining the benefits of this technology could help them overcome their resistance, and that they would very likely change their minds if it were for the ultimate benefit of patients and the healthcare organization overall. Two other challenges we faced were the following: the time constraints of busy physicians, and appropriate questionnaire design, in terms of length and content.

Our research involved the use of the model we developed here in the design, development, and testing of a mobile user interface for healthcare apps.

In the next chapter, we examine how context can be characterized for developing context sensitive user interfaces for smartphone apps in a hospital environment.

Chapter 6 Context Model and Methodology

6.1 Introduction

In a dynamic user environment, user efficiency can be improved with a context-aware device that responds to fast-changing situations. The hospital is an example of such an environment. Currently, mobile computing is a supplemental service in hospitals, which complements existing clinical information systems; for instance, it provides an alternative means to access medical information and supports interpersonal communication.

One of the main contributions of this thesis is to characterize the hospital context as a favorable foundation for designing and developing a context-sensitive mobile medical app. Computer systems for healthcare present a number of usability challenges (Ash, Berg, & Coiera, 2004), among them the number of medical errors that occur related to entering and retrieving patient information, which can be caused by: (1) human-computer interfaces unsuited for the highly disruptive use context of hospital work; and (2) cognitive overload resulting from the number of steps required to retrieve correct information. The designer of an MUI can incorporate context-based modifications into the appearance or the behavior of the interface, either at design time or at runtime. It is important for us to point out that our proposed model is based on separating how context is acquired from how it is used, by adapting the MUI features to the user's context. The adaptation of MUI sources, like features, widgets, and sensors, in the recording and accessing of information is based on both our context model and the user stereotype. We provide a brief exposé, along with an example, of an approach for developing an MUI based on context. The research reported in this thesis builds on earlier work carried out for the author's Master's degree (Alnanih, 2008), which focused on the design of a desktop UI in the healthcare domain based on adaptation of the principles of HCI. Our present research, in the same domain, moves into the area of mobile design, focusing on human-mobile interaction. We hope that meeting our objectives in this latest research will enhance this interaction with an MUI that responds intelligently to contextual changes. Our goal is to match the cognitive ability of the user to the cognitive responses of the MUI, and for the mobile device, serving as an extension of the user's tool set, to be able to adapt through an interface model that accommodates change. In other words, as it interacts with the user, keeping in mind the user stereotype, the MUI will change, depending on the situation or the location, for instance, and respond intuitively.

The chapter is organized as follows: In section 6.1, we provide a summary of the literature on context characterization. At the end of this section, we show how our views are similar to those of others and how they are different. In section 6.2, we present our model, which will provide a clear understanding of context as it applies in our app. In section 6.3, we introduce that context, which is the hospital domain, and, for the sake of brevity, we present an overview of a representative example from a sub domain of the healthcare system. Note that we do not address the important issues of security, privacy, and reliability in this section, with regard to healthcare apps. In section 6.4, we explain our use of the context descriptor to extract the characteristics of the user based on the interviews conducted. Section 6.5 concludes this chapter.

6.2 Context Model: Related Work

The commonly used English word *context* can be intuitively understood, but its formal or technical meaning in the area of context-aware computing or in the design of adaptive software systems is quite an elusive one. There is no single definition that suits all purposes. However, a clear and unambiguous understanding of the term will help software engineers to achieve their development goals more efficiently and effectively, whatever the domain of application. A thorough review of the extensive literature in this area would be challenging. In this section, we consider some critical issues, but we do not claim this review to be comprehensive.

Some researchers define the term *context* as the user's physical, social, emotional or informational state, or as the subset of physical and conceptual states of interest to a particular entity (A. Dey & Abowd, 2000). The authors in (A. Dey & Abowd, 2000) have presented the definition or interpretation of the term by various researchers, including Schilit and Theimer (Schilit & Theimer, 1994), Brown et al. (Brown, Bovey, & Chen, 1997), Ryan et al. (Ryan, Pascoe, & Morse, 1998), Dey (A. K. Dey, 1998), Franklin & Flaschbart (Franklin & Flaschbart, 1998), Ward et al. (Ward, Jones, & Hopper, 1997), Rodden et al. (Rodden, Cheverst, Davies, & Dix, 1998), Hull et al. (Hull, Neaves, & Bedford-Roberts, 1997), and Pascoe (Pascoe, 1998).

In Dey and Abowd (A. Dey & Abowd, 2000), the authors are interested in context-aware systems, and so they focus on characterizing the term itself.

In Pascoe (Pascoe, 1998), the author's interest is wearable computers, and so his view of context is based on environmental parameters as perceived by the senses. He also studies the difficulties encountered in building context-aware systems, and introduces the notion of the Context Information Service (CIS). Subsequently, in Pascoe et al. (Pascoe, Rodrigues, & Ariza, 2006), the authors introduce the notion of a universal context model which can enable apps to sense the environment surrounding them, so that they can react or adapt using objects and relationships. Two parameterized relationships are developed: the "is-in" relationship, which links objects in a relationship and suggests a location; and the "is -near" relationship, which the authors have re-labeled as the "as with" relationship and characterizes the difference between a location inference and the question of ownership.

In Gwizdka (Gwizdka, 2000), the author considers what is internal (to the user) in a context and what is external, in dealing with contextual information. Internal context describes the state of the user. It can be composed of the work context (e.g. current projects and their status, the status of to-do items, project team), personal events (i.e. events experienced by the user, which are internalized external events), the communication context (i.e. state of interpersonal email communication), and the emotional state of the user. The external context describes the state of the environment, and can be composed of location, proximity to other objects (both people and devices), and temporal context.

In Winograd (Winograd, 2001), the author introduces the term from a linguistic point of view, and explains how context plays a role in communication or dialogue among people. In our opinion, he also points out three important issues: universal definition is unlikely; and there is a lack of conceptual models, and lack of tools and tool boxes.

In Flanagan et al. (Flanagan, Mantyjarvi, & Himberg, 2002), the authors are interested in the recognition of context based on the features of multidimensional data from different sensors. The authors use a running example. An example of context at the higher order level is: "the user is walking in the evening on a busy street and talking to a friend" (on his mobile phone). The authors of that paper show how this higher order concept is generated from multiple sensor measurements based on an unsupervised clustering algorithm. The sensors used are an accelerometer, a light sensor, a temperature sensor, a humidity sensor, a thermometer, and a microphone.

Schmidt (Schmidt, 2000) suggests that an implicit HCI model based on a situational context needs to be created for both input and output information. In his work, basic mechanisms of perception for acquiring context are discussed, such as context based on external sensors. Our work depends on the internal sensors of a mobile device, and adaptation of the MUI features for both entering and accessing data, such as widgets.

Dourish (Dourish, 2004) considered context as both a representational problem and an interaction problem. He argues that context arises from the activity, instead of taking context and content (activity) to be two separate entities. He considered context to be indivisible, while our model is based on separating how context is acquired from how it is used, by adapting MUI features to the user's context.

Tools like those proposed by Damask (Lin & Landay, 2002) are intended to simplify the process of designing UIs for several platforms, but he does not take into account user categories. MIContext (Savio & Braiterman, 2007) is a context model for mobile interaction that stays at the conceptual level, and details on characterizing spheres of interaction are not provided. The Supple system (Gajos & Weld, 2004), which is concerned with the automatic generation of customized UIs, takes a different approach, which is to treat interface generation as an optimization problem: the rendition of UI widgets is optimized under the device's constraints, while the estimated cost of the user activity is minimized. The Supple study focuses on creating specific algorithms and locating user patterns, but does not take into account characterization of the context related to a mobile device. It only explores mechanisms that would allow users to customize the functionality of the interface, rather than the presence and placement of pre-specified UI widgets. By contrast, Korpipää et al., in their ontology for mobile device sensor-based context awareness (SBCA) (Korpipää & Mäntyjärvi, 2003) propose an interface capable of adapting to the user's current activity. SBCA focuses on modeling at the sensor level, but does not take into consideration adaptation of the input, such as user interactions involving voice or touch. The authors bring up another important issue in the sensing of context, namely the confidence level. CAMob (Martinez-Garcia & Favela, 2003) is a mobile handheld system designed for communication and collaboration between healthcare professionals in a hospital environment, in which user roles facilitate the communication and collaboration process. However, contextual elements, such as User Category and MUI elements, are not considered. Project TEA (EspritProject26900, 2000) developed a sensor board equipped with eight sensors which supplies contextual information. The application described is a mobile phone that recognizes its context (in a user's hand, on a table, in a suitcase, outdoors, etc.). However, this work did not consider MUI features or the User Stereotype, focusing on completely sensor-based context awareness, and adaptation was only addressed for the output mechanisms (ringing mode) and not the input mechanisms. Finally, in (Yamabe & Takahashi, 2007), adaptability is examined in terms of a user's continuously moving state as a necessary component of MUI design, although the number of MUI parameters this research considers is limited.

Most of the research in this area has been based on analyzing context-aware computing that uses sensing and situational information to automate services, such as location, time, identity, and action. More detailed adaptation has been generally ignored, input data based on context, for instance. In this chapter, we have focused on categories such as location, time, environment, user, task, and object. These abstract notions of context need to be incorporated into MUI design, so that mobile context-aware tools can be provided to healthcare personnel in a hospital environment. One of the ways of achieving this is to consider what context means in this application domain.

6.3 Our Context Model (CON-INFO)

In this section, we characterize the term *context* from the MUI designer's perspective, with a view to implementing an adaptable MUI. Let us look at a scenario in a hospital ward, in which a task is performed by an actor (a physician) using different objects (such as patient data). We define the vector <L, M, E> as the *basic context (BC), where L* indicates a *location, M* indicates the *time* when the task was performed and *E* characterizes the *physical environment.* Zero or more of the basic context parameters might be relevant to the set of tasks performed in the domain of application considered. Each of these parameters in the basic context can be characterized by one or more features at a varying level of granularity; for example, the location can be a patient's room; the patient's bed; the emergency room; a transplant ward, etc.

We define the vector <A, T, O> as the *domain-based context (DBC) where A is an actor, T is the task performed, and O indicates the target object.* Each of these parameters in the DBC can be characterized by one or more features. For example, the actor belongs to a user stereotype (a young physician or a senior physician, for example), and is characterized as such because we are interested in adapting the MUI to user stereotypes. The actor (a physician in this case) is playing a role (health promotion for patients, in this case), but that role may involve several tasks. The task *T* is obtained from the task model, where different methods could exist to perform the same task in a context with different context descriptors. The *informational object set* is made up of all the data objects on the MUI, e.g. patient data.

Separating the basic context parameters from the rest of the context-defining parameters creates the possibility of creating a CIS (Pascoe, 1998) to facilitate the sharing of context recognition among multiple applications running in different places in a large hospital. With our definition of context, we can create many levels of application (Figure 6-1).

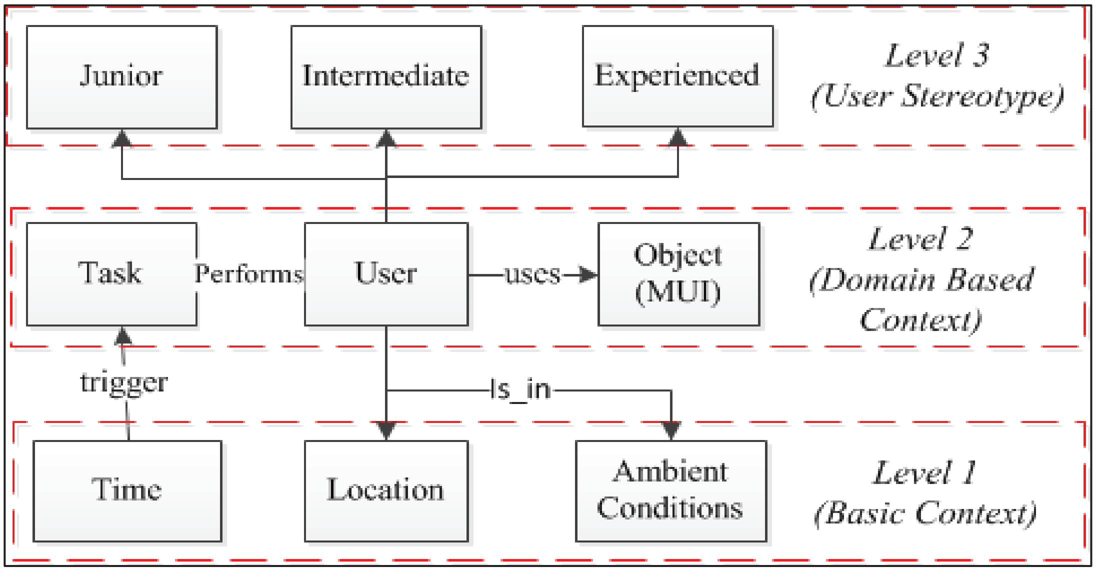


Figure 6-1 Context characterization.

Level 1 is the BC, which is independent of the DBC, because it includes similar applications, whether or not they are mobile, that run in a particular location, at a particular time, and under particular environmental conditions, light and audio, for example. The BC is acquired from the sensors installed in the smartphone. Level 2 is the DBC, which is dependent on the BC, clearly meaning the adaptation of mobile-based data entry rather than computer-based data entry, because the BC changes when the application is executed on the mobile device, while in the stationary case the BC does not. For example, when the user performs a task using a mobile device, the MUI will adapt the features based on the available BC, which is the handling context. Level 3 is the user stereotype, and represents the user preferences and profile. The MUI will also adapt the features to the user stereotype, as well as to the BC.

6.4 Hospital Domain

Our app context is the hospital domain. A hospital is a sub unit of a larger healthcare system which can have an emergency unit, several hospital wards, specialized diagnostic or service divisions, and outpatient clinics. It can be viewed as a coordinated set of services, including physician consultation, examination, imaging, testing (blood, pulmonary function, etc.) and nursing, among others. Owing to the large volume of information and varieties of operations to be handled in hospital automation processes, the control and data management procedures that support information classification for decision-making purposes can become quite complex. In our research, we are focusing on the particular area of the hospital ward that involves physician's services. The relevant actors are: nurses, patients, physicians, pharmacists, orderlies, nutritionists, and PCAs (personal care assistants). An actor can play many different roles, and so the relationship between roles and actors is many-to-many. All role playing in hospitals involves both information-based actions (e.g. record keeping) and skill-oriented physical actions (e.g. physical exams).

A mobile smartphone has been shown to be of significant benefit to these workers (at no cost to the user for extra sensors or devices), considering both their context and user stereotype, and the ability to use the device anytime and anywhere in the hospital as a recording and display medium for patient information.

It is important to provide hospital workers with point-of-care access to information by bridging the gap between caregivers and the information they need using computer and network technologies. The EU-funded Ward-In-Hand project (Ancona, Dodero, Minuto, Guida, & Gianuzzi, 2000) was one of the earliest research efforts to investigate the use of mobile computers and wireless access to medical information at the point of care. More recent work has focused on the use of contextual features, like hospital workers' physical locations, to provide point-of-care access to relevant medical information, such as the MobileWard prototype (Høegh & Skov, 2004). In addition, there are many hospital scenarios in which the UbiComp technology may prove helpful, such as (Bardram, 2004), where the author focuses on designing a context-aware pill container and hospital bed, both of which react to what is happening in medical terms at the hospital, and adapt to it. The paper concludes that clinical work requires context aware clinical applications. However, that paper's focus on the context of patient beds in a defined zone, which constitutes the location entity, is limited, because the display of patient information will only function in the presence of a nurse within the zone. Our work, by contrast, will move beyond examining the infrastructure of a single zone by making it possible to adapt MUI features to the various input and output contexts of a physician.

A 'Sub Domain' Scenario in a Hospital Ward

Consider one of the many tasks carried out by a physician, examine the patient physically, for example. This task may involve taking the patient's temperature; checking his or her pulse, oxygen saturation, or pain level; examining the healing status of a wound; etc. The frequency of recording the results of such examinations in a single day depends on the patient's status. In a traditional system, the physician notes down the important lab test that the patient needs on a piece of paper and gives it to the nurse. The nurse keeps it in her pocket, and later enters in the patient's file. This paper file serves as a medium of communication with other actors who need to share that patient information. Typically, a physician on duty in a hospital ward looks after two to six patients during one shift. Noting and copying the data of multiple patients can be a possible source of error, and sharing a paper document can be less reliable than sharing information electronically. In a modern hospital ward, there is usually a nursing station, along with a desktop computer accessible to all nurses for entering vital signs into shareable electronic records. An improved method would be to give a mobile device to every physician and nurse on duty, which would allow them to enter the patient information immediately, ideally at the point of care. This mobile device can also be used by a nurse to communicate instantly with other actors and to verify specific details about medications and medical care protocols in complex situations.

As explained in the section below, context descriptors contain the information that we extracted from the interviews with the members of the three categories defined in chapter 5.

6.5 Context Descriptors of Users in Healthcare Applications

Our design goal is to develop an MUI for healthcare apps that is adaptable to the user stereotype (chapter 5) and to the context model (Figure 6-2). Based on the characterization of healthcare app contexts and the model of user stereotypes that we develop here, we have conducted a thorough requirement analysis and drawn up a set of context descriptors through interviews with 30 physicians (Controlled Experiment III). A context descriptor (CD) at a higher level of abstraction can be viewed as a statement or a sentence fragment (Si) that forms the basis for MUI adaptation. An example of an Si could be: {*I am in an isolation ward; on my way to pick up something; my hands are not free; ambient light is too low; etc.*}. In general, context descriptors can arise from a variety of sources, such as constraints to be satisfied, the physical state of an individual, etc. The user model can add context descriptors to the Si set. For example, the user stereotype Uj may add {I need large letters; I don't like being disturbed when my hands are not free; etc.}

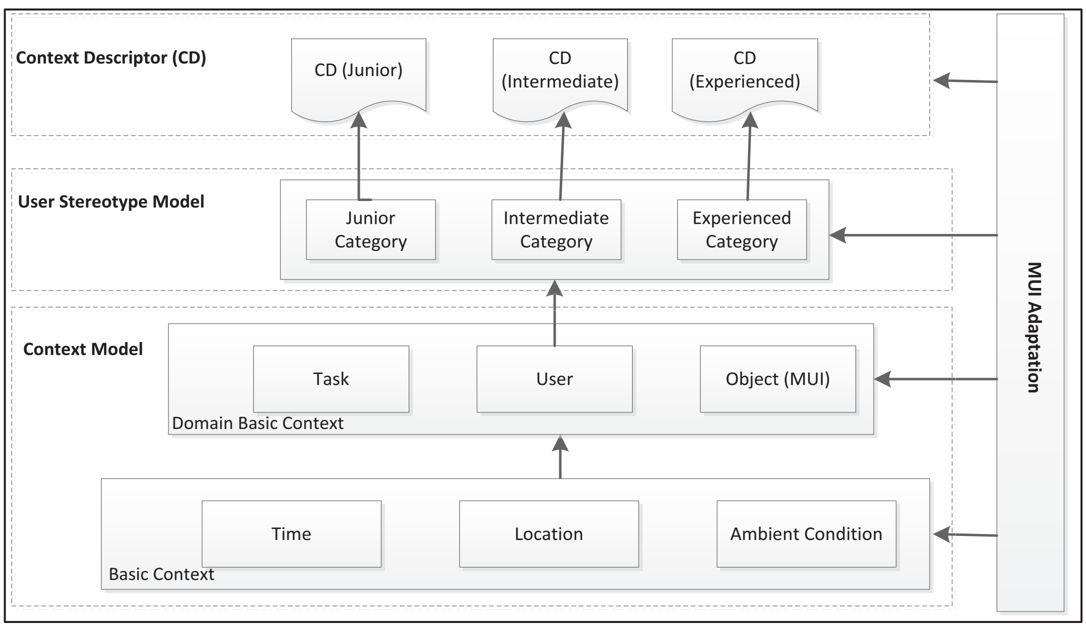


Figure 6-2 MUI adaptation approach.

From our interviews with 10 physicians for each of the 3 user stereotypes (chapter 5), we collected different CDs for each user stereotype. For the sake of clarity, we provide a sample of the CDs collected from each user stereotype in Table 6-1.

Table 6-1 Sample of context descriptors

User Stereotype	Context Descriptor
Junior	1. I want conversations taking place in meetings to be recorded automatically. 2. When I am in a meeting, if someone calls twice, mark the call as urgent and vibrate. 3. When I am calling another physician, inform me if he or she is in a meeting. 4. I am in an isolation room, and so my smartphone is not with me; store messages or notify callers. 5. I am on the way to pick up something; I don't like being disturbed when my hands are not free. 6. It is late at night. So, do not make a noise; enable voice input. 7. I am gone for a pick-up (on a break). So, hold messages except those from enabled sources.
Intermediate	8. I am in a patient's room which has multiple beds. So, no voice output. 9. I am in a meeting (with a physician). So, vibrate and do not ring; indicate urgent messages. 10. I want to be able to adjust the screen's brightness in daylight. 11. I want to be able to adjust the volume when my phone is in speaker mode. 12. I want different access levels for patient profiles. 13. I am at the nurse's station (free to interact), so, resume normal operation.
Experienced	14. I have difficulty reading small fonts, so I need large letters. 15. My environment is noisy most of the time. So, speak louder or confirm the input given. 16. I do not want to be disturbed during a meeting. 17. Ambient light is too low (too high). So, increase (decrease) the brightness level.

From the results of the interviews and the questionnaire responses, we found that the choices made by physicians follow differing patterns. We can clearly see that young physicians at the beginning of their career have, like the medical residents, had more exposure to the smartphone platform and are more familiar with its apps and their use, and the use of a smartphone in general. They also show little or no resistance to being introduced to new apps, especially if they will assist them with their daily work in the healthcare environment. These individuals easily fall into a single stereotype category. Intermediate physicians are very similar to junior physicians, in terms of their exposure to smartphones overall, but have much more domain expertise and stricter time constraints. They do not spend much time exploring and playing with technical gadgets like the young physicians, but they also don't show resistance to using a new app that could assist them in their work. We can readily place these physicians in a second stereotype category. Experienced physicians are likely to have had less exposure to the smartphone platform and show more resistance to using a new app that could support them in their daily work. We have no hesitation in placing them in a third stereotype category.

6.6 Conclusion

A universal definition of context would not be viable. Consequently, every software engineer needs to clearly understand and characterize the context in question, and the goal of its development. We have introduced a new characterization of context for MUIs in the healthcare domain, based on identification of the principles of concern that are relevant to MUI designers. The adaptations built into an MUI can enhance the acceptability of the technology in the healthcare domain, where diversity is inherent. We introduced the term *context descriptor* to contain extracted contextual information related to the various daily tasks that need to be taken into account in adapting the MUI.

In the next chapter, we propose a context-based and rule-based approach for designing adaptable MUIs in a healthcare app and use our quality-in-use model to evaluate the adaptable MUI design.

Chapter 7 Rule-based Adaptation of the MUI in Healthcare

7.1 Introduction

In mobile health (mHealth) applications, the adaptability of the MUI is considered one of the most important issues to address. According to the World Health Organization (WHO) MUI adaptability is a major problem in eHealth tools and services, including mHealth (Organization, 2005). The prevalence of smartphones, such as Apple's iPhone™, which functions as a computer, and the latest sensor technologies (e.g. GPS, proximity sensors, accelerometers, and ambient light sensors) offer more flexible HCI possibilities and enable information to be adapted more sensitively, responsively, and powerfully. Context-based UIs offer a new approach to automated UI adaptation in a mobile device, and can include many interesting features, such as input/output channels, innovative widgets, etc. Furthermore, every feature has variables, such as text, voice, color, etc. From the designer's perspective, feature adaptability is planned either at the time of design or during runtime. During runtime, the UI is generated by applying a set of rules to the set of context descriptors, which, in healthcare applications, matches the set of user stereotypes.

The MUI adaptation approach proposed in this thesis is based on the user stereotype models – see chapter 5, (Alnanih et al., 2013b), our context model – see chapter 6, (Alnanih et al., 2012), and the context descriptors. The benefit of using a context model is that it can provide physicians with UIs that are simpler and more efficient. Although the stereotype model tailors the UI to their basic needs, preferences, and priorities, the context descriptors, in fact, allow for even more adaptability and versatility, in the form of real-life logic and individual user information. In this chapter, we propose the use of decision tables to represent the adaptation rules for MUI, using the iPhone™ specifically. The proposed approach for MUI adaptation is validated on a case study, which we refer to as Controlled Experiment IV. In this experiment, our quality-in-use measurement model, developed specifically for MUIs – see chapter 3, (Alnanih et al., 2013c), is used to compare the quality-in-use of MUIs with and without adaptation for the same healthcare app to be used by the same users in the same context.

This chapter is organized as follows. Section 7.2 summarizes the literature on UI adaptation and explains how our approach is similar to the methods found there and how it differs from them. In section 7.3, our MUI adaptation approach is introduced, and a decision table based on our approach is defined. Section 7.4 describes Controlled Experiment IV, in which we use our MUI in simulated, but realistic, scenarios, to validate our approach. In section 7.5, we discuss our results, and, finally, in section 7.6, we conclude the chapter.

7.2 Motivation and Related Work

MUI adaptation, which is at the forefront of innovative research today, focuses on improving the user experience by adapting the UIs in mobile applications to different contexts of use. Most existing designs of medical apps do not consider the context of the user's task, and few of today's hospital apps are designed with empirical knowledge of physician activity (Battisto, Pak, Vander Wood, & Pilcher, 2009). In recent years, a number of adaptive MUIs have been developed for healthcare apps.

In Vogt and Merier (2010) (Vogt & Meier, 2010), the authors use a model-based approach and the user interface markup language (UIML) to create mobile and adaptive UIs for the healthcare apps. Their models are adapted during runtime, and context is taken into account to simplify input choices and limit the scope for error. However, they do not consider output choices. Their approach differs from ours in three ways: (1) our MUI adaptation is based on a context model, a stereotype model, and a set of context descriptors that captures the essence of the dynamic context; (2) our adaptation technique is based on decision-making rules; and (3) our adaptation model consists of all the parameters of the MUI that adapt either at the time of design or at runtime, and takes into account the features and sensors embedded in the smartphone to help the physician accurately input his needs to the app and obtain meaningful output by offering various choices.

Frohlich et al. (2009) (Fröhlich, Meier, Möller, Savini, & Vogt, 2009) presented their LoCa approach to context-aware monitoring apps in homes equipped for digital technologies. They focused on the users (patients) who are mobile. The LoCa project differs from this research in that Frohlich et al.'s adaptation is based on external sensors: the patient is given a *smart shirt* equipped with several external sensors that monitor physiological parameters, like ECG and blood glucose level. In addition, the patient receives a smartphone with a GPS sensor and camera. Our adaptation technique is based on internal sensors, and UI adaptation as a design issue.

The process of developing context-based UIs has been explored in a number of projects. Clerckx et al., (2005) (Clerckx, Winters, & Coninx, 2005), for example, discuss various tools to support the model-based approach. Model-based UI development (MBUID) is an important technique for designing interactive systems, and is aimed at reducing the complexity generated when developers have to design many models that will be interconnected. MBUID creates a UI from an abstract model automatically, verifies and validates the UI, and checks its consistency with other UIs. However, this technique is not being used, because of the lack of commercial tools available to support it.

Yuan and Herbert (2012) (Yuan & Herbert, 2012) presented a fuzzy logic-based context model and related context aware reasoning middleware which provides a personalized, flexible, and extensible reasoning framework for a context aware real time assistant (CARA). They then attempted to adapt the technology to elderly patients living at home. They performed simulations, but were unable to devise and carry out field experiments that yielded results with enough diversity to justify scientific application. In our research, we adapt the MUI to physicians in a hospital setting to assist them in their daily tasks. Unlike previous research, our work indexes all the information gathered by sensors from the basic context level as conditions. We then combine these conditions in a user stereotype model and collect all the possible adaptation rules using a decision table.

Many studies have been conducted on adaptation using a decision table. In (Shi-wei & Shou-Qian, 2009), an approach is proposed for modeling adaptive 3D navigation in a virtual environment. In order to adapt to different types of users, they designed a system of four templates corresponding to four types of users. Our work differs in that our adaptation technique is based on a context model which extracts values from sensors in smartphones and maps them to a user stereotype model, and then to the MUI using a decision table.

In the following section, we present our MUI adaptation approach.

7.3 Proposed MUI Adaptation Approach in Healthcare Applications

Our approach to adaptation is to change the MUI parameters based on user stereotype – chapter 5, (Alnanih et al., 2013b), and context model – chapter 6, (Alnanih et al., 2012). This allows users to easily handle information on the MUI, which shakes with the user's motion, and various environmental effects, such as different combinations of ambient conditions (i.e. light and noise levels). The UI features, which can be changed on a modern smartphone, are listed in Table 7-1.

Table 7-1 MUI Features

MUI Features	Value (conditions)
Font size*	Small, medium, large
Font color	An RGB color, black & white
Font format	Times New Roman, Tahoma, etc.
Background color	Auto adjust, change manually
Data entry*	Typing, tapping, voice
Display information*	Text, sound
Message delivery	Text, voice, alert, silent, pre answer
Brightness level*	Increase/decrease
Ring volume	Low, medium, high, alert, vibration
Sound level	Mute, regular, loud

*A feature implemented in the case study (see section 7.5).

Physicians work in a dynamic user environment, in which they must respond to rapidly changing contexts. Adapting MUIs to these contexts, such as location, time, and ambient conditions, will enhance the user's experience of a context-aware device (Alnanih et al., 2012).

Categorization of the contexts in which a physician works in a hospital setting depends on the location in which the medical tasks are performed; for example, primary care, the emergency room, an operating room, inpatient care, an inpatient ward, an outpatient clinic, the intensive care unit, etc. (Battisto et al., 2009).

The time limitations of this doctoral research made it impossible to consider the entire range of hospital services provided and their many context descriptors. To develop a prototype to validate the adaptable features of the MUI presented in Table 7-1, we focused on a particular area of the hospital, an inpatient ward, where physicians' services are provided. Our MUI is based on this area and on the junior healthcare providers required in this context, along with the context descriptors for these types of users. This step is performed from the MUI designer's perspective, and is a continuation of our previous work in this domain (Alnanih et al., 2013b; Alnanih et al., 2012). Transformation rules were created by designers (after receiving a great deal of knowledge from the healthcare provider in the analysis phase when collecting a set of context descriptors), who codified the best practice patterns for predefined situations, as well as fall back rules, as suggested by de Melo et al. (de Melo, Honold, Weber, Poguntke, & Berton, 2009). Rules at different stages can be developed independently of each other. At each stage, different contextual information is included – see the example below.

A 'Sub Domain' Scenario in a Hospital Ward

Consider a junior physician who wants to: 1) check the most recent CBC test result, and 2) order medication for a patient at night in the patient's room, which is dark and quiet. The MUI can: display the result by sound, rather than by text; accept the order given by the physician by voice, rather than typed on the keyboard of a smartphone; decrease the brightness level of the screen.

In this scenario, the MUI features adapted or modified (temporarily) are: change to speech output while the physician is in the room; change to speech input while the physician is in the room; and increase the level of brightness of the display.

Decision tables (DTs) provide a schematic view of the inference process in decision-making. The advantage of the DT is that it provides a more compact visual presentation, and so contributes to a better understanding of the selection problem. But probably the most important advantage of using a DT is that the completeness, correctness, and consistency of the information entered are easier to check.

All the condition attributes presented in our work are founded on our basic context, which consists of the location (i.e. the inpatient ward), the time (morning and evening shifts), the ambient conditions (light level: bright or dark), and the noise level (low or high) (see Table 7-2) (Alnanih et al., 2012). For the domain-based context, which consists of the user and the tasks, we defined three user stereotype models for healthcare apps based on experimental work: junior, intermediate, and expert (chapter 5, (Alnanih et al., 2013b). Also, context descriptors for each type were collected. The tasks that physicians most frequently perform on a daily basis on the inpatient ward can be summarized in two categories, input tasks and output tasks, as follows:

- Output tasks: review laboratory, radiology, and pharmacy results.
- Input tasks: submit new pharmacy, laboratory, and radiology orders.

In the following subsections, all the possible rules for adapting MUI actions to the physician's tasks that belong to different user stereotypes are presented in (Tables 7-2, 7-3, and 7-4). In each table, the set of actions A1 to A9 belongs to the group of input and output tasks, A10 and A11 belong to the group of output tasks, and A12 and A13 belong to the group of input tasks. For example, for the junior stereotype, when conditions C2 and C3 hold (see "Y" in Rule 1, Conditions, Table 7-2) for output tasks, then the corresponding set of actions adapting the MUI will be: A1, A3, A5, A7, A10, and A12 (see "X" in Rule 1, Actions, Table 7-2). These rules can be specified by design experts and a healthcare provider. For practical reasons, in terms of our experimental work and the time available for conducting the research, the pilot test focused on the *junior* stereotype initially, since this type performs most of the work on the inpatient ward, and is the most familiar with using a mobile device.

Table 7-2 Decision table for output and input tasks (Junior Stereotype)

	Conditions	**Rules**			
		1	2	3	4
C1	The physician is in the patient's room during the morning shift.	-	-	-	-
C2	The level of light in the room is bright.	Y	Y	N	N
C3	The level of noise in the room is low.	Y	N	Y	N
	Actions				
A1	Adjust the font size for displaying information to "user default".	X	X		
A2	Adjust the font size for displaying information to "large".			X	X
A3	Adjust the display brightness to "user default".	X	X	X	
A4	Adjust the display brightness to "high".				X
A5	Receive text message alerts by a ring tone.	X	X		
A6	Receive text message alerts by "vibration".			X	X
A7	Adjust the ring tone volume to "user default".	X			
A8	Adjust the ring tone volume to "high".		X		X
A9	Adjust the ring tone volume to "silent".			X	
A10	Receive information via a headset.	X		X	
A11	Receive information via a text.		X		X
A12	Input the data using a microphone.	X		X	
A13	Input the data by typing on a keyboard.		X		X

Table 7-3 Decision table for output and input tasks (Intermediate Stereotype)

	Conditions	Rules			
		1	2	3	4
C1	The physician is in the patient's room during the morning shift.	-	-	-	-
C2	The level of light in the room is bright.	Y	Y	N	N
C3	The level of noise in the room is low.	Y	N	Y	N
	Actions				
A1	Adjust the font size for displaying information to "user default".	X	X		
A2	Adjust the font size for displaying information to "large".			X	X
A3	Adjust the display brightness to "user default".	X			
A4	Adjust the display brightness to "high".		X	X	X
A5	Receive text message alerts by a ring tone.	-	-	-	-
A6	Receive text message alerts by "vibration".	X	X	X	X
A7	Adjust the ring tone volume to "user default".	X			
A8	Adjust the ring tone volume to "high".		X		
A9	Adjust the ring tone volume to "vibrate".			X	X
A10	Receive information via a headset.			X	
A11	Receive information via a text.	X	X		X
A12	Input the data using a microphone.			X	
A13	Input the data by typing on a keyboard.	X	X		X

Table 7-4 Decision table for output and input tasks (Experienced Stereotype)

	Conditions	**Rules**			
		1	2	3	4
C1	The physician is in the patient's room during the morning shift.	-	-	-	-
C2	The level of light in the room is bright.	Y	Y	N	N
C3	The level of noise in the room is low.	Y	N	Y	N
	Actions				
A1	Adjust the font size for displaying information to "user default".	X			
A2	Adjust the font size for the display information to "large".		X	X	X
A3	Adjust the brightness to "user default".	-	-	-	-
A4	Adjust the brightness to "high".	X	X	X	X
A5	Receive text message alerts by a ring tone.	X	X	X	X
A6	Receive text message alerts by "vibration".	-	-	-	-
A7	Adjust the ring tone volume to "user default".	X			
A8	Adjust the ring tone volume to "high".		X		X
A9	Adjust the ring tone volume to "vibrate".			X	
A10	Receive information via a headset.			X	X
A11	Receive information via a text.	X	X		
A12	Input the data using a microphone.	X		X	X
A13	Input the data by typing on a keyboard.		X		

To verify the completeness of the rules, we simply calculate how many are represented in each column and compare the result with 2^m, where m is the number of conditions. Rule analysis also indicates that the proposed rules are consistent.

In the next section, we empirically validate our MUI adaptation approach.

7.4 Controlled Experiment IV: Validating the High-Fidelity Prototype

In a high-fidelity prototype, the focus is on the interaction between the PHIS2-MA app and the physician in relation to the adaptable features that have been implemented (see Table 7-1) and the principles of HCI. In addition, the recommendations and changes recorded on the paper fidelity test are added to the prototype before the high-fidelity test is carried out.

The proposed context-based and rule-based approach for MUI feature adaptability was implemented on an iPhone 4.1™ and subsequently updated to an iPhone 5™ loaded with the PHIS2-M app. The result was a new version of PHIS2-M, which is PHIS2-MA.

The high-fidelity prototype of PHIS2-MA was designed and implemented using Xcode version 4 and 5.1. It also uses the special framework for speech recognition and text to speech, iSpeech iOS SDK.

All the components used for GUI implementation are standard components provided by default in iOS SDK:

Tab Bar Item, Table View, Navigation Item, View, Data Grid View, Label, TextBox, Round Style Text Field, Image View, Bar Segmented Control, Switch, Horizontal Slider, Top Layout Guide, Bottom Layout Guide, UIDataPicker, Groups, and Panel.

The background of the Login view and the graphics used for the main buttons presented on the Tabs (Laboratory, Radiology, and Pharmacy) were designed in Adobe PhotoShop 10.0. The Database tables were designed using the SQLite Manager application.

Figures 7-1 to 7-5 describe the design solutions applied to achieve the rule-based adaptation of the features to the context and the user stereotype.

The names of the navigation controllers are listed below:

1. Login View – PHIS
2. Sign Up View Controller
3. User Info View Controller (for Administrator access to collect the user profile for each physician).

4. PHIS Master View Controller (to consult the patient list retrieved from the FinalPhis.Sqlite database).
5. Patient Menu Controller
6. Result Tab Controller (View Results)
    1. Lab Results View Controller
        6.1.1. Detail Results View Controller
            6.1.1.1. Results Comparison View
    2. Radiology Results View
        6.2.1. Detail Results View Controller
    3. Pharmacy Results View Controller
7. New Order Tab Controller (New Order)
    1. Lab New Order View Controller
        7.1.1. New Order Form Controller
    2. Radiology New Order View
        7.2.1. New Order Form Controller
    3. Pharmacy New Order View
        7.3.1. Second Order Form Controller
8. Summary View Controller
9. Profile View Controller
10. Option View Controller (for Settings)

Figure 7-1 shows all 21 of the implemented views to prove our concept related to the physician tasks.

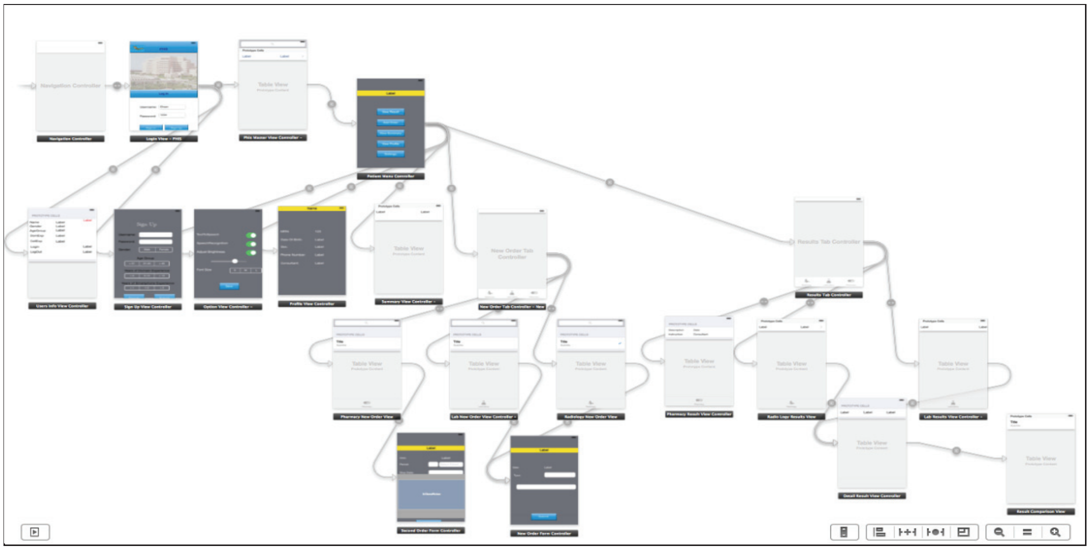


Figure 7-1 PHIS2-MA: all views.

Figure 7-2 shows the navigation for the Login View, Sign Up View Controller, User Info View Controller, PHIS Master View Controller and Patient Menu Controller.

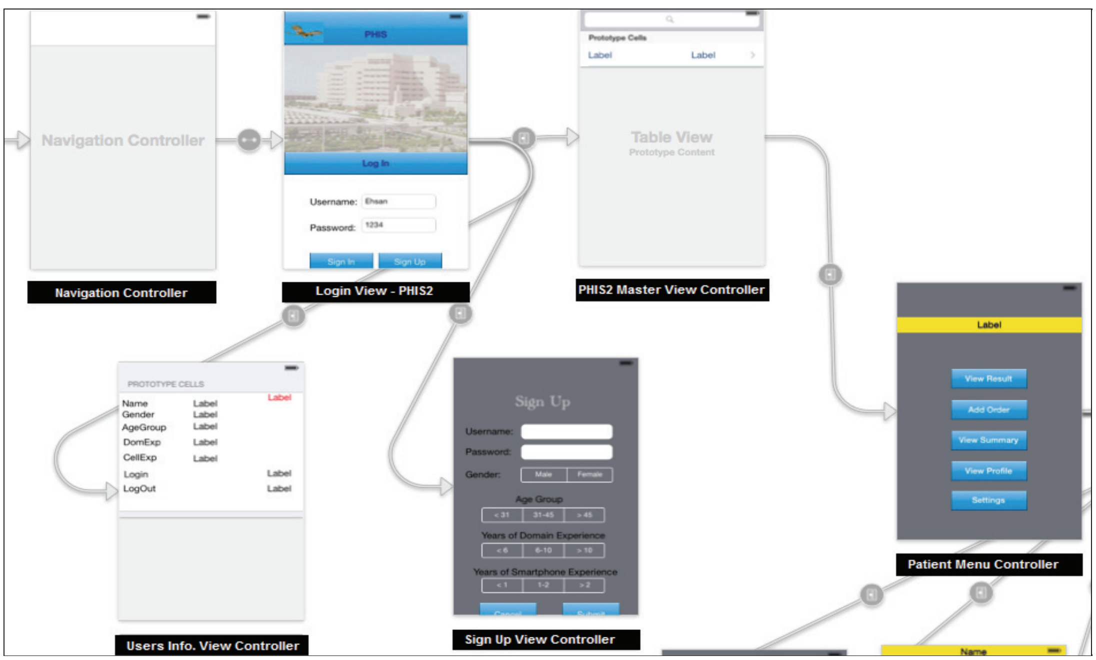


Figure 7-2 Login view.

Figure 7-3 shows the navigation for the Patient Menu Controller and the sub choices including: Results, Order, Summary, Profile, and Option View Controller.

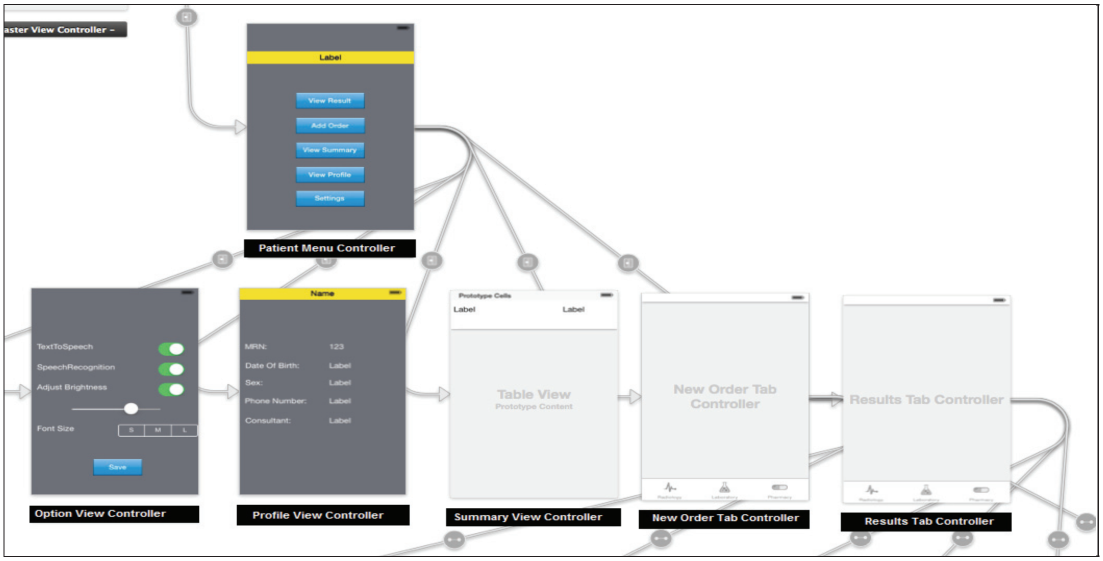


Figure 7-3 Main menu.

Figure 7-4 shows the navigation for the New Order Tab Controller and the sub choices including: Pharmacy, Radiology, Lab New Order View Controller, the New and Second Order Form Controller.

Figure 7-4 New order type view.

Figure 7-5 shows the navigation for the Results Tab Controller and sub choices including: Pharmacy, Radiology, Lab Results View Controller, Detail Results and Results Comparison View Controller.

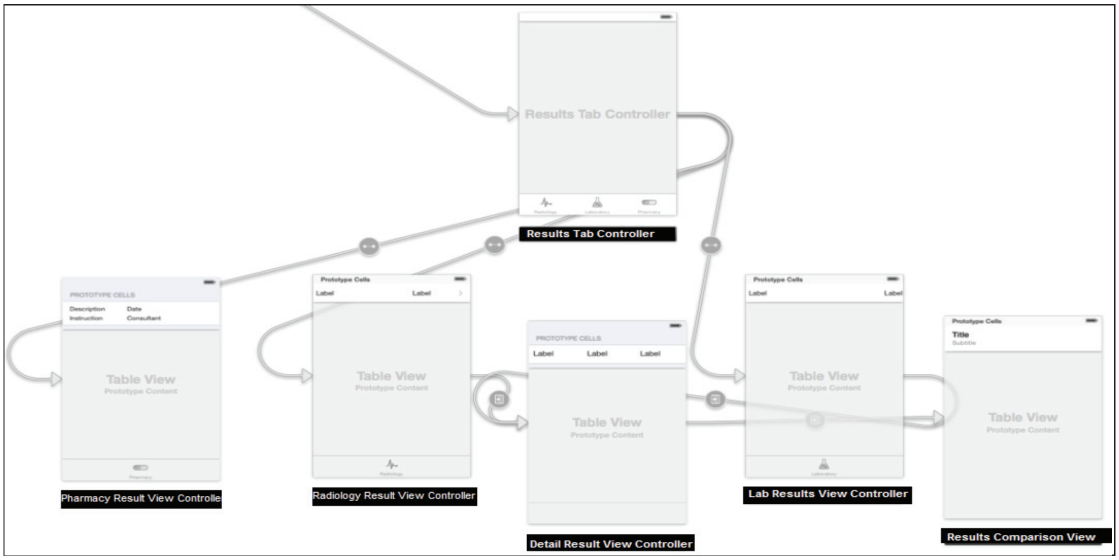


Figure 7-5 Results type view.

We used the iSpeech framework, which includes the text-to-speech (TTS) feature, and the automated speech recognition (ASR) feature. iSpeech offers mobile developers a way to integrate speech into their apps free of charge. The iSpeech cloud combines powerful (TTS) and (ASR) technologies with easy-to-use APIs and SDKs. Human quality TTS is available for iPhone™ and BlackBerry™ with the Android, Windows Phone 7, and other platforms. iSpeech SDKs are simple to use, fast, and reliable. In 2010 alone, the iSpeech cloud powered over half a billion text-to-speech conversations and speech recognition utterances for our developer partners. iSpeech recognition is able to recognize free-form dictation and defined grammar in most major languages, as well as enabling voice search, a voice user interface (VUI), voice control, voice commands, voice note dictation, messaging, SMS, emailing, and more. Moreover, it works online with a connection to the internet and it is provided to us free.

We chose the iSpeech framework after testing other frameworks. We compared the results of three available SDKs, and iSpeech seemed to us a better choice overall.

We started with the OpenEars framework, which is a shared-source iOS offline framework for iPhone™ voice recognition and speech synthesis (TTS). It lets the user easily implement round-trip English language speech recognition and TTS on the iPhone™, it uses the open source CMU Pocketsphinx, CMU Flite, and CMUCLMTK libraries, and it is free to use on an iPhone™. However, after integrating it into PHIS-MA, we found the framework to be slow, in terms of recognizing commands compared with iSpeech.
The second framework we tested was the CeedVocal, which is a multi-locutor, isolated word and keyword spotting speech recognition SDK for iOS. It operates locally on the device (no network connection required), and supports six languages. However, when we integrated CeedVocal into PHIS-MA, we found that it had the following limitations: its fault tolerance was not acceptable, and it crashed easily. Its speech recognition capability was limited to words and numbers which were not defined by default. Finally, TTS was not supported by the CeedVocal framework.

The features in Table 7-1 that are not implemented in our app, such as message delivery, ring volume, and sound level were incompatible with iOS. The Apple iOS SDK restricts apps to the ability to change device settings only within the application environment. This means that the sound level of an app can be changed, but as soon as the app quits or goes into background mode (when a message or phone call is received), the default device settings apply and the phone rings with the sound level set for the device, and not the level set by the user in the app. However, a work-around exists for this, which is to use a private API (an application programming interface intended for use by Apple developers only). The behavior of a private API is not guaranteed, however, as it is not recommended for use by Apple, nor is it documented by Apple, and it is very likely to change with each version update.

Conducting a pilot test is important for debugging the equipment, software, materials, and procedures that will be used during the formal test. Also, once PHIS2-MA was ready for formal evaluation in the hospital with the various user stereotypes, a pilot test was needed to check whether or not any final changes to PHIS2-MA or the testing process were required.

7.4.1 Pilot Test: Evaluating PHIS2-MA

At this stage, a pilot test was conducted to ensure that all the necessary changes were made and that we were as prepared as possible for testing in a real environment. In this test, the adaptable features of the MUI were evaluated on PHIS2-MA. Twelve junior physicians at the KAUH aged 25 to 35 who were experienced users of PHIS, 4 males and 8 females, were selected randomly, and then interviewed and observed. By "experienced users", we mean that they have been working at the KAUH and using PHIS for more than six months. The participants were observed one at a time in a simulated environment, in which prerecorded background noise from the real hospital setting was played to them. In order to determine the effect of using the adaptable MUI features, each participant was asked to perform the same set of tasks, once using PHIS2-M (without adaptable features) and once using PHIS2-MA (with adaptable features). The test environment included the participant, the iPhone™ device loaded with the two apps, a timer, and a tester (Figure 7-6).

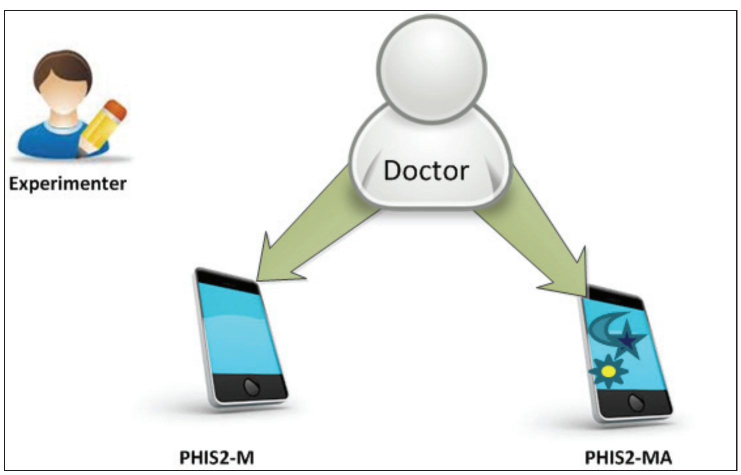


Figure 7-6 Test environment.

Based on the context model, the device recognizes the context at runtime by extracting values from sensors installed in the smartphone (time, levels of brightness and noise), mapping them to the user stereotype model and then to the MUI widgets using a decision table. For example, based on the user stereotype, these values will increase or decrease and the font size will be adapted accordingly.

7.4.2 Validation Method

In order to compare the results of using the adaptable application with the non-adaptable one for the same participant, the quality-in-use measurement objective and subjective characteristics of the UI were evaluated by those twelve physicians (see chapter 5). The objective quality-in-use characteristics quantify effectiveness, productivity, efficiency, error safety, and cognitive load, and the subjective characteristics describe the level of satisfaction of the twelve physicians. Four attributes were collected during the tests: the number of correct actions and the number of incorrect actions performed for each task, the number of views, and the time required to complete each task. Note that the total number of actions performed equals the sum of the correct and incorrect actions for each task. The measurement formulas and the interpretations of the measurement data for the objective characteristics are listed in chapter 5.

7.4.3 Task List

A list of tasks (see Table 7-5) was prepared for the twelve participants in the test. Five tasks correspond to the ways in which the physicians interact with the application: search, choose, select, read, and write, because these are the functionalities most frequently used by the physicians on a daily basis.

7.4.4 Objective Measures (Paper Login Form)

A paper login form was used to collect the attributes for each task, and a satisfaction questionnaire was handed out at the end of the test to the participants to gather their feedback about using the health information system application on a mobile with adaptable features and on one without those features (Appendix D).

Table 7-5 List of tasks

Tasks	MUI Feature	Type
1. Search for the patient from the patient list	Text or speak	Input
2. Choose from the main menu: Laboratory result	Tap or speak	Input
3. Select the most recent Complete Blood Count (CBC)	Tap or sound	Output
4. Choose from the Order Radiology main menu (chest X-ray)	Tap or speak	Input
5. Review the patient summary	Increase or decrease brightness	Output

7.4.5 Hypotheses

In order to empirically investigate the effect of MUI with and without adaptable features for the same healthcare application used by the same user in the same context, hypotheses related to PHIS2-M and PHIS2-MA were formulated as follows:

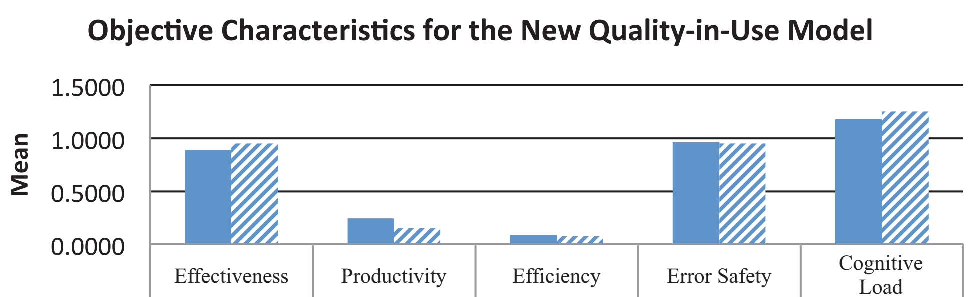
HYP^0: there is no significant difference (according to the QiU-4-MUI objective characteristics) between using the PHIS2-M and the PHIS2-MA.

HYP^a: there is significant difference (according to the QiU-4-MUI objective characteristics) between using the PHIS2-M and the PHIS2-MA.

The following materials were prepared before conducting the formal experiment test.

7.4.6 Analysis of Objective Characteristics

For the objective analysis, the measurement data required to evaluate the objective characteristics: effectiveness, productivity, efficiency, error safety, and cognitive load, were collected during usability testing performed with 12 physicians performing 5 tasks on PHIS2-M and PHIS2-MA. The raw data for the empirical study were tabulated in MS Excel for each of the participants. We compared the means of all the quality-in-use measurement data collected for PHIS2-M and PHIS2-MA for all tasks and all participants in the test case. Figure 7-7 shows clear improvement for PHIS2-MA compared to PHIS2-M for the effectiveness and cognitive load characteristics (see the interpretation in chapter 3). The fact that the MUI of PHIS2-MA is more effective than that of PHIS2-M indicates that the MUI is better in this respect for mobile adaptation. Level of cognitive load for the MUI of PHIS2-MA is also better than for the MUI of PHIS2-M, which indicates that this MUI is better for mobile adaptation in this respect as well. The productivity and efficiency features of the MUI in PHIS2-M score higher than they do in PHIS2-MA, indicating that the MUI without adaptability is better in these respects for the mobile app. The mean for safety in PHIS2-MA is virtually the same as that in PHIS2-M.



	Effectiveness	Productivity	Efficiency	Error Safety	Cognitive Load
PHIS2-M	0.8924	0.2440	0.0897	0.9609	1.1771
PHIS2-MA	0.9506	0.1511	0.0773	0.9506	1.2550

Figure 7-7 Objective characteristics for all PHIS2-M and PHIS3-MA tasks

The hypotheses are verified for each QiU-4-MUI quality characteristic, in order to investigate the statistical significance of the observed differences in the objective quality characteristics: effectiveness, productivity, efficiency, error safety, and cognitive load, and, since we have two conditions (using PHIS2-M, and using PHIS2-MA) for the same participants, we can pair the data. Consequently, we used the paired t-Test for the data analysis. Based on the critical value approach of the t-Test at 11 degrees of freedom, $\alpha = 0.025$ for the two-tailed test, the critical level of t being ± 1.79. Our decision rule is to reject the null hypothesis if the computed test statistic is less than -1.79 or more than 1.79. Table 7-6 shows the t-Test value and the P-value for all the characteristics for statistical validation.

Table 7-6 Paired t-Test and P-values for all the objective characteristics

	Effectiveness	Productivity	Efficiency	Error Safety	Cognitive Load
t-value	2.26	-4.86	-1.79	5.44	-1.48
P-value	0.022	0.0002	0.05	0.0001	0.083

Since the t-Test values obtained for effectiveness and error safety fall into the positive critical region, we reject the null hypothesis for these characteristics. From the P-value approach of the t-Test, the P-values for effectiveness and error safety are less than α. Therefore, we reject the null hypothesis HYP^0 and accept HYP^a for these characteristics. Our conclusion is that there is a significant difference between PHIS2-M and PHIS2-MA for the effectiveness and error safety characteristics. Since the t-Test values for effectiveness fall into the positive region, we can conclude that PHIS2-MA is better than PHIS2-M based on this characteristic.

Our t-Test values for productivity and efficiency fall into the negative critical region, and so the null hypothesis is rejected. The P-values of the t-Test for productivity and efficiency are less than α, and so the null hypothesis HYP^0 is rejected. Since the t-Test values for productivity and efficiency fall into the negative region, we can conclude that PHIS2-M is better than PHIS2-MA based on these characteristics.

Since the t-Test values obtained for cognitive load do not fall into the critical region, we failed to reject the null hypothesis for this characteristic. From the P-value approach of the t-Test, the P-values for cognitive load are greater than α. Therefore, we failed to reject the null hypothesis HYP^0. Our conclusion is that there is no significant difference between PHIS2-M and PHIS2-MA for the cognitive load characteristics.

7.4.7 Analysis of Subjective Characteristics

For the subjective characteristics, the users were asked to complete our User Satisfaction Questionnaire to give us their impressions and opinions on the apps, and to enable us to evaluate their interaction with PHIS2-M and PHIS2-MA (Appendix D). The subjective analysis was measured on a Likert scale. The results of the questionnaire reveal that all the users gave both PHIS2-M and PHIS2-MA ratings of ‘easy’ and ‘very easy’ to the features related to:

F1: Easy to learn
F2: Easy to use
F3: Performing a particular task
F4: Finding the features
F5: Navigation
F6: Recovery from error
F7: Performing a task anywhere and at any time

For the brightness features, more than sixty percent of the participants preferred to increase the brightness level at night. This indicates that PHIS2-MA is best suited to their needs. Thirty percent preferred to decrease the brightness level at night. This indicates that PHIS2-M is best suited to their needs. Ten percent preferred the default user preference level.

These results confirm HYP^0, which means that the level of satisfaction experienced by the participants when using the PHIS2-M MUI is no different from that using the PHIS2-MA MUI.

In order to investigate the differences observed in the subjective characteristics, and since the measurement data are arranged on an ordinal scale and the characteristics are paired, we used the Wilcoxon Rank Sum Test, which looks at the size of the differences in rank orderings. The Wilcoxon Signed-Ranks Test is similar to the Rank-Sum Test, except that it takes into account the magnitude as well as the direction of the differences. It gives more weight to a pair that shows a large difference than to a pair that shows a small difference. Sample hypotheses for the above subjective characteristics features are listed below:

HYP^0: There is no difference between PHIS2-MA and PHIS2-M on the subjective feature ‘Easy to learn’.

HYP^a: There is a difference between the PHIS2-MA and PHIS2-M on the subjective feature ‘Easy to learn’.

The hypotheses for each subjective feature listed are verified based on the Wilcoxon Test for the z-statistic and the P-value (see Table 7-7) that shows the positive and negative ranks of each feature, F1...F7.

Table 7-7 Ranks

		N	Mean Rank	Sum of Ranks
F1_MA - F1_M	Negative Ranks	4a	2.50	10.00
	Positive Ranks	0b	.00	.00
	Ties	8c		
	Total	12		
F2_MA - F2_M	Negative Ranks	5d	4.00	20.00
	Positive Ranks	2e	4.00	8.00
	Ties	5f		
	Total	12		
F3_MA - F3_M	Negative Ranks	3g	2.00	6.00
	Positive Ranks	0h	.00	.00
	Ties	9i		
	Total	12		
F4_MA - F4_M	Negative Ranks	3j	2.67	8.00
	Positive Ranks	1k	2.00	2.00
	Ties	8l		
	Total	12		
F5_MA - F5_M	Negative Ranks	2m	3.00	6.00
	Positive Ranks	2n	2.00	4.00
	Ties	8o		
	Total	12		
F6_MA - F6_M	Negative Ranks	4p	3.00	12.00
	Positive Ranks	1q	3.00	3.00
	Ties	7r		
	Total	12		
F7_MA - F7_M	Negative Ranks	0s	.00	.00
	Positive Ranks	2t	1.50	3.00
	Ties	10u		
	Total	12		

The data obtained were used to analyze the effect of the new mobile application in terms of each of the 7 features. The statistical significance was set at $P<0.05$. All the statistical analyses were performed with SPSS 18 software (IBM, New York, USA).

The Wilcoxon Signed-Ranks Test verifies that two related medians are equal. It is based on the absolute value of the difference between two test variables.

For the first feature, 'learning the application', our null hypothesis H^0 states that the median of feature 1 for mobile users is equal to a median of feature 1 for mobile adaptation users. The Wilcoxon Test gives no evidence against the null hypothesis of the difference between medians for feature 1 for mobile users and mobile adaptation users (we note that the P-value=0.059, which is larger than 0.05) (see Table 7-8). Consequently, we conclude that the new application does not change the level of user satisfaction. Moreover, the satisfaction of the user with both PHIS-MA and PHIS-M is the same.

Table 7-8 Wilcoxon signed-ranks test

	F1_MA - F1_M	F2_MA - F2_M	F3_MA - F3_M	F4_MA - F4_M	F5_MA - F5_M	F6_MA - F6_M	F7_MA - F7_M
Z	-1.890[a]	-1.134[a]	-1.633[a]	-1.134[a]	-.378[a]	-1.342[a]	-1.414[b]
Asymp. Sig. (2-tailed)	0.059	0.257	0.102	0.257	0.705	0.180	0.157

a. Based on positive ranks
b. Based on negative ranks
c. Wilcoxon Signed-Ranks Test

We conclude from the above table that we cannot reject the null hypothesis, HYP^0, for all the features based on the z-statistic value and the P-value. We can, however, conclude that the subjective quality-in-use of PHIS-MA is the same as that of PHIS-M.

7.5 Discussion

Controlled Experiment IV, conducted with 12 physicians, showed that productivity and efficiency favor PHIS2-M over PHIS2-MA. Since productivity and efficiency depend on time, while the other characteristics do not, entering input verbally using the voice recognition functionality on a mobile platform is slower than typing or tapping, and most of the tasks evaluated in the experiment for the adaptable version depend on the vocal features.

In terms of effectiveness and error safety, the results show a trend toward the superiority of the adaptable version based on the t-Test for these characteristics. This indicates fewer actions in each view in a specified context of use, and the incorrect actions are all in the best range. For cognitive load, our results show that there is no significant difference between PHIS2-M and PHIS2-MA. It may be that the less sophisticated design is what is needed in the medical field. However, it is very early to come to this conclusion.

A possible explanation for this negative observation is the unfamiliarity of the physicians with the new adaptation tools, e.g. perhaps most of the physicians forgot to press the microphone button to activate voice recognition in task 2. (This was considered a design issue and has now been fixed, so that for the formal test of PHIS2-MA – see chapter 8 –the voice button will activate automatically.)

Generally, mobile applications in healthcare in a hospital context are the most useful when physicians are moving from one location to another. From the interviews conducted with the physicians, some inherent drawbacks were identified. One of them is the lack of privacy when using the voice features. These features are the most useful when privacy is not an issue, or when the physician is alone, his hands are not free to tap or type, and his eyes are busy looking elsewhere. So, the ability to talk to the device when the other faculties are occupied is a useful option.

7.6 Conclusion

The term *mobile health* (*mHealth*) is becoming more and more widely used in the research community and among those interested in providing and receiving healthcare services. Adaptation of MUI features based on a decision table can introduce systematization in the design of adaptive systems and enhance the acceptability of the technology in the healthcare domain, where diversity is an inherent characteristic.

Our purpose in this chapter was to achieve one of our thesis objectives and to demonstrate the use of the decision table technique as a basis for designing complete and consistent MUI context adaptation rules for mobile application with one category of users. MUI adaptation is tested by means of both objective and subjective quality-in-use characteristics.

In the next chapter, we present the empirical evaluation of the final adaptable MUI in a real environment and with all the user stereotypes, including junior, intermediate, and experienced physicians.

Chapter 8 Empirical Evaluation of MUIs in Healthcare

In this chapter, our MUI adaptation approach is validated empirically through a carefully controlled experiment in a real environment from the point of view of the three user stereotypes.

The rest of the chapter is organized as follows: Section 8.1 describes the PHIS2-MA, including the PHIS users, physicians' tasks, and the final PHIS2-MA design before the experimental test is conducted. Section 8.2 presents the experimental evaluation of the MUI adaptation for our healthcare app, Controlled Experiment V. The results of the experiment are provided in section 8.3. We present our summary in section 8.4.

8.1 Description of PHIS2-MA

Our goal is to enhance communication between the user and the MUI features adapted to their needs. This app was developed to help physicians interact with patient information at any time during the performance of their daily tasks, such as reviewing patient results, ordering tests, and prescribing medications.

8.1.1 Identifying PHIS Users

PHIS is divided into three parts at KAUH: administrators, nurses, and physicians. Physicians obviously play a critical role in the health of the patient. The actions they take depend on the patient's test results displayed via the UI and on the expertise of the physician who interprets those results. It is vital, therefore, that patient information be displayed clearly and accurately. A cluttered view can cause confusion and frustration among users, which can affect their performance and even endanger a patient's life. Consequently, we have chosen to focus on the PHIS subsystem used by KAUH physicians and the tasks they perform on a daily basis. There are three categories of physician at the KAUH: interns, residents, and consultants. These categories fit the user stereotype model for healthcare apps (see chapter 5) (Alnanih et al., 2013b).

8.1.2 Identifying Physicians' Tasks

The daily tasks performed by the physicians registered in the PHIS system are the following:

1. Check the patient profile;
2. Check patient laboratory, radiology, and pharmacology results;
3. Order tests, e.g. a lab test or an X-ray, and prescribe medications;
4. View the Patient Summary.

8.1.3 Final Design of PHIS2-MA

The final design of the adaptable mobile app that is ready for formal testing is PHIS2-MA. This is the app on which the adaptable features of the MUI that match the various user stereotypes were evaluated. Below is a summary of the functionalities of the MUI features implemented on PHIS2-MA for different user types:

- Patient selection by voice;
- Main menu option selection, such as View Result, Add Order, View Profile or Summary, by voice or tapping;
- Output display (e.g. patient results) by audio or text;
- Font size adaptation of (small, medium, or large);
- Brightness level adaptation (increase/decrease) based on user type and time of day (morning shift, evening shift, night shift);
- Audio feature adaptation (on/off) based on user type;
- Settings (newly added functionality), to allow the user to change preferences;
- MUI setting adaptation at user sign-in;
- User preferences saved (only accessible by the administrator).

8.2 Controlled Experiment V: Usability Testing of the Adapted MUI

The experimental evaluation presented in this section involves the usability testing of the PHIS2-M and PHIS2-MA apps on a mobile device used by a single participant in one environment.

8.2.1 Participants

Our participant samples were made up of physicians from different specialties in the Pediatrics department during the fall of 2013 at KAUH. Forty-five physicians were selected at random for the study. The test was conducted one participant at a time during the physicians' working hours. The duration of the experiment was over 50 hours for all the participants. They were interviewed and observed in a real environment in different contexts at the hospital. In order to determine the effectiveness of the adaptable MUI features, each participant was asked to perform the same set of tasks, once using PHIS2-M (with no adaptable features) and once using PHIS2-MA (with adaptable features). The two versions of the PHIS-based mobile app were loaded onto the same iPhone™ device. To ensure the accuracy of the experiment, half the participants started with PHIS2-M and the other half started with PHIS2-MA. The characteristics of the participants are presented in Table 8-1.

The test environment included the participants, the iPhone™ device, a timer installed in the app, and a tester. Prior to conducting the formal experimental evaluation, a list of materials to be used during testing was prepared, as suggested by Dumas and Redish (Dumas & Redish, 1999) and explained below:

Table 8-1 Characteristics of participants

Physicians	Junior	Intermediate	Experienced
Experience	Intern, 1st and 2nd year of residency	More than 2 years of residency and a fellow	Consultant
Total number of participants	20	15	12
Female: male ratio	13:7	7:8	4:8
Timeframe	8:00 am to 6:00 pm		
Test environment	Pediatric ward, ER, NICU, PICU, physician's office, meeting room on the Pediatric floor, conference room in the Pediatric building		
Age group	<31	31- 45	>45
Domain experience	<6 years	6-10 years	>10 years
Smartphone experience	>3 year	1-3 years	<1 years

8.2.2 Task List

Three tasks corresponding to the functionalities most frequently used by the physicians on a daily basis, and to the ways in which the physicians interact with the app are: Select, Order, and Review (Table 8-2). These tasks include various adapted MUI features that help researchers extract the effectiveness of MUI adaptation for different types of user. For example, the UI features that are adapted or modified temporarily are as follows: in Task 1, change to voice input while the physician searches for the patient Medical Record Number (MRN); in Task 2, change to voice input while the physician adds a new order; and in Task 3, change to audio output while the physician reviews patient results.

Table 8-2 List of tasks

Tasks	MUI Features	Task Functionality	Task Type
1. Select the patient with the MRN (100)	Tap, text, or speak	Select, search, or speak	Input-write
2. Order a chest X-ray	Tap or speak	Choose, select, or speak	Input-write
3. Review the most recent complete blood count (CBC) results	Tap or sound	Choose, select, or speak and listen	Output-read

8.2.3 Objective Measures (Paper Login Form)

Since the aim of testing the adapted MUI features was to make sure that they are easy to use and easy to learn for all the user stereotypes, compared with using the non-adapted MUI features, we used the quality-in-use model for MUIs (see chapter 3) (Alnanih et al., 2013c) as the testing method. The objective characteristics of the quality model are defined in terms of effectiveness, productivity, efficiency, error safety, and cognitive load. To collect these measures a paper login form was provided to each participant which gives criteria for the independent variables, such as the time taken to complete each task, along with the number of correct actions, the minimum number of correct actions, the number of incorrect actions, and the number of views (Appendix E).

8.2.4 Subjective Measures (User Satisfaction Questionnaire)

A user satisfaction questionnaire was handed out at the end of the test to gather the impressions of the participants of the two applications, for subsequent evaluation. The questionnaire contained two types of questions: general questions, which focused on personal information and were applicable to any app; and specific questions, which only applied to the apps tested. We have to point out here that the responses to the questions relating to personal information, such as age group, gender, domain experience, and smartphone experience, were collected during the formal testing of PHIS2-MA, when the participant signed in to create a user name before starting the test (Appendix E).

8.3 Discussion of the Results

In order to empirically investigate the effectiveness of the MUI with and without adaptable features for the same user belonging to a specific user stereotype in a real-world environment, two sets of hypotheses were formulated, one relating to the objective characteristics of the new quality-in-use model and the other to its subjective characteristics (see chapter 3). Sample hypotheses for both the subjective and objective characteristics are listed below:

HYP^0**:** There is no significant difference between the effectiveness of PHIS2-M and the effectiveness of PHIS2-MA.

HYP^a: There is a significant difference between the effectiveness of PHIS2-M and the effectiveness of PHIS2-MA.

HYP^0: There is no significant difference between the user satisfaction of the participants using PHIS2-M and using PHIS2-MA, in terms of the median of each feature for mobile app use.

HYP^a: There is a significant difference between the user satisfaction of the participants using PHIS2-M and using PHIS2-MA, in terms of the median of each feature for mobile app use.

8.3.1 Objective Characteristic Results for All User Stereotypes

The objective characteristics of the new quality-in-use model provide measures of the effectiveness, productivity, efficiency, error safety, and cognitive load relating to 45 physicians belonging to three user stereotypes performing three tasks on PHIS2-M and PHIS2-MA (Alnanih, Ormandjieva, & Radhakrishnan, 2014a).

Since our data are normal, and we have two conditions (using PHIS2-M and using PHIS2-MA) for the same participants, we have paired them. Consequently, for statistical validation, the hypotheses are verified for each characteristic, based on the t-Test value and the P-values for the data analysis for all the participants in each stereotype (Table 8-3).

i) **Junior stereotype:** with the critical value approach of the t-Test at 19 degrees of freedom, α =0.025 for the two-tailed test and the critical level of t = ± 2.09. Our decision rule is to reject the null hypothesis if the computed t statistic is less than -2.09 or more than 2.09.

ii) **Intermediate stereotype:** with the critical value approach of the t-Test at 12 degrees of freedom, α =0.025 for the two-tailed test and the critical level of t = ± 2.17. Our decision rule is to reject the null hypothesis if the computed t statistic is less than -2.17 or more than 2.17.

iii) **Experienced stereotype:** with the critical value approach of the t-Test at 11 degrees of freedom, α =0.025 for the two-tailed test and the critical level of t = ± 2.20. Our decision rule is to reject the null hypothesis if the computed t statistic is less than -2.20 or more than 2.20.

The results of the hypotheses for each characteristic, based on the t-Test value and the P-value approach, are presented in Table 8-3. The symbol ↑↓ means that we can't reject the null hypothesis, HYP^0, so there is no significant difference between using the mobile app without adaptable features and using the mobile app with adaptable features for the objective characteristic. The symbols ↑ and ↓ mean that we reject the null hypothesis, HYP^0. So, there is a significant difference between using the mobile app with and without adaptable features, and, based on the value of the t-Test that falls either in the positive or the negative region for each characteristic, we accept the alternative hypothesis, HYP^a. If the symbol is ↑, one app type (either the app with adaptable features or the app without adaptable features) is better than the other, and if the symbol is ↓, one app type is not better than the other.

Table 8-3 Results of the hypotheses evaluation for the QiU-4-MUI objective characteristics for all the user stereotypes

New Quality-in-Use Objective Characteristics	Application Type	User Stereotype		
		Junior	Intermediate	Experienced
Effectiveness	Mobile	↓	↓	↑↓
	Adaptable mobile app	↑	↑	↑↓
Productivity	Mobile	↑	↑	↑
	Adaptable mobile app	↓	↓	↓
Efficiency	Mobile	↑↓	↑↓	↑↓
	Adaptable mobile app	↑↓	↑↓	↑↓
Error Safety	Mobile	↑↓	↑↓	↓
	Adaptable mobile app	↑↓	↑↓	↑
Cognitive Load	Mobile	↑↓	↑↓	↑↓
	Adaptable mobile app	↑↓	↑↓	↑↓

Legend	↑	↓	↑↓
	Better than HYP[a]	Not better than HYP[a]	No difference HYP[0]

8.3.2 Subjective Characteristic Results for All User Stereotypes

The subjective characteristics of the new quality-in-use model measure how easy it is for the user to interact with the MUI app in a specified context of use in terms of the eight features listed below:

F1: Easy to learn

F2: Easy to use

F3: Navigation

F4: Recovery from error

F5: Performing a task anywhere and at any time

F6: Use of the whole application

F7: Performing a particular task

F8: Finding the desired features on the screen

The results of the questionnaire reveal that all the participants gave all the features of both types of mobile app ratings of ‘easy’ and ‘very easy’. For statistical validation, HYP^0 and HYP^1 are verified for all the features, based on the Wilcoxon Signed-Ranks Test method for all the participants in each stereotype. The statistical significance was set at $P < 0.05$. All the statistical analyses were performed with the SPSS 18 software (IBM, New York, USA). The Wilcoxon Signed-Ranks Test method tests whether or not two related medians are equal. It is based on the absolute value of the difference between the two test variables.

i) **Junior stereotype group:** The Wilcoxon Test provides sufficient evidence against the null hypothesis for the difference between the medians of features 1 and 4 for mobile users and mobile adaptation users (the P-value is less than 0.05). Consequently, we can conclude that learning and recovery from errors is easier with PHIS2-M than with PHIS2-MA. For the other features, there is no significant difference between the medians for mobile users and mobile adaptation users (Table 8-4). However, this result is expected from this group. Since the junior users are young physicians and they don’t want to spend time learning how the adaptation features work, or waiting until the speech recognition feature recognizes their voice input. They speak quickly, which causes errors in speech recognition, and so they repeat the input many times until the correct view is displayed. In this respect, the habits of these users don’t match the framework that we used in the application.

ii) **Intermediate stereotype group:** The Wilcoxon Test provides no evidence against the null hypothesis of the difference between the medians of all the features for mobile users and mobile adaptation users. Consequently, we can conclude that the new application results in no significant difference for all 8 features (Table 8-4). We can interpret this as meaning that the habits of these users do not match our framework, but, at the same time they don't have any problems with learning or using the application. It may be that by increasing the sample size of this group in the future, the results for using the adaptation application will be significant.

iii) **Experienced stereotype group:** The Wilcoxon Test gives sufficient evidence against the null hypothesis for the difference between the median of feature 5 for mobile app users and adaptable mobile app users (P-value < 0.05). Consequently, we can conclude that PHIS2-MA is easier to use anywhere and at any time than PHIS2-M. For the rest of the features, there is no significant difference between the medians for mobile app users and adaptable mobile app users (Table 8-4). This result is expected from this group, since these users are mature individuals who move and talk more slowly than those in the other groups. For them, the adaptable app works very well, as their habits match our framework.

Table 8-4 Results of the hypotheses evaluation for the QiU-4-MUI subjective characteristics for all the user stereotypes

New Quality-in-Use Subjective Characteristics	Application Type	User Stereotype		
		Junior	Intermediate	Experienced
F1: Easy to learn	Mobile	↑	↑↓	↑↓
	Mobile Adaptation	↓	↑↓	↑↓
F4: Recovery from error	Mobile	↑	↑↓	↑↓
	Mobile Adaptation	↓	↑↓	↑↓
F5: Context	Mobile	↑↓	↑↓	↓
	Mobile Adaptation	↑↓	↑↓	↑

The participants were asked to evaluate each of the new features implemented in the mobile-based adaptable app, such as data input by voice, audio data output, brightness level adjustment, font size change, and user preference customization under Settings, following the Dumas and Redish (Dumas & Redish, 1999) classification. The performance levels for the tasks were the following:

i) Excellent: The application is very easy to use.
ii) Acceptable: The users are satisfied with the application on this task.
iii) Unacceptable: The users are not satisfied with the application on this task, because they have problems using the adaptable features on the MUI.

Table 8-5 presents the results for the adaptable features of PHIS-MA for all the user stereotypes, and, although the results were not statistically significant, the trend is toward the superiority of the adaptable version in terms of the adaptability of the latest iPhone™ features (note the higher percentages in the Excellent column).

Table 8-5 Adaptation feature results

MUI Feature	Junior			Intermediate			Experienced		
	Exc.	Acc.	Unacc.	Exc.	Acc.	Unacc.	Exc.	Acc.	Unacc.
Voice	60%	40%	-	62%	38%	-	83%	17%	-
Audio	85%	15%	-	77%	23%	-	83%	17%	-
Brightness	90%	10%	-	77%	23%	-	83%	17%	-
Font size	85%	15%	-	92%	8%	-	100%	-	-
Customization	75%	25%	-	85%	15%	-	92%	8%	-

8.4 Summary

Our goal in this chapter was to propose a design approach for an adaptable MUI which has been evaluated empirically on a real-world case study from the healthcare domain. The results of our research show that the adoption techniques may well offer a practical approach to the effective realization of the MUI features that exist in the latest smartphones. The adaptability of MUI features has the following benefits: i) it allows considerable changes to be made to apps, in terms of personal preferences and customization, as defined by the content being adapted; and ii) it allows information to be presented that is tailored to the user.

From this case study, we conclude the following: first, the proposed methodology for designing adaptable MUIs is sound and useful. Second, in this particular case study, where users are categorized into junior, intermediate, and experienced physicians, we found that it is much easier for the experienced physicians to use the adapted version of the system in terms of effectiveness and safety, since they have patience and take the time to learn about its features. However, the other two groups (junior and intermediate physicians) either tend not to spend their time adapting or do not need the adapted features.

It is interesting to compare the results of this work with those of our previous experimental work carried out for the junior user stereotype (Controlled Experiment IV) (Alnanih et al., 2013b). In that research, 12 junior physicians with previous experience using the PHIS application at KAUH participated in a controlled experiment conducted in a simulated environment in a university lab, which was different from a real hospital environment. The connection to the network was a fast one, the environment was not crowded as it is in a hospital, and the physicians were not under the usual pressures of work. The results of that experiment support PHIS2-M for productivity and efficiency, while for the remaining characteristic there is no significant difference between PHIS2-M and PHIS2-MA. However, when the sample of junior users was increased to 20 in the experiment reported in this chapter, which was conducted in a hospital, PHIS2-MA scored higher in terms of the effectiveness characteristic, but not the productivity or efficiency characteristics. A possible explanation for this might be the time available: entering input by voice using the voice recognition functionality on a mobile platform is slower than typing or tapping, and most of the tasks evaluated in the experiment for the adaptable version depend on voice input. Since productivity depends on time, PHIS2-M scores better on this feature than PHIS2-MA, because of a slower Internet connection in the hospital.

To solve this problem, we plan to use a different speech recognition modality, which does not require a network connection. We will then be able to examine the effect of using the new framework with junior users.

In the next chapter, we summarize our thesis research and its major contributions, and suggest future avenues for research in this area.

Chapter 9 Conclusion

In this thesis, we proposed our CON-INFO methodology for designing specialized context-driven MUIs for healthcare applications, along with an associated process, to support designers in deriving a conceptual design from the user's context. In current practice, there is no clear definition or understanding of the nature of the healthcare context.

The CON-INFO methodology covers the definition, development, and validation of a practical method to support MUI designers and developers in the design of adaptable MUIs based on the user's context model. The CON-INFO methodology consists of five phases: a user stereotype model, context descriptors, a context model, adaptation rules, and a quality-in-use model. We introduced a new characterization of the notion of context, based on the principle of isolating the concerns that are relevant to the application domain. We coined the term *context descriptor*, that captures the notion of context at the end-user level, while the characterization of the context incorporates the user model via the user stereotype. We bridge the context model, context descriptors, and the user stereotype to the MUI adaptation features using the decision table technique to represent the adaptation rules for MUIs on the iPhone™ technology. The proposed approach for MUI adaptation has been validated on four empirical studies (controlled experiments). The quality-in-use measurement model that we developed specifically for the MUI was used to compare the quality-in-use of MUIs with and without adaptation for the same healthcare app to be used by the same user in the same context.

Following this method, we proposed an approach for designing an adaptable MUI which rigorously defines a set of steps from context creation to the composition of a set of adaptable rules that match the right user and the right context variables. The details of each process were partly derived from our experience gained during Controlled

Experiment IV, where a design prototype and implementing using the adaptable features was built as the primary design directives and tested with one type of users as a pilot test. The application of interest was the medical software of Phoenix Health Information System (PHIS) that is used at King Abdulaziz University Hospital (KAUH). The new design with adaptation was compared to the design without adaptation, and resulted in significant improvement in terms of quality-in-use characteristics.

Based on the knowledge elicited from the domain experts in Controlled experiment IV, we proposed more representation rule for adaptable features that matched all the user stereotypes with considering both the input and the output tasks. As part of Controlled Experiment V, we tested our adaptable application in real hospital in the pediatric department with different user stereotypes (junior, intermediate, and expert). Our design was compared to the original design of mobile application without adaptation using our new quality-in-use model.

We have analysed the benefits of context-based MUI adaptation, and find that they are the following:

- Increased usability, e.g. if the MUI only supports one interaction model, such as typing or voice input/sound output, the usability of the service would be drastically decreased.
- Increased awareness of social ethics, e.g. in a quiet, multi-bed unit after midnight, when the patients are asleep and the nurse needs to enter vital signs, sound based interaction, such as voice input and sound output for call notification, could be an obstacle and disturb other patients. In this case, the sound could be turned off automatically.
- Improved workflow productivity, by reducing redundancy and repetitive tasks, like noting data on paper to be entered later onto a PC. Enabling nurses to enter vital signs immediately, preferably at the point of care, would allow them to be more efficient and productive.

- Improved level of attention, e.g. if the patient's result is outside of the normal range, this would be recognised by the MUI adaptation and the nurse notified based on his or her context.

9.1 Major Contributions

In this section, we summarizes the major contributions of this thesis as follow:

1. Characterization of a Context Model for healthcare application (Alnanih et al., 2012).

The novelty of this contribution consists in proposing multi-variables and multi-layers model. The variables determine variations in interaction behavior and needs, and are amenable for analysis and design support. The model consists from three layers, layer A (Basic context) share the physical characteristic of the context that match different types of hardware such as mobile devices, or desktop computer. Layer A maps to the layer B (Domain basic context) that define the specific characteristic of the user when performing several list of tasks in specific domain, and the characteristic of the object that used in order to perform the tasks in terms of MUI. Layer B maps to layer C (User Stereotypes), which defines the characteristic of the user in terms of his profile and preference.

2. Definition of a user model for healthcare application (Alnanih et al., 2013b)

The research contribution lies on the novelty of the user model for healthcare professionals that help the designer to understand the user needs and the user preferences in this critical environment.

The user model clearly benefits from the use of the context descriptor, introduced in this thesis to define a set of sentences or sentence fragments in the user's language describing the user's domain experience to match the context model from the designer's experience. Context descriptors help the designer extract the various values from the context model that affect the user in the real context, and adapt them to the design of the MUI. They also allow for more adaptability and greater versatility in the form of real-life logic and individual user information.

3. Derivation of Adaptation Rules (Alnanih et al., 2013a)

We used the Decision Tables technique to adapt the MUI features. The benefit from using this technique is to define a systematic process that involves moving from context model represented as conditions to adaptation rules represented as column. This process is traceable since the designer can add any number of row or column. The decision table is used to model all the adaptation rules in a logical way, so that the consistency and completeness of this incremental table can be easily checked.

4. Propose a new Quality-in-Use Measurement Model for MUI (Alnanih et al., 2013c)

The quality-in-use of the MUI designed with this methodology is monitored using a measurement model inspired by the ISO 9126-4 international standard and adapted to UI types in healthcare domain. The measurement model is validated both theoretically and empirically. The benefit of this model is to effectively evaluate the quality-in-use of mobile healthcare applications, and helping the designer evaluate the designing mobile application to execute work-related tasks more effectively and efficiency.
Our quality-in-use model is flexible enough to be applicable to other domains such as social, and educational. This allows for more practical research and industry feedback to improve the model.

5. Empirical Evaluation of Intelligent Mobile User Interfaces in Healthcare (Alnanih et al., 2014a)

This contribution presents the results of the formal evaluation of the CON-INFO methodology that conducted with 45 end-users in KAUH belongs to different user stereotypes.

6. A New Methodology (CON-INFO) for Context-Based Development of a Mobile User Interface in Healthcare Applications (Alnanih, Ormandjieva, & Radhakrishnan, 2014b)

This is a chapter of book presented in the Pervasive Health State-of-the-art and beyond. This chapter summaries the whole CON-INFO methodology including all the process that presented in this thesis. The benefits of the proposed methodology for healthcare professionals include i) increased effectiveness, productivity, efficiency, and satisfaction, through the ability to use the application anywhere and at any time; and ii) increased safety, through error reduction by restricting the use of the MUI to the options available, as well as making the UI more pleasant to use and easier to manipulate; and iii) reduced cognitive load on the user, through a reduction in the number of navigation-related actions.

7. HCI Principles and the Quality-in-Use Measurement Model for a Mobile User Interfaces in Healthcare (Paper submitted to a Journal in April, 2015).

This paper presents a result for assessing the impact of HCI principles on the quality of MUIs in terms of effectiveness, productivity, efficiency, error safety, and cognitive load in the healthcare context. The empirical investigation results confirmed that considering the HCI principles Mental Model, Metaphor, Visibility, Affordance, and Feedback in the MUI design leads to increased quality-in-use of the MUI.

8. Improving Software Quality-in-Use Model for Mobile Applications (Paper submitted to the Software Quality Journal, under review).

This paper aims to meet the mobile users need by proposing a software quality-in-use model for mobile applications. This paper is an improvement work for our previously published quality-in-use measurement model motivated by the recent international standards ISO [25010: 2011, 25021: 2012, and 25022: 2014] and adaptable to various applications. The measurement model was validated both theoretically (using the representational theory of measurement) and empirically (using controlled experiments).

9.2 Future Work

Its recommend to continue on with the development of the nursing and outpatient parts of PHIS2-MA which were not included in this study.

Our future works includes testing the final adaptable design with different types of smartphones and find a way to automatically convert the adaptable design to any mobile platform. Extensive testing of the remaining adaptable features has to be carried on with other categories of users within the hospital-ward domain. More specifically, we will test the effectiveness of the adaptable mobile healthcare application for use by physicians with a speech impediment.

It is important to automate our quality-in-use measurements for mobile applications. We also anticipate the further development of a tool to assist the developers in their quality-in-use evaluation of new MUIs. The tool will simplify the collection of the future experimental data required to explore and analyze the needs of the remaining user stereotypes in depth. Measuring software quality-in-use automatically is considered as a challenge due to the anticipated complexity of designing MUIs; this will be tackled in our future research work

We plan to use a different speech recognition modality, which does not require a network connection. We will then be able to examine the effect of using the new framework with junior users.

Also, we recommend using our CON-INFO methodology in different domains by applying the following steps (Figure 9-1):

To this end, to advance the state-of-the art, it is necessary to develop more rigors adaptable representation and models for different medical devices that can be linked to our conceptual design of CON-INFO methodology within a systematic and traceable process.

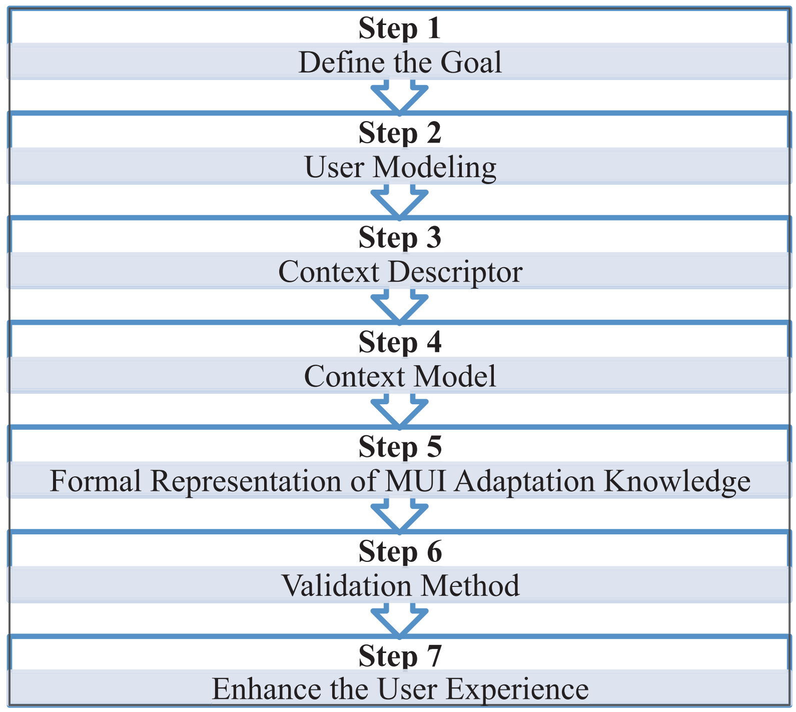


Figure 9-1 Our approach for MUI development in different domains

Finally, and after being involved in what affects peoples' health and safety, I felt that all ill peoples are living with compromises that affect their quality of life. When they approach a healthcare facility, it is their expectation, that they are in the caring and devoted hands of health practitioners using state of the art technology to help ease their pain. Neither they or their relatives nor caring health professionals are willing to give their precious lives to machine or not prove technology that could harm them.

Medical professionals spend many years in education, and apply strict roles of responsibilities and enforce strict ethics to avoid harming their patient. When they use devices, they share this burden with the developers of the interfaces of the machines or software they are trusting. They were not highly trained in technology and its ever-changing state.

In the medical practice, especially emergencies, seconds counts, the brain dies in five minutes, the heart could change its rhythm in seconds and a drugs could show its enormous effect in seconds. In this environment, healthcare professionals are well trained to respect the seconds and to maintain their accuracy and efficiency to help patients. During these types of struggles, there is no need to add an extra challenge or an extra burden to handle a device that could be used erroneously.

Since many studies have shown that most of the device-related errors in hospitals are due to inappropriate designs for user interactions rather than mechanical failure, therefore, it is extremely important for designers and developers of a medical software or devices to assume the responsible role of considering HCI principles as a major part of their responsibility. User-friendly interfaces will free up healthcare professional to spend more time caring for their patients rather than trying to learn how to use the interface to perform a given task. It may actually save lives.

Bibliography

Abrams, D., & Hogg, M. A. (2006). *Social identifications: A social psychology of intergroup relations and group processes*. Routledge.

Acharya, C., Thimbleby, H., & Oladimeji, P. (2010). Human computer interaction and medical devices. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp. 168-176). British Computer Society.

Albinsson, P. A., & Zhai, S. (2003). High precision touch screen interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 105-112). ACM.

Allport, G. W. (1979). *The nature of prejudice*: Basic books.

Alnanih, R. (2008). *The Application of Human Computer Interaction Principles on Medical Software. A Case Study: The Phoenix Health Information System in King Abdulaziz University Hospital.* (Master Dissertation), King Abdulaziz University, Jeddah, Saudi Arabia.

Alnanih, R., Alnuaim, H., & Ormandjieva, O. (2009). *New Health Information Systems (HIS) Quality-in-Use Model Based on the GQM Approach and HCI Principles.* Paper presented at the the proceeding of the 13th International Conference on Human-Computer Interaction. Part IV: Interacting in Various Application Domains.

Alnanih, R., Ormandjieva, O., & Radhakrishnan, T. (2011). *Challenges of Designing User Interfaces for Mobile Health Information Systems*. Paper presented at the IEEE Canada Women in Engineering National Conference (IEEE WIENC), Toronto, Canada.

Alnanih, R., Radhakrishnan, T., & Ormandjieva, O. (2012). Characterising Context for Mobile User Interfaces in Health Care Applications. *Procedia Computer Science, 10*, 1086-1093.

Alnanih, R., Ormandjieva, O., & Radhakrishnan, T. (2013a). Context-based and Rule-based Adaptation of Mobile User Interfaces in mHealth. *Procedia Computer Science, 21*, 390-397.

Alnanih, R., Ormandjieva, O., & Radhakrishnan, T. (2013b). Context-based User Stereotype Model for Mobile User Interfaces in Health Care Applications. *Procedia Computer Science, 19*, 1020-1027.

Alnanih, R., Ormandjieva, O., & Radhakrishnan, T. (2013c). A New Quality-in-Use Model for Mobile User Interfaces. *Proceedings of the 23nd International Workshop on Software Measurement, IWSM-MENSURA , pp. 165-170. IEEE.*

Alnanih, R., Ormandjieva, O., & Radhakrishnan, T. (2014a). Empirical Evaluation of Intelligent Mobile User Interfaces in Healthcare. In M. Sokolova & P. Beek (Eds.), *Advances in Artificial Intelligence* (Vol. 8436, pp. 23-34): Springer International Publishing.

Alnanih, R., Ormandjieva, O., & Radhakrishnan, T. (2014b). A New Methodology (CON-INFO) for Context-Based Development of a Mobile User Interface in Healthcare Applications. In A. Holzinger, M. Ziefle & C. Röcker (Eds.), *Pervasive Health* (pp. 317-342): Springer London.

Alvarado, C. J., Cao, C., Klein, G., Weinger, M. B., Patterson, E., Cook, R. I., & Carayon, P. (2004). *Panel: The Role of Human Factors in Healthcare–2020.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Ancona, M., Dodero, G., Minuto, F., Guida, M., & Gianuzzi, V. (2000). *Mobile computing in a hospital: the WARD-IN-HAND project.* Paper presented at the Proceedings of the 2000 ACM symposium on Applied computing-Volume 2.

Archer, N. (2009). Mobile E-Health: Making the Case1. *Handbook of Research on Advances in Health Informatics and Electronic Healthcare Applications: Global Adoption and Impact of Information Communication Technologies: Global Adoption and Impact of Information Communication Technologies*, 446.

Ash, J. S., Berg, M., & Coiera, E. (2004). Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *Journal of the American Medical Informatics Association, 11*(2), 104-112.

Bardram, J. E. (2004). *Applications of context-aware computing in hospital work: examples and design principles.* Paper presented at the Proceedings of the 2004 ACM symposium on Applied computing.

Bardram, J. E., & Hansen, T. R. (2004). *The AWARE architecture: supporting context-mediated social awareness in mobile cooperation.* Paper presented at the Proceedings of the 2004 ACM conference on Computer supported cooperative work.

Basili, V. R., Caldiera, G., & Rombach, H. D. (1994). Experience factory. *Encyclopedia of software engineering.*

Battisto, D., Pak, R., Vander Wood, M. A., & Pilcher, J. J. (2009). Using a task analysis to describe nursing work in acute care patient environments. *Journal of Nursing Administration, 39*(12), 537-547.

Bevan, N. (1999). Quality in use: Meeting user needs for quality. *Journal of Systems and Software, 49*(1), 89-96.

Briand, L., El Emam, K., & Morasca, S. (1995). Theoretical and empirical validation of software product measures. *International Software Engineering Research Network, Technical Report ISERN-95-03*.

Brown, P. J., Bovey, J. D., & Chen, X. (1997). Context-aware applications: from the laboratory to the marketplace. *Personal Communications, IEEE, 4*(5), 58-64.

Buranatrived, J., & Vickers, P. (2002). *An investigation of the impact of mobile phone and PDA interfaces on the usability of mobile-commerce applications.* Paper presented at the Networked Appliances, 2002. Liverpool. Proceedings. 2002 IEEE 5th International Workshop on.

Clerckx, T., Winters, F., & Coninx, K. (2005). Tool support for designing context-sensitive user interfaces using a model-based approach. In *Proceedings of the 4th international workshop on Task models and diagrams* (pp. 11-18). ACM.

Cooper, A. (1999). *The inmates are running the asylum:[Why high-tech products drive us crazy and how to restore the sanity]* (Vol. 261). Indianapolis: Sams.

Dahl, Y. (2007). Ubiquitous Computing at Point of Care in Hospitals: A User-Centered Approach.

Dantin, U. (2005). *Application of personas in user interface design for educational software.* Paper presented at the Proceedings of the 7th Australasian conference on Computing education-Volume 42.

de Melo, G., Honold, F., Weber, M., Poguntke, M., & Berton, A. (2009). Towards a flexible UI model for automotive human-machine interaction. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 47-50). ACM.

Dey, A., & Abowd, G. (2000). *Towards a better understanding of context and context awareness*.

Dey, A. K. (1998). *Context-aware computing: The CyberDesk project.* Paper presented at the Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments.

Desharnais, J. M., Abran, A., & Suryn, W. (2011). Identification and analysis of attributes and base measures within ISO 9126. Software Quality Journal, 19(2), 447-460.

Dix, A., Finlay, J., Abowd, G. and Beale, R. . (2004). *HUMAN-COMPUTER INTERACTION*.

Dourish, P. (2004). What we talk about when we talk about context. *Personal and ubiquitous computing, 8*(1), 19-30.

Dumas, J. S., & Redish, J. (1999). *A practical guide to usability testing*: Intellect Books.

Easterbrook, S., Singer, J., Storey, M.-A., & Damian, D. (2008). Selecting empirical methods for software engineering research *Guide to advanced empirical software engineering* (pp. 285-311): Springer.

Ebner, M., Lienhardt, C., Rohs, M., & Meyer, I. (2010). Microblogs in Higher Education–A chance to facilitate informal and process-oriented learning? *Computers & Education, 55*(1), 92-100.

EspritProject26900. (2000). Technology for enabling Awareness (TEA). Retrieved 02-2012, from http://www.teco.edu/tea/

Fenton , N., & Bieman, J. (2014). *Software Metrics: A Practical and Rigorous Approach.* (3rd Edition). CRC Press.

Fenton, N. E., & Kitchenham, B. (1991). Validating software measures. *Softw. Test., Verif. Reliab., 1*(2), 27-42.

Flanagan, J. A., Mantyjarvi, J., & Himberg, J. (2002). *Unsupervised clustering of symbol strings and context recognition.* Paper presented at the Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on.

Floyd, I. R., Cameron Jones, M., & Twidale, M. B. (2008). Resolving incommensurable debates: a preliminary identification of persona kinds, attributes, and characteristics. *Artifact, 2*(1), 12-26.

Food, U., & Administration, D. (2013). Mobile medical applications: guidance for industry and Food and Drug Administration staff. September 25, 2013.

Franklin, D., & Flaschbart, J. (1998). *All gadget and no representation makes jack a dull environment.* Paper presented at the Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments.

Fröhlich, N., Meier, A., Möller, T., Savini, M., & Vogt, J. (2009). *LoCa – Towards a Context-aware Infrastructure for eHealth Applications*. Paper presented at the 15th International Conference on Distributed Multimedia Systems, San Francisco, USA.

Gajos, K., & Weld, D. S. (2004). *SUPPLE: automatically generating user interfaces.* Paper presented at the Proceedings of the 9th international conference on Intelligent user interfaces.

Gwizdka, J. (2000). *What's in the context.* Paper presented at the Computer Human Interaction.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological review, 103*(2), 336.

Hinton, P. R. (2013). *Stereotypes, cognition and culture*: Psychology Press.

Hodgson, P. (2010). Usability for Medical Devices: A New International Standard. Retrieved 05/2011, from http://www.blueprintusability.com/topics/articleiec62366.html

Høegh, R. T., & Skov, M. B. (2004). *MobileWard: A Context-Aware Mobile System Supporting Work in a Hospital Ward.* Paper presented at the IADIS International Conference-Applied Computing 2004.

Holzinger, A. (2004). Rapid prototyping for a virtual medical campus interface. *Software, IEEE, 21*(1), 92-99.

Holzinger, A. (2005). Usability engineering methods for software developers. *Communications of the ACM, 48*(1), 71-74.

Holzinger, A., & Errath, M. (2007). Mobile computer Web-application design in medicine: some research based guidelines. *Universal Access in the Information Society, 6*(1), 31-41.

Huang, K.-Y. (2009). *Challenges in human-computer interaction design for mobile devices.* Paper presented at the Proceedings of the World Congress on Engineering and Computer Science.

Hull, R., Neaves, P., & Bedford-Roberts, J. (1997). *Towards situated computing.* Paper presented at the Wearable Computers, 1997. Digest of Papers., First International Symposium on.

Hwang, T. K. P., & Yu, H.-Y. (2011). Accommodating Both Expert Users and Novice Users in One Interface by Utilizing Multi-layer Interface in Complex Function Products. In P. L. P. Rau (Ed.), *Internationalization, Design and Global Development* (Vol. 6775, pp. 159-165): Springer Berlin Heidelberg.

IEC 62366:2007. (2007). Medical Devices—Medical Devices. Geneva: IEC.

IEEEStd1061. (1998). IEEE Standard for a Software Quality Metrics Methodology.

ISO 9241-11. (1998). Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability.

ISO/IEC TR 9126 -2. (2003). Software Engineering – Product Quality - Part 2: External Metrics , Geneva: International Organization for Standardization.

ISO/IEC TR 9126 -3. (2003). Software Engineering – Product Quality - Part 3: Internal Metrics , Geneva: International Organization for Standardization.

ISO/IEC TR 9126 -4. (2004). Software Engineering – Product Quality – Part 4: Quality in Use Metrics, Geneva : International Organization for Standardization.

ISO/IEC 15939:2007. (2007) Systems and software engineering – Measurement process. International Organization for Standardization, Geneva.

ISO/IEC 25000: 2014. (2014). Systems and software engineering —Systems and software Quality Requirements and Evaluation (SQuaRE) — Guide to SQuaRE.

ISO 9241-210:2010(E). (2010). Ergonomics of human–system interaction — Part 210: Human-centered design for interactive systems, International Organization for Standardization (ISO). Switzerland.

ISO/IEC 25010:2011(E). (2011). Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models. Geneva, Switzerland: International Organization for Standardization.

ISO/IEC FDIS 25021: 2012. (2012). System and Software Engineering - Systems and software product Quality Requirements and Evaluation (SQuaRE) – Quality measure elements ISO/IEC JTC1/SC7/WG6.

ISO/IEC CD 25022: 2014. (2014, July 20). Systems and software Engineering –Systems and software Quality Requirements and Evaluation (SQuaRE) –Measurement of quality in use ISO/IEC JTC1/SC7/WG6.

Jahns, R.-G. (10 November 2010). Reasearc2Guidance the mobile research specialists. Retrived from http://research2guidance.com/500m-people-will-be-using-healthcare-mobile-applications-in-2015/.

Jones, M., & Marsden, G. (2006). *Mobile Interaction Design*. West Sussex, England: John Wiley & Sons Ltd.

Junco, R., Heiberger, G., & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning, 27*(2), 119-132.

Jung, C. G. (1981). *The archetypes and the collective unconscious* (Vol. 9): Princeton University Press.

Kinkade, S., & Verclas, K. (2008). Wireless technology for social change: trends in mobile use by NGOs.

Korpipää, P., & Mäntyjärvi, J. (2003). An ontology for mobile device sensor-based context awareness *Modeling and Using Context* (pp. 451-458): Springer.

Lin, J., & Landay, J. A. (2002). *Damask: A tool for early-stage design and prototyping of multi-device user interfaces.* Paper presented at the In Proceedings of The 8th International Conference on Distributed Multimedia Systems (2002 International Workshop on Visual Computing).

Manickam, Y. (2011). Basic Guidelines for Developing Mobile Healthcare Applications. Retrieved 04/2012

Martinez-Garcia, A. I., & Favela, J. (2003). Context-aware mobile communication in hospitals. *Computer, 36*(9), 38-46.

Massanari, A. L. (2010). Designing for imaginary friends: information architecture, personas and the politics of user-centered design. *New Media & Society*, *12*(3), 401-416.

McCracken, D. D., & Wolfe, R. J. (2004). *User-centered website development: A human-computer interaction approach* (pp. 1-305). Englewood Cliffs: Prentice Hall.

Nielsen, J. (2005). Medical usability: how to kill patients through bad design. 2013, from http://www.nngroup.com/articles/medical-usability/

Nielsen, L. (2012). *Personas-User Focused Design* (Vol. 15): Springer.

Norman, D. A. (2002). *The design of everyday things*: Basic books, Inc. New York, NY, USA.

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotyping and social reality*: Blackwell Publishing.

Organization, W. H. (2005). eHealth tools and services: Needs of Member States. *Retrieved November, 12*, 2008.

Orwat, C., Graefe, A., & Faulwasser, T. (2008). Towards pervasive computing in health care–A literature review. *BMC Medical Informatics and Decision Making, 8*(1), 26.

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). *Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Pascoe, J. (1998). *Adding generic contextual capabilities to wearable computers.* Paper presented at the Wearable Computers, 1998. Digest of Papers. Second International Symposium on.

Pascoe, J., Rodrigues, H., & Ariza, C. (2006). *An investigation into a universal context model to support context-aware applications.* Paper presented at the On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops.

Pruitt, J., & Adlin, T. (2010). *The persona lifecycle: keeping people in mind throughout product design*. Morgan Kaufmann.

Pruitt, J., & Grudin, J. (2003). *Personas: practice and theory.* Paper presented at the Proceedings of the 2003 conference on Designing for user experiences.

Roberts, F. (1979). Measurement theory, encyclopedia of mathematics and its applications (Vol. 7). *Massachusetts: Addison-Wesley Publishing Company*.

Rodden, T., Cheverst, K., Davies, K., & Dix, A. (1998, May). Exploiting context in HCI design for mobile systems. In *Workshop on human computer interaction with mobile devices* (pp. 21-22).

Rönkkö, K., Hellman, M., Kilander, B., & Dittrich, Y. (2004, July). Personas is not applicable: local remedies interpreted in a wider context. In *Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices-Volume 1* (pp. 112-120). ACM.

Ryan, N. S., Pascoe, J., & Morse, D. R. (1998). Enhanced reality fieldwork: the context-aware archaeological assistant. In *Computer applications in archaeology*. Tempus Reparatum.

Savio, N., & Braiterman, J. (2007, June). Design sketch: The context of mobile interaction. In *Mobile HCI* (pp. 284-286).

Schilit, B. N., & Theimer, M. M. (1994). Disseminating active map information to mobile hosts. *Network, IEEE, 8*(5), 22-32.

Schmidt, A. (2000). Implicit human computer interaction through context. *Personal Technologies, 4*(2-3), 191-199.

Schneidewind, N. F. (1992). Methodology for validating software metrics. *Software Engineering, IEEE Transactions on, 18*(5), 410-422.

Shi-wei, C., & Shou-Qian, S. (2009). *Adaptive 3D navigation user interface design based on rough sets.* Paper presented at the Computer-Aided Industrial Design & Conceptual Design, 2009. CAID & CD 2009. IEEE 10th International Conference on.

Tang, P. C., & Patel, V. L. (1994). Major issues in user interface design for health professional workstations: summary and recommendations. *International journal of bio-medical computing, 34*(1), 139-148.

Tavasoli, A., & Archer, N. (2009). *A proposed intelligent policy-based interface for a mobile eHealth environment.*: Springer.

Thimbleby, H. (2007). User-centered methods are insufficient for safety critical systems *HCI and Usability for Medicine and Health Care* (pp. 1-20): Springer.

Thimbleby, H., & Thimbleby, W. (2007). *Internalist and externalist HCI.* Paper presented at the Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2.

Turner, P., & Turner, S. (2011). Is stereotyping inevitable when designing with personas? *Design Studies, 32*(1), 30-44.

Versloot, C. (2005). Information Filtering: Personal Profile Usable as Stereotype Profile?

Vogt, J., & Meier, A. (2010). An Adaptive User Interface Framework for eHealth Services based on UIML. *BLED 2010 Proceedings*.

Ward, A., Jones, A., & Hopper, A. (1997). A new location technique for the active office. *Personal Communications, IEEE, 4*(5), 42-47.

Winograd, T. (2001). Architectures for context. *Human-Computer Interaction, 16*(2), 401-419.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.

Yamabe, T., & Takahashi, K. (2007). *Experiments in mobile user interface adaptation for walking users.* Paper presented at the Intelligent Pervasive Computing, 2007. IPC. The 2007 International Conference on.

Yao, M. (2013). Mobile design: Avoid these 4 common user experience fails. Retrived February 2014.

Yin, R. K. (2013). Case study research: Design and methods. Sage publications.

Yuan, B., & Herbert, J. (2012). A fuzzy-based context modeling and reasoning framework for CARA pervasive healthcare. In Impact Analysis of Solutions for Chronic Disease Prevention and Management (pp. 254-257). Springer Berlin Heidelberg.

Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L., & Kubose, T. (2003). Using usability heuristics to evaluate patient safety of medical devices. *Journal of biomedical informatics, 36*(1), 23-30.

Appendix A: (Controlled Experiment I)

List of Tasks

Using desktop and mobile for Twitter Application

Please conduct the following tasks on *Twitter*:

Task 1: Search for a friend

Task 2: follow a friend

Task 3: write a comment “Hello”

Task 4: Unfollow a friend

Name:

Paper Login Form

Desktop Collecting Data

P #		**V**= # Of Views			
Date:		**X**= # of Incorrect actions			
Time to finish the test:		**A**= # of Actions (clicks)			
Task	**Time**	**V**	**A**	**X**	**P. Comments**
Task1 Search for friend on Twitter					
Task 2 Follow a friend					
Task 3 Write a comment "Hello"					
Task 4 Unfollow a friend					

Mobile Collecting Data

Mobile Collecting DataDate:		**X**= # of Incorrect actions			
Time to finish the test:		**A**= # of Actions (clicks)			
Task	**Time**	**V**	**A**	**X**	**P. Comments**
Task1 Search for friend on Twitter					
Task 2 Follow a friend					
Task 3 Write a comment "Hello"					
Task 4 Unfollow a friend					

P #…

Posttest Questionnaire

This questionnaire is designed to tell us how you feel about the PHIS you used today on desktop and mobile platforms. Please circle the number that most clearly expresses how you feel about a particular statement. Write in any comments you have below each question. Your cooperation is highly appreciated

1. Gender:
    - Male
    - Female

2. Age Group
    - 19- 25
    - 26-30
    - 31-35

3. Education
    - College
    - Bachelor Student
    - Master Student
    - Other ……….

4. How many years have you been using a smart phone?
    - Never used a smart phone
    - Less than one year
    - More than one year and less than two years
    - More than two years

5. Do you use your smart phone during your working hours?
    - Yes
    - No
6. If 'Yes', please specify what type of application do you use the most?
    - Basic application such as send/ receive message, email
    - Social application such as Facebook, Twitter, WhatsApp, LinkedIn
    - Browser
    - Only calls
    - Other……………………………….

7. How well can you do each of the following on computer and mobile devices?

		1	2	3	4	5
Learning the Apps application was in	Desktop	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Using the PHIS application was in	Desktop	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Your satisfaction with doing the particular task was in	Desktop	Very satisfied	Satisfied	Can't judge	Not satisfied	Not at all satisfied
	Mobile	Very satisfied	Satisfied	Can't judge	Not satisfied	Not at all satisfied
Your satisfaction with finding the features you wanted in the screen was	Desktop	Very satisfied	Satisfied	Can't judge	Not satisfied	Not at all satisfied
	Mobile	Very satisfied	Satisfied	Can't judge	Not satisfied	Not at all satisfied
Understanding the navigation in the prompts was:	Desktop	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Recovering from error was:	Desktop	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Performing the tasks anywhere, and anytime	Desktop	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult

Comments: ____

8. What type of application you prefer to do your tasks?
  - Desktop computer application
  - Mobile application
  - Other ………………….

9. Do you think this type of application will help you to manage your work better than the existing way?
  - Yes
  - No

10. Thinking about your experience with mobiles: To what extent do you agree with the following statements?

	Strongly agree	Agree	N/A	Disagree	Strongly disagree
Using (Twitter) on **Mobile** will help me to manage my work more than the existing way	☐1	☐2	☐3	☐4	☐5
Using (Twitter) on **Desktop** will help me to manage my work more than the existing way	☐1	☐2	☐3	☐4	☐5
I would prefer to use **Mobile** to perform my tasks	☐1	☐2	☐3	☐4	☐5
I would prefer to use **Desktop** to perform my tasks	☐1	☐2	☐3	☐4	☐5
Using **Mobile** save time	☐1 bn	☐2	☐3	☐4	☐5
Using **Desktop** save time	☐1	☐2	☐3	☐4	☐5

Please list any comments you had about the designing of application?

……………………………………………………………………………………………

Appendix B: (Controlled Experiment II)

List of Tasks

Using PHIS2-Desktop

Scenario:

You have been asked by your senior to review the following result for the patient: MRN **(2006876)**

Task 1: Check the most recent Complete Blood Count (CBC)

Task 2: Compare the most recent HB level with previous results

Task 3: Check the most recent medication prescribed

Your patient MRN (**2007435)** in the ward needs to order the following investigation:
Task 4: Order a "Chest X-Ray"

Task 5: Order a "Captopril" prescription

Using PHIS2-Mobile

Scenario:

You have been asked by your senior to review the following result for the patient: MRN (100**)**

Task 1: Check the most recent Complete Blood Count (CBC)

Task 2: Compare the most recent HB level with previous results

Task 3: Check the most recent medication prescribed

Your patient MRN (104) in the ward needs to order the following investigation:
Task 4: Order a "Chest X-Ray"

Task 5: Order a "Captopril" prescription

Name:

Paper Login Form

Collecting Data for PHIS2-D

P #		V= # Of Views			
Date:		**X**= # of Incorrect actions			
Time to finish the test:		**A**= # of Actions (clicks)			
Task	**Time**	**V**	**A**	**X**	**P. Comments**
Task1 Check the last HB					
Task 2 Compare the last HB level with previous results					
Task 3 Check the last medication					
Task 4 Choose another patient and Order a radiology					
Task 5 Order a "Captopril" prescription					

Collecting Data for PHIS2-M

P #		V= # Of Views			
Date:		**X**= # of Incorrect actions			
Time to finish the test:		**A**= # of Actions			
Task	**Time**	**V**	**A**	**X**	**P. Comments**
Task1 Check the last HB					
Task 2 Compare the last HB level with previous results					
Task 3 Check the last medication					
Task 4 Choose another patient and Order a radiology					
Task 5 Order a "Captopril" prescription					

Posttest Questionnaire

This questionnaire is designed to tell us how you feel about the PHIS you used today. Please circle the number that most clearly expresses how you feel about a particular statement. Write in any comments you have below each question. Your cooperation is highly appreciated

1. Gender:
    - Male
    - Female

2. Age Group
    - 20- 25
    - 26- 35
    - 36 -45
    - 46 +

3. How long have you been in this line of work?
    - Less than 5 years
    - More than 5 years and less than 10 years
    - More than 10 years and less than 15 years
    - More than 20 years

4. How many years have you been using a smart phone?
    - Never used a smart phone
    - Less than one year
    - More than one year and less than two years
    - More than two years

5. Do you use your smart phone during your working hours?
    - Yes
    - No

6. If you answers "Yes" in the previous question, please specify what application do you use the most?
    ....
    ……………………………………….
    ……………………………………….

7. How well can you do each of the following on computer and mobile devices?

Features		1	2	3	4	5
Learning the PHIS application was in	Computer	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Using the PHIS application was in	Computer	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Your satisfaction with doing the particular task was in	Computer	Very satisfied	Satisfied	Can't judge	Not satisfied	Not at all satisfied
	Mobile	Very satisfied	Satisfied	Can't judge	Not satisfied	Not at all satisfied
Your satisfaction with finding the features you wanted in the screen was	Computer	Very satisfied	Satisfied	Can't judge	Not satisfied	Not at all satisfied
	Mobile	Very satisfied	Satisfied	Can't judge	Not satisfied	Not at all satisfied
Understanding the navigation in the prompts was	Computer	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Recovering from error was:	Computer	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Performing the tasks anywhere, and anytime	Computer	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
	Mobile	Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult

Comments:____

8. What is the existing type of medical application that you used in the hospital to follow your patients?
   - Computer medical application
   - Mobile medical application
   - Other ……………………

9. What type of application you prefer to do your tasks at the hospital?
   - Medical desktop computer application
   - Medical mobile computer application
   - Other ………………….

10. Do you think this type of application will help you to manage your work better than the existing way?
    - Yes
    - No

11. Thinking about your experience with computers and mobiles: To what extent do you agree with the following statements?

	Strongly agree	Agree	N/A	Disagree	Strongly disagree
Using PHIS based Mobile will help me to manage my work more than the existing way	☐1	☐2	☐3	☐4	☐5
Using PHIS based Computer will help me to manage my work more than the existing way	☐1	☐2	☐3	☐4	☐5
I would prefer to use Mobile to perform my tasks	☐1	☐2	☐3	☐4	☐5
I would prefer to use Desktop to perform my tasks	☐1	☐2	☐3	☐4	☐5
Using PHIS on Mobile save time	☐1	☐2	☐3	☐4	☐5
Using PHIS on Computer save time	☐1	☐2	☐3	☐4	☐5

Please list any comments you had about the program?

Appendix C: (Controlled Experiment III)

Questionnaire

Dear Sir,

I would like to take few minutes of your time to help me answer few questions that are related to a research I am conducting on the use of Smart Phones in the domain of Health Care. Your cooperation is highly appreciated.

Thank you

1. General Information

*Name

*Hospital

*Department

2. Gender:

- Male
- Female

3. Age group:

- 25- 34
- 35- 44
- 45+
- 60+

4. How long have you been in this line of work?

- Less than 10 years
- More than 10 years but less than 15
- More than 15

5. How many years have you been using a smart phone?

- Less than 1 year
- More than 1 year and less than 2 years
- More than 2 years

6. Do you use your smart phone during your working hours?

- Yes
- No

7. If you answered yes to the previous question, the please specify which application do you use the most.

8. Which is your favourite functionality of your smart phone?

- Web browsing
- Medical
- Email
- Messaging
- Facebook and Twitter
- Other (Please Specify):

9. Usage of my Smart Phone

	Strongly Disagree	**Disagree**	**Agree**	**Strongly Agree**	**Not Applicable**
I often use my smart phone to assist me in my daily activities	◉	○	○	○	○
I often search for the location of the an item, person or place with the help of my smart phone	◉	○	○	○	○
I often use social networking tools to communicate with other people	○	○	○	○	○

10. Do you consider the security and privacy specifications of an application before using it to exchange work-related information?

○ Yes
○ No

11. Have you ever participated in a real-time, online event from a remote location?

◉ Yes
○ No

12. How often do you connect your smart phone to the Internet?

◉ Never
○ Daily
○ Weekly
○ Always connected

Appendix D: (Controlled Experiment IV)

List of Tasks

Using PHIS2 based Mobile and PHIS2 based Mobile Adaptation

Scenario:

You have been asked by your senior to do the following tasks:

Task 1: Search form the patient from the patient list- MRN (100)

Task 2: Choose from the main menu: Laboratory Results

Task 3: Select the most recent *Complete Blood Count (CBC)*

Task 4: Choose from the Order Radiology main menu "Chest X-Ray"

Task 5: Review the patient summary.

Name:

Paper Login Form

Collecting Data for PHIS2-D

P #		**V**= # Of Views			
Date:		**X**= # of Incorrect actions			
Time to finish the test:		**A**= # of Actions (clicks)			
Task	**Time**	**V**	**A**	**X**	**P. Comments**
Task1 Select the patient MRN: 100					
Task 2 Choose Lab Result					
Task 3 Check the last CBC					
Task 4 Order Chest X-Ray					
Task 5 Review the Patient Summary					

Collecting Data for PHIS2-MA

P #		**V**= # Of Views			
Date:		**X**= # of Incorrect actions			
Time to finish the test:		**A**= # of Actions (clicks)			
Task	**Time**	**V**	**A**	**X**	**P. Comments**
Task1 Select the patient MRN: 100					
Task 2 Choose Lab Result					
Task 3 Check the last CBC					
Task 4 Order Chest X-Ray					
Task 5 Review the Patient Summary					

P #…

Posttest Questionnaire

This questionnaire is designed to tell us how you feel about the PHIS you used today. Please circle the number that most clearly expresses how you feel about a particular statement. Write in any comments you have below each question. Your cooperation is highly appreciated

1. Do you use your smart phone during your working hours?
   - Yes
   - No
2. If you answers "Yes" in the previous question, please specify what type of application do you use the most?
   - Basic application such as send/ receive message, email
   - Social application such as Facebook, Twitter, WhatsApp
   - Medical Application such as Medscape
   - Browser
   - Other ………………..
3. What is the existing type of medical application that you use in the hospital to follow your patients?
   - Computer medical application
   - Mobile medical application
   - Other ……………………
4. What type of application you prefer to do your tasks at the hospital?
   - Medical desktop computer application
   - Medical mobile application
   - Medical adaptable mobile application
   - Other ………………….
5. Do you think this type of application will help you to manage your work better than the existing way?
   - Yes
   - No
6. Thinking about your experience with smart phones: to what extent do you agree with the following statements?

	Strongly agree	Agree	N/A	Disagree	Strongly disagree
Using PHIS based **Mobile** will help me manage my work better than the existing way	☐5	☐4	☐3	☐2	☐1
Using PHIS based **Mobile Adaptation** will help me manage my work better than the existing way	☐5	☐4	☐3	☐2	☐1
I would prefer to use **Mobile** to perform my tasks	☐5	☐4	☐3	☐2	☐1

I would prefer to use **Mobile Adaptation** to perform my tasks	☐5	☐4	☐3	☐2	☐1
Using PHIS on **Mobile** save time	☐5	☐4	☐3	☐2	☐1
Using PHIS on **Mobile Adaptation** save time	☐5	☐4	☐3	☐2	☐1

7. How well can you do each of the following on computer and mobile devices?

Features		Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Learning the PHIS application was in	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Using the PHIS application was in	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Understanding the navigation in the prompts was	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Recovering from error was	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Performing the tasks anywhere, and anytime	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Features		**Very satisfied**	**Satisfied**	**Can't judge**	**Not satisfied**	**Not at all satisfied**
Your satisfaction with doing the particular task was in	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Your satisfaction with finding the features you wanted in the screen was in	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1

Comments:____

8. Please evaluate the adaptation features on the PHIS application

Interface Features	Morning Shift	Evening Shift	Night Shift	Other
Use order speaking				
Listen to the result				
Brightness level				
Font Size				

Appendix E: (Controlled Experiment V)

List of Tasks

Using PHIS-M and PHIS2-MA

Scenario:

You have been asked by your senior to do the following tasks:

Task 1: Select the Patient with the MRN (100).

Task 2: From the main menu, order radiology “Chest X-Ray”.

Task 3: Review the most recent Complete Blood Count (CBC) results.

Task 4: Adjust your setting preferences.

Name:

Paper Login Form

Collecting Data for PHIS2-M

P #		**V**= # Of Views			
Date:		**X**= # of Incorrect actions			
Time to finish the test:		**A**= # of Actions (clicks)			
Task	**Time**	**V**	**A**	**X**	**P. Comments**
Task1 Select the patient of MRN:100					
Task 2 Order Chest X-Ray					
Task 3 Choose Lab Result and Review the CBC					

Collecting Data for PHIS2-MA

P #		**V**= # Of Views			
Date:		**X**= # of Incorrect actions			
Time to finish the test:		**A**= # of Actions (clicks)			
Task	**Time**	**V**	**A**	**X**	**P. Comments**
Task1 Select the patient MRN: 100					
Task 2 Order Chest X-Ray					
Task 3 Choose Lab Result and Review the CBC					

Setting Features	
Speech Recognition	
Text to Speech	
Font Size	
Brightness Level	

P #…

Posttest Questionnaire

This questionnaire is designed to tell us how you feel about the PHIS you used today. Please circle the number that most clearly expresses how you feel about a particular statement. Write in any comments you have below each question. Your cooperation is highly appreciated

1. What types of the following applications you may use during the work?
   - o Basic application such as send/ receive message, email
   - o Social application such as Facebook, Twitter, WhatsApp
   - o Medical Application such as Medscape
   - o Browser
   - o Other ……………………………….

2. What is the existing type of medical application that you use in the hospital to follow your patients?
   - o Computer medical application
   - o Mobile medical application
   - o Other ……………………

3. What type of application you prefer to do your tasks at the hospital?
   - o Medical desktop computer application
   - o Medical mobile application
   - o Medical adaptable mobile application
   - o Other ………………….

4. Do you think this type of application will help you to manage your work better than the existing way?
   - o Yes
   - o No

5. Thinking about your experience with smart phones: to what extent do you agree with the following statements?

	Strongly agree	Agree	N/A	Disagree	Strongly disagree
Using PHIS based **Mobile** will help me manage my work better than the current way	☐5	☐4	☐3	☐2	☐1
Using PHIS based **Mobile Adaptation** will help me manage my work better than the current way	☐5	☐4	☐3	☐2	☐1
I would prefer to use **Mobile** to perform my tasks	☐5	☐4	☐3	☐2	☐1

I would prefer to use **Mobile Adaptation** to perform my tasks	☐5	☐4	☐3	☐2	☐1
Using PHIS on **Mobile** save time	☐5	☐4	☐3	☐2	☐1
Using PHIS on **Mobile Adaptation** save time	☐5	☐4	☐3	☐2	☐1

6. How well can you do each of the following on Mobile and Mobile adaption?

Features		Very Easy	Easy	Neither Easy Nor Difficult	Difficult	Very Difficult
Learning the PHIS application was in	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Using the PHIS application was in	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Understanding the navigation in the prompts was	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Recovering from error was	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Performing the tasks anywhere, and anytime	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Features		**Very satisfied**	**Satisfied**	**Can't judge**	**Not satisfied**	**Not at all satisfied**
Your satisfaction with the whole application	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Your satisfaction with doing the particular task was in	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1
Your satisfaction with finding the features you wanted in the screen was in	Mobile	☐5	☐4	☐3	☐2	☐1
	Mobile Adaptation	☐5	☐4	☐3	☐2	☐1

Comments: ---

7. Please evaluate the adaptation interface features on the PHIS application

Interface Features	Excellent	Acceptable	Unacceptable
Enter data spoken			
Receive data listening			
Adjust Brightness level			
Font Size			
Customize your preferences in the application			