Utilising a Human Centered Design (HCD) approach for Enhancing Government Systems in Saudi Arabia through Usability Evaluation from the user's perspective

A thesis submitted for the degree of Doctor of Philosophy

By

Faisal Baglin

Department of Computer Science
Brunel University London

May 2015

ABSTRACT

When aiming to successfully improve an existing software system, usability evaluation methods (UEMs) and user experience (UX) are key aspects for consideration. The UEMs identify the level of usability of the system through assessing: (1) the extent to which it is easy and pleasant for the user (Cockton, 2012); (2) the specific effects of the system user interface (UI) on the user; and (3) any other problems that the system may have (Dix et al., 2007). On the other hand, considering UX places usability in context through providing a comprehensive understanding of the users' perceptions during and after their interactions with a specific system (Kuniavsky, 2010).

Undoubtedly, in most countries, there is a wide range of services, activities, and procedures that are supported by government systems (Buie and Murray, 2012). However, because of the lack of consideration of the usability requirements in addition to the limited attention given to the involvement of UX in system development (Downey and Rosales, 2011), a significant number of these government systems were designed without taking into account human-centred design guidelines (Johnson et al, 2005). Consequently, the success of these systems varies widely in terms of their usability (Downey and Rosales, 2011). In some cases, they fail to provide effective, efficient, and generally positive UX to people who interact with government systems from the outside, such as the citizens, or for those who work for the government on the inside, such as the employees (Buie and Murray, 2012).

The research problem in this thesis addresses how UEMs, techniques, and tools can be integrated and developed to support the redesigning and enhancement of current government systems (Legacy Systems) in a developing country. More specifically, the main aim of the research work reported in this thesis is to develop a way of proposing appropriate methods for evaluating the usability of the current internal systems in the Saudi government context. In this regard, three studies were conducted to achieve the aims of the research.

As a general approach for the thesis, Human-Centred Design (HCD) was adopted due to the fact that HCD is concerned with the integration of the users' opinions into the software development process in order to achieve a usable system (Spencer, 2004). In addition, a mixed method of quantitative and qualitative approaches was used in all of the studies. In the development of this project, the first study was aimed at evaluating the usability of a current internal system of a governmental organization in Saudi Arabia, the Visa Issuance (VI) system, from the actual users' points of view in order to identify the strengths and weaknesses of the system. A usability evaluation query technique was employed for collecting data via a survey method by targeting 135 participants who were the users of the VI system. The survey used both qualitative and quantitative

instruments, namely a questionnaire and semi-structured interviews. In the second study, an experimental approach was applied and a comparative usability test was conducted between the current VI system and a suggested prototype design which was developed based on the outcomes of the usability evaluation in the first study. The results of this study showed improvements in the quality of the system (usefulness), the information, and the interface. After analysing these results, the iterative method was used in the third study to redesign the suggested prototype. Therefore, another comparison test was conducted between the two versions of the prototype and the results indicated enhancement in the UX by using the new version.

This research developed a methodological framework for the usability evaluation of the current government systems which involved query techniques and user testing methods. It was formulated by combining different methods for guiding the redesign process, and testing was conducted throughout the entire research project. The results indicated that the involvement of query techniques as a preliminary step provides a quick, simple, and cost-effective way of identifying the usability problem areas in the VI government system. Furthermore, the usefulness of this developed framework could be beneficial in raising awareness and acceptance of the established methods among governmental organisations in other contexts in order to enhance their software systems effectively and improve the UX. It is hoped that this awareness of the fundamental usability methods could lead to developments in Information Communication Technology (ICT) for all communities (Holzinger, 2005) so that the advantages of making certain improvements could be shared with others.

In addition, the outcomes of the two experiments conducted in this research provide some lessons that are considered valuable in the usability testing domain. In this regard, the results are expected to assist and support the usability practitioners and system developers who are concerned with improving the usability of existing internal software systems, and in planning and conducting usability testing sessions in government organisations through utilising such guidance about UEMs.

Dedicated to the soul of my Mother,
who has always waited for this moment of achievement to be reached.
Indescribable how much I wish she was still with me, as indeed she will remain in my heart and soul forever.

Acknowledgements

First and foremost, I thank Allah the Almighty from the bottom of my heart for guiding and inspiring me. All great and good things that I have had in my entire life are due to His assistance, mercy and love.

I would like to express my deepest gratitude and appreciation to my academic supervisor Dr. Steve Love for his superb supervision, guidance, invaluable feedback, and encouragement during hard times.

My sincerest thankfulness to my father, who has supported me since I was a child until now. Without his unlimited support, endless prayers and patience, this mission would not have been possible.

I owe my sincere thanks to my beloved wife, Nahla, who recently finished her PhD fruitfully. I could not find suitable words to express how essential her inspiration and support were. They meant a lot to me as a source of continual encouragement. Thank you very much my Love for everything you have done for me and our kids; it is indeed unforgettable moments we have had in our journey.

To my great sons, Bader and Basil, my two eyes, and my lovely new daughter, Bushra, the sun of my entire life, you are all a constant source of inspiration and motivation; I see my future through your loving eyes. This thesis is a small gift for you, GOD bless you all!

Many thanks go to my wonderful sisters and brothers and all family members for their unconditional love, encouragement, prayers and solid support.

Special appreciation and thanks go to my mother- and father-in-law for their great support and kind prayers.

Additional thanks go to all colleagues and staff members in the Department of Computer Science, who have helped me and shared their knowledge during my study.

I extend my gratitude to the administration of the Ministry of Labour in Saudi Arabia for providing approval and assistance and facilitating a lot of matters related to the data collection. In particular, I would like to thank all participants from the recruitment department, most especially those who were selected for and agreed to participate in the usability test sessions.

Lastly, I would like to express my thanks to the Ministry of Higher Education in Saudi Arabia, represented by the Saudi Cultural Bureau in London, for sponsoring my PhD study.

Table of Contents

List of Figures

List of Tables

Declaration

From the research conducted for this PhD thesis, I'm intending to propose three papers for publication, as the following topics:

1- The Evaluation of System Usability of a Governmental Organization Using a Query Technique.

2- Designing the User Interface of a Government System Using Iterative Prototyping Method.

3- A Methodological Framework for Enhancing the Usability of Government Internal Systems in Developing countries

These papers will be submitted namely in the following journals:

- *International Journal of Human-Computer Interaction*
- *Ergonomics in Design - SAGE*
- *Applied Ergonomics - Elsevier*

Abbreviations

AADM	Automation Assistant Decision Making
AF	Archiving Function
ATM	Automated Teller Machine
ATM1	Air Traffic Management
CDSS	Clinical Decision Support System
CI	Confidence Interval
CS	Current System
CUSQ	Computer Usability Satisfaction Questionnaire
DICE	Diagnoses of Intensive Care Evaluation
DPD	Developed prototype design
HCD	Human-Centred Design
HCI	Human-Computer Interaction
ICT	Information Communication Technology
INFOQUAL	Information Quality
INTFQUAL	Interface Quality
ISO	International Standardization Organization
MATIS	Multimodal System for Train Timetable Information
MS	Microsoft
MUSiC	Metrics for Usability Standards in Computing
MUMMS	Measuring Usability of Multimedia System
NIC	National Information Centre
RPA	Remotely Piloted Aircraft
PD	Participatory Design
RQ	Research Question
RQMT	Requirement
SAB	Scientific Advisory Board
SDLC	System Development Life Cycle
STBI	Scenario Task Based Interviews
SUMI	Software Usability Measurement Inventory
SYSQUAL	System Quality
UE	Usability Engineering
UEMs	Usability Evaluation Methods
UEV	Usability Evaluation
UI	User Interface
UX	User Experience
V1	Version of prototype design number 1 (old version)
V2	Version of prototype design number 2 (new version)
VI	Visa Issuance

Chapter1: Introduction

1.1 Overview

This chapter begins by presenting a brief background of the research, followed by discussing the motivation behind the research work conducted for this thesis. This discussion will help in explaining the development of the primary research questions. Afterwards, the research methodologies used to investigate the research questions are then introduced. Finally, an outline of the thesis structure is provided, with a brief description of the contents of the remaining chapters.

1.2 Background

Recently, the demand to develop a high-quality system that affords appropriate effectiveness, efficiency, and satisfaction to the end users is growing rapidly (Madan and Duby, 2012; Agenr et al., 2011; Kumar and Dubey et al., 2010) among the various organisations in both the private (Bias and Mayhew, 2005) and government sectors (Buie and Murray, 2012). Accordingly, the focus on the system usability has increased and achieved significant popularity due to being considered a key success factor of software systems among developers and users (Bygstad et al., 2008; Stary and Eberle, 2008) and an essential quality factor of the overall software system (Jayaletchumi et al., 2014; Bevan, 2009; Seffah et al., 2006; Catarci et al., 2004; Xenos, 2001). In addition, performing the usability evaluation has become an essential step towards the success of a software system through improving the user experience (UX) (Casalo et al. 2010; Maguire, 2001; Nielsen, 1994).

In terms of ease of use, usability determines the usability evaluation for developing a system’s quality characteristics (Odeh and Adwan, 2009). In the discipline of usability engineering (UE), this view can lead to the design or redesign of the UI of a software system through involving users’ participation (Gulliksen et al, 2007) to enable an enhanced productivity and experience, to reduce the error rate in the workplace (Macleod, 1994; Hix et al., 2004), and to guarantee the efficacy of the target system. According to Metzker and Offergeld (2001), the usability engineering (UE) lifecycle is considered an attempt to redesign the entire system development process around (UE) knowledge, methods, and activities. In addition, it focuses on accomplishing the defined usability goals using an iteration of UE methods, such as prototyping, usability testing, and user interface mock-ups. Although the 'UE' highlights providing a variety of techniques to analyse users, specify usability goals, and evaluate designs, it does not address the entire development process.

One of the most important factors in the usability evaluation is focusing on the system end users' perspectives and their experiences in using the system and how they will interact with it and perform the actual tasks using it. Nebe and Zimmermann (2007) stated that one of the most important issues in the design of usable applications is to learn about the people who will be using the application'. Accordingly, the users' views, particular knowledge and needs about the target system become significant issues that need to be involved within the system development process to improve the system usability (Nielsen, 1993). In addition, this importance of considering the users' opinions to improve the system usability through measuring it as a key quality factor requires the adoption of a Human-Centred Design (HCD) approach to software (Ferre and Medinilla, 2007) due to the fact that HCD is concerned with the integration of users' opinions into the software development lifecycle to achieve a usable system (Maguire, 2001) and as a significant aspect that is directed to the improvement of successful interfaces (Rubin, 1994; Preece et al., 1994; Costabile, 2001). This in turn would lead to helping users interact and communicate with the system (Yee et al., 2010) and then perform and finish their tasks easily.

Despite the fact that the perception of the importance of system usability has grown considerably in the last twenty years (Juristo, 2009; Ferre and Medinilla, 2007; Folmer el al., 2004), there are still a number of systems with poor design and useless capabilities (Bias and Mayhew, 2005; Maguire, 2001; Och Dag et al., 2001). This could mostly be due to the fact that the projects of the development of systems have not constructed the system ease of use and fitted usability well, and they also fail to address the users' requirements properly (Juristo, 2009).

Accordingly, this may result in a complicated design of the system user interface (UI) or even in poorly designed primary tasks that need to use the system to be performed, which results in procedures of targeted tasks that consume considerable time (Freiberg and Baumeister, 2008). This would have been conveyed by users with frustrating experiences, as they would have found these systems difficult to learn and complicated to use (Bias and Mayhew, 2005: Maguire, 2001). Bygstad et al. (2008) pointed out that although there is an emphasis upon the importance of usability by organisations, in reality, they are not willing to use the resources of usability evaluation in their projects, and it is apparent that the importance given to it is much less than that of usability requirements.

The government sector would embrace these issues more than the private sector for several reasons, such as government systems having the largest user base of any technology. In addition, these systems provide enormous services and activities to the public (Buie and Murray,2012). For example, the US General Accounting Office, a main supporter of software engineering, found that 98% of software designed for the US government was 'unusable as delivered' (Smith, 1993 cited in Johansen and Zhang, 2004).

Thus, it is crucial for government systems to be easy to use for users who intend to interact with their government, such as citizens using government websites, employees performing tasks on a specific government system for specific services, or even for the specialists in the system's domain (Quesenbery, 2011). Although the government sector has been involved in different aspects of UX for over 80 years, which reflects a good progression of human consideration in system design for the benefit of public servants or citizens, there is a substantial variation of the involvement of government organisations in UX. This includes the different usability levels that have been developed or refers to a complete lack of attention to UX in government systems (Downey and Rosales, 2011).

In addition, as mentioned, government sectors have already made progress regarding human consideration in the system design. Accordingly, it would better to use the legacy of the 'current' systems as the benchmark for future ideas in redesigning considerations. In addition, this beneficial progress should be continued for enhancing and obtaining more usable systems rather than replacing them with completely new systems, which would be considered a waste of human resources as well as economic recourses (Johnson and Zhang, 2005). According to Bianchi et al. (2003, p225), ‘in order to preserve the asset represented by the legacy system, the familiarity with it gained by the system’s maintainers and users, and the continuity of execution of current operations during the reengineering process, the system needs to be reengineered gradually’.

1.3 Research Motivations

Currently in Saudi Arabia, which is one of the entrepreneurial developing countries, the government has put considerable effort into the attempt to provide the best quality services to the citizens; however, the demand for government services continues to grow with the population increase, and the number of applicants for these services has increased, which leads to a burden on the government systems and a heavy workload on the employees who are considered the users and operators of these government systems. This affects their performance as well as decreases their productivity. For instance, according to the report issued by Ministry of Labour, the number of issued work permits to foreign workers who received Visas to enable them to officially work for the citizens in Saudi Arabia has continuously increased from 2009 to 2013, and it archived 6.87 million work permits. Table (1-1) presents the growing number of work permits issued during the past five years.

Table 1.1 The issued work permits

Years	Numbers	Percentage of the changing
2009	3222809	(% 10)
2010	3345630	% 4 +
2011	4088093	% 22 +
2012	4811095	% 18 +
2013	6874799	% 43 +

- This information was provided to the public by a statistical website (Argaam,2014), which provides reliable economic and financial information [1]

Furthermore, columnists in local newspapers have published a number of reports reflecting several problems that occurred within a government organization regarding the Visas Issuances System. One of these reports mentioned that there was overcrowding due to the slow procedure (Al-harthi, 2008; Al-edwani, 2010; Al-baydani, 2011; Al Kabli, 2013). These reports reflect the general view of the members of society regarding the applied government systems. Obviously, there were several factors that could have caused this issue, but the lack of attention pertaining to the system usability in general and specifically to UX would be a critical factor in this frustrating situation. Seffah et al. (2006, p.159) stated that 'unusable user interfaces are probably the single largest reason why encompassing interactive system-computers plus people fail in actual use'. According to this imperfection of the current government systems in Saudi Arabia and its implications and the delay in completing the applicants' requirements, a general negative impression regarding the internal software systems of governmental organisations has been growing and has become ingrained among the public. This issue has motivated conducting related research for developing the existing government system in a developing country and providing guidance for a solution to this issue.

[1] ARGAMM " investment number", is a company specializing in the dissemination of economic and financial information on the Internet of interest to investors and decision-makers in the Arab world. Available online at: http://gulf.argaam.com/article/articledetail/445610

In addition, from reviewing the existing literature and related sources of the research in the fields of Human-Computer Interaction (HCI) and Usability Engineering (UE), it was found that there were several studies that provided validated human-centred design methodologies; however, none addressed the methods required in the process of re-designing and enhancing the existing government system. In addition, it was difficult to find research that paid enough attention to evaluating the usability of current government systems in developing countries, that provided a validated methodological framework as guidance for the improvement process or that added potential value to the system for users, such as the improvement of their performance in finishing their tasks using a particular system. Therefore, conducting a research study in an attempt to fill the gap in the existing literature is needed.

In considering the explanation of all of the issues that have been mentioned, the author took the opportunity to discover through research what significant contribution could be made in regards to knowledge and management in these matters. The research work was therefore conducted regarding the evaluation of system usability in respect to identifying an appropriate methodological framework for the purpose of enhancing and upgrading the current internal systems in a developing country context such as Saudi Arabia. On the other hand, the research contributes by filling the gap in the existing literature in the domain of usability engineering (UE), user experience (UX) and human-computer interaction (HCI) in relation to the usability in government systems.

Regarding the benefits of this research and after achieving useful findings, it is an endeavor to change the attitude towards system development in governmental organizations in Saudi Arabia through presenting this research work as an introduction of involving the users' viewpoints as an integral part of this improvement process of the current government systems, and in the end, to obtain usable government systems.

1.4 Research aims and questions

Given that a system usability evaluation along with implementing the HCD approach and interconnected parts as key factors would impact users' interactions with a government system, it is important to consider the usability evaluation of a current system in a government context to achieve a good understanding and deep insight of the research work of this thesis and to provide a sound initiative to the government system developers to be able to produce more usable internal systems through re-designing the current system. According to Cockton (2011), 'unusable software could be made usable through re-design'.

Therefore, the main aim of this research was to explore the possible and applicable usability evaluation methods (UEMs) to be formulated as a methodological framework that can be applied in improving the existing government systems in Saudi Arabia through identifying usability problems and challenges as well as their sources from the users' viewpoints. This would be accomplished by employing different measurements, techniques and instruments and then working to find solutions to these problems to obtain usable systems.

This also could be used as an answer to a key research question, which motivated the thesis: How can the existing usability methods, techniques and tools be utilized, improved, and integrated to enhance the design of internal government systems? According to Durbin (2004), 'having a questioning attitude is the first step in the research process', and 'most quality research consists of comparisons. By carefully selecting a comparison group or condition, the quality of the research project can be improved'.

Through conducting the research work in chapter four, this thesis aims to carry out a summative usability evaluation of a current internal system in a government context in Saudi Arabia as a case for understating and presenting evident from the real world. The system that has been used in this research work is the Visa Issuance (VI) system, which will be defined in the next chapter. The assessment here is based on the employees' perspectives as users who have access to use the (VI) system. Moreover, this evaluation may lead to identifying issues that reflect the system's need for improvements to meet its users' needs and requirements. In addition, through their feedback, an analytical approach and discussion were conducted to provide and approve suggestions to improve the ease of use of this government system. Thus, through the research work provided in chapter five, this thesis aims to develop a prototype design (DPD) based on the obtained recommendations from the first study for the purpose of conducting a comparison usability test between its characteristics and abilities and the current VI system (CS). Furthermore, the actual users validated the proposed design to determine whether it fulfils their quantitative and qualitative usability requirements that were provided earlier in the previous study of usability evaluation of the VI system and whether they are able to use it more easily and effectively. The previous research work reported in chapter six aimed to assess and validate the new version of the developed prototype design (DPD) from the employees' point of view as target users and participants of the study. The design that was developed relies on the outcomes selected from the former experiment in the second study as a stage of the VI system development cycle.

To fulfil the aims of the research, a number of sub-questions that motivated the research programme had to be attained and answered:

RQ1: What are the usability problem areas of a current VI system that have an impact on the UX?

RQ2: What is the usability state of a current VI system based on its users' perspectives?

In addition, to fulfil a thorough primary usability evaluation study according to the usability problems discovered, a prototype design (DPD) was developed, and the effects of these proposed design solutions were examined via conducting a comparison test between the two systems: the current VI system and the suggested design. The following research questions frame this part of the research work:

RQ3: How should the current VI system be re-designed to produce a prototype design?

RQ4: What are the effects of the proposed usability design solutions on the UX?

Lastly, to thoroughly fulfil the first experiment study outcomes, in this research, a developed prototype design was re-designed through using an iterative technique, and the effects of these proposed re-designing solutions were examined via conducting a comparison test between the two versions: the old version (PDV1) and the new version (PDV2). The following research questions shaped this part of the research work:

RQ5: How should the UI of prototype design of a VI system be re-designed for the actual users of the VI system?

RQ6: What are the effects of the proposed re-designed prototype with additional features on the UX?

1.5 Research Methods

The research work reported in this thesis was conducted in three stages which are represented by the three studies. Thus, as Human Centred Design (HCD) was followed as the main approach of the research, it would help to ensure that a software system is designed with high levels of usability. The three basic principles of HCD are as follows: To involve users as early as possible during the design process, so that users' cognitive, social and attitudinal characteristics are identified and understood; to measure performance and attitude by utilizing suggested interfaces and simulations of the system; and then to design iteratively, so that a comparison testing and usability evaluation can be conducted to check that the design meets the user requirements, and to subsequently increase the validity of the outcomes (Gould and Lewis,1985).

In this thesis, a mixture of quantitative and qualitative methods was applied in the research work conducted for each study. A number of techniques and instruments were used to gather data for the three studies of this thesis, such as: a survey (questionnaires and interviews), usability testing, thinkaloud, observation, and free textboxes. Chapter three provides a detailed discussion of the methodologies used.

1.6 Research Structure

This section presents a brief summary of the seven chapters that make up the thesis.

Following the **introductory chapter (one),Chapter two** reviews the related literature, and discusses specific areas associated with the research field in previous research. The chapter begins by reviewing the main sources related to system usability and outlining which would be suitable within the research paradigm.

Chapter three provides and reviews the general approach, methodologies, techniques, and data collection instruments that are used for the research work conducted in this thesis. The chapter is divided into three sections; in the first section, a description and justification is provided for selecting and using HCD as the main research approach. The second section presents the main stages of the research work of the thesis, which included the three studies. The last section provides the data analysis methods and procedure, the design of the experiment, and the planning and conducting of the experiments, which are explained with references.

Chapter four presents the first study undertaken for this thesis. The purpose of this study was to evaluate the usability of the current VI system in order to identify the problems with the system and find solutions by addressing them in a prototype design in the next study. A query evaluation technique was adopted via utilizing a survey method for gathering data. Finally, the results are reported, analysed, and discussed, and a conclusion to the chapter is given.

Chapter five describes the second study, which is conducted as an experimental approach. A comparison usability test was applied with 32 participants, between the current VI system and the developed prototype design which was built based on the outcomes of the first study. The research techniques, test procedures, and instruments used to collect data for subsequent analysis are provided. Finally, the results of the study are analysed and discussed, and finally a conclusion with the main findings of the chapter is reported.

Chapter six presents the third study, it follows the same experimental approach settings applied in the first experiment reported in chapter five but with 26 participants. In this study, the iterative design was applied to involving new features such as automation assistant decision making (AADM) to the suggested prototype based on the main findings and requirements of the previous experiment.

A comparison usability test was conducted between the new and old versions of prototype design for the purpose of validating the improvement of it.

Chapter seven introduces the main findings and conclusions of the research from chapters four, five, and six. Afterward, a discussion of the findings is provided after the outcomes of the research work are analysed, and then the contribution to the addition to knowledge in the scope of this thesis topic. This chapter also clarifies the limitations of the research work conducted, which is observed through this study, and then provides the recommendations for future study to provide a possible improvement on the current research findings.

1.7 Chapter Summary

This chapter has presented and explained the motivation behind the research work of this thesis. The background to the research was described briefly and discussed; also the aims and objectives of the research were presented. Finally, the research structure of the thesis and a brief description of it were provided.

The following chapter is concerned with providing a detailed background and relevant sources for the research. This will provide a rich backdrop to the aims and objectives of conducting the research.

Chapter 2: Literature Review

2.1 Introduction

This chapter presents a review of preliminary background literature and related work from previous research in order to achieve a demonstration of the main aims of the thesis. Generally, the literature is associated with the (HCI) domain, and specifically with the User Experience (UX) discipline which encompasses several areas such as Usability Engineering (UE), information system (IS), and sub-discipline such as computer science (CSE), and cognitive science (CGS) (Downey and Rosales, 2011 cited in Buie and Murray, 2012).

The chapter begins by presenting early stuff related to making system UI design less usability, besides the characteristics of potential problems that could interfere the developing processes. In addition, it provides an overview of HCI, usability, and the main sources of its common themes, besides an overview of some state of the art usability evaluation methods, techniques, and tools; moreover, it surveys the previous works and studies. Then, the selected main methods of usability testing and the associated techniques for conducting it, are reviewed in specific detail in the second section, which focuses also on the related aspects of designing a usable system User Interface (UI) and a suggested prototype. Following on from that, the last section emphasises the iterative process literature as a complementary stage of system development, together with focusing on the involvement of Information Communication Technology (ICT) in the design via utilising the Automation Assistant Decision Making (AADM) feature. At the end, this chapter attempts to conclude by indicating how it would form the research questions for the thesis, relying on the perceived gap in the existing literature, and how this research study would be able to fill this gap.

2.2 What makes a system user interface (UI) less usable?

This sub-section, presents and discuss the several reasons that are likely to lead products or systems to be a challenge for users to use (Rubin and Chisnell,2008), and this needs to be taken into account in the development process:

a. Since the development of a system or product is aiming to enhance the human (users) performance generally, the developers and designers should take into account the three major components as presented in Bailey's Human performance model for considering any type human performance situation as follow (Bailey, 1993): The human, the context, and the activity. These three factors have an affect the ultimate results of how well humans (users) lastly perform of using the system. However, the designers, developers and programmers have traditionally placed the greatest focus on the activity factor, and much less focus on the human and the context factors. Therefore, the machine or system has become the main concentration at the development stage rather than emphasizing on the end users during the designing and developing the process of this machine or system.

b. Nowadays, it is almost difficult to find the average person not to use such electronic or computer-based equipments in either workplace or in personal life, whereas in the past it was unusual for a non-expert person to use these equipments. So, the development organization does not react adequately to this dramatically changing in the target users compared with the past in several aspects like improved skills, increased aspiration towards using technology, increased knowledge of computers and mechanical machines, and increased pride in their ability to identify and solve any problem;

c. The designing of a usable product or system is considered as a hard task and an unpredictable endeavour, nevertheless, many organizations deal with it as a basic ability to perceive, understand and judge things in practical matters, or in other words treat it like if it was "common sense".

d. Team specialists in the organisation do not always work in an integrated manner due to the fact that often the product or system development proceeds separately. Because of the various elements that the software product is comprised of, i.e. the UI, the help system, and the written materials, so these components would be developed by separate individuals or teams. After that the challenge would arise when these separate elements have little integration with weak communication among the different development teams.

e. The last reason is that the design of the UI and the technical implementation of this UI are different processes with different required skills. Furthermore, with technological development and the emergence of new programming languages and instruments for automatically improving program code, this leads to the obstacles to technical implementation being decreased. On the other hand, the challenge of design has increased with increasing expectation of ease of use, besides the demand to achieve a broader and less sophisticated user.

Rubin and Chisnell (2008, p.13) stated that *"These five reasons merely brush the surface of how and why unusable products and systems continue to flourish"*

2.3 Characteristics of usability problems of UI

The usability problems can be categorized in four different ways: as a single part in the UI, as two or more parts that need to be compared to find the problem, as a problem with the entire structure of the UI, or as something is currently missing and it needs to be included in the UI. Although there is a difference between these four categories to define the usability problems of UI, Nielsen (1992a) after analyzed 211 usability problems, he found that this difference was small and it is not statistically significant. In addition, the evaluators were approximately equally well at identifying all four types of usability problems, however, *"the interaction effect between these categories and interface implementation was significant and had a very large effect"* (Nielsen and Molich, 1991,p57).

2.4 Emergence of Human Computer Interaction

The term Human Computer Interaction (HCI) has been in widespread use since the early 1980s (Dix *et al.*, 2004), and in the middle of the same decade. It was adopted as a way of describing the new field of study concerned with different aspects that relate to the interaction between users and computers (hardware and software as system) (Preece *et al.*, 1994). This field of research was originally known as Man-Machine Interaction or interfacing (Dix *et al.*, 2004). However, the definition of HCI relies on *"the situational context and the referent discipline being considered"* (Zhang *et al.*, 2004, p 359). So, taking into account that the terms user, computer, and system are often used in this context, HCI can be defined as a design that should create a fit between the user, the machine (computer), and the required services, in order to achieve a certain performance for both in quality and optimality of the services. This fit is based on "the understanding of human physical constraints, limitations, and potentials" (Zhang *et al.*, 2005, p520), as well as how the functions of a system can assist in achieving the purpose of the system (Karray *et al.*, 2008).

Although there is no single agreed-upon definition for HCI, the following definition represents the common agreement upon by several experts, so according to Hewett *et al*. ,2009c, p5), *"Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them."*

In line with the one of main focuses in this research, HCI can be described as the ways that humans interact with information, technology, and tasks, especially in business, managerial organizational, and cultural contexts (Zhang *et al.*, 2005). This definition highlights the fact that HCI and concerns engage in all possible interaction between an end user and a software or system during its development process. .

2.4.1 The importance of HCI

There is no doubt that HCI draws on many disciplines and has its beginnings with them (Dix *et al.* 2004), such as computer science, computer engineering, cognitive psychology, social psychology, system design, management science, and, most recently, consumer psychology and marketing (Zhang, 2002). However, Dix *et al.* (2004) point out that computer science and system design must be accepted as a central concern, and that other disciplines will fit into important parts of the design process. Therefore, with this perspective in mind, HCI involves the design, implementation, and evaluation of an interactive system in the context of the user's task and work. In addition, Preece *et al.* (1994, p19) state that "one way of demonstrating the importance of HCI is by showing tangible benefits that can be talked of in cash terms", and that means by providing clear examples of case studies showing, for instance, cost reduction, improvement in the performance of work, and reduced absenteeism. Furthermore, a study conducted by Chapains (reported in Preece, 1994) shows that HCI has a significant influence on improvements in installing systems in some organization i.e IBM. The study discovered that the main problems related to the installation resulted from it being a labour-intensive task that took a long time, which in turn caused severe disruption to customer services. After HCI specialists were included in the team, indirect benefits were seen through improvements in productivity and customer satisfaction. There was an additional benefit in employee satisfaction because of the reduction in hard physical activities.

2.5 Usability

The term “usability” emerged in the early 1980s, and was adopted in the 1990s by the software industry (Nielsen,1990; Lewis, 2006; Bygstad et al., 2007). It is an aspect that can be approached from numerous perspectives, which is why multiple disciplines are concerned with it (Ferre et al., 2001). The most common view of usability in the field of Usability Engineering (UE) is that regarding the User Interface (UI), and is mainly related to the ease of use and learnability of a given software system (Juristo et al., 2007; Gonzalez et al., 2008). Scholz (2004) pointed out that UE is the discipline that delivers structured approaches for attaining usability in UI design during the system development lifecycle. In other words, it can be viewed as a process of achieving usability goals in a specific software system through employing a set of methods and techniques in different development stages, as it aims to improve the UI of this target system (Lecerof and Patern,1998; Nielsen,1993). However, Ferre et al. (2001) stated that usability is not just about the appearance of the UI, but is likewise associated with how the system interacts with the user, as the field of Human Computer Interaction (HCI) has traditionally dealt with the usability of the software system (Juristo et al.,2007), aiming to measure how easy and efficient it is for users to perform tasks, to achieve their goals, when using a piece of software or product (Han et al., 2001). In addition, Lewis (2006) indicated that Usability is an emergent property which represents one of the most significant consequences that rely on the interactions among users, systems, tasks, and environment.

2.5.1 What is Usability?

The earliest attempt to define usability as measures for "ease of use" was made in 1971 by a researcher called RB Miller (Shackel,1990). Folmer and Bosch (2002) stated that the term "usability" was initially generated from the term "user friendly", however, they clarified that, as with many other software engineering terms, usability has many definitions (Shackel,1990). This agrees with Dubey et al.(2012) and Baven et al. (1991), who mentioned that there are many studies that have provided different definitions of usability based on a number of views. It relates to a set of notions, such as the time taken to perform the action, performance, learnability, and user satisfaction. The common term could be usability, referring to a quality attribute that is found in the greatest number of classifications (ISO 9126-1, 2000; Juristo, 2007). Nielsen (1993) provided an exhaustive description of the usability as one of the issues that "*characterizes a global feature of a system, that is acceptability by the end users*, *reflecting whether the system is good enough to satisfy the needs and the requirements of the users*". (Nielsen,1993 cited in Costabile, 2000, p.3). In spite of the difficulty of finding a standard definition of usability (Ferre et al., 2001; Abran et al., 2003;Juristo et al., 2007), a number of studies have proposed several appropriate definitions of usability (Casalo, 2010), especially those by the International Standardization Organization (ISO) (Abran et al., 2010). For instance, ISO 9241-11 (1998), in part 11, presented one of the most generally accepted definitions, which identified three aspects of usability related to the user's "effectiveness, efficiency, and satisfaction", and defines Usability as *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*". (Arban et al., 2003; Stone et al., 2005 ; Beckert and Grebing, 2011). Furthermore, the ISO standard for software qualities (ISO 1991b, cited in Baven et al., 1991) took a product- and user-oriented view:

*"A set of attributes of software which bear on the effort needed for use and on the individual assessments of such use..."*A study by Brinck et al. (2002) indicated that usability reflects the extent to which people can fulfil several required tasks. Similarly, Dumas and Redish (1993, p.4) define usability as*people who use the product can do so quickly and easily to accomplish their task."* Lewis (2006) defined it as a budding possession that relies on the interactions between users, products, tasks and environments (Lewis,2006). Furthermore, in brief, it would define usability as "quality in use" (Arban et al,2003; ISO14598, 1999, cited in Juristo et al,, 2007). In addition, according to a study by Bevan (1995a), it distinguished between a broad and narrow view of usability, with the broad viewed as a high level of "quality of use", which should be considered as a major part of the design for a software system, if it enables the service users to complete the targeted tasks, whilst the narrower view is more about focusing on the design of features of a system which are a pre-requisite for quality of use. This view would connect the usability as a quality of use to the concept of usefulness. For example, it is inadequate to have just a well-designed UI system, it should also comprise high utility by having the right system for the right users and the right task. In addition, in respect to the reference to the term usability as "quality", Usability.gov[2] defined it as the quality of a user's experience during interaction with products or systems, including websites, software, devices, or applications, and it was determined to be about effectiveness, efficiency and the overall satisfaction of the user.

2.5.2 Why is it Important?

It is widely believed that usability is considered a substantial quality aspect of a software product (Abran et al., 2003;Stary and Eberle, 2007;Juristo et al., 2007, Mazumder and Das,2014). This quality of product or system can be described as a documented set of procedures proposed to certify that a product or system will meet primarily stated requirements (Bevan,1995). Furthermore, initially the importance of usability derives from the consideration that it has become a major aspect that can detect the qualities of a product and the needed functionality, which

[2]Usability.gov is the leading resource for user experience (UX) best practices and guidelines, serving practitioners and students in the government and private sectors. The site provides overviews of the human-centered design process and various UX disciplines. It also covers the related information on methodology and tools for making digital content more usable and useful. Access online at http://www.usability.gov/what-and-why/usability-evaluation.html.

commonly emerge throughout a product development lifecycle, and specifically, during the design stage (Hub and Zatloukal,2008).

2.5.2.1 Usability and User Experience (UX)

As many studies and scholars have pointed out enormous advantages gained from having concern for the usability of a system (Mayhew, 1999; Landauer, 1995; Juristo,2009, the study by Saffeh et al., (2006) pointed out several benefits of usable user interfaces, that can be classified based on the beneficiary. For instance, in respect of end users, the usability is important to increase the speed and accuracy of selected tasks to be accomplished using a certain system, it would lead to generating a high level of user experience UX as an indirect outcome. Furthermore, it would also gain several added values to the users, such as increasing productivity and improving performance, i.e. the number of occurrence errors by users would be reduced, and the experience of learning and using the system would become much easier. Besides this, it is important to ensure the safety of the user, in the case that the system is used to monitor unsafe processes. Although, the usability and UX relate to each other in terms of how well system is designed, yet there is still an important distinctions between them (Church,2013). Thus, the usability is regarding the system ease of use, and for enabling the users to achieve their target goal efficiently and quickly with minimum frustration and less errors. Whilst, UX comprises usability, but also could includes further emotional dimension. For instance, the desire, meaning, joy, , reflection, value or frustration that a user experiences.

Therefore, the UX can be defined as "*a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service"* (ISO 9241-110:2010, clause 2.15). This can be interpreted as meaning that UX seeks to find out how a person would feel towards using a product or system in terms of affective, meaningful, and value-related issues associated with the product usage (Vermeeren et al.,2010). Furthermore, it is commonly believed that UX is inherently dynamic, given the ever-changing internal and emotional state of a person and variances in the conditions throughout and after an interaction with a product (Law et al.,2009; Hassenzahl,2008;). Therefore, UX should be considered not only after a person interacts with a product, but even before and during interaction (Vermeeren et al.,2010). Rogers et al. (2011) stated that the UX is considered to be the main focus of the interaction design. From the explanation of these points, it becomes clear that UX is important as a factor and needs to be seen as something evaluable.

In relation to the research study reported in this thesis, according to Buie and Murray (2012) *"the history of UX in government illustrates the progression of human consideration in system design for the benefit of public servants and citizens"*. Thus, UX professionals are constantly seeking, with creativity and flexibility, to enhance UX in government system, as well as many members of the public sector, comprising decision makers, employees, and contractors, take part in UX activities with the goal of developing UX for both internal systems and public-facing technology, regardless of the existence of challenges.

2.5.2.2 Usability and Stakeholders

For the developer, usability would increase productivity and work standardization, in terms of achieving more usable systems. Lastly, for the organisation, usability decreases development and maintenance costs and increases sales and the organisation's profit, so they would become economically feasible. For an example of the positive financial implications, Juristo (2009) mentions that the IBM Corporation was persuaded that usability would make a business more successful, through having an improvement on its own website after it was found to be difficult to navigate and ineffective. So, the IBM Corporation tackled this issue by spending much effort to redesign the site over a ten week period, which involved more than 100 employees at a cost estimated to be "in the millions". The reported consequence was that the sales increased by 400% in the first week (Juristo, 2009; Rauterberg, 2003, and the net savings for IBM during one year were estimated as $554,840. This is in fact what usability expert Jakob Nielsen (1993) suggested; that organizations should expend 10% of the improvement budget on usability. Moreover, in a study conducted by Donahue (2001) it was stated that each dollar spent by an organization would offer a return on the investment of $30.25. The study of Bias et al. (2005) supported this view and reported the benefits of evaluating the usability of a software system, as it achieved results which stated that $10 invested in usability will make a return of $100. Furthermore, improving productivity and increasing team morale is one of the significant benefits of the interest in usability, so the usable system will help the users to operate effectively and save time, and enables them to focus on the task rather than the tool (Maguire,2001). Also, a well-designed and usable system enables them to have less training and reduces the need for human support (Macleod,1994). These aspects are likely to have a positive effect on efficiency and productivity within the organization (Jordan, 1998). Also, he indicated in his book a field study conducted by Allwood (1984) which showed that difficulties faced by users (employees) in using a system could cost organizations between 5% and 10% of total working time (Jordan,1998). Additionally, if the system has an appropriate usability, the users will need to be less computer literate, which would otherwise oblige them to spend a long time in learning how such a system works (Ferre et al.,2001). Mazumder and Das (2014) stated that the system users will not like to accept the system if it is not usable enough.

2.5.3 What makes a software system usable?

There are different studies that have provided a number of factors that would have an impact on the development process to provide a usable software system. These factors comprise several aspects, such as, to what extent the functionality meets the user's needs, to what extent the inflow via the system fits with user tasks being performed effectively, and to what extent the response of the system adequately meets user expectation. For example, Bruno and Al Qaimari (2004) stated that there are four factors that need to be understood and considered for enabling the development of a usable system. These are, "*the targeted users, the selected tasks, the type of technology, and context of use or environment"*. Similarly, Spencer (2004)[3] pointed out that the definition of usability as a quality aspect which is provided by ISO 9241-11, as mentioned earlier contains four important elements for producing a usable system:

a. "The represented users of the system have to be placed at the top of the list of priorities;

b. These users have to be knowledgeable with a set of specified goals;

c. The system should allow its users' goals to be achieved (effectively) in an efficient way and the users will then be satisfied with the process or outcome;

[3] Donna Spencer is an alumni of Step Two Designs, and is a specialist in information architecture. Donna has presented widely on Information Assurance (IA). Available at: *http://www.steptwo.com.au/papers/kmc_whatisusability/*

b. The system should be used in a particular context (e.g. a physical location, and a business or service process)".

Additionally, Rubin and Chisnell (2008) pointed out that although the developers of a product or system design must be concerned about the technology in the first place and then about the features that will be helpful for designing a usable product or system, the user should be at the centre of this process, and the developers and designers must consider what the user’s experience will be during the system usage. This would explain the importance of applying a Human-Centred Design approach which begins throughout its development process by focusing on the user as a base, and considering the capabilities and limitations of the fundamental technology and features that the organisation has intended to propose.

2.5.4 Specifying and understanding usability requirements

Defining the usability requirements is considered as most difficult task in development and modernisation (re-designing) systems, due to it related to processing three activities: analysing the context of use, defining task scenarios that can be test, and specifying requirements for effectiveness, efficiency, and satisfaction for each scenario (Serco Ltd. ,2002). Initially, the usability requirements divided into four classifications (Robert et al, 2003,p30):

1- User requirements are capabilities that must be provided by the system. These requirements are often expressed as tasks or activites that must be supported by the system.

2- System requirements describes the capabilities of the system and the system itself

3- Constraints include decisions that have already been made, such as interactions with other system, development standards, and cost.

4-Non functional requirements include behavioural properties, such as performance, usability, and security, that the system must have.

2.6 User Interface (UI)

The User Interface (UI) is the part of a computer that people (users) can interact with through software, so they can see, hear, touch, and even talk to it.

Fundamentally, it includes two components: input and output. Input is about how users communicate their requirements to the computer through using some common input devices like the keyboard, mouse, trackball, and screen touch, while output is how the computer transfers the outcomes of its computations and requirements to the users. The most popular computer output mechanism is the display screen (Galitz,2002).

2.6.1 Design Principles of User Interface (UI)

For obtaining better knowledge of how to design a good UI, learning design principles and design guidelines would be a way to this end. A well designed UI should provide an easy, natural, and good interaction between a user and a system, and help users to perform their tasks effectively (Rogers et al.,2011). Furthermore, the proper UI should offer a mix of well-designed input and output mechanisms (Galitz, 2002) that are compatible with the needs, experience, skills, and anticipation of users (Sommerville, 2004). Fundamentally, the design principles of UI are generated from different theories that are based on knowledge, experience, and common sense (Rogers et al.,2011). It would assist the insight of designer and developer into interaction UI when they design for the user experience. Although it is commonly believed that there are no particular UI design principles that are applicable to all UI designs (Lauesen and Wesley,2005), Stone et al. (2005) suggested four crucial design principles, as follows:

a. Simplicity

The designers of UI should focus on the importance of producing the UI to be as simple to its users as possible and to be communicated clearly with language they can easily understand. For instance, the UI design should utilise action, icons, words, and UI controls that are familiar to the users in order for them to be able to achieve their goals effectively. This principle of simplicity in UI design can be constructed by relying on four factors; context of use, what is the main purpose of the system, what is the task that users claim to do using the system, and, keep to core functionality (Little,2009).

b. Structure

A design standard concentrates on the importance of organizing UI in meaningful and useful ways, through providing features that users expect to have on the UI or at least reflecting closely their understanding of the domain and of how the UI should be structured.

c. Consistency

This design principle significantly affects usability, because it emphasizes the importance of uniformity in appearance, placement, and behaviour within the UI, in order to generate a system which is easy to learn by users. Otherwise, it would be very confusing not only for an individual computer system, but across all the systems in an organization. Therefore, using such a customized style guide and having a similar look, arrangement, format and labels of most commands and menu items would be aids to gaining an appropriate level of consistency of UI. In addition, reuse of UI design techniques within and between systems can be beneficial to users for the reason that they need to learn and remember fewer things. For example, the consistent UI design across applications like the Microsoft Office software- PowerPoint, Word, Excel, Access, and so on -allows users to switch their knowledge and skills easily from one application to another.

d. Tolerance

The design principle of Tolerance focuses on the importance of providing a UI design that helps users to avoid making errors, or provides some resilience to these errors and allows the users to resolve them. It may consist of an undo feature, displaying error messages that provide the important information for recovery, and requiring confirmation of destructive action, which might enable users to understand what went wrong and how to correct the errors, and avoid committing them again.

2.6.2 The UI design process

The UI design is a part of the stages of the development process for a system (Hix and Hartson, 1993). It has been mentioned before that the evaluation of what has been already developed is very much at the core of interaction design, through applying a human centred design HCD approach by involving users in the design process and understanding what they do with this interaction design. For instance, evaluating the usability of an existing software system in terms of whether it is easy to use provides feedback that specific improvements must be done or that particular requirements have not yet been met.

The process of UI design includes four basic activities that are intended to inform one another and to be iterated (Rogers et al., 2011):

a. Determining users' requirements, goals, abilities, emotions, besides identifying what causes them to be frustrated or annoyed. This is through understanding what they do with the system;
b. Designing alternatives;
c. Prototyping;
d. Evaluating.

2.6.3 Prototyping as design solutions

Prototyping is deemed to be a major activity within the process of the system user interface (UI) design (Buchenau and Suri, 2000). Walker et al., (2002) stated that the prototypes have usually been tested by usability professionals through observing the users while they perform tasks, and look like the actual working of the system. It fundamentally can be defined as a limited representation of a final interactive system. It is employed to achieve various goals, such as, to clarify the scope of different solutions and users' requirements. According to (Rogers et al., 2011; Szekely,1995) a prototype design can be defined as a technique that involves the developing of a small scale version which represents a product design of a final software system that enables stakeholders to interact with it to explore its appropriateness. Also, prototyping can be defined as a group of design activities that enable the evaluators to test real users through allowing them to interact with it and performing typical tasks (Beaudouin-Lafon and Mackay, 2000; Buchenau and Suri , 2000; Sharp et al.,2007), in order to achieve a better understating of the UI design (Nielsen,1994). In addition, the prototyping technique of software system development consists of the production of at least one version of the UI design of the system, which represents the important features of the actual system, not all of them. According to Rubin and Chisnell (2008,p.31):*"When developing a prototype, one need not represent the entire functionality of the product. Rather, one need only show enough functionality to address the particular test objective*". Furthermore, the initial UI prototype can be defined as a simulation model of the original system, which is developed by relying either on an initial selection of functions or on users' requirements that have been recognized earlier (Carr and Vener,1997), which present concrete representations that can be compared.

Consequently, a prototype is a useful support in several aspects; i.e. in the early stages of system development, to present and discuss concepts with stakeholders and to test them, and to obtain initial users' needs; in the middle stages, to validate system specifications, while in the later stages to identify solutions to particular usability or design problems. What is more,, the prototypes would assist in answering the research questions, besides supporting designers in choosing between alternative designs (Rogers et al., 2011).
Although, there are various methods in the development of prototyping; however, the following is a summary of the basic process:

1. Figure out the primary users' needs and requirements for specific products or systems through utilising different tools such as focus groups, survey, interviews etc;
2. Develop a suggested prototype via appropriate techniques such as simple sketches, dialogue design, sorting cards, and navigational guidelines;
3. Evaluate the developed prototype using different methods like heuristic evaluation, testing users, walkthrough, etc.;
4. Use and evaluate the prototype with the aid of user tests, walkthroughs etc;
5. Revise the prototype using the evaluation stage again (c) until achieving the satisfactory one.

2.6.3.1 Types of Prototyping

According to different studies, there are many ways to categorize the methods for prototyping with various purposes (Jordan,1998; Martin and Gaver,2000; Righetti,2006). The processes of developing prototype designs can vary from paper-based, such as a storyboard, to complex computer-based forms, and as Walker et al., (2002) stated, "*prototypes more similar to the final product are 'high-fidelity' while those less similar are 'low-fidelity.'*", and "*choosing either paper or computer as the medium for a prototype has implications for the realism of the representation, the types of usability testing methods available, and the ability of users to participate in the design process"*.

Therefore, the most appropriate type of prototype to use relies on different problems For instance, one of the main classifications of prototyping is Low/Medium/High Fidelity. A **Low fidelity prototype** such as a 'paper prototype' is a simplified model of the UI, which does not look very much like the actual UI design. It is cheap, simple, and fast to draw, so it can be easy to find out the alternative designs and visions (Righetti, 2005; Sharp et al.,2007). The main purpose of conducting it is to link early design concepts to users in the cheapest and most efficient way (Nielsen-Norman Group, 2007), and to explain design ideas (Stone et al., 2005). A **Medium fidelity prototype** is often used after early design, for the purpose of providing more details and a fairly complete design, and usability validation (Engelberg and Seffah,2002). It can provide simulated interactive functionality and dynamic features, and full navigation in the design, but its objects are presented in schematic or approximate form. In addition, the users are able to interact with computer-based prototypes rather than paper-based ones, for example, the selection of an icon that controls proper functions for the iteration design (Peuple and Scane,2003). The benefit of applying this kind of prototype is when the design requires more than a low-fidelity prototype. A **High fidelity prototype** is an approach that uses materials that would be very similar to the final system in terms of form and feel. It is useful to assess technical and operational issues.

Another main classification is Horizontal / Vertical. A **Horizontal prototype** is a technique that models several common features of the system that user is expected to perform regularly but with few details. It is useful in terms of giving an overall perception of the final system from the users' perspective. What is more, mainly horizontal prototypes are expanded to ultimately develop into the final system in later stages, while a **Vertical prototype** is often employed to measure the feasibility of features described in a horizontal prototype. Such a technique, with high precision, mocks-up little features of the system but with rich detail in order to validate a concept at system level.

In related to the usability testing, the participants who are involved in the usability test research have mostly preferred to interact with computer-based prototypes (Leavitt and Shneiderman, 2006). In addition, Yee et al (2010) stated that the research devoted to developing UI is rapidly increasing in order to gain more advantages from technological development. Therefore, in this PhD thesis, the key findings and issues which were obtained from analysing the data on the evaluation of the VI system reported in the first study, have been incorporated into a prototype design using the appropriate technology with the required fidelity for the potential UI design. Furthermore, Preece et al. (2002) stated that HCD should involve both low and high fidelity prototyping, although most of the researchers have pointed out that most of the low fidelity prototypes would provide results equivalent to high fidelity prototypes. Thus, the approach of developing the prototype UI in the first study consists of a medium-fidelity prototype and high-fidelity prototype with iterative technique.

2.6.3.2 Prototyping tools and techniques

This sub-section describes a variety of tools that are used to develop an interface prototype, which can be either paper-based or computer-based. A UI prototype can be constructed with a number of different instruments, ranging from the use of paper and pencil to draw models of displays, to employing complex toolkits or programming languages prototyping tools. Low-fidelity to medium-fidelity prototyping methods are commonly used for designing the UI of a system in the earlier stages (Szekely, 1995), because they provide numerous advantages such as saving time, resources and cost by allowing the developers and designers to construct a UI faster and modify it easily to test it before starting the stage of development of the final design with the higher fidelity of prototype. It includes sketches, screen mock-ups, and storyboards. They can be developed either using a hand drawing, or a drawing package like Paint or Microsoft PowerPoint (Stone et al.,2005). Computer-based or software (high-fidelity) prototyping is used to assist in the rapid improvement of prototyping by using different tools such as Visio, HTML, and Adobe Dreamweaver applications. Although it may seem obvious that both of these approaches would be equally effective in discovering most usability issues, some researchers have pointed out that participants in usability testing usually prefer to interact with software prototyping. In recent times, use of these computer based techniques have achieved much popularity due to their being able to afford much valuable information for Human Computer Interaction (HCI) purposes (Porta, 2002; Turk and Kölsch, 2005).

2.6.4 The benefits of re-designing (re-engineering) the existing system

Until the last decade, software engineering (SE) projects focus exclusively on the defining and enhancing of the software development process, which lead to obtain a lot of crucial findings, such as a range of structured analysis, object oriented analysis, case environments, etc. This results is considered as a very useful in developing new systems, besides re-engineering the existing systems, which will maintained and documented through the systematic transformation into new improved form in order to realize quality in system capability, functionality, performance, or the capacity of a system for adaptive evolution at a lower cost.

Therefore, the decision regarding the re-designing (re-engineering) of the current (legacy) systems is an important factor that directed to several benefits and a greater return of investment (ROI), as follow (Atlantis Technologies, 2002):

a. **Lower costs.** For example, there is a number of US schemes suggests that re-engineering an existing system costs significantly less than new system development. Ulrich (2002),reports on a re-designing project that cost $12 million, compared to estimated redevelopment new system which costs of $50 million. What is more, since the one of main problems of existing systems is that it often runs on outdated hardware or other system platforms, that may not longer be available or they require extensive maintenance, re-designing the current system can be selectively targeted to rework problem area within existing system code which have a history record of maintenance issues in order to reduce future maintenance costs. (UniqueSoft LLC,2013).

b. **Lower risk**, re-engineering the current system is rely on progressive improvement of systems, rather than radical system replacement. Therefore, the risk either, of losing critical knowledge, which may be embedded in a legacy system, or of developing a system that does not meet the users' requirements and needs, is dramatically reduced. On the other hand, the re-engineering of the current system would enable the end users to increase and have a better understating of the critical parts of the current system due to this process of re-engineering translates the current system into series of highly updated forms and documentation.

C. **Incremental development** the incremental nature of re-engineering mans also that existing staff skills can improve as the system capabilities improves. Also, the end users will be able to gradually adapt to the reengineered as it is delivered in increments.

D. **Add new features easily**: Considering the re-designing of the current system lead to software with features that are well-defined in documentation and test cases and obvious separated. This in turn enables new features to be added without risk of interrupting a critical link within the current system code, for both enhancing speed of delivery and decreasing the cost of adding new features (UniqueSoft LLC, 2013).

2.7 Usability Evaluation (UEV)

Usability evaluation started to be an interest of research virtually with the beginning of the Human-Computer-Interaction domain (Harston et al.,2000). As mentioned earlier , usability is an essential quality factor in the improvement of successful software applications (Madan and Dubey,2012; Abran et al.,2003); moreover, defining it in terms of quality of use would open the path to evaluation, which could inform some issues that need to be either improved or redesigned to grant the enhancement to users in the workplace (Macleod,1994). Accordingly, the measurement of this quality attribute with the involvement of actual target users performing actual tasks within a specific software system, logically becomes a crucial technique for how efficiently and effectively the users have been able to use this system. Furthermore, Quesenbery (2004) stated that usability means evaluation since it relies on assessing users' point of view rather than crediting the capability of the designer. Mainly, the usability evaluation aims to find out the strengths and weaknesses of a system, and then to provide suggestions for developing the system usability (Hamborg et al.,2004). According to Nielsen (1993) it could save money, time and effort if it was considered through a proper process and at the right time. In fact, the importance of usability evaluation has started to rise rapidly over the last twenty years (Nielsen,1999). It is well accepted as a technique which can provide views on whether the system design and development meet the users' requirements (Rubin 1994). In addition, it affords a robust and rich basis for understanding and developing the design of user interaction with software systems (Hoegh et al.,2006). Preece et al. (1994) and Sharp et al.(2007) defined usability evaluation as a methodological process of gathering data with the purpose of gaining better insight about specific users and how they perform a specific task using a specific product under specified conditions. In a study by Rosson and Carroll (2002, p.227), UEV was defined as "*any analysis or empirical study of the usability of a prototype or system*").

2.7.1 Why Usability Evaluation?

The main aim of usability evaluation is to develop the usability of a product or software system for users (Abdul Rauf et al. 2010; Trivedi and Khanum,2012), through measuring its UI design to see if it is effective, efficient, satisfying, error tolerant, and learnable, or not, and then identifying the possible weaknesses and various problems which were encountered by the users during performance of the tasks, and that have affected its usability. Essentially, usability evaluation emphasises identifying to what extent a system supports users in meeting their goals (Buie and Murray, 2012). Furthermore, the usability evaluation of the UI would help to obtain a better understanding of the user experience with the system, and determine its flaws, to discover ways of developing it (Stone et al.,2005). Dix et al. (2007) mentioned that the goals of conducting the usability evaluation are to assess the extent of system functionality, assess the effect of the UI on the user, and identify specific problems. Also, Rosson and Carroll (2002) indicated that the main purpose of the usability evaluation is to afford feedback in software development, and highlight the problem and its causes, which help an iterative development process to correct this problem. As a consequence, usability evaluation can be considered as the whole, systematic process of deciding the usability of a specific software system (Hoegh et al.,2006).

ISO 13407 (1997) summarized the purpose of usability evaluation as follows:

- To present feedback to development of design;
- To evaluate both user and organizational goals if both have been targeted;
- To monitor the use of product or system in the long term.

In addition, Usability Evaluation has three chief objectives: to evaluate the extent and acceptability of the system’s functionality, to evaluate the users’ experience of the interaction, and to identify any specific problems with the system. Furthermore, for an appropriate activation of the general procedure and standards of the evaluation scheme, (Gediga et al., 2002,p.131) the following regulations need to be considered in the evaluation process:

a. The characteristics of a software product’s users such as, age, gender, experience or other more specific features;
b. The type of representative activities or tasks that the user will intend to perform using a software system;

c. The environment of the research study itself, varying from controlled laboratory conditions to largely unstructured filed studies;
d. The nature of the evaluation object, which can be a paper prototype, a software mock-up, a partially functional prototype or an accomplished system.

2.7.2 What are the goals of Usability Evaluation?

A usability evaluation can be conducted aiming to answer numerous different questions during a particular study or project. Each of these questions is a variation on the ultimate goal of becoming knowledgeable on how good the system is at assisting users in doing their work or performing specific tasks utilising their own experienced process. The following Table (2.1) presents the types of usability evaluation with some suggested questions adopted from the study by Quesenbery (2011) cited in Buie and Murray (2012,p.319).

Table 2.1 Type of Evaluation and Suggested Questions

Type of evaluation	Questions answered
Benchmarking	Can we measure the usability of the current system, so that we will know if the new system or redesign is an improvement?
Comparison	How do two systems used for similar tasks or functions compare?
User research	How do people interact with the current system? What can we learn about their goals, behaviour, or preferences that will help us design a new system or improve the current one?
Diagnostic, or formative	How well does a system in development work for users? What usability problems can we find early and fix before the design is complete? What is working well and should be kept as part of the design?
Measurement, or summative	Can we measure the overall usability of a system? A summative test at the end of a project can be used as the benchmark for a future redesign.
Compliance	Does this system meet the requirements of a standard or a regulation?

2.7.3 Factors for differentiating evaluation methods

There are some factors that should be taken into consideration when choosing evaluation methods, which would also provide a hint for classifying the variety of methods and then enabling developers to compare and select between them. According to a study by Dix et al. (2007,p.357), eight factors have been identified for choosing appropriate methods at the correct point in the development process, which are *"the stage in the cycle at which the evaluation is carried out, the style of* *evaluation, the level of subjectivity of the technique, the type of measures provided, the information provided, the immediacy of the response, the level of interface implied, and the resources required".*

a. The impact of the first factor on the determination of evaluation method depends on the stage in the design process the evaluation is required. The distinction is mainly between evaluation of a design stage and evaluation of implementation, where an artefact exists that could be anything from a paper prototype to a full implementation of target product or system, whilst the evaluation of a design before the implementation stage, seeks to afford information and feedback to inform the improvement of the physical artefact.

b. In regards to the second factor, it is about choosing the style of evaluation, either as laboratory or field studies. So, this factor would affect the choice of the appropriate methods, which would gain the evaluation some advantages and disadvantages. For example, the laboratory research creates suitable conditions for conducting a controlled experimentation and observation; however, it would lose some of the natural characteristics of the user's environment. In contrast, the field studies take the evaluator out into the user's environment and enable him to observe the system in action, but they do not provide the ability to exercise control over user activity.

c. The third factor is about the influence of the diversity of the evaluation methods according to their objectivity. For instance, the more subjective methods are significantly dependent on the evaluator's interpretation and knowledge, for example cognitive walkthrough and think aloud. On the other hand, the more objective method should generate repeatable results, such as controlled experiments.

d. The fourth factor is the type of measurement, qualitative or quantitative, and this is considered to be an important determination of the evaluation technique. Principally, the quantitative measure is usually determined from numbers, and it can be analysed by statistical techniques. In contrast, the qualitative is non-numeric, so it is hard to analyse; however, it could provide valuable information that cannot be gained from numbers. Furthermore, these types of measures are associated with the subjective or objective techniques. The subjective techniques usually tend to produce qualitative measures, whilst the objective techniques provide quantitative measures.

e. The fifth factor is about the level of information that is needed from an evaluation process. For instance, the required information at the design stage may vary from low-level information for the purpose of making a design decision, like distinguishing which font colour is most visible, to higher-level information which can be collected as a more general impression of the user's view of the system through employing questionnaire and interview techniques.

f. The impact of factor six on the determination of evaluation method is in respect to the immediacy of the user's response while using the system. For example, in some methods, like a think aloud protocol, the user's behaviour would be recorded directly at the same time as his interaction with the system, which means the evaluator will gather the response from the user immediately. However, this immediate technique would be sometimes a challenge due to the fact that the measurement process might affect and change the way that users use the system.

h. Factor seven is about the interference of the method which is linked to the immediacy of the user's response. So, the specific techniques which provide immediate measurements are apparent to the user during the evaluation process, which may generate the risk of influencing the way that the user behaves.

i. The last factor is about the influence of resources availability for considering the evaluation method. The resources to consider comprise tools, available representative participants, time, financial cost, experience and knowledge of evaluator and context. For instance, the source of participants, time and money might be limited, which would force the decision between two possible evaluation methods; however, the evaluator should bear in mind for his decision which evaluation method would mostly provide effective and useful information for the target system.

2.7.4 The classification of evaluation methods

The introduction of usability evaluation methods to support and develop the usability of a specific system or application has resulted in a range of different approaches (Hartson et al.,2000; Blandford and Green, 2009). Zins et al. (2004) mentioned that the categorisation of usability evaluation methods varies from one author to another. This has led to the lack of understanding, the limitations of each method, and the confusion in determining which method is more useful in particular ways for a specific product (ibid). However, some studies have tried to present a standardized set of usability metrics to be able to compare the different Usability Evaluation Methods (UEMs), and then to have a proper understanding of this variety of UEMs. For example, the study by Riihiaho (2000) divided the usability evaluation into two wide classifications: user testing and usability inspection. User testing comprised contextual enquiry, usability testing, walkthroughs, and pluralistic, whilst usability inspection comprised heuristic evolution, cognitive walkthroughs and GOMS (goals, operators, methods, and selection rules). Another published study by Harms and Schweibenz (2000) identified two methods: the heuristic evaluation and usability testing. In addition, it is essential to distinguish between the two basic approaches, formative and summative, which are relying on the evaluation goals (Hartson et al., 2000).

a. Formative evaluation

This type of evaluation takes place when the evaluation aims to identify the weaknesses of a software system through gathering information and users' feedback on the concepts or designs of a software system to provide suggestions for developing it, particularly when typically prototypes or re-designing are used for a certain system. Thus it is used during the system design development process at different points in the developing of prototypes or system versions, in order to improve a design. As was stated before, the purpose of utilising the formative evaluation is usually to figure out usability problems and issues of design that need be improved through setting the priorities, aiming generally to produce guidance on how to conduct the important changes to a design (Rosson and Carroll, 2002). According to Hix and Harston (1993) the formative evaluation aims to determine the drawbacks of a particular software system, and then come up with some recommendations and suggestions for more development.

b. Summative evaluation

In general, the summative evaluation is assessment done after development to evaluate a design, and to compare the level of usability evaluation of software design; however, it will not offer constructive information for altering the design of the system in a direct way (Hix and Harston,1993). Hartson et al. (2000) described the summative evaluation as a technique applying either to evaluating the efficiency of the final design or to comparing the usability of two alternative designs. Furthermore, Dix et al. (2007) stated that the summative evaluation is usually applied at the end of the development cycle in order to verify if the system has met the requirements sufficiently. Furthermore, Gediga et al. (2002) stated that the summative evaluation utilities to evaluate prototypes are needed for controlling the effect of design changes in comparison to a preceding version of the system.

In the following (Figure 2.1) the diagram explains the two types of evaluation, formative and summative evaluations, which simply shows that the formative evaluation is conducted during the design stage and the summative evaluation after the design stage.

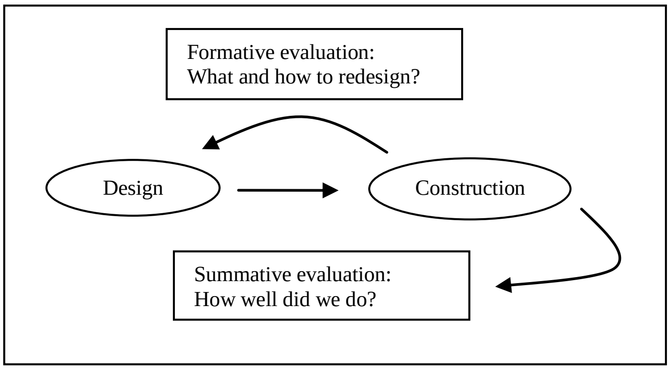


Figure 2.1 Formative and Summative Evaluations

In another common categorisation of evaluation, Fernandez et al., (2011), Otaiza et al.(2010) and Holzinger (2005) pointed out that the usability evaluation methods are generally divided into Usability Analytical (Inspection) methods and Empirical methods. In addition, according to the study conducted by Dillon (2000), it is worth making the distinction between usability evaluation methods depending on available resources, evaluator experience, ability, and the stage of improvement of the method under review. Accordingly, the usability evaluation applying analytical methods will be through expert analysis or expert based, whilst the usability evaluation conducting empirical methods will be through user participation or user based.

2.7.4.1 Analytical Methods (Expert Based Methods)

In this type of method for usability evaluation of a system or application it is examined by HCI or usability expert evaluators only, in order to figure out possible usability problems and then present judgement relying on their knowledge. After that they provide recommendations for fixing the problems and developing the usability of the system (Abran et al., 2003). Although it is useful for reshaping design, it would not replace usability testing with target users. Additionally, with respect to user based evaluation, which will be presented in the following section, there is a substantial attention to this type of evaluation because the results can be accessed faster, and it is likely to be cheaper than the user based methods (Dillon,2001). Furthermore, this type of approach can be called inspection methods, and it is more subjective, since it is heavily dependent upon the evaluator's skills (Nielsen and Mack,1994; Abran et al.,2003). It has two common methods, Heuristic evaluation, and Cognitive Walkthrough.

2.7.4.1.1 Heuristic evaluation

This is the most popular informal method in the field of usability evaluation (Nielsen, 1993). It involves usability experts who judge if each individual interactive element follows implemented usability principles (Holzinger,2005; Abran et al., 2003; Dag et al.,2001; Nielsen,1994), to make some improvement iterations shorter, and to perform more iterations in the development stages (Ferre et al.,2001;Dillon,2001).The Heuristic method presents a simple list of design guidelines which enables the evaluators to assess the interface justby following a standard path through a given task. Through a single evaluation process, the evaluator goes over the interface many times to inspect the different interactive elements and compares them with a list of standard usability basics. After all assessments have been completed, the usability experts are allowed to communicate and collectively describe their outcomes. The choice of implementing this method must be studied carefully because it reflects the specific system being examined, in particular for Web services where additional heuristics have become increasingly crucial. The study mentioned earlier conducted by Dag et al. (2001) was in Market-Driven Packaged Software Development where it is important for the industries to produce the best product in order to obtain customer satisfaction. This study has been approached through two methods, a questionnaire to obtain users' opinions about the software, and a heuristic evaluation to find the usability problems. They used a standard heuristic evaluation which was presented by Nielsen (1994). It is a set of ten guidelines focused on the most significant aspects of usability, such as "user control and freedom" and "flexibility and efficiency in use".

2.7.4.1.2 Cognitive Walkthrough

There are various types of Walkthrough technique in HCI (Lewis,1997). The cognitive walkthrough is a common branch that can be described as a task-oriented method through which the user interface expert or usability evaluator determines the exact sequence of correct task performance, and then estimates step by step user behaviour for a given task. Additionally, according to Wharton et al (1993), mentioned that cognitive walkthrough is it concerns evaluating a design for the learnability as usability aspect via exploration, which is many users' preferred method of learning a system (Fischer, 1991). Furthermore, in this method a more explicitly detailed review of a sequence of the user's actions is required, like the problem solving process at each stage through the dialogue, checking if the simulated user's goals and memory content can be estimated to lead to the subsequently accurate action (Nielson, 1994;Polson et al.,1992). Furthermore, this method focuses on cognitive issues, such as learnability, by examining the mental process required of the users (Holzinger, 2005; Dillon, 2001). Accordingly, the role of usability expert in this method must be to produce an informed estimate of the expected reactions of users, and explain why certain interface attributes are likely to cause users challenges.

2.7.4.2 Model Based Approaches, is another common method that has been reported in a study by Dillon (2001), it is called **Model based approaches** to usability evaluation. These are the least popular form of evaluation, but several methods have been planned which can accurately predict certain aspects of user performance with an interface such as time to task completion or difficulty of learning a task sequence. In such cases, the expert determines the exact sequence of behaviours a user will exhibit through detailed task analysis, applies an analytical model to this arrangement and calculates the index of usability**.** According to Dillon (2001)the most common model-based approach to assessing usability is the GOMS method of Card et al. (1983), which describes a cognitive psychology derived framework that breaks down user behaviour into a sequence of fundamental units. So, by using this method, any interface design can be analysed to enable the evaluator to estimate the user's time to complete a task.

2.7.4.3 Empirical methods (User Based Methods)

This test technique is considered as an empirical usability test where users are invited to participate directly to perform target tasks, and interact with specific interface designs (Bastien,2010). So, this type of method concerns situations where a sample of real target users is asked to attempt to perform a set of tasks in order to examine a system or application. After the users success at completing these tasks, they are often requested to provide data, through numerous approaches, such as surveys, which represent quantitative methods, and interviews, thinking aloud, field observation, focus groups, and observation, which represent qualitative methods. Dillon (2001) stated that conducting a testing method on an application or system with target users who are requested to perform a number of specific tasks is "*generally considered to yield the most reliable and valid estimate of a system or application's usability*". Therefore, the evaluators are able to assess the usability by actual users using software prototype design (Catarci et al.,2005), which in turn helps them to obtain information regarding potential usability problems and user preference in an interactive system (Abran et al.,2003; Freiberg and Baumeister,2008). According to Nielson (1994,p.165) user testing is considered to be the most essential usability method, or in another explanation is "*irreplaceable*". This is due to the significant direct information it provides regarding how real users perform tasks and what exact challenges are encountered when using a specific interface which is being tested. Furthermore, since in our case the system is already implemented, in other words is available, so user-based evaluation is often recommended (Costabile, 2000). In addition, the user-based method is considered to generate the most reliable and valid assessment of a system's usability (Dillon,2001). Thus it is the approach most commonly adopted for evaluating the usability of a system (Abran et al.,2003). It depends on the experimental design tradition of human factors, which mainly involves user testing. Therefore, in this way, the effectiveness, efficiency and satisfaction of the system can be measured, the problems and errors can be recognised, and re-design advice can be determined (Arban et al, 2003)).

2.7.4.2.1 Controlled experiments

The experimental method is commonly used in the HCI area for developing and assessing a system's user interface (McGuffin and Balakrishnan, 2005), finding answers to several critical research questions, such as those regarding technology adoption, modifying the user or task model, and evaluating various suggested design solutions (Lazar et al.,2010). Thus, the experimental data is considered to be the gold standard for usability evaluation (Rosson and Carroll,2002). More details regarding the usability testing will be provided in this chapter - section 2.9. Rubin and Chisnell (2008) have divided the experimental approach into four types of testing that would be fitted into different stages of the development cycle.

a. Exploratory

The exploratory study is a proper test to be done quite early in the system development cycle, when a system is in the initial phases of being developed and designed. The main aim of conducting the exploratory test is to measure the effectiveness of the preliminary design concept and how well it is represented in the system. Usually in this type of study a few participants are involved in the test using a prototype UI, and designer and developer will collect and analyse the data to identify problems encountered by those participants.

b. Assessment

This type of test is usually conducted either in the early or midway stages, as it is maybe the most simple and conventional usability test to conduct and design. The main goal of employing this assessment method is to expand the outcomes of the exploratory test by asking the users to perform specific tasks using a developed prototype of the target system. The qualitative and quantitative measures are used for gathering and analysing the data.

c. Validation

The validation testing usually takes place quite late in the development cycle, close to release of the system or product. The purpose of this test is to assess the system or product with a fully functional prototype by using actual tasks and scenarios to determine if a specific set of standards have been met, and if any modifications are required.

d. Comparison

This test can be conducted suitably at any point in the development life cycle. It can be called competitive usability testing to compare various designs, as candidates for developing the target system, or it can be employed to compare a new interface to current versions or to a similar system from competitors (Shneiderman and Plaisant,2010). The test can be an exploratory test where different designs are compared to determine which design is easier to use and learn, and what are the benefits and drawbacks of different designs. It is employed either informally as an exploratory test, or more formally by conducting a typical controlled experiment with different user groups.

2.7.4.2.2 Observational evaluation

The observational evaluation methods can be described by stating that the evaluator views and records users' interactions with a certain system.

a. Think Aloud

This method is considered to be a formative evaluation attempt to learn which detailed aspects of the interaction are good and how to develop the design (Nielsen, 1993). It helps to a proper implementation of formative evaluation in usability testing, through asking the participant to think aloud while they are using the system in the usability test, in order to verbalize his or her action or thoughts; then it would be possible to collect the observations. This method was used in a study conducted by Holzinger (2005), and he reported that thinking aloud may be the most significant usability engineering method (Nielsen, 1994).By verbalizing the users’ thoughts, it helps them to think through the design problem and then they provide suggestions for recovering from it (Rubin and Chisnell,2005). In addition, it enables the researchers to understand how they view the system, which in turn helps to recognize the users’ main misconceptions. There are some advantages of using this method, including disclosing why users do something, and providing a close estimation of how individuals use the system in practice, which could provide significant data from a small number of users. Moreover, due to the users being focused and concentrated, gathering the results is faster. On the other hand, disadvantages include the fact that this learning style is often perceived as distracting, unnatural, and straining by the users. In addition, this method is time consuming due to trailing the users being an important part of the preparation.

b. Observation techniques and note taking

This technique is considered to be the simplest and most common of all usability methods (Nielsen,1994;Holzinger,2005). According to Rogers et al. (2011) the observation is considered a useful technique for collecting data at any stage of developing a product. It is mainly concerned with observing the real users while they perform typical tasks on the interactive system in the workplace. Furthermore, it is considered to be the most reliable and accurate approach for gathering data about users (Costable, 2000). Besides, it enables the evaluator to recognise significant factors, such as social pressure, that might have a negative impact on user behaviour when they work with the interactive system in the field. The main advantage of utilising this technique in the laboratory is that since several users can perform the same task, so it is easier to repeat the process. Moreover, "*the observer could also be more objective because he was more of an outsider"* (Rogers et al.,2011, p.248).

Observational methods can be divided into two types; the first type is the direct method where the evaluator is physically in attendance during the task. This type also divides into two methods. a)- Unobtrusive observation where the evaluator can just concentrate on observing how the user is performing with the system, and avoid as much as possible interacting with the user by explaining any points or asking questions. b)- obtrusive observation where the evaluator is allowed to engage with the test user in discussion, so he can for example explain design decisions, ask questions, and answer the user's questions. The second is indirect, where the task is recorded on video by the evaluator to analyse it later. Direct observation is very expensive due to the evaluator needing to evaluate each user individually. Thus, reducing the number of observations is enough to generalize behavioural expectation. Ultimately the number of participant users relies on three factors: how many different demographics there are, budget, and time.

2.7.4.2.3 Query techniques (Interviews and Questionnaires)

Query techniques comprise asking the users directly about their experience of using the system under evaluation; this may be done face to face as an interview, or in writing as a questionnaire. It affords subjective information from the user, also in case the objective information is needed, the user's physical response to the system can be captured. According to (Amelia and Carvalho, 2001) stated that "*Surveys are employed to know user's opinions or to understand their preference about an existing or potential product through the use of interviews or questionnaire"*.

a. Interviews

This instrument is a commonly used tool where users, stakeholders and domain experts are asked questions by a practitioner in order to obtain information about their requirements regarding the system (Maguire,2001). The interview method can be divided into two parts, structured and semi structured. Structured interviews can be carried out when the respondents' variety of answers is well known, and there is a requirement to measure the strength of each single opinion (Macaulay, 1996). Semi structured interviews rely on a series of fixed questions giving a domain for the user to expand on their responses. It is considered as useful in cases where general issues are likely well understood, however the variety of responses to these issues is not fully known. Green (2001) presented in his study a method to evaluate usability called Scenario and Task Based Interviews (STBI). He stated that the introduction of a technology can transform the context of an interview to collect statistical data, contributing to the feeling that this is an essential event for the usability expert. In some cases, curiosity about the technology can draw more people to observe or participate in the interviews, so it can be easier to talk with them. However, in some rare cases, the term technology can inspire fear or anxiety. More details about utilizing the interview as a tool for data collection in this research will be presented in chapter three.

b. Questionnaire

There are numerous characteristics of usability that can be best researched by simply asking the users. Questionnaire is considered particularly as a common and proper tool for exploring issues that are linked with the users' subjective satisfaction and prospective concerns, which are difficult to evaluate objectively (Shneiderman and Plaisant, 2010; Nielsen,1993). Besides, it is a well established method for collecting demographic data (Rogers et al., 2011). Thus, a survey of existing users includes administering a set of written questions to a target sample of users, as it can be helpful to determine their needs and requirements, current work practices and attitudes regarding new system ideas. This method is useful for gathering quantitative data from a large sample of users about existing tasks or the current system (Preece et al., 1994). Using a survey method is considered to be one of the main approaches within user based methods. The quality of the information relies on the quality of the questions, and whether the questionnaire has been built properly (Ferre et al., 2001). Kiraakoswski (2000) mentioned that *"The biggest single advantage is that a usability questionnaire gives you* feedback from the point of view of the user. *If the questionnaire is reliable, and you have used it according to the instructions, then this feedback is a trustworthy sample of what you (will) get from your whole user population"*. Many studies concerning the field of evaluating usability have adopted this method. In a study conducted by Lewis (1993) to measure usability of an IBM application, the focus was on the application of psychometric methods to the improvement and evaluation of standard questionnaires to evaluate subjective usability. The subjective usability measures in this study were generally responses to Likert-Scale questionnaire items that measure user attitude regarding features such as ease of use, learnability, and interface likeability (Alty,1992). In another study published by Dag et al. (2001) for evaluating usability of a system in a large software development company, they employed both quantitative and qualitative methods by utilising two known usability evaluation methods, a questionnaire was conducted as one major tool in this study. In order to obtain end users' opinions concerning the system they used a commercially available questionnaire called the Software Usability Measurement Inventory (SUMI) (Kirakowski and Corbett, 1996). SUMI is "*a standard questionnaire specifically developed and validated to give an accurate indication of which areas of usability should be improved"*(Dag et al., 2001). It was tested in industry, and it is mentioned in ISO-9241 as a method of measuring user satisfaction, through responses to 50 statements to which each end user answers if he or she agrees, disagrees or is undecided.

The related questionnaires which are considered to be as tools for collecting data in this research will be presented in chapter three and in more details in related chapters, four and five.

b.1 Types of questions

There are different types of questions that can be included within the questionnaire for collecting both qualitative and quantitative data (Adams et al., 2008). It is crucial for the questionnaire's designer to distinguish between the types of questions that can comprise a questionnaire as these questions need various levels of interpretation by the reader, which in turn would have an influence on the level of attention needed by the participants for completing the questionnaire easily (ibid). There are two main types of questions, closed and open-ended questions. The closed questions have three subtypes of question:

- Simple factual type questions: for objective data and they require yes/no responses, e.g. ‘Do you have internet access at home?’
- Complex factual questions: requiring some interpretation or analysis, e.g. ‘How many times have you used this application today?’
- Opinion and attitudinal questions: directing the respondents' thoughts outwards and inwards, and requiring more alternatives and deeper concentration.

Open-ended questions require full concentration by participants (Adams et al., 2008; Griffiths, 2004). Overall, effective questionnaires include a mix of both closed and open questions (Stone et al.,2005).

2.8 Usability Attributes

Principally, it would be an ideal way to identify and measure system usability, in purposing to specify the features and attributes which are required for producing a usable system. According to Nielsen (1993), determining the usability attributes is considered the first step for achieving a successful HCD development. Additionally, the usability attributes are a conceptual aspect which specifies the significant area in order to determine the state of usability of an existing system. So, as the usability attributes have become usability requirements for quantifying its specifications, they would have influence on the development process which in turn would affect the design outcome directly (Bruno and Al-Qaimari,2004). Since the important of usability is widely identified, and it has a multidimensional aspect, several studies have developed different views on several key attributes of usability, e.g. Effectiveness, Efficiency, Learnability (easy to learn), Memorability, Error Tolerance, and Engaging (satisfaction) (Nielsen,1993; Dix et al.,1993; ISO 9241-11, 1998; Shneiderman,1998; Quesenbery,2004; Holzinger,2005). However, Abran et al. (2003) claimed that the determination of the usability characteristics and attributes isin fact challenged due to the nature of these characteristics and attributes and relies on the context in which the system is utilized. The following Table (2.2) presents the main usability attributes that have been found in previous research.

Table 2.2 Similarity between the attributes of system Usability

Source	Easy to Learn (Learnability)	Efficiency	Effectiveness	Error tolerant	Engaging (Satisfaction)	Control	Helpfulness	Memorability
Kirakowski and Corbett(1992)	√	√	√			√	√	
Dix et al (1993)	√	√	√		√			
Nielsen (1993)	√	√		√	√			√
ISO 9241-11 (1998)		√	√		√			
Quesenbery (2001)	√	√	√	√	√			
Oulanov and Pajarillo (2001)	√		√			√	√	
Shneiderman (1998)	√	√		√	√			
Holzinger (2005)	√	√		√	√	√		

From (Table 2.2) above, it can be seen that five usability attributes have the highest frequency in each study undertaken by researchers, so they have been considered and selected based on this fact, and because they are appropriate and important to evaluating such a software system. These attributes are Effectiveness, Efficiency, Easy to Learn (Learnability), Error Tolerance, and Engaging (satisfaction), while Memorability has less frequency than the other attributes.

In addition, the recent study by Quesenbery (2012) indicated that identifying the usability in respect of these five dimensions is useful for evaluating systems used in government, since they can be utilized equally for evaluating systems in any other, different, context e.g. business administration in the private sector.

a. Effectiveness

Measuring the attribute of effectiveness is linked to the goals or sub-goals of using the system, and it addresses the issue of ensuring that the software system is useful and helps the users to complete specific tasks accurately. It can be defined as the completeness and accuracy with which the system users can achieve certain goals. Furthermore, the attribute of effectiveness consists of two components: the quantity and quality measures. The quantity measure can be defined as the amount of the task completed by a user, and represented in the results of the task, while the quality measure is about measuring the degree to which the results achieve the task goals. Furthermore, because these two components are both measured as percentages, the task effectiveness can be calculated as a percentage (Bevan,1995). In usability testing, to assess this attribute, for most tasks, we would be able to utilise a pass/fail measure of whether the user was able to complete the task successfully. Accordingly, calculating the percentage of users who could manage to complete the task successfully becomes a measure of the attribute of system usability (Fidgeon,2011).

b. Efficiency

According to Quesenbery (2004) the efficiency can be described as the speed of performing accurately by users. Another explanation is that, once those users have learned how to use the system, how quickly can they perform the intended tasks (Neilsen, 2003), and have the ability to obtain a high level of productivity (Holzinger,2005). A proper technique for measuring this attribute would be through calculating the time taken by users to perform the test tasks(Fidgeon,2011).

c. Error tolerance

The error tolerance attribute is about how well the software product would support the user by either preventing errors occurring or making errors easily rectifiable, which might let users make fewer errors during use of the system.

d. Easy to learn (Learnability)

The learnability attribute as we can call it (easy to learn), is about how easyit is for the users to finish a certain task at the first time of interacting with the system design (Nielsen,2012), so they can rapidly start performing the tasks with the system (Holzinger, 2005). Furthermore, it should be kept in mind that a software product might be used for just one time, once in a while, or on a daily basis. Also, it might assist users in a task which could be easy or complex and even the user himself may be an expert or a novice in using the system. Therefore, the system UI interface must be remembered or relearned when it is explored over time (ibid).

e. Engaging (Satisfaction)

This attribute of satisfaction or engaging (Quesenbery,2004) can be simply defined as how pleasant and satisfying the system design is to use. Accordingly, the software system has an emotional influence on users that helps them to engage with the UI, and then to work on the system with confidence. So, it reflects their personal opinions of systems as an important part of usability evolution. This is done through various aspects such as the visual displaying of the required information in a way that is easy to read. The general way of measuring users' satisfaction is using a questionnaire (Fidgon,2011), which is usually conducted after users have interacted with a system and completed actual tasks which were sufficient to have enough experience of using it to provide useful feedback on it.

f. Memorability

The memorability attribute concerns how easy it is for users to remember performing targeted tasks using a particular system after a period of not using it. In other words, a system would have high memorability when its user returns to use it and does not need to relearn how to use it again (Thomas, no date). However, in regards to the systems which are used on a daily basis by the employees or users of this system, this attribute would have an insignificant influence due to the reason mentioned before.

2.8.1 Demographic factors' influence on user performance

There are quite possibly factors that might be consistent predictors of user performance, such as users' previous Experience of using either a system or task, and in addition, their Age, Educational level, and Gender. According to Aziz and Kamaludin (2014, p.137) *"the perception of usability is also influenced by user profile such as gender, age, and educational level and technology skills"* .

2.8.1.1 Age

Age-specific usability requirements for software system interfaces are inspected from various disciplinary and methodological viewpoints, such as psychology, ergonomics, computer science, economic studies, and engineering (Schneider et al.,2008; Ziefle and Bay, 2008). As various information system (IS)studies have discovered, age is an important factor that has effects on several aspects ofsystems use,behavioural and cognitive (Al-Shafi and Weerakkody, 2010; Dwivedi and Lal, 2007a; Bigne, et al., 2005). Furthermore, some research has been conducted to find possible age related differences in skill, between younger and older, in using ICT, e.g. computer and Internet. In regards to computer usage, Wirtz et al., (2009) reported that people who were aged 55 and over hada low rate of computer usage compared to younger age groups, but now this rate is continuously increasing. Furthermore, the using of new ICTs is becoming important in the daily lives of older adults. In addition, Freudenthal (2004) demonstrated a different aspect of the age factor, in relation to the capacity to retrieve information.

2.8.1.2 Educational level

It is commonly known that a person who has a lower level of education, and associated cognitive abilities, is likely to display less proficiency in using the ICT then one who has a good level of education with more efficient abilities and skills to do so (Al-maskari and Sanderson, 2011; Van Deursen and Van Dijk, 2009; Johnson, 2008). However, Kim (2001) indicated a good point, which is that it is obvious that the differences between a particular group of people might reduce as some gain more experience with the use of ICT (Kim, 2001). In another angle on using the advanced technology, Burgess (1986, cited in Dwivedi and Lal,2007) believed that people who have obtained educational qualifications are more likely to have a better profession, and therefore to adopt more new innovations of ICT.

2.8.1.3 Gender

The factor of Gender differences is one of the relevant demographic characteristics. It is one of the overall cultural differences between human beings. According to Mayhew and Media, (2015) *"Culture is defined as a set of values, practices, traditions or beliefs a group shares, whether due to age, race or ethnicity, religion or gender. Other factors that contribute to workplace diversity and cultural differences in the workplace are differences attributable to work styles, education or disability"*. Additionally, it has been considered in several studies which aimed to investigate whether or not there are differences between the genders (men and women) in regard to technology usage. A number of researchers have provided different academic examinations of the roles adopted by the two genders in using and implementing technology (Al-Shafi and Weerakkody, 2010; Jackson et al., 2001).

In the context of this research study, the factor of Gender has been excluded, due to the fact that all the participants were male. So, the effect of the female gender was not considered, and it has been one limitation of this study.

2.9 Usability testing

The basis of usability testing arose in the classical experimental approach (Rubin and Chisnell, 2008). It is considered to be one of the most useful research methodsfor evaluating product design (Lewis,2006), because it provides direct input and information on how representative users (participants) interact with the system UI, and illustration of certain challenges that they faced during their interaction (Nielsen and Mack, 1994). According to Barnum(2002), the objectives of the usability test link to the measurable goals of the test. Additionally, conducting usability testing in a laboratory, or a temporarily controlled site in an informal laboratory, would enable the practitioners to have control over what the participants need to perform using a particular system, and to control environmental and social influences that might impact on the participants' performance (Rogers et al.,2011). Furthermore, Spencer (2004) and Rogers et al. (2011) mentioned that the usability testing is the main method for gaining knowledge of whether a system is usable or not. In addition, it is applied in a Human-Centred Design (HCD) approach at different stages of the development and design process in order to evaluate system design (Nielsen,1993; Preece,1993; Rubin,1994). However, chiefly it has been most effective when it has been carried out as a part of the system development process (Rubin,1994). Burmeister (2001) stated that the importance of usability testing has grown as a part of the development of quality software products, in as muchas it contributes to a proper Human-Centred Design (HCD) (Rubin and Chisnell, 1994).

The term usability testing is commonly referred to as a technique used to evaluate a product or system with regard to in what degree it meets specific usability measures through testing targeted participants (Rubin and Chisnell, 2008). Ferre et al. (2001) explain the usability testing as the experimental activity of conducting laboratory usability tests on a specific group of users and recording the outcomes for analysis purposes. Microsoft Corporation (2000) describes the usability testing as the gold standard for enabling the practitioners to determine whether the design of a system meets the requirements of its intended users, which in turn can enable them to perform their tasks more productively. Furthermore, Wichansky (2000) defined usability testing as "*any of those techniques in which users interact systematically with a product or system under controlled conditions, to perform a goal-oriented task in an applied scenario, and some behavioural data are collected*"

2.9.1 Principles of usability test Methodology

There are five essential aspects that need to be taken into account when conducting the usability test. Firstly, a question or hypothesis must be formulated based on the expectation of what is going to happen when the usability test is conducted. For instance, "version 1 of the design will support improving the speed and error rate of experienced users more than this will be supported by version 2". It is important that the questions be specified as precisely as possible. Secondly, determine the characteristics of the target population in order to select a representative random sample of particular participants who must be assigned to experimental conditions. The third aspect is regarding the experimental controls which need to be tightened and employed within the test procedures.In addition, "the amount of interaction with the test moderator must be controlled". The fourth aspect concerns the creation of a control group in order to validate the results. Lastly, the sample size needs to be sufficient to measure statistically significant differences between the groups (Lazar et al., 2010; Rubin and Chisnell, 2008).

2.9.2 The usability test plan

According to Rubin and Chisnell (2008) the test plan is considered the foundation for the whole test, as it provides answers to the how, when, where, who, why, and what of the usability test. So, for example, the test plan explains how the procedure will go regarding testing the particular system, and set the stage for all that is to follow. In addition, since the test plan purposes to describe the required internal and external resources, and define clearly what will happen and when, this would make it easier to anticipate what is required to be accomplished with the test. Coolican (2004) mentioned that there are some features that need to be taken into account when planning the experiment, i.e. the sample of participants that will perform the test tasks, the design of the experiment, and lastly the statistical tests that will be utilized to analyse the collected data from the experiment. Furthermore, Rubin and Chisnell (2008,p.67) provided the parts of the test plan which will be varied according to the type of test and the level of formality which is required for conducting the usability test, as follows:

- Purpose, goals, and objectives of the test;
- Research questions;
- Participant characteristics;
- Method (test design);
- Task list;
- Test environment, equipment, and logistics;
- Test moderator role;
- Data to be collected and evaluation measures;
- Report contents and presentation.

These highlighted elements of the experimental plan need to be considered when planning the test by the experimenters (Robson, 2002). In the work reported in this thesis, these parts will be discussed in detail in the relevant studies, in chapters five and six.

3.9.3 Experiment design

Having a well-designed experiment is important to obtain reliable and significant results (Lazar et al.,2010), what is more it could save time and effort , besides answering the research questions clearly (Tulls and Albert, 2008). The main focus in experiment design is to decide which participants are to be recruited for particular conditions in an experiment (Rogers et al., 2011). There are two main types of experimental design: between-subjects and within-subjects (Lazar et al., 2010; Rubin and Chisnell, 2008; Robson, 2002). A between-subjects design is called an independent group design, because each part of the software product is tested by a particular set of participants (Lazar et al., 2010; Rubin and Chisnell,2008). There is a disadvantage of employing this design, which is that more participants are required for the test, so the differences between groups and users could have a negative impact on the final results (Dix et al., 2004). Another type of experiment design is a within-subject model in which each group of participants attends the test in more than one condition, or in other words they perform all tasks in the test (Lazar et al., 2010; Rubin and Chisnell, 2008). However, there is a challenge of transfer of learning influence that might arise in this type of experiment design; when each participant switches from one condition to another; he could have gained some learning on how to perform the tasks. In order to solve this problem, a technique of "counterbalancing" should be applied, whereby the order of tasks is mixed up or balanced out. This would mitigate the effects of learning transfer (Rogers et al., 2011; Rubin and Chisnell, 2008; Robson, 2002).

2.9.4 Usability testing Metrics

Different usability studies have provided varying points of view on how to define and measure usability (Folmer and Bosch, 2004) and user experience (Bevan,2008). Sauro (2011) stated that *"there isn't a usability thermometer to tell you how usable your software or website is"*. Furthermore, Finstad (2010) demonstrated that measuring and tracking usability continues to create obstacles for organisations that intended to enhance their users' experience. However, Nielsen (2001) pointed out that the usability measures are most fundamental when usability is defined as a quality of use; he had already mentioned (1994) that the technique for measuring the quality of use is one of the two techniques for improving the usability of products besides the technique for diagnosing usability problems. The IEEE[4] metrics standard defined a usability software metric as *"a function whose inputs are software data and whose output is a single numerical value that can be interpreted as the degree to which the software possesses a given attribute that affects its quality"*. (IEEE,1998, cited in Seffah et al., 2001). Additionally, considering and measuring usability as quality of use would generate several advantages (Bevan,1995), such as: there will be some incentive to allocate resources to the design process for system usability, since usability is an objective measurement for the requirements specification; obtaining results which would help in judging how much more work in usability (if any) is left to be done in order to achieve the project objectives; and also the means can be established benchmarks and making comparisons with previous versions of a system, or with competitive designs, and with alternative designs. Moreover, increasing numbers

[4]The IEEE (Institute of Electrical and Electronics Engineers) is one of the largest technical professional societies in the world that is prompting the development and application of electrotechnology and allied sciences for the benefit of humanity, and the advancement of the profession.

of these published studies have addressed the challenges of a way to measure the system usability in respect to users' performance of a selected set of tasks. Fundamentally, usability metrics are utilised for the purpose of measuring the quantitative usability aspects of the existing system (Dix et al., 2003), and theycan be divided into two sets of criteria: **Objective operational criteria,** which are usually directly observable data on user behaviour while using the interface of the application system, and **Subjective operational criteria**, which concern the point of view of the user within the consideration of the usability of the system UI. In addition, Bevan (1995a) mentioned the following three views that are agreed commonly for measuring the usability:

a. In the product view, the ergonomic perspective of the product will be considered within the evaluating of usability;

b. In the user-oriented view, the usability will be evaluated based on the user's mental effort and attitude perspectives;

c. In the user performance view, the usability can be evaluated by measuring the users' interaction with products and systems.

2.9.5 MUSiC Methods

The MUSiC methods were developed by the European MUSiC (Metrics for Usability Standards in Computing) scheme to produce valid and reliable means of determining and measuring usability, and then to provide supportive feedback based on 87 papers out of a survey of 500 papers (Lewis, 2006). It has developed a package of evaluation methods based on users' perspectives, which enable the evaluator to be selective in adopting methods individually or in a mixture for measuring the important aspects which have been considered by the developer. Furthermore, MUSic methods enable measurement of the rate at which people can learn to use a system, and they also provide valid and reliable means of measuring usability, as well as the fact that the methods grant diagnostic information which enables the design to be modified to improve usability.

In addition, there are some common metrics that were identified by different researchers. For example, according to Nielsen (2001) the following have been given.

2.9.5.1 Time on task

This metric concerns the amount of calculation time that it takes a participant to finish a selected test task.

2.9.5.2 Successful task completion rate (whether users can perform the tasks at all)

In this metric, each participant will be given each scenario that requires him to gather particular data that would be used in a representative task. Thus, the scenario is successfully completed when the participant is able to find an answer to finish the given tasks.

2.9.5.3 Error rate

This metric concerns the number of errors that occurred during a participant's performance of the test tasks, and it includes three levels of committing errors, critical and non-critical errors.

2.9.5.4 Users' satisfaction

Measurement of usability by utilising the metric of satisfaction will be through asking the participants to fill out a questionnaire

The following figure 2.2 presents the model of using the average of the four standardized metrics proposed by Sauro and Kndlund (2005), which sufficiently simulate the relationship that is built from the principal components analysis, and represents "*the equal weighting of the standardized component matrices to summarize the construct of usability"*. Furthermore, this standardised metric can be used for quantitative analysis besides the comparisons across test tasks and studies, and generally to evaluate the impact of design changes.

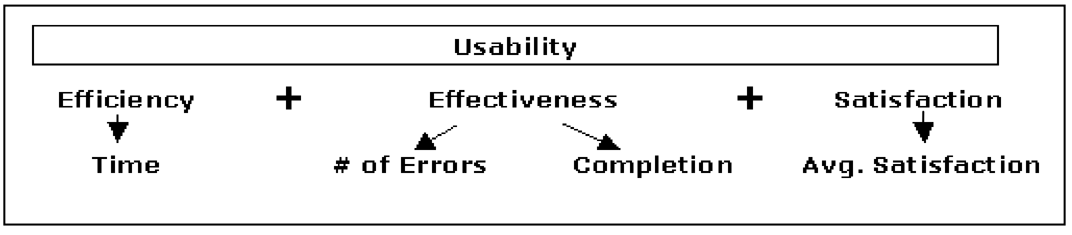


Figure 2.2 Quantitative Model of Usability by Sauro and Kindlund (2005)

2.10 Main Sources and Related Work

This sub-section presents the literature in relation to the usability evaluation studies in order to acquire the best knowledge, lessons learned, and possible understanding that would be beneficial to produce a comprehensive insight of the process leading to systems with better usability. In addition, it lays out some of the groundwork that has been leading to more human-centered design, and explain the justification of each stage.

Moreover, some case studies will be presented, described, and followed up by providing related versions with guidance on which enables us to utilize the valuable lessons learnt from previous experience to resolve some current problems.

Although the main aim of the study is to evaluate the usability of the current internal system in a governmental organisation which provides such a service to the public, due to the lack of literature concerning the government and usability together (Buie and Murray, 2012), this part of the study includes the usability evaluation of the internal software systems, web based systems, and system applications, in both profit and non-profit organisations, with the purpose of gaining more knowledge, as we mentioned earlier. Keeping in mind, considering and proposing only the relevant literature, which is consistent with the characteristics of this research.

2.10.1 Human-Centred Design (HCD)

It is widely recognised that the requirement for involving the end users actively within system development has increased (Ferre and Medinilla,2007), as it is a critical factor for generating a system that is more efficient, effective, and safe (Kahraman,2010). Furthermore, this in turn lets the users' needs become increasingly the focus of design research (Zhang and Dong, 2008). Although there are several advantages obtained from developing the system generally, such as the improvement of the users' performance and increasing productivity, human-computer interactions are changing logically, and accordingly complex computerization is being folded into existing systems (Daouk and Leveson,2001). Furthermore, there are many unsuccessfully designed and unusable systems, which they are likely to be *"under used, misused or fall into disuse with frustrated users* *using the system, and harmful to the reputation of the company which developed and supplied it"* (Maguire,2001, p.587)

This leads to introducing a new source of errors, risks, and accidents through human intervention and limitations. Therefore, the requirement to decrease these problems and issues has led to the need to apply the concept of Human-Centred Design (HCD), which is extensively confirmed as a major factor for leading to improvement of ease of use, performance, and to generating a well-designed interface through considering for imporvment of its users' experince.. (Costabile,2000). Thus, HCD evaluation to be complete is required to identify three aspects: who will use the system (representative users), to do what (representative task), and how (improving a procedure for recognizing the problems that the users might face while using the system (Scholtzk,2004).

2.10.1.1 What is a Human-Centred Design (HCD)?

There are a number of studies that have proposed several definitions of HCD. For example, the ISO 13407 standard defines HCD as "*multi-disciplinary activity, which incorporates human factors and ergonomics knowledge and techniques to enhance effectiveness and productivity, while improving human working conditions*" (ISO-13407,1999), while ISO 9241-210 standard describes HCD as "*an approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics and usability knowledge and techniques*". Furthermore, Zhang and Dong (2008) pointed out that HCD is additionally known under different designations as "user-centred design (UCD)", "people-centred design", and "user/client-oriented design". So, the term Human-Centred Design can be described as a contemporary design approach in which the users have a significant impact on the design process (Abras et al., 2004). It is focused on putting the users' viewpoint at the centre of the design process; conducting testing and evaluation with those target users, and designing iteratively (Preece et al., 1994). Furthermore, in this process, the developer designs products and services for a specific purpose based on the user's request for operations and tasks (Rubin, 1994, cited in Kahraman, 2010). In another sense, Human-Centred Design is "*the HCI approach to the development process, and it has traditionally introduced* *uncertainty when labelling itself as iterative*" (Ferre and Medinilla, 2007, p.73). Gill (1991) describes HCD as an innovative technology which considers human requests, skill, creativity, and potentiality as the core of the activities of the technological system (Gill, 1991, cited in Gasson, 2003). Accordingly, Zhang and Dong (2008) propose that the characteristics of HCD in the domain of product design are as follows:

a. The central place of human beings;
b. Understanding people holistically;
c. Multi-disciplinary collaboration;
d. Involving users throughout the design process;
e. Making products or services useful, usable, and desirable.

2.10.1.2 Why Human-Centred Design?

There are a number of benefits that can be gained from employing an HCD approach; according to Costable (2000) they are mainly associated with *"completeness of system functionality, repair effort saving, as well as users' satisfaction"*. In addition, Abras et al. (2004) mention that the major advantage of the HCD approach is about understanding precisely several factors that have an impact on the use of computer technology, which arises from including users in each stage of the design and evaluation of the system, such as users' psychology, and organizational, social, and ergonomic factors. This involvement of users would help ultimately in the development of systems that are more effective, efficient and safe, and assist in enabling the designers to manage the users' expectations about this system, as at an early stage the users' feedback and suggestions have been considered during the process, which leads to achieving higher users' satisfaction (Preece et al.,2002). Thus, in order to prevent the usability problems occurring, the designers need to consider employing a human-centred approach with their design project (Norman, 1998; Shneiderman, 1998). In addition,Singari, (2006), Hirasawa et al. (2010) and Usability.gov (no date) reported other possible benefits such as decreased cost of technical support and training, increased productivity as the system becomes easy to use and satisfies its users, and reduced time and costs of the development process of the system. Although some possible advantages of implementing an HCD approach have been stated previously, and how involving the users' requirements analysis and specification would make the system successful, yet there are many government systems from all over the world that can be described as a failure in these terms, having experienced the provision of poor requirement analysis (Sutcliffe and Gulliksen, 2012). This has in turn led to these systems that do not fulfil the goals of their users, who let these systems be assigned to a faulty functioning requirements specification process . For instance, a software system for managing national patient records for the health sector in Sweden has a negative reputation for not performing its users' requirements analysis. Similarly in the UK, the London Ambulance Service system has become an obvious case of government systems that do not fulfil their users' needs, which has led to the system becoming more sophisticated (Sutcliffe and Gulliksen, 2012).

2.10.2 HCD Development Process

According to the ISO 13407 standard on HCD (ISO,1999), there are five important processes that should be undertaken in order to incorporate usability requirements into the software development process, as follows :

1. Plan the human-centred design process;
2. Understand and specify the context of use;
3. Specify the user and organisational requirements;
4. Produce designs and prototypes;
5. Perform user-based assessment.

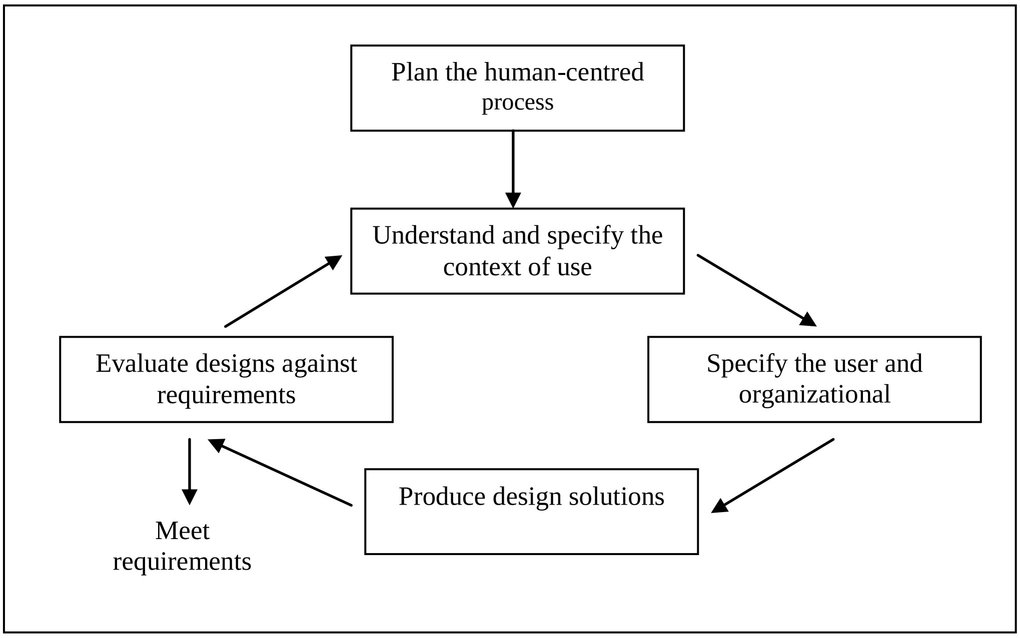


Figure 2.1 Key human-centred design activities (ISO 13407,1999)

Figure (2.1) shows that after the initial planning stage, the remaining four stages are carried out in an iterative way, as represented here, and the cycle is repeated until the particular usability objectives have been achieved.

1- Plan the human-centered design process. HCD plays a crucial role in a project by reducing the risk of system failure through maintaining the effective flow of information about users to all associated parts of a project team. Thus, in order for this application of a HCD approach to be successful, it must be properly planned and managed throughout all parts of the system development process. For example, engaging usability expertise in some specific parts but not others will be inadequate. In this stage it is important to ensure full integration of the HCD activities as part of the system strategy for the whole of the project (Earthy et al., 2001, cited in Maguire, 2001). In the same study Maguire (2001) proposed that this stage should be concerned with collecting high-level information regarding the following aspects:

a. The purpose of the system development, and the overall objectives;

b. The target users, their experience and capabilities, and the selected tasks that the users intend to perform;

c. The functionality required to assist the users;

d. How and why the system will be used;

e. The usability goals;

f. Guidelines that might be utilized;

g. User support;

h. Primary design ideas.

These aspects were embodied as an outcome of the targeted VI system and in the research background, related work, and discussion with other domains of identified stakeholders from experts to users. It is apparent that the literature review helped in the discovery of the challenges and weakness areas in the current VI system, besides the potential usability advantages that can be obtained through conducting the usability evaluation and how it works to obtain a usable internal software system. Furthermore, the literature presented a demonstration of the functionality of the VI system, including useful guidelines.

2- Understand and specify the context of use; when a system or product is developed, it will be utilized by end users with certain characteristics, as they will have certain objectives and tasks. The system itself will be used within a certain variety of technical, physical and social or organizational conditions that will affect its use (Maguire, 2001). The quality of use of a system, including usability and user’s wellness, are based on having a superior understanding of the context of use of the system. For instance, the design of a cash machine (or ATM) will be more usable if it is designed to be used in various conditions, night or day, by all people whether healthy or disabled. The context of use aspects for this thesis were discussed throughout the planning stage of the design process, as background information to which appropriate design techniques and evaluation methods should be applied. Moreover, at this stage we should keep in mind that the target system is already implemented within the government context.

3- Specify the user and organizational requirements; it is widely believed that the success of a software development programme is based on how well this activity is carried out, because the requirements elicitation and analysis is the most decisive part of software development. In addition, there are general methods that can be used to support user and organizational requirements specification, such as, stakeholder identification and analysis, the user cost-benefit analysis method, user requirement interviews, focus groups, scenarios of use, exiting system or competitor analysis, task or function mapping, allocation of function, user, usability and organization requirements, and comparison of all methods for specifying user and organizational requirements. Therefore, in order to be certain that the suggested system design would assist the employees as the users of the current system; the requirements were collected from employing two main sources. Semi-structured interviews with five participants from among those employees in the government department were conducted in the first study. Besides this, qualitative responses through the participants' comments on the statements or in the section asking what they like or dislike, and why, within the questionnaire were used. The highlighted themes from analyzing these qualitative methods were utilized to form the foundation for developing the suggested prototype design as a solution of VI system issues. Furthermore, the usability requirements were initially based on the metrics mentioned in the ISO 9241-11 definition, as improvements in effectiveness, efficiency and subjective satisfaction.

Thus, effectiveness refers to how good the target system is at meeting what it is supposed to do, while efficiency refers to the way a system assists users in performing their intended tasks, and the satisfaction or engagement refers to the users' viewpoints on how easy the system is to use.

4- Produce design solutions and prototypes; there are many ways that the design solutions can arise, from copying and development by logical progression from previous design, through to innovative creativity. All design ideas may consist of a series of user interface screens and a partial database allowing possible users to interact with the future design., initially using prototypes, followed later by more sophisticated prototypes. In addition, Preece et al. (1994) stated that there are a variety of alternative designs for any system that could meet the system's specification, while the role of the designer is to find these alternative designs that could make the system's goals possible. The solutions and ideas designs start to be formed by utilizing low (or medium) fidelity prototypes, and then following these with high fidelity prototypes, as Preece et al. (2002) stated that HCD should comprise both low and high fidelity prototyping.

5- Carry out user-based assessment; and designs should be evaluated throughout the development process. This is a very crucial activity within the system development lifecycle, which can be through confirming how user and organizational goals have been met so far as well as providing further information for refining the design. Thus, after the suggested prototype has been developed as a solution, the target users can be involved in testing the proposed prototype by conducting the experimental approach for validating the suggested ideas within the simulation design. Carrying out re-design of this developed prototype and applying further usability testing is an important feature of iterative HCD, which generally allows developers to evaluate their ideas and collect feedback from the users. The two main reasons for conducting usability evaluation were picked in line with the objectives of this thesis as follows (Preece et al., 1994):

a. To improve the system usability as part of the development process (by identifying and fixing usability problems), and testing if it has reached the usability target: “formative testing”;

b. To compare different designs and ensure that the system or suggested prototype can be used by the users successfully: “summative testing”.

After the stages of HCD have been explained and highlighted and the related aspects of the research conducted for this thesis, it is necessary to ensure that a system is designed with high levels of usability. Gould et al. (1987) indicated that there are three main principles required to be considered in the development process of a system, which are: to involve users as early and as much as possible during the design process, so that users' cognitive, social and attitudinal characteristics are understood and accommodated; to measure performance and attitude by using interfaces and simulations of the system; and to design iteratively, so that testing and evaluation can be conducted to make sure that the design meets the user requirements. The following Table (3.1) presents a summary of the methods and activities that were used for the research conducted for this thesis, which can support each stage of the HCD process. Fundamentally, it was adopted from Preece et al. (2002), as he suggested ways to involve users in the design process and the development of a system or prototype.

Technique	**Purpose**	**Stage of the Design Cycle**
Background questionnaires and interviews (survey of existing users)	Collecting data related to the needs and expectations of users; evaluation of design alternatives, prototypes and the final artefact .	At the beginning of the design project (brainstorming)
Sequence of work interviews and questionnaires	Collecting data related to the sequence of work to be performed with the artefact	Early in the design cycle (design guidelines and standards)
On-site observation	Collecting information concerning the environment in which the artefact will be used.	Early in the design cycle
User observation, think aloud protocol and simulations (software prototyping- medium to high fidelity)	Evaluation of alternative designs and gaining additional information about user needs and expectations; prototype evaluation.	Early and mid-point in the design cycle
Usability testing	Collecting quantitative data related to measurable usability criteria.	Final stage of the design cycle
Post-study questionnaires	Collecting qualitative data related to user satisfaction with the artefact.	Final stage of the design cycle

Table 2.3 Involving users in the design process (adapted from Preece et al., 2002)

2.10.3 Participatory Design (PD)

Participatory design (PD) can be defined as a set of theories, practices and studies related to the end users as full participants in activities towards software and hardware computer products and computer-based activities (Greenbaum and Kyng, 1991; Muller and Kuhn, 1993; Schuler and Namioka, 1993). PD is characterised as a maturing field of research with evolving users of computer products among the developments of design professionals (Kensing and Blomberg,1998). In essence, users are co-designers (Abras et al., 2004). PD emerged earlier in Europe, particularly in the Scandinavian workplace, and it originated in strong labour unions, acting as advocates for workers to have more democratic control in their work environment (Ehn, 1989) and for the importance of the social dimension of work with technology. It provided a fruitful environment for those who were concerned about the workplace (Muller et al,1993). Thus, PD focuses on system development at a design period by bringing developers and users together to visualize the contexts of use (Ficher and Ostwald, 2002). In addition, PD schemes often concentrate on specific challenges in real world projects. Greenbaum (1993, p.28) stated that 'PD implies that workers as users of computer products should take part in the decisions that affect the system and the way it is designed and used. Because technology is not developed in isolation, participation in decision about technology also involves decisions about work content and job design. Seen in this broader context participatory computer system design needs to be part of an integrated design that looks at work organization'.

2.10.4 Previous Research

There are many important research studies that have been conducted to evaluate the usability of a software system using several evaluation methods.

2.10.4.1 Usability evaluation in business organization context

For instance, Catarci et al. (2005) conducted an evaluation of usability of the Sewasie (SEmantic Webs and Agents in Integrated Economies) system, as an essential step in the Human-Centred Design (HCD) aims to demonstrate whether the system is usable for the end users in the domain of their own experience, followed by the aspiration to improve it. In another study conducted by Sulaiman et al. (2009), it observed the usability of a system which is currently used in the main control room of an oil refinery plant by conducting a survey among the users (the control operators) in order to figure out if there are any usability aspects required for the system to be improved. Owing to the meeting of users' requirements, and after obtaining their feedback, several advantages would be achieved; one of them is the economic gains. In addition, Sturm et al. (2002) have published their project to evaluate the usability of a Dutch Multimodal System for Train Timetable Information (MATIS) in order to gain guidelines for improving the system interface. They measured three main attributes with the system, the efficiency and effectiveness of the interaction between the users and the system and their satisfaction in using it. According to an informative study by Bakhshi-Raziez et al. (2012), the evaluation of the usability of a compositional interface terminology of an intensive care admission system, which is used for data entry in Patients Data Management, should be implemented to find out if it meets its intended objective and to come up with issues for development. In conducting this usability evaluation of DICE (Diagnoses of Intensive Care Evaluation) system, they applied a mixture of methodologies both quantitative and qualitative for collecting data, through measuring five aspects: effectiveness, efficiency, learnability, overall user satisfaction, and usability problems encountered by the users. The results from qualitative and quantitative analyses enabled the researchers to observe and analyse clinicians' interaction with a large controlled compositional system user interface to carry out data entry, and provide a detailed view regarding this interaction.

In another study evaluating the usability of system UI in business administration, Odeh and Adwan (2009) proposed a study that was implemented on two computerized business administration systems, so they could determine the differences between the user interfaces. Initially, the first target system was a commercial software system which was dealing with organising supplies, tendering procedures, issuing purchase orders, and controlling and tracking the movement of the stocks to make sure they have balanced and they are on their tabulations. The second system was a suggested prototype which was developed with modern technology for the purpose of covering the problems and errors of the first system. In their study they applied the usability test approach through enrolling different techniques and methods in two stages. For example, in the first stage they applied experimental tasks by conducting several techniques, such as a questionnaire for subject evaluation. In the second stage, they applied a comparison usability test with 6 users who had received training to perform three scenarios using both systems. After the collected data was analysed using different statistical methods, some usability differences between the two systems were highlighted, and the negative aspects of the commercial system were covered through using different modern technologies of computer graphics and multimedia. However, they recommended at the end of their study, that further experimental studies need to be carried out on the role of Human Centred Design (HCD).

Dag et al. (2001) presented a project on an industrial case study of Usability Evaluation in Market-Driven Packaged Software Development without using any usability experts, and they obtained valuable results from the usability evaluation performed by the organisation's employees, which were useful and meaningful. The researchers carefully selected two mixed methodologies, qualitative and quantitative, by utilizing a commercially available questionnaire 'SUMI' and conducting the Heuristic evaluation technique. The results guided them to identify the particular issues that needed to be improved and considered in the development process. Beckert and Grebing (2012) have performed a usability evaluation for the Key Program Verification System. They presented a questionnaire aiming to investigate in which direction the target system should be improved in its usability. The questionnaire was developed by a combination of standard questionnaire, the System Usability Measurement Inventory (SUMI) and Green and Peter's Cognitive Dimensions, due to the fact that the SUMI method only has a "score" reflecting in which dimension the system requires to be improved, so including the cognitive dimensions would provide a channel to discuss cognitive issues that also allows the evaluator to assess the system without an expert assistant. Similarly in a study by Zins et al. (2004) which presented an outline of the experimental evaluation of a travel recommendation system, a user questionnaire was used for measuring different aspects such as ease of use, design and layout, and functionality. In this study also an overview of various methods concerning usability evaluation was introduced, and in addition it started by reviewing theoretical perceptions focusing on the influencing factors for HCI, system usage and satisfaction. This was followed by the evaluation of the the legacy system that was under the development process, and the prototypes were tested and evaluated by representative users in order to identify the system weaknesses and problems so that these could be tackled in the further development process. The main findings that came up from analysing the data indicated that although the collected data significantly indicated a preference for using the original system rather than the developed prototype, the responses on the satisfaction ratings of the three dimensions (i.e. system usefulness, information quality and interface quality) proposed by Lewis (1995) were higher for the prototype design, with more recommendation functions which could confirm the right direction for the aspirations.

Additionally, Abdul Rauf et al. (2010) in their study employed a usability evaluation for a popular application software, Microsoft Word 2007. This application was chosen to be evaluated due to it having a large user base at academic and professional level. They conducted their study in two stages; the first stage was a comparison test between MS Word 2007 and the previous version of MS Word, Word 2003. 75 participants were surveyed, and asked to respond to the questionnaire, which consisted of 40 questions, besides including 10 open ended questions to collect the qualitative feedback towards the application from 2003 to 2007. The second stage was aiming to evaluate the features of the last version of MS Word 2007 through utilising also a shorter questionnaire which consisted of 10 multiple-choice questions. The overall findings from analysing the participants' responses showed that the new version of MS Word of 2007 had better efficiency and effectiveness than MS Word 2003, and they were satisfied with the product in respect to its usability; however, they only needed the particular features that were essential in their daily business. In addition, the study verified that using a survey could be a best guide to obtaining users' viewpoints towards any products.

2.10.4.2 Usability evaluation in government organization context

A study of usability evaluation by Maybury (2011) was conducted in a governmental organisation, it was focused on the usability of its sensitive systems, represented by defence systems which were implemented to train and support the military forces. One of these systems was the Remotely Piloted Aircraft (RPA) system, which had grown rapidly for assisting several divisions, such as border security, humanitarian relief, and surveillance. After analysing the system usability, it was found to be poorly designed, which failed to produce effective, robust, and safe mission management (Board, 2010). Additionally, another professional institution, the SAB[5], identified that "*insufficient and inflexible platform and sensor automation increases operator workload and limits mission effectiveness*" (Maybury, 2011 cited in Buie and Murray,2012). Therefore, Thunberg and Tvaryanas (2008) mentioned that a human-centred design was considered necessary for enhancing RPA operator UI, and to satisfy the requirement of addressing the broad range in which defence systems need to operate, and the problems and challenges which were manifested through analysing the system. The study highlighted several lessons learned from considering human factors in military fields. Similarly, another military system that focuses on usability was in regards to air traffic management (ATM1) which operates military airspace, besides national airspace and international airspace, via two methods of communication, synchronous[6] and asynchronous[7]. Thus, in the towers, the controllers have interdependent roles and need to be carefully controlled from their side, while the system has to evolve to allow for the controllers moving from a highly automated process to visually tracking aircraft and keeping records manually. So, the complexity and severity of overlap of the roles of human and machine forced defence systems essentially into applying the most advanced developments in system usability..

In the health division of the public sector, enormous research studies have been published aimed at improving the system usability in order to provide better medical services to the public. For example, Carrroll et al. (2002) published a study which was aimed at designing and evaluating a clinical decision support system (CDSS) to assist the decisions that need to be taken upon particular medical problems. At the beginning, after the initial requirements and specification had been collected, three prototype CDSS inference designs were designed, and then this was followed by conducting usability tests with fourteen participants, seven patients and seven clinicians. Five different tasks were designed which the

[5]The Air Force Scientific Advisory Board (SAB) is a Federal Advisory Committee that provides independent advice on matters of science and technology relating to the Air Force mission, reporting directly to the Secretary of the Air Force and Chief of Staff of the Air Force.
[6]This type of communication is e.g. radio and face-to-face communication.
[7]This type of communication is e.g. emails, and chats.

participants were asked to perform using the developed prototypes, to test their efficiency, effectiveness, and ‘user-friendliness’. Furthermore, structured qualitative questions were developed for exploring the participants’ preferences from among the different designs, besides their overall opinions of clinical usefulness. The results in regards to the designing and evaluating of the CDSS revealed that the participants were enthusiastic to use this system technology after they have had a short training period. However, some participants from among the patients had come up with some challenges in interpreting clinical data, but still they showed their interest in using the CDSS as it would help them to understand their case when a clinician explained their results. On the other hand, the participants from the clinicians emphasised how involving the CDSS would increase their consultation times.

Another study, by Britto et al. (2010), evaluated the usability of a medical system which was concerned with patient portals, intended to improve a particular illness’s outcomes. They conducted a usability testing method through using a particular scenario with a think-aloud protocol as a technique for collecting data. Besides this, the Computer Usability Satisfaction Questionnaire (CUSQ) was used for measuring the users' satisfaction by asking the participants to complete it after they finished the tasks. Sixteen novice participants were recruited to perform 14 selected tasks using a suggested prototype which was iterated three times. The results indicated that many problems and challenges were found related to different aspects such as terminology, task completion, ease of use and satisfaction. However, regardless of these encountered problems the participants could have enormously positive perceptions about the system.

Another study in the same field, by Hamborg et al. (2004) presented a usability evaluation of a hospital information system, which faced a major disruption, as its daily users complained about wasting time each day by *“filling in forms, reviewing medical inspection results and handling an amount of information for administration needs*”. The reason for this issue were found to be that the recent evaluation was concentrating mainly on financial issues and considering the patients’ interests, without taking into account the employees as users of the system, including nurses, physicians and other healthcare staff. Therefore, for evaluating the system and gathering data from the participants, the ISO metrics inventory questionnaires (Gediga et al., 1999) which afford user-oriented summative and formative approaches to evaluating software on the basis of ISO 9241 part 10, has been used extensively to assess the practical usefulness of the targeted system. The main finding was that the evaluation of the Hospital Information System was useful in terms of supporting the clinical work of health care employees in public through incorporating their requirements into the system software development. In addition, it was determined that using the questionnaire tool for summative evaluation and to identify general challenges facing usability aspects was not sufficient. It needed to be conducted with the assistance of formative evaluation techniques such as user tests for discovering the most significant usability aspects and the main causes of the lack of usability (Dumas and Redish, 1999). This consideration of systematic evaluation of hospital information systems will support the clinical work of health care employees in public by adapting the software to user requirements, improving its functionality continuously and avoiding errors and stress reactions as well as the costs associated therewith.

As shown before there are several studies focused on utilizing various of UEMs for measuring the usability as a vital quality factor in the success of a software system design (Shackle, 1991; Nielsen, 1994; ISO,1997b), and in order to produce a usable system, a Human-Centred Design (HCD) methodology has been considered an appropriate approach to achieve system usability (Preece et aI., 1994; Maguire, 2001a). According to Ferre and Medinilla (2007, p.68), since usability has become an important quality factor in software systems, *"it requires the adoption of a human-centred approach to software development"*. In addition, Spencer (2004) mentioned that the HCD can be used to develop a usable system, and its principles could be involved during the design and development process, from concept development to final testing. Fundamentally, a usable system should allow its users to find it easy to use and learn and to perform particular tasks effectively with low rates of errors, this would lead to enhancing their acceptance of it (Preece et al., 2002).

2.11 Iterative Design

Undoubtedly, it is difficult to achieve well-designed UIs at the first attempt, even by the best usability experts (Nielsen, 1993), and in relation to this difficulty, there continues to be significant evidence that the usability test, conduct modifications, and employing refining or iterative design method are clearly successful (Bailey, 2005). Therefore, the system UI design should be developed around the perception of iterative design, the development process of which comprises stable design refinement relay on user testing and other evaluation methods. Robin and Chisnel (2009,p.28) stated that *"An iterative design and testing approach also allows one to make steady and rapid progress on a project, to learn through empirical evidence, and to 'shape' the product to fit the end users' abilities, expectations, and aptitudes".* Furthermore, the iterative design enables the designer to rectify any problem they might come across during the development lifecycle, due to the repetition process granting several opportunities for them to correct the mistakes.. However, Nielsen (1993) argues that the iteration design will not always tackle a usability problem, because it is well known that an attempt at improving one part of the system interface might have an opposite impact on another part. For instance, work on the improvement of data validation, for reducing recurring errors, might affect the users' performance of data entry.

2.11.1 System Development Lifecycle (SDLC)

In relation to human-centred design and typical software engineering views, there are various approaches to computer system design. However, the system developers traditionally should take each stage of the system design lifecycle as an independent part of system development, which is required to be fully satisfied before moving on to another stage. Most organizations have found that using a traditional standard of steps, a "systems development methodology," is beneficial for improving and supporting their information system (Hoffer et al., 2011). This development of an information system consists of many processes and often follows a lifecycle, which concentrates on the activities of system development (Dix et al., 2004). It is a circular process in which the end of the useful life of one system points to the beginning of another scheme that will lead to "*a new version or replacement of an existing system altogether*" (Hoffer et al., 2011, p 35). The most common methodology for systems development is the classic system development lifecycle. However, Stone et al., (2005) indicated that there is an important difference between this type of development lifecycle and human-centred design (HCD), which is that user interface (UI) design and development relies on involving users during the design life cycle. Another type of the development life cycle is the star life cycle (Hix and Harston, 1993) (mentioned previously in chapter 2, section 2.8.1.). Kay (2002) defines the System Development Lifecycle (SDLC) as an overall process of developing information systems, moving through different stages or a set of processes from investigation of primary requirements through analysis, design, implementation, and maintenance. Different lifecycle models have been created for developing systems, ranging from three to almost twenty stages, such as waterfall (classic life cycle), spiral, rapid prototyping, incremental, and build and fix (Kay, 2002).

2.11.2 Making the design process iterative

The iterative design is considered a way of ensuring that the users are able to be involved in design and other different types of knowledge acquisition. For this purpose, mention should be made of an approach adopted from a model called the star life cycle, which is presented by Hix and Harston (1993).This model encourages following the iteration design approach. Figure 2 shows the star life cycle and the research's main interest (adopted from Hix and Hartson,1993)

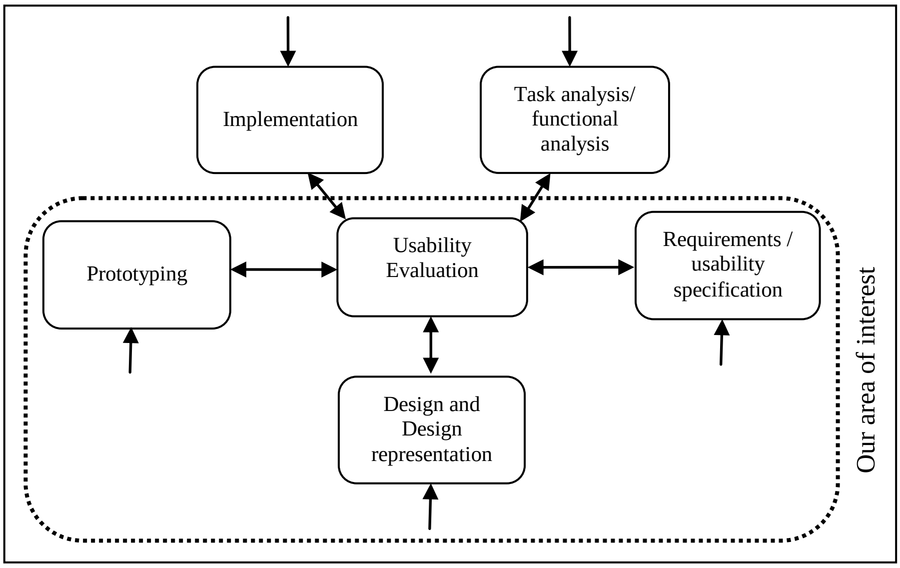


Figure2.3 The star life cycle for user interaction development. Adopted from (Hix and Hartson,1993)

It is obvious from (Figure 2.3) above presenting the star life cycle that the usability evaluation is at the central point of the star model, which is viewed as being linked to all stages in the life cycle. So, firstly, the evaluation is focused on collecting data in relation to the usability of a design or system by a selected group of users for a specific task within a specified environment or work context (Stone et al.,2005). The presence of different evaluation techniques is required to support the various stages of design. These techniques include interviews, observing users in their workplace, and gathering their views by conducting questionnaires. Secondly, the star life cycle model is *"intended to be equally supportive of both top-down and bottom-up development, plus inside-outside in development"* (Hix and Harson,1993). Therefore, a system design process could be begun with any process in the star life cycle. In addition, as the users' interaction development activities were shown in (Figure 2.3) above, in the part labelled "our area of interest", the included activities are represented by the three studies reported in this thesis, in chapters four, five and six respectively.

2.11.3 Consideration of information communication technology (ICT)

The rapid advancement of information communication technology (ICT) into all spheres of human activity has made this aspect a moderator that deeply affects people's interaction with their environment. Therefore, the ICT evaluation has direct effects on economic and political aspects of society and social and cultural activities all around the world (Ogunsola,2005), which in turn has changed the role of governments, organizations, and citizens, and strengthened the link between each of them by presenting powerful new instruments of communication within the private and government sectors. According to Narasimhadevara et al.(2008, p.137), "*Computer-based support has become essential in almost every work environment". *Furthermore, as Buie and Murray (2012) have said, "*arguably, the government systems have the largest user base of any technology*", therefore, the ICT should be considered a key factor in developing the government systems. Practically, the effects of ICT in the public sector could appear in terms of governments' investments, which have been directed towards developing efficiency and rule effectiveness as well as achieving broader democratic values (Kamarck, 2007; Gil-Garcia and Pardo, 2005; Gronlund and Horan, 2004).

There is no doubt that the users' requirements of a particular software system, whether they use it, or intend to use it, would have "no limits"[8] due to the fact that it could emerge and be regenerated according to the changing of several concerns that might have influence on a way or purpose of using this system, speed of decision making, or on the goals that need to be achieved through it. For instance, technological advance is increasing continually, which prompts the users to improve their productivity, and ability to cope with any issues that they may encounter during the use of the system. On the other hand, it has been already encouraging different sectors, such as government and service organizations in the public sector, and commercial and manufacturing firms in the private sector (Narasimhadevara et al.,2008), to conduct a large amount of research directed towards obtaining possible benefits through determining the measurements of aspects within their system to be considered successful (Zviran and Ehrlich, 2003). These studies stated that in practice some benefits are tangible and can be quantified, such as, lower inventories, increased sales, reduced costs, and shorter reaction times. Other benefits can be intangible, and therefore hard to quantify, nevertheless, a number of studies have pointed out the importance of the intangible benefits as making a major contribution to the organization's success. For example, to the performance of employees through improved decision-making capabilities, improved information by easier cross-checking of it (Thong and Yap,1996), and improved work process (Gibson et al.,2004). Consequently, considering a computer-based support within an organization's system has become a crucial element for almost each ergonomics study (Narasimhadevara et al.,2008), which would help to eliminate the influence of negative factors on the employees' performance, and then to meet the right requirements (Aguinis,2009; Gusst,2002). This might lead especially to benefits for the organization and the nation as a whole (Asabere and Gymafi,2013).

2.12 Implications for this research

This section has presented the factors associated with system usability and introduced the importance of it being measured as a crucial technique for producing a useable system. In addition, it has discussed the common Usability

[8]An expression that has been used to describe the difficulties of expectation in which a point of developing the system design would satisfy the users, and the process should be stopped at that stage.

Evaluation Methods (UEMs) to highlight the appropriate techniques and tools for evaluating a software product to achieve particular aims in connection with developing its usability. The presentation has demonstrated the importance of conducting the evaluation of system usability to obtain the possible advantages in the interests of the users and the organisations alike. Hence, the current guidelines for the usability evaluation of systems have been reviewed, and previous studies were examined to gain possible benefits from previous experiences. Having arrived at a conception of the best methods for evaluation of the usability of an internal system in a governmental organisation with a large sample allocated in different departments, we will now proceed to take the first step towards proposing a framework for the entire usability evaluation process. The following section focuses on the experimental approach to usability testing in detail and reviews the ways to produce a prototype of the target system by designing a usable and appropriate UI, which is the subsequent stage for developing the system usability.

In addition, this section has presented usability testing as an important method for evaluating a system to identify whether the previously defined usability goals have been achieved as well as to assist in confirming that the system and design are effective. Furthermore, usability testing could identify problems, enabling the government administration to allocate the essential resources to sort out these problems, and it showed that there are two main aspects that need to be considered when conducting usability testing (Leavitt and Shneiderman, 2007). The first aspect involves ensuring the utilization of the most appropriate and feasible method for usability testing. Generally, the best method for conducting a usability test can be identified as a classical method for conducting a controlled experiment in which certain participants (users) perform representative tasks while the tester collects relevant data on the participants, such as their success in finishing the task, the task time (speed of performance) and their level of satisfaction. In other words, Usability testing can rely on its origin in the classical approach for conducting a controlled experiment. After that, the outcomes consist of both quantitative data and qualitative data from observation techniques, which are presented to be used within the design process in further stages. The second aspect is associated with the determination to ensure that an iterative approach is applied. Therefore, reviewing the characteristics of the usability test, suggesting how the prototype can be measured and tested, and determining how the provided results would enable the developers to make changes and design a usable UI should be followed by employing a stage of iterative design, which will be presented in the next section.

2.13 Chapter Summary

This chapter discussed the main sources and literature related to the research area to highlight the best method for proposing a usable design system. Different usability methods were considered to be used in the usability evaluation process of the target system, based on the purpose of the evaluation in each phase of system development in the research study. In the next chapter, the methods that have been selected to conduct these issues in the research work reported in this thesis will be discussed and justified.

Chapter 3: Research Methodology

3.1 Introduction

The following chapter presents the methodologies that were utilised in the conduct of the research for this thesis. It consists of three main sections. In the first section the chapter begins by providing an overview of the research programme followed by presenting the description of the VI system and the research participants. In the second section, the methods have been used for conducting the research work which used (HCD) as a base for the entire research approach, and in order to involve the target users' perspective in the system design (Abras et al., 2004), and development process to introduce a usable system (Maguire, 2001). This was represented by conducting the three studies with employing different UEMs and techniques in different stages through undertaking usability evaluation by query technique, and then by conducting an experimental methodology. Thus, the planning of the conduct of the first study and the two experimental studies are described with reference to the sample of participants, the variables, and the design of the experiments. This at the end would help to produce a framework that could be used for developing the implemented internal software system in the governmental organisation. In addition, in the last section, the data analysis methods and procedure, the design of the experiments, and the planning and conducting of the experiments are explained with references.

3.2 Description of the Current Visa Issuance System (VI)

The target internal software system is concerned with issuing visas, mainly to the citizens or other residents[9] for individual purposes, or organizations (in public or private sectors) for services and business purposes in Saudi Arabia to enable the citizens to recruit workers from abroad. It was designed as web-based application

[9] According to the rules for issuing visa to the foreign residents (Ministry of Labour), only the foreign investor and the highly qualified e.g. doctors and engineers are eligible to obtain visas for recruiting workers from outside.

and implemented at the Recruitment Department in the Ministry of Labour in 2008.

3.2.1 The targeted task of the Visa Issuance (VI) system

The main task that is the focus in this study was related directly to the main purpose of the implemented VI system, which is issuing Visas to the applicants, as described previously. This task is performed frequently on a daily basis by the system user (the employee in our case). Employees have their own discs and PCs where the VI system is installed and ready to be used for providing the service of issuing Visas. The workflow of using the VI system principally begins when an employee (as a user of the system) logs into the system and enters the user name and password. After that, the required documents are received from an applicant, which includes an application, a copy of ID and a letter approval of personal income. Then, the ID number (ten digits) of an applicant is entered in the specific column in the main screen of the VI system (Appendix B - Screen 1) for displaying the applicant's details. After that, the employee needs to check other applicant details by viewing other screens, such as the number of sponsored workers (Appendix B - Screen 2), and the number of issued Visas (Appendix B - Screen 4). After the employee has made sure that the applicant is eligible to receive a Visa, then the screen is opened for processing the issuance of a Visa (Appendix B - Screen 5). Therefore, all required information needs to be entered into the system, which will end by clicking on the enter button to issue a Visa, get a reference number for the request and print it out. Accordingly, the process of finishing the task requires 5 steps in typical cases; however, in some cases, it can require 6-8 steps depending on the particular case of an applicant, which can be carried on through several screens until the task is completed. Regardless, the variation of the employees' capabilities for using the VI system and the average time that the target task consumes for a normal case to be completed is about four minutes. After that, the visa will be handed out directly to the citizens, and others will be sent to another department to be delivered to the recipients. Pair tasks may involve operating the system to issue one or more visas depending on the application submitted by either an individual citizen or an institution.

3.3 The research participants

For any research project, considering the appropriate user sample and how many are to be recruited is a crucial factor in its success (Cairns and Cox, 2008). For example, if the experiment is intended to test the effect of a changed display structure for a specialist task of a particular system, such as a new air traffic control system, it becomes necessary to recruit participants who are familiar with that task, and are experienced air traffic controllers. Similarly, for the research reported in this PhD thesis, the target sample was comprised of the employees working in a government organisation in Saudi Arabia. Since the current VI system was chosen to be the system targeted in the study, only employees were selected who had access to and experience of using the VI system, which is mainly implemented to issue work visas to the citizens. However, the number of participants was varied at each stage, based on different determinations such as the study aims and objectives, and methodology, the details of which will be reported in the relevant study. For example, the first study was targeted to include the whole sample or the total number of users of the VI system, which was 135 users (according to the Department of Electrical and Computers in the government organisation, March-2012[10]), and so it was conducted as a survey study. The number of participants in the experimental studies, however, was varied, which contributed to achieving these studies' goals. So, in the first experiment, reported in chapter five, 32 participants were invited from one branch of the government organisation to perform the actual tasks using both systems, using the current VI system and the developed prototype. In the second experiment 26 participants were involved in the test; all of them had participated previously in the former experiment, due to the consideration that if these participants had been excluded, it would have affected the whole sample size at the end, and might have resulted in too small a sample size (Teijlingen and Hundley, 2001). All the participants involved in this research study were male, their ages ranged between 28 and 50 years, and all of them had experience of between 1 and 6 years of using the actual VI system.

[10] This information regarding the total number of users was identified via the counting of all their accounts of logging in to the VI system - at shown date.

3.4 The research work programme, methods and mapping

Three studies were conducted for the research work reported in this thesis, with the application of a mix of qualitative and quantitative methods, in order to achieve the best understanding of the research aims, and to create confidence in the research findings (Wong and Blandford, 2003). According to Tashakkori and Teddue (2002), applying mixed methods within the research would be useful to afford better opportunities to find the answers to the research questions; what is more, it would afterwards assist the researcher to meet the criterion for evaluating the quality of these answers. In addition, these three studies, one a survey study and two experimental studies were conducted in succession to form the human-centred design (HCD) process for achieving the responses of different research questions.

Based on the findings of the first study, the second study reported in Chapter 5 was conducted as a laboratory study which aimed to develop a prototype design by addressing the problems and challenges that emerged, and then to conduct the comparison usability test between the current system and the developed prototype in order to assess and validate the improvement. After we came across more challenges and additional requirements, the last study reported in this thesis was also an experimental study that employed the iterative design approach to improve the prototype design with more additional features of ICT, and then conducted comparison usability testing between the two versions of the developed prototypes.

3.4.1 Stage One: Query Usability Evaluation techniques

The query evaluation techniques was considered at the first stage in this research to evaluate the usability of the Visa Issuance (VI) system through applying a survey as the basic method for gathering the users' point of views and suggestions on the current VI system, and in order to have a full insight into the state of its usability by ranking specified items from amongst alternative responses provided and answering a set of written questions. So, fundamentally, the query techniques in this research, require users to response to number of questions either by an evaluator directly in semi-structured interview, or in form of questionnaire (1A - Figure 3-1). Hence, additional information from those users could be gathered, "*which makes query techniques a powerful supplement to other techniques*" (Freiberg and Baumeister,2008,p5). According to Lazar et al. (2010)

surveys are one of the most widely accepted research methods used across all fields of research. Initially, this technique is concerning with questioning the user directly face to face at an interview, or indirectly in writing a questionnaire, in respect to their experience of using the target system that is under evaluation (Dix et al.,1998).This will also generate some interesting feedback (Spencer, 2004) which will help to obtain new concepts for enhancement the system. According to Maguire (2001) surveys could help in determining the users' needs, current work practices and attitudes to new system ideas. Moreover, the users would be able to report on issues and problems with the system, or any suggestion for changes to any aspects of the system that they need. A mixed method of quantitative and qualitative research was used by employing questionnaires and interviews as instruments. Miguire (2001) also indicated that a survey method is useful for collecting quantitative and some qualitative data from a large number of users about existing tasks or the current system. Furthermore, Covey (2002) stated that a survey through using questionnaires and interviews is considered the best method to be used for gathering large quantities of data on users' preferences, attitudes, motivation, and satisfaction. This strategy of a query usability evaluation through conducting a survey was taken as an initial step for conducting the research in this thesis according to several considerations. For instance, in order to include the whole targeted sample, which consists of the exact number of employees as the only users who have an access for using the current VI system (135 users), and with less time, effort and cost. Additionally, these users were working in different branches of the government organisation, where they were allocated to various areas. Preece et al.(1994) reported some conditions to applying a survey of existing users, e.g. when there is difficult access to the target users due to their location, role or status, or when there are different user populations, or when quantitative data is required, e.g. functional preferences. According to Dix et al (2004) one of the main advantages of applying query techniques is that the flexibility they provide. Furthermore, Pace (2004) indicated that with the rapid changing of technologies and associated human interaction issues, the requirement for evaluation of systems with distributed users in different contexts is increased, which in turn has led to the rise in using questionnaires, interviews, and focus groups in commercial usability and academic research contexts (Adams and Cox, 2008). In this stage, before the final questions of query technique are conducted to the targeted sample, a pilot study is employed for verifying and validating both instruments, the study questionnaire and semi-structured interviews (1B - Figure 3-1) - the details of this pilot study is provided in chapter 4). After that the feedback which collected from the participants of the pilot is considered into the final versions of the query technique instruments to be ready for conducting a large - scale survey with all targeted sample (1C- Figure 3-1). The collected results of evaluating the usability of VI system using query technique is taken for data analysing, interpretation and evaluation statistical information via enrolling different statistic techniques such as, analysis of variance (one way ANOVA), and multiple liner regression (1D - Figure 3-1) - (The details of analysing collected data is presented in chapter 4). The final results enclosed the usability problems and issues of the current VI system that need to be addressed for the enhancement process of the current system (1E- Figure 3-1). The outcomes of this stage is transferred to be fitted within the next stage of redesigning the current VI and for developing a new realise of the current system (1H- Figure 3-1).

3.4.2 Stage Two: Experimental methodology (Usability testing)

The system usability has commonly been tested in a controlled laboratory setting (Rogers et al.,2001). Thus, the experimental method or controlled testing is the most formal evaluation approach for evaluating a particular system design and any aspects related to it (Dix et al., 2004), and identifying, in a straight line, how the real users will interact with the full working system, performing actual tasks in a series of prepared trials (Li et al., 2006), through employing several measures. Some of these measures that are used in usability tests are efficiency, effectiveness, and satisfaction. Furthermore, it is generally accepted that experiments are conducted for proving a known fact, or examining the validity of a question that explains a relationship between two or more variables by controlling one or more variables (Preece et al., 2002). One of the important advantages that would be gained from applying this method is that the target system can be tested in similar conditions to those where it will actually be used, and then reliable and useful data will be gathered rapidly. However, the experimental design and statistical analysis are considered to be complex issues (Lazar et al., 2009; Cozby, 2006; Elmes et al., 2005), and this would be considered as one of the disadvantages, leading to requiring much time, cost and effort from the practitioners and users (Cairns and Cox, 2008).

Based on the outcomes, identified problems and challenges from the first study, a design solutions is developed via utilizing medium fidelity prototype (2A Figure 3-1), the details is presented in chapter 5. A pilot study is applied for obtaining the participants' point of views on the initial prototype design (2B Figure 3-1). This medium fidelity prototype is adjusted and improved once the feedback collected from the participants regarding its design (2C Figure 3-1). After consolidating the improvement, the medium fidelity - developed prototype design DPD, it upgraded into high-fidelity prototype (2D Figure 3-1) for the purpose of involving functional options into the design of DPD. Thus, it developed with required functional for enabling to conduct the usability test (2E - Figure 3-1) . Also, a pilot study is conducted for validating and conforming the improvement of the design and making sure that it is capable for performing the actual task of the VI system (2F Figure 3-1). More details of both types of prototype design are presented in chapter 5. So, inthe second study of this research, which was applied as a controlled user testing study, the targeted participants are recruited and invited to attend the test in an informal lab to perform selected actual task tests. Then the developed prototype design DPD was used for conducting a comparison usability test between the two systems, the current VI system and developed high fidelity prototype, with 32 participants, in order to confirm the design improvement (2I Figure 3-1). Additionally, before conducting the final usability test a pilot test is conducted to ensure that the proper usability test procedure and tools are in the right place and chosen correctly (2J Figure 3-1). Thus, after getting the feedback from the participants regarding the final test and other problems and issue that come cross while piloting the test, it is considered for revising and fixing the final usability test sittings (2K Figure 3-1). The collected data from the comparison usability test is analysed using different statistical techniques for considering the improvement in the final DPD (2L - Figure 3-1). The outcomes of this study with additional users' requirements and suggestions of is shaped the developed prototype design as an improvement of the existing VI system (2M - Figure 3-1). After that according to the outcomes and the additional users' requirements and needs on the previous DPD, it is taken to next stage of the research to be iterative (2N Figure 3-1).

3.4.3 Stage Three: Iterative Design Methodology

Since the HCD process was adopted for the work to be conducted in this research, it provided an iterative means of developing the suggested prototype of the current VI system with more involvement of ICT to meet high levels of system users’ participation in the first experiment, in order to produce a more realistic user interface for the VI system (**3A** Figure 3-1). To make sure that the re-designing of DPD is developed with consideration of the additional users' requirements, needs, a pilot study is conducted for this purposes (**3B** Figure 3-1). After receiving the feedback from the participants regarding the design, the DPD is edited and adjusted with required information to be ready for conducting the usability testing. There are two main concerns when performing usability testing provided by Leavitt and Shneidermen (2006), the first concerns the need to make certain that the best possible method for testing is used. The second main consideration is to make certain that the iterative design approach is used. Therefore, in the second experiment, reported in chapter six, after the results of the primary usability test were obtained, iterative design was used, and consequently the developed prototype was re-designed based on these findings. Additionally, the experimental methodology with the same setting as the former experiment was applied all over again. It is targeted to include 26 of the actual users at this stage, in order to provide a controlled means of comparing the two versions of prototype design, before and after the re-designing process , in order to evaluate and confirm the improvement of the new release of DPD (**3D** Figure 3-1). A pilot test for this the usability testing' sittings is also conducted to ensure that the right procedure, proper materials, and recording mechanism are employed properly in the experiment and understandable by the participants (**3E** Figure 3-1). The gathered feedback from the participants regarding the usability test sitting, is taking into account for conducting the final usability test with actual participants (**3F** Figure 3-1). After that the collected data from the participants is analysed and measured using several statistic methods (**3G** Figure 3-1). The outcomes of this test is considered as requested solution represented in the new release of DPD with additional functions (**3K** Figure 3-1). Consequently, the final results of the research work is provided as value added solutions as a enhancement for the current VI system users and as a guidance to practionars, designers and to system developers (**3L** Figure 3-1). Lastly, for the future work which would cover the limitation of this research work, another iteration approach could implement again and the re-designing technique upon the DPD could happened for more enhancement (**3M** Figure 3-1which presents mapping of the whole thesis).

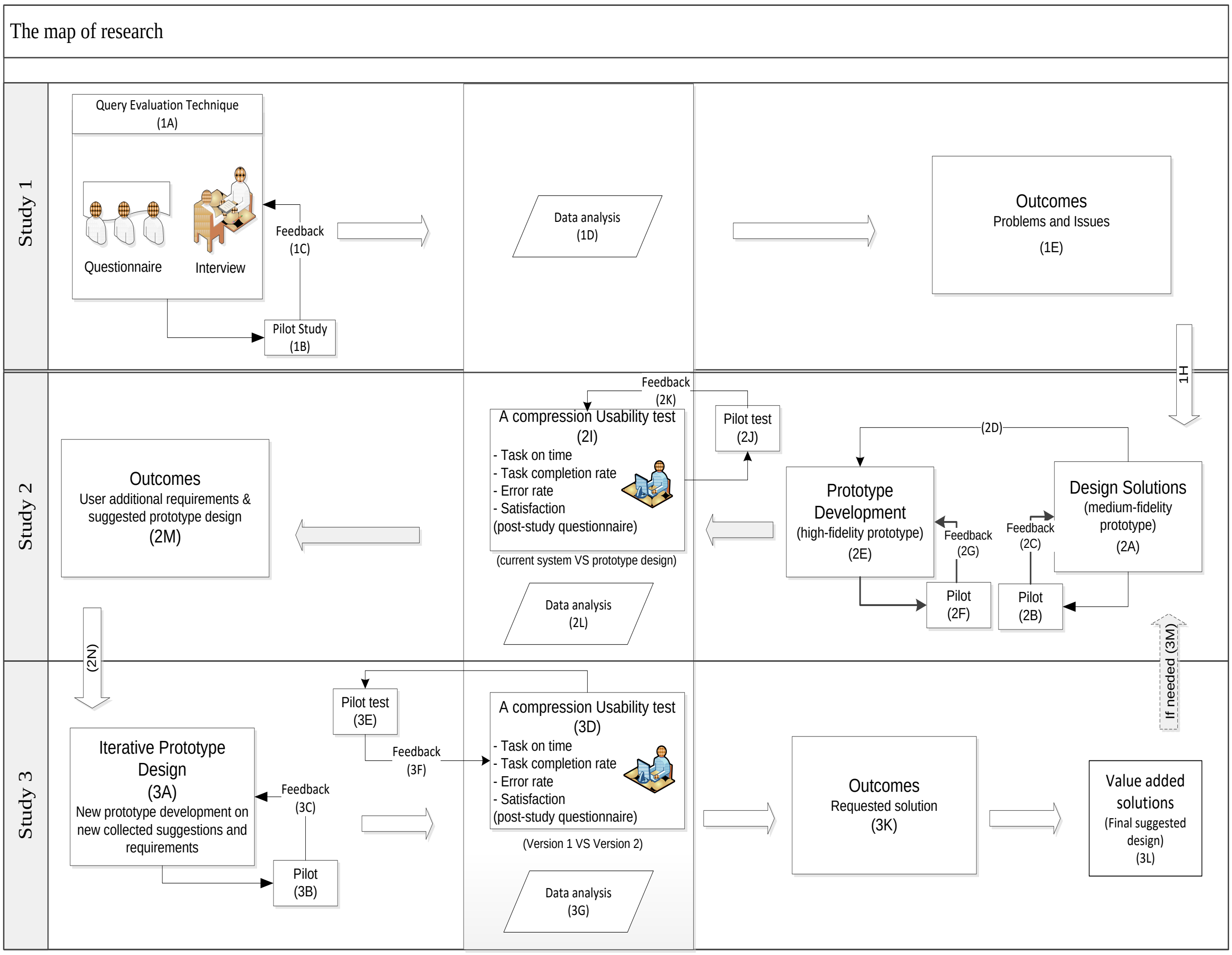


Figure 3.1 The map of research

The type of within-subject experiment design was considered for the two experimental studies. Initially, the participants are divided into groups A and B, each group had specified number of participants, and each group is asked to perform all the tasks by using both systems. In the first experiment, the current VI system and the developed prototype design (DPD), while in the second experiment between the two versions of DPD However, the two groups have a different order of starting to use the two systems to perform the test tasks, as a "counterbalancing" technique is applied. So, the participants in the first group A started the test session by using the DPD, while the others in group B started the test session by using the current VI system. Similarly, in the second experiment the participant are swapping between the two versions of DPD. More details about both experiments is provided in the related chapters.

3.5 Measurement of the study variables

A variable means any properties that can differ among people, or situations that can be of diverse levels or types (Breakwell and Schaw, 2000). Primarily, variables are divided into independent and dependent (Rogers et al., 2011). An independent variable is what the investigator decides to manipulate in the test, for instance, the two versions of the suggested prototype design. Another type of variable is called dependent, for example, the time taken to finish an actualtask. For the work reported in this thesis, the independent variables manipulated within each of the two experiments are covered in experimental chapters five and six.

The comprehensive definition of usability that has been provided by the international standard ISO 9241-11 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. This would explain if the system is effective in terms of enabling the users to finish a particular task, efficient in allowing them to do this task, and lastly if they were satisfied with the system. Therefore, in the experimental studies reported in chapters five and six, the participants' subjective preferences regarding the target system usability were used to measure UX and satisfaction, via a post study questionnaire, whilst objective measures of user performance were used to measure the effectiveness and efficiency of the target system, through the common measurements, task completion rate, time taken to complete a task, and error rate.

3.6 Data collection instruments

A number of instruments were utilized to gather important data for the research reported in this thesis. Data was collected through using questionnaires, interviews, usability test, direct observation technique, think aloud protocol; and each of these instruments is described in the following section.

3.6.1 Questionnaires

In this research, two questionnaires were designed carefully and used for achieving several objectives of the study. Both questionnaires were designed based on the previous standard questionnaires and some modification and editing have been made which are commensurate with the nature of the actual tasks needing to be performed using the VI system that is intended to be evaluated or that has been recently tested.

The details of designing each questionnaire will be presented in the related study. One questionnaire was specifically chosen to be used as a main instrument for evaluating the usability of the current VI system subjectively, in the first study in chapter 4 (Appendix C). Another questionnaire was designed for the two experimental studies reported in chapters five and six (Appendix H). Sheneiderman and Plaisant (2010) indicated that applying this type of instrument for usability test and expert review is widely accepted. For the experimental studies, it was determined that the questionnaire should be completed after the participants finished all the usability test session of performing the selected test tasks. Therefore, the study used the post-study questionnaire, which was constructed with three common factors of the PSSUQ with a spirit of USE questionnaire statements, which are System Quality (SysQual) factor; Information Quality (InfoQual) factor ; and Interface Quality factor (IntfQual). Furthermore, the questionnaire utilise the counterbalancing of positive and negative statements, besides, a free text box has been added to each statement for enabling a participant to provide his own comments upon that particular statement or any issues related to it, after he is encouraged to do so. Furthermore, in this questionnaire another type of question, an "open ended question" has been added within its design, by asking the participant to list the particular aspect(s) of the system he just tested that he liked or disliked, and explain why. (Appendix G).

In regards the details of designing processes of the questionnaires, it will be provided in the related chapter.

3.6.1.1 Questionnaire Translation

Because the research study was concerned to evaluate the usability of a software system implemented in a government organisation in Saudi Arabia, the first language of the targeted sample, the users of this system, was Arabic. Owing to the fact that the research questionnaires were designed in the English language, the translation of the questionnaires into the Arabic language became an essential process for conducting the survey among the selected participants. The translation of the questionnaire has been validated and reviewed by two authorities, a member of the academic staff in IS, and an expert in a certified translation office.

However, after the data was collected from the participants, the responses were translated back to the English language for analysis and completion of the study, as the back translation strategy was highlighted by Lewis et al. (2007). According to Brislin (1970), this strategy can gather a further in-depth understanding, and the version encompassing the target language can be better.

3.6.1.2 The reliability and validity of the questionnaire

Measuring the reliability is about the consideration of whether a questionnaire measure was developed consistently across time (Babbie,1990; Ozok, 2007). Cronbach’s alpha coefficient was used as a measurement of the reliability of the two questionnaires that were administered in this research study, and then to make sure that these questionnaires were reliable survey instruments, and thus that we can have confidence in the findings.

3.6.2 Interview

The use of interviews is considered to be the second major method for collecting data (Tashakkori and Teddlie, 2010). So, as an instrument they have become a widely used technique for gathering information from users to identify their requirements in relation to particular aspects of the system (Myers and Newman, 2006; Maguire, 2001).

In this PhD thesis, the interview was used as a technique for collecting more comprehensive and accurate data from the participants in the first study reported in chapter four. One-to-one semi-structured interviews were conducted with eight participants among the target sample who held different positions and were from different department within the government organization, this would help to achieve a good understanding of some issues that were raised by the natural social context, according to Marshall and Rossman (2006). Basically, the interviews were held after the data collected by the questionnaires was analysed in order to have more information in depth about their requirements and experiences with the system (Nielsen,1993), and more explanation could be gathered regarding some of the quantitative results obtained. Shneiderman and Plaisant (2010) stated that conducting interviews with users can be productive because it can follow up particular issues of concern.

3.6.3 Observation and note taking

This technique was employed in the two experimental studies which are reported in chapters five and six. Its purpose was to observe the users when they were interacting with the system, and record their reaction towards any issues that they might come across during their performing the selected tasks for the usability test. In addition, note-taking was utilised during observation of the users while they performed tasks, and this is considered the most flexible method of recording data (Rogers et al., 2011). However, utilising this instrument has some challenges, for instance, when the practitioner goes to write down one usability issue, this would prevent him from observing the next one. Thus, it is obviously hard to write down or record all usability issues at the same time. Therefore, James's approach was used as a solution to this issue; one of his solutions is note taking that focuses just on the key issues after each participant has finished., besides other assessments like the explanation of identified problems, filled-in data concerning the accuracy of task completion, and the amount of time taken on finishing a task. Sauro (2012) stated that conducting lab-based usability testing would provide excellent qualitative data via observation of both interface and users' reactions, so the practitioner can easily pick out issues. Therefore, the participants in each session were observed, and the notes were recorded.

3.6.4 Think aloud protocol

Essentially, the procedure of think aloud protocol was for helping to evaluate the functionality, strengths and weaknesses of the system, besides its usability (George, 2005). So, as think aloud protocol was applied in this experiment to collect a useful participant's feedback as qualitative data, each participant was informed that he was being observed directly during the running of the experiment (Hertzum and Jacobsen, 2003), and his reactions and comments were being recorded and noted down for later analysis, such as when he mentioned his feelings, faced challenges, or got confused during or after performing the tasks.

3.7 Data Analysis

This section of chapter presents the data analysis test used in the three stages of this thesis. Since the mixed methods of quantitative and qualitative was adopted for the work reported in this thesis, different data analysis techniques were employed based on the gathered, either quantitative or qualitative.

3.7.1 Statistical Analysis of Quantitative data

The analysis of quantitative data primarily used the following techniques for the statistical analysis:

1. For all the studies reported in this thesis, the descriptive simple statistical analysis of the general variables was provided for helping to illustrate and understand the data. It presented the various figures such as the average (mean), standard deviation, percentages, diagrams and frequency distributions for each independent variable.

2- For analysing the data collected using the questionnaire in the first study, one-way analysis of variance (ANOVA) was applied, as it is commonly appropriate statistical analysis in many research situations, especially when the study has three or more groups, and it is required to discover if those groups vary significantly on any outcome (Acock, 2010). Brace et al. (2012, p.209) stated, *"In ANOVA we are trying to determine how much of the variance is accounted for by our manipulation of the independent variables (relative to the percentage of the variance we cannot account for)."* Therefore, the main purpose of applying this analysis technique is to determine whether there is a significant difference between the means of responses of more than two independent populations. This statistical analysis was followed by Scheffe's test which is used for post hoc comparison analyses. So, it was used after the results of the analysis of ANOVA indicated that there is a significant difference, a Scheffe's test was performed to compare the average of several samples, and then determine the main cause of the significant differences between each two classes in the independent variables, (such as Age, Experience, and Level of Education), and to determine if these factors have any effect on the user's view regarding the usability of the system. In addition, a multiple linear regression analysis was used in this research study to illustrate in depth the statistical dependence of one variable on other variables, because having more than one independent variable is useful when estimating users' behaviours. Consequently, it was an extension to the previous analysis of ANOVA when it was not clear and the data was needed for predicting one variable on the basis of a number of other variables.

3- In the experimental studies, reported in chapters five and six, since the comparison usability test was applied in the two experiments, a paired t-test was used to compare means in the different systems or versions. The Wilcoxon test was an appropriate one to be applied as a data analysis as a dependent means or matched pairs t-test (Hole, 2011), due to the within-subject technique being used for conducting the test, so, the same participants performed in both conditions of the targeted systems.

3.7.2 Statistical Analysis of Qualitative data

For analysing the qualitative data gathered in the three studies, whether from the comments on the free text box within the questionnaire or from the responses from the interviews, thematic inductive analysis was conducted by focusing on the summative approach to qualitative content analysis. This mainly has a descriptive and exploratory orientation. Although producing the outcomes from qualitative content analysis is challenging due to the difficulties of generating the sense of enormous amounts of data, it is considered a common practice to utilize typical quotations to justify conclusions (Schilling, 2006). The general inductive content method provides a productive way of analysing qualitative data for several research aims (Thomas, 2003). Therefore, a classification technique has been applied to analysing the qualitative data for this study (Blandford et al., 2008). Silverman (2006) pointed out that the content analysis is a well-accepted technique of textual inquiry, and it has advantages for qualitative research, one of them is that utilising this type of analysis allows the results to be summarized while retaining the raw data. It is important regarding classifying the participant's responses into meaningful classes or coding data through creating words that are used to categorise responses, and then counting the number of instances that are linked to each class. This would help to stimulate original insights. Thus, in this research study, a coding technique was developed inductively to generate categories from the data.

3.8 The developed Methodological Framework

In respect to the methods adopted for conducting the three studies in this thesis, which were reported previously, they are considered to be represented in this section as the overall process framework for evaluation and improving usability of current systems in government organisation. The following Figure (3.2) shows the usability methodological framework.

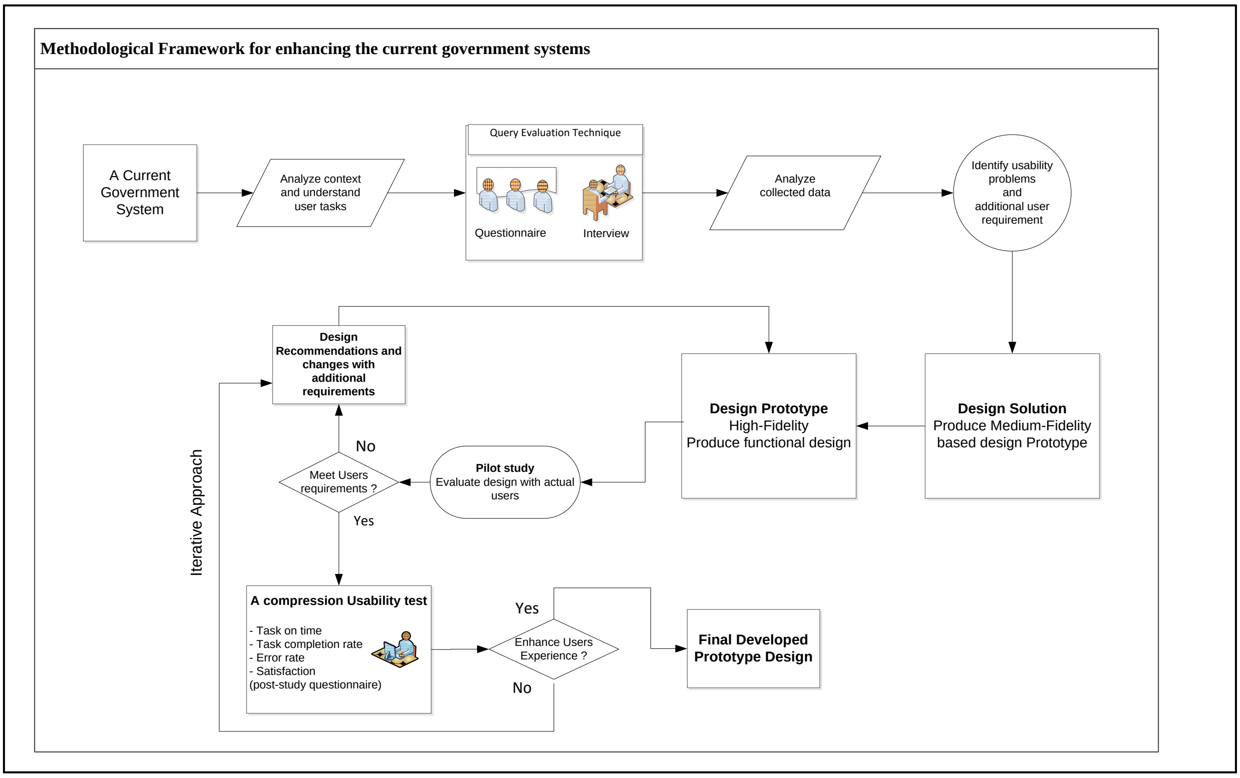


Figure 3.2 Methodological Framework of Government Systems Usability Enhancement

3.9 Chapter summary

This chapter on methodology has explained the general methodologies and techniques utilised for the work conducted in this thesis, and presented a justification for those chosen, besides how the overall system development process was planned according to an HCD approach. Firstly, an overview of the programme of research was presented and described, besides the map of the study. It was followed by a section that explained the targeted current system, which was intended to be evaluated though out proposed the three stages which represented each study within the overall research study. After that, the next section presented the data collection procedures and instruments, in addition, a description of the data analysis procedure and the tests conducted to draw conclusions from the studies were offered. Finally, the methodological framework created for evaluating and developing the system usability in government was provided.

Chapter 4: Query technique based Usability Evaluation of a current system in a government organization from a user perspective

4.1 Introduction

With regards to the agreement on the importance of usability as a key quality factor of software systems among the developers and users (Jayaletchumi et al., 2014; Bevan,2009; Harston et al. 2000), performing a usability evaluation has achieved much popularity as an essential factor in the success of a system (Nielsen 1994; Maguire 2001; Casalo et al 2010), that will help in developing it based on seeking to meet the users' needs and additional requirements (Liu, 2008). Initially, the usability evaluation can be defined as a systematic process of gathering data for the purpose of having a better understanding of users' attitudes and how they use the software product to perform a particular task under specified conditions. (Sharp et al.,2007; Preece et al.,1994).

Since the main object of a software system is to allow a user to perform different tasks in achieving specified goals within a certain application domain, therefore, the user view is the most common perspective apropos of quality as fitness for purpose. According to ISO 9126 (cited in Al-Badreen et al, 2010),the major concern that the users have is the software's usability, and the effects of using it, without the need to know the internal components of the system, how it works, or how it was developed. In addition, the users who will handle and work with a system, their attributes, specific knowledge and requirements, have become very crucial components in increasing user friendliness (Nielsen,1993) and overcoming challenges in designing and improving a system (Costabile 2001; Abran et al. 2003). Therefore, in software development, the users' requirements must be taken into account and have serious attention (Al-Baderen et al.,2010). Furthermore, in relation to the users' point of view and referring to any interactive software system, the user interface (UI) is considered the most crucial component of the system due to it being such a evident front-end part of the system that users can see and work with while using the system, and also can do a primary evaluation on. It is defined by Norman and Sproull (1979) as an essential part of the software system that links the user to the computer and then enables him to control it. Consequently, taking into account the users' requirements in the system development would create successful user interfaces and usable interactive systems (Costabile,2001). On the other hand, a poor UI design could have contributed to the challenges that users might face. Although there is no direct evidence in recent years to propose that poor UI design was to blame, it is very likely that UI problems contributed to the challenges that users had with the systems (Stone et al.,2005).

The issue of the lack attention to usability can be found in the government sector, and maybe it related to the lack of literature concerning the association of usability and government systems. Buie and Murray (2012) ratified this statement and found a shortage of search results on "usability and government",with only 89 titles, which is less compared to other figuresfor search results in several terms, e.g. " user experience" with 4275 results, and "government systems" with 106.957 results. However, this is very apparent in the developing countries, this could be due to several reasons, for instance, lack of involving the system users within the developing process.

Due to the fact that the term Usability deals with the entire user-system interaction, the user interface and the coordination of reciprocation information between the user and the system (or interaction) (Juristo et al.,2007) the research in this study is situated within the fields of Usability Engineering (UE) and Human-Computer Interaction (HCI), so that UE is concerned with the systematic integration of methods and techniques for producing a developed and usable system (Gediga and Hamborg, date?), while the HCI is "the *study of how people interact with computers and to what extent computers are or are not developed for successful interaction with human beings"* (Bell College, 2004, p.2).For achieving more tangible findings from real world, which related to the interest of this research, it endeavours to present a usability evaluation of the current VI system, which has been defined in chapter 3, by utilizing mixed methods of quantitative and qualitative in order to achieve the aims of the study. Through a query technique (the questionnaire and the interview) more information can be gathered from actual users by requesting them to fill out a form of a questionnaire or by direct communications (Maguire, 2001) and to answer a number of questions, which makes it “a powerful supplement to other techniques” (Freiberg and Juristo, 2008, p.5). Nielsen (1993) stated that the questionnaire and interviews are beneficial instruments for measuring how users use the system and what attributes they are especially concerned to improve (Nielsen, 1993). In addition, the questioning can be conducted online with the advantage of targeting potentially more diverse users and a superior number of users at lower cost.

This assessment will be from the actual users' perspective, that is employees who have access to the VI system. Moreover identifying challenges within the system would reflect the requirement to conduct the improvements process in order to meet its users’ needs and additional requirements. This in turn will enhance the quality of services which are presented to the public via issuing Visas to citizens. In addition, through their feedback an analytical approach and discussion part will be conducted to provide and approve the points to improve the ease of use of this system. Thus, the outcomes of this study aim to response to RQ1 and RQ2 which reported in chapter (1).

4.2 Aim and Objectives

As it considered earlier in chapter (1) for conducting the research work of this study, initially, it aims to employ a summative usability evaluation of a current internal system, which is implemented in a governmental organization in Saudi Arabia, in order to detect the weaknesses and strengths of the VI system for the purpose of improving its ease of use.

In order for the aim of the study to be achieved, some specific objectives are followed:

- For applying a survey approach to evaluate the VI system’s usability through utilizing different tools, such as, a survey, free textbox, and interviews. The aim of using these methods is to:
    - Identify the existing usability problems of the VI system;

- Identify the existing additional user requirements that could help to makeup a suggested design in a simulation system that would represent the new developed UI of the VI system.

4.3 A consolidated usability dimensions

4.3.1 The baseline usability dimensions

This section of the research presents a proposal for improving the dimensions of the 5E's of Usability, which include Effective, Efficient, Engaging, Error tolerance , and Easy to learn) (Quesenbery,2001), as the preferred usability model as the baseline for the research work of assessing the usability of the VI system. Bevan (2001) stated that the usability is not an intrinsic aspect of the system, but it rather such a capability to be utilised in the particular context of "users, tasks, environments". Therefore, Quesenbery (2001) developed her model originally through selecting and expanding the ISO 9241 characteristics of the usability model as the basic architecture for the consolidated model. This helped her to understand in which context particular qualities that bear on the effort are needed for use. In other words, it helps to explain how to recognize the information that has to be considered when evaluating usability in relation to measuring user performance and satisfaction (Abran et al.,2003). After that, a model which "*explained the requirement of an interface design that must be easy to learn, remember, and use, with few errors for its implied users and the tasks that it is assigned to use*" (Madan and Dubey, 2012,p590)

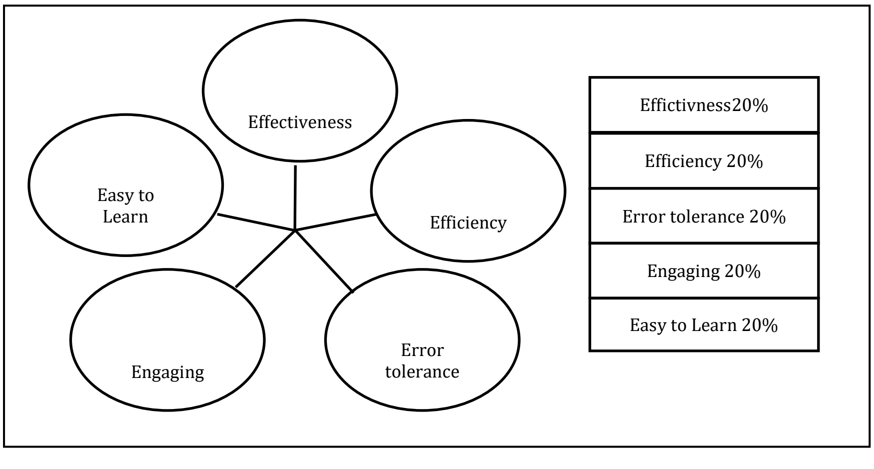


Figure 4.1 The five dimensions (5Es) describe different aspects of usability (Quesenbery,2003)

4.3.2 Improved usability dimensions

Although the dimensions described by Quesenbery were selected and adopted for this study as measurable criteria of users' experience, performance, and use of the system, however in this assessment, it is obvious that some issues have not been taken into account that would help during usability evaluation of an interactive system. For example, the lack of measuring the degree of the user's feeling that he gains ability to control the system adequately, and the degree of determining to what extent the interactive system is self-explanatory, besides the suitability of the help system (Kirakowski and Corbett, 1993).

Based on the instrument of the System Usability Measurement Inventory (SUMI) questionnaire standard, the study questionnaire has been adopted and modified in order to be compatible with this study for the purpose of collecting the data. It includes five attributes of perceived usability (efficiency, effect, learnability, control, and helpfulness). These dimensions are primarily providing a measure for users' perceptions of the quality of usage of a system (Macleod, 1994). Thus, from the previous clarification mentioned in chapter 2 section , and the chosen Quesenbery model, control and helpfulness are the two characteristics that would be associated with the system, and then should be included in the research model.

In order to provide a comprehensive model of usability evaluation a model should comprise dimensions that are linked to both the process and the system (Abran et al.,2003). After the selection of the E5 dimensions of Quesenbery was made to be used as the baseline for the research dimensions, the other two relevant usability dimensions, namely Control and Helpfulness, added and integrated into the selected baseline model, to be suited to the nature and aim of the study. The following Figure (4.4) presents the enhanced usability dimensions model for the study.

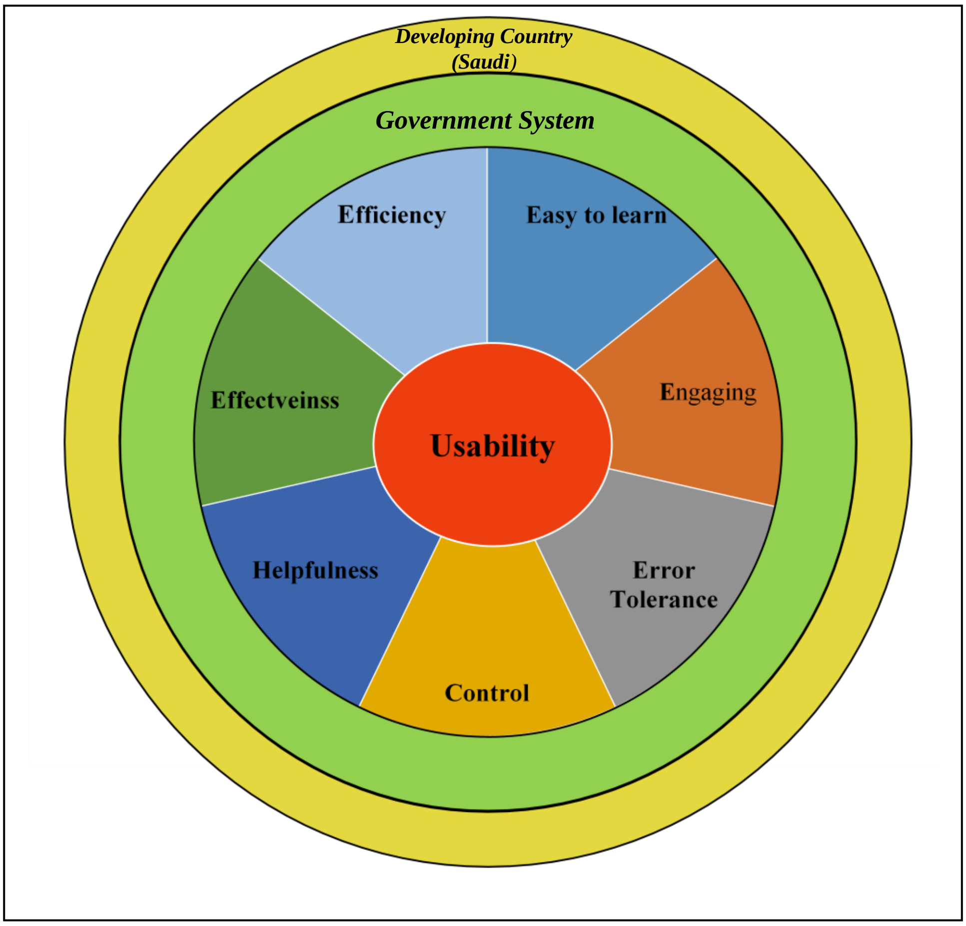


Figure 4.2 Consolidated Usability Model for the study

Although it is important to specify usability dimensions within the evaluation process in order to evaluate the UX of the quality of the system, it is obvious that there is confusion and intersection between some usability dimensions and quality attributes of the system. Accordingly, this issue would rise the importance of adopting a comprehensive quantitative and qualitative evaluation methods which can deal with both usability and quality evaluation techniques (Nikov et al., 2006).

4.4 The study instruments, techniques and procedures for collecting data

As it presented earlier in chapter three, in this study, a query evaluation technique was conducted with a number of instruments to evaluate the usability of VI system in order to obtain information from the actual users of the system. Although, the qualitative and quantitative data are equally important because they present unique insights into a UI design's strengths and weaknesses (Hix et al, 2004), the quantitative data individually presents limited constructive information for the researcher to understand. On the other hand the qualitative data are not always collected as objective data, so combining both methods might be a way for them to complement each other very well (Dag et al., 2001). Zazelenchuk (2008) pointed out that a mixture of quantitative and qualitative methods is widely accepted by the user-centred design community.

In general terms, quantitative collection techniques are utilized to gather measurements of performance, beliefs and subjective attitude information. In our study the main purpose of applying this method was to collect the actual users' opinions of the current VI system in order to quantify the information obtained.

4.4.1 Questionnaire

According to Hamborg et al. (2004) *"Questionnaires are well suited for the summative evaluation of software applications, especially in larger organisations like hospitals, public administrations etc".*

4.4.1.1 Developing the Questionnaire Design

The questionnaire design was adopted based on the Software Usability Measurement Inventory (SUMI) questionnaire, which is a common instrument and a global measurement of usability to extract the user's opinions on specific statements based on his experience of using the system. Furthermore, this questionnaire standard is recommended to any organisation which is attempting to assess the perceived usability as a quality of the use of a system. It can be carried out by either a system developer, a consumer of the system, or as a purchaser/consultant. What is more, it is increasingly being utilized to set quality of use requirements by system producers (Kirakowski and Corbett, 1993; McSweeney,1992). Precce et al.(1994) stated that SUMI is a standard method for evaluating users' attitudes. Essentially, it includes a validated 50 statements attitude questionnaire (Macleod et al,1997), and each statement is ranked on a three scaled Likert measure: agree, undecided, or disagree (Stone et al.,2005). These statements concern the following attributes: efficiency, effectiveness, learnability, control, and helpfulness.

In respects to the research objectives and the nature of the existing VI system, the questionnaire was slightly re-worded and modified to cover all the aspects of the VI system for the project, and the related information that was desired in the same area. Furthermore, Cairns and Cox (2008) indicated that it is widely believed a questionnaire should not be over long, due to participants being likely to complete it with less accuracy if they feel rushed to finish it; also they would skim reading it rather reading it carefully (Cairns and Cox, 2008). Therefore, the initial questionnaire has been designed to consist of 31 statements; some were taken from the original context of the SUMI questionnaire and the rest were taken from relevant studies, it is intended to provide relative information on each dimension of the study. In addition, as this research aims to gather users' opinions, attitudes, and behaviours, it is considered as a pathway within the human-computer interaction (HCI) field, therefore, a five-point Likert scale was chosen as this type of questionnaire design (Rogers et al.,2011), is commonly used in this area of HCI research (Love,2005). Furthermore, Taylor and Heath (1996) indicated that the Likert scale has become one of the leading methods of measuring social and political attitudes, because it would naturally maintain the direct involvement of the respondent, and has shown a high degree of validity and reliability (Coolican,2004). In order to ensure that the questionnaire had checks and balances and was an unbiased evaluation, some negative types of questions were included within the survey. Love (2005) reported that it is vital to mix-up the order of the positive and negative statements in the Likert scale within the questionnaire, in order to reduce the risk of "constant error" which is caused by the acquiescence effect. In other words, including a mixture of positive and negative statements within the questionnaire would help to control for any possible acceptance effect from the participants when they were filling out the questionnaire. Furthermore, Rogers et al., (2011) mentioned that some questionnaires were designed with a mixture of positive and negative statements, in order to check whether a user has good intentions or not for responding to the questionnaire. Therefore, some re-organizing has been done of the statements in order to obtain the final questionnaire. In addition, one type of question was used within this questionnaire, factual opinion and attitude questions with rating scales (Adams et al., 2008), and it also involved a free text box following each statement, to encourage a participant to record details regarding a particular statement or any related issues that need to be reported. Thus, the questionnaire was composed of mixed techniques for collecting quantitative and qualitative data (Appendix C).

4.4.1.2 A Pilot Study for the study questionnaire

A simple definition of a pilot study is that it is a small study for helping to design a further confirmatory study (Arnold et al., 2009). Thabane et al. (2010) define it as *"a trial study carried out before a research design is finalised to assist in defining the research question or to test the feasibility, reliability and validity of the proposed study design".*

In this study, after the primary questionnaire was built and designed, it was reviewed by three experts, one in the HCI area, and two from the UE field, in order to find any parts of the questionnaire that needed to be changed and improved. Next, the questionnaire was translated into the main language that the sample study spoke, which was Arabic. Afterwards, a pilot study was adapted in order to improve the quality of it, spot difficulties that the participant might face when they were filling it out, and verify the clarity of its appearance in the context of being understandable by the participants for completion (Moore and Benbasat, 1991). Hence, this feasibility study was prepared and conducted with 15 participants (employees from different departments who have an access to use the VI system). The questionnaire was built using online, and sent as an attachment to them through emails. After that their feedback has been received by emails, phone calls, and through the Skype software application. According to their responses some questions were edited and modified as most of the participants suggested that it ought to be clearer. For example, the question number (2) has been modified because they mentioned that it had a similar meaning to another question number (1); they were respectively "Learning to operate this system is difficult" and "It is easy to learn how to use this system". So the question number (2) became "I quickly became skillful at operating this system". Furthermore, some participants explained that the questionnaire needed to cover a major aspect of the system related to the amendment ability, so a statement number (16) which says that "the amendment ability of the system affects my ability to complete a task", has been added for this purpose. Also, in relation to another question number (24) regarding whether the users have an opportunity to recover any errors or mistakes while they are using the system, the participants queried if this would be before or after finishing the task. Therefore, the question has been edited to be, "When I make a mistake before I finish a task while using the system, I can recover it easily and quickly”.

4.4.1.3 Measure of Reliability and Validity

In order to examine the reliability and validity of the questionnaire a Cronbach’s Alpha coefficient, which normally ranges in value from 0 to 1 to evaluate the reliability of the scale for the statements, was used after conducting the pilot study. It is an examination of reliability technique that requires a single test administration aiming to provide a unique estimate of the reliability for a given test (Glim and Glim,2003). George and Mallery, (2003,p.231) presented the following rules of thumb: “_ > .9 – Excellent, _ > .8 – Good, _ > .7 – Acceptable, _ > .6 – Questionable, _ > .5 – Poor, and_ < .5 – Unacceptable”.
Thus, we used a Cronbach’s Alpha coefficient to evaluate the reliability of the scale for the included items (statements), and we obtained the following results.

Table 4.1 Cronbach's Alpha Value before pilot study

Construct	No. of Items	Cronbach’s Alpha
Effectiveness	3	0.796
Efficiency	4	0.871
Easy to learn	5	0.687
Engaging “Satisfaction”	6	0.824
Error Tolerance	5	0.709
Helpfulness	4	0.762
Control	4	0.793

According to the results obtained after applying the pilot study and calculating Cronbach’s Alpha, it can be seen that the value of Cronbach’s Alpha of each of the dimensions was 0.7 or more, except one dimension was less than 0.7. Afterward, the procedure of testing the internal consistency of each statement was applied. This procedure is applied as follows: we remove one statement (item) and calculate Cronbach’s Alpha for the remaining statements. If the calculated Alpha is more than the Alpha for all statements, this means that the reliability has increased and therefore we remove this statement. But if the calculated Alpha is less than the Alpha for all statements, this means that the reliability has decreased, therefore we let the statement remain. We repeat this procedure for each statement. Consequently, some statements have been removed because they were inconsistent and invalid, so their correspondence Alpha was more than the Alpha for the whole scale and the correlation was very small and insignificant. For example, we removed the statement number (16) from the diminution, which stated "Getting data files in and out of the system is not easy". This had a Corrected item-total Correlation (-0.120) with Cronbach's Alpha (0.635). Another question was also deleted for the same reason, one that stated "Working with this system is not mentally stimulating", whereas other statements have been retained.

4.4.1.4 Final Questionnaire Design

The final version of the questionnaire was initially divided into two sections; the first section was concerned with the classes of demographic independent variables (age, level of education, and experience). It should be noted here that, as was mentioned before in the section on sample characteristics, all the selected users were male, thus, the question regarding gender has been omitted. The second section of the questionnaire comprises all the dimension statements. In line with the previous explanation and information, the questionnaire has been improved and edited to include only 29 questions (Appendix B). So, we repeated the collecting of the Cronbach's Alpha. Table (4.2) presents the reliability for each dimension, along with its interpretation. A high Cronbach alpha value for all seven dimensions indicates that they are internally consistent and measure the same content of the dimension (Al-Shafi and Weerakoody,2010).

Table 4.2 Cronbach's Alpha Value for the final Questionnaire

Construct	No. of Items	Cronbach's Alpha	Type
Effectiveness	3	0.796	Good reliability
Efficiency	4	0.871	Excellent reliability
Easy to learn	4	0.711	Good reliability
Engaging "Satisfaction"	6	0.824	Excellent reliability
Error Tolerant	4	0.735	Good reliability
Helpfulness	4	0.762	Good reliability
Control	4	0.793	Good reliability

4.4.1.5 The procedure for collecting data by questionnaire

Due to the fact that the sample was diffused in different areas of Saudi Arabia, and for the purpose of obtaining quick access to the whole targeted population of users of the system, the final questionnaire was employed as the proper tool designed to measure users' attitudes toward the system usability. It was targeted at all the users. Thus, at the beginning of this stage and after the approval was obtained for conducting the survey among the users, the final questionnaire in this study was built and distributed online though creating a link by using QULITRICS online survey software. Rogers et al. (2011) pointed out that using an online questionnaire would be a good way of accessing a large and geographically dispersed sample quickly. Furthermore, an online web based questionnaire would avoid several issues, such as the high cost and effort of printing hard copies, and then distributing them, and lastly collecting the completed paper forms. This QULITRICS software enables the users (employees) to answer the questionnaire via their Smartphone besides PCs and laptops. However, due to the difficulty that was faced in reaching some of the participants' contact email addresses; we needed to distribute some of the questionnaires using a conventional method. Therefore, via the formal postal service of the Ministry of Labour the questionnaire was sent and delivered as a hard copy to those users (employees) whose contact emails could not be obtained, who represented approximately less than half of the users.

4.4.2 Free Textbox

Furthermore, in order to gather more details about users' opinions, the free textboxes (Comments in Survey) were added below each question within the questionnaire, so the respondents would be able to write their own answers in natural language (Nielson,1993), and provide their opinions in the form of free text (Dix et al, 2004). Furthermore, Stone et al. (2005,p.36) indicated that generally questionnaires which contain a mixture of both closed and open questions are more effective. However, in relation to this issue Nielsen (1993) mentioned that users usually do not write their specific responses, thus, in order to tackle this matter a statement has been added at the beginning of the questionnaire showing that providing their answers in a detailed form will be worthwhile for the research and they would thus add more value to the final results.

4.4.3 Interviews

As the main object of this study is to evaluate the usability of a system, and gather as much comprehensive and accurate data as possible, interviews were performed as they are considered to be a crucial tool for collecting data (Myers and Newman,2006), and to yield flexible, in-depth information regarding the users' attitudes and experiences with a system (Nielsen,1993). They were also aiming to obtain more information that could shed more light on the interpretation of the questionnaire's results. In addition, conducting interviews with different users who were holding different occupation and positions within the organization would be helpful for gaining a proper understanding of the subject and some issues that are raised by the natural social context. (Marshall and Rossman, 2006).

4.4.3.1 The procedure for collecting data by Interview

In this study, semi-structured interviews were carried out with eight employees in different departments; however, all of them had access to the VI system and used it in their daily business. The five interviews were conducted by attending physically in the work place with the two heads of the departments, and three employees served as inspector assistants. Field notes were taken during the interviews for recording important information. The other three interviews were conducted with ordinary employees; two of them were in the section for dealing directly with the applicants. Of these interviews, two were conducted through utilising one of the social network (Skype) applications, and one interview was conducted by phone, as in some cases this can be successful. The interview questions comprised open-ended questions, which were designed to obtain further rich information and details on the practice and Usability of the VI system; moreover, using this approach would decrease the ambiguity of some issues which were raised within quantitative analysis, and other requirements we might come across for enhancing the user's usage of the system. The following presents an example of the open-ended questions that were asked in the interviews in this study; some emerged from the questionnaire results (Appendix D) :

- Do you think a task takes more time than expected? If yes, why?
- Where are users likely to get stuck?
- What mistakes do users make "particularly in data entry"?
- Do you mind telling us, what is your suggestion to enhance the performance of the system through its interface for doing the task?
- Does ease of use aspect of the VI system make an impact on the employee performance?
- Do you agree for using this system to issue visas?

4.5 Analysing and Interpreting Data

This chapter section discusses the data analysis methods used for the research. In the light of Usability Evaluation, analysing the data for this study can be described as altering raw (qualitative and quantitative data), that was collected from the employees as users of the system, into outright results which can be used to provide recommendations for improving the usability of the system. Accordingly, quantitative and qualitative techniques were employed for analysing the data collected.

4.5.1 Statistical Analysis of Quantitative data

The analysis of quantitative data primarily used the following techniques for the statistical analysis:

1. Descriptive statistical analysis of the general variables is provided for helping to illustrate and understand the data. It makes use of average (mean), standard

deviation, percentages, diagrams and frequency distributions for each independent variable (the biographical data).

2. Descriptive statistics for the questionnaire's statements in the form of a Likert scale are applied for the purpose of creating a score for each dimension in the research in order to identify the state of it. Scoring of the negative statements' responses was reversed, thus, a high score revealed a more positive opinion about an individual dimension. This was represented through an average rating score across all the statements.

3. The main quantitative analysis tool used in this research is multiple regression analysis which is applied to estimate a continuous dependent variable from several independent variables. The general aim of applying it is to learn more about the relationship between a number of independent variables and a dependent variable, and which has more influence on them. In addition, the one-way analysis of variance (ANOVA) is employed earlier as it is commonly appropriate in many research situations, in particular when the study has three or more groups, and it is necessary to figure out if those groups differ significantly on any outcome (Acock, 2010). . Brace et al. (2012, p.209) stated, *"In ANOVA we are trying to determine how much of the variance isaccounted for by our manipulation of the independent variables (relative to thepercentage of the variance we cannot account for). In multiple regressions, we do notdirectly manipulate the IVs but instead just measure the naturally occurring levelsof the variables and see if this helps us predict the score on the dependent variable".* The ANOVA one-way was followed by Scheffe's test which is used for post hoc comparison analyses.

a. Multiple linear regression analysis

The regression analysis is the statistical method for identifying the relationship between two or more variables; a dependent variable, and an independent variable or more than one variable which can be defined here as multiple regression (Plallant,2007). It is fundamentally an extension of predicting one variable on the basis of a number of other variables. Moreover, the purpose of this technique is to illustrate the statistical dependence of one variable on other variables. Having more than one independent variable is useful when estimating users' behaviours.

Furthermore, it provides an equation that predicts one dependent variable from two or more independent variables, it is as follows:

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \varepsilon$$

where *y* is the dependent variable, representing a quantity that varies from individual to individual throughout the population, and is the primary focus of interest. X_1,..., X_k are the independentvariables(or they are called "predictor variables"), which also vary from one individual to the next, and are thought to be related to *y*. β0, β1, . . . , βk: unknown constants ("the coefficients"). Finally, ε is the error of observation, in other words the residual term, which represents the composite effect of all other types of individual differences not explicitly identified in the model.

b. *F*- testAnalysis of Variance (ANOVA), For more than two independent samples.

This is used to test and determine whether there is a significant difference between the means of the responses of more than two independent populations (k). In the case of this study, the results were calculated for whether there was a significant difference between responses according to all independent factors (Age, Experience, and Level of Education), and if these factors had any effect on the user's attitude regarding the usability of the internal VI system.

c. The Scheffe test was used for Post Hoc Comparison analyses.

Owing to the analysis of variance (ANOVA) results indicating that there is a meaningful difference attributable to the treatment levels, and due to the fact that, the application of this analysis between the averages of the samples would lead to increasing the possibility of falling into error, therefore the Scheffe test was used for comparing the averages of several samples, in order to prevent falling into several errors, of which one was mentioned above. It is several t-test between each two classes in the independent factors. In addition, in the case of this study where the samples are unequal, Scheffe's test would be a proper way of doing this.

4.5.2 Content Analysis of Qualitative data

The analysing of the qualitative results obtained from the responses on the free textbox within the questionnaire and the interviews were conducted as thematic inductive analysis by focusing on the summative approach to qualitative content analysis, which mainly has a descriptive and exploratory orientation. According to Graneheim and Lundman (2004) the content analysis concerns the manifest content, which means the collected transcription text is analysed to explain the apparent and obvious components. Although presenting the outcomes from qualitative content analysis is challenging due to the difficulties of generating a sense of enormous amounts of data, it is considered a common practice for utilizing typical quotations to justify conclusions (Schilling, 2006). The general inductive content method provides a productive way of analysing qualitative data for several research aims (Thomas,2003). Therefore, a classification technique has been applied to analysing the qualitative data for this study (Blandford et al.,2008). Silverman (2006) pointed out that the content analysis is a well accepted technique of textual inquiry, and it has advantages for qualitative research, one of them being that utilising this type of analysis allows the results to be summarized while retaining the raw data. It is essentially about classifying the participants' responses into meaningful classes or coding data through creating words that are used to categorise responses, and then counting the number of instances that are linked to each class. This would help to stimulate original insights. In this research, a coding technique was developed inductively to generate categories from the data.

4.6 Results

This section contains the findings from both the methods, quantitative and qualitative, which were used in the research. For easy referencing and clarity, the research analysis and results are presented in different sections. Part 1 covers the quantitative data from all the demographic questions (independent variables), and similarly, from the dimensions (dependent variables) of the study. Likewise, the qualitative data obtained will be presented in Part 2 as a basis for findings to justify and complement the quantitative data. Then, the usability issues are discussed and presented in Part 3 with exceptional consequences.

4.6.1 Quantitative data (Questionnaire results)

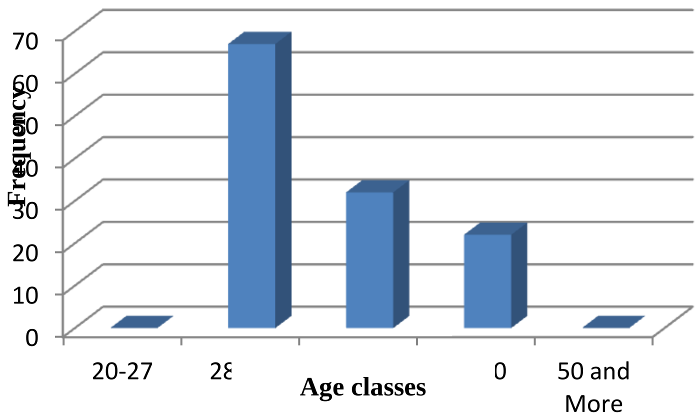
A total number of 122 out of 135 questionnaire responses were received from the users of the system (76.57% response rate). More than half of these responses, 65%, were completed online, and the rest of them were filled out manually, because the respondents received it as hard copy by the formal post. As mentioned previously each Usability attribute was ranked using a five scale; '1'indicates 'Strongly Disagree', '2' is for 'Disagree', '3' is for ' Neutral', '4' is 'Agree', and '5' indicates 'Strongly Agree'.

4.6.1.1 Descriptive statistics for the general variables

This section shows the distribution of the sample according to all independent variables in the demographic section.

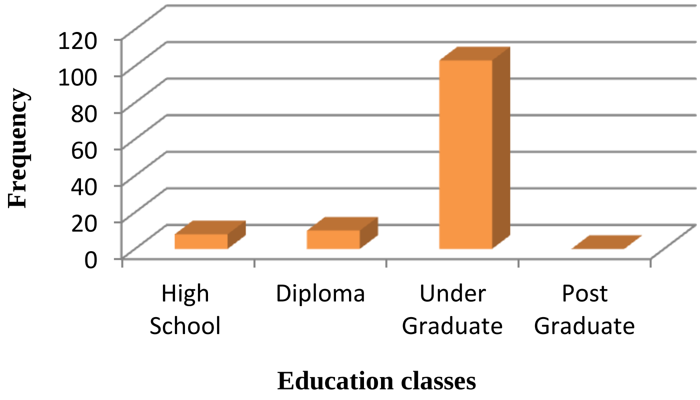
a) - Age

Figure 4.3 Frequency of the sample according to Age classes



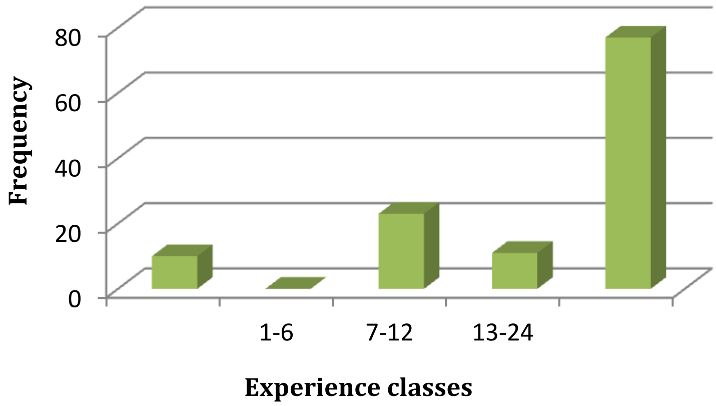
b)- Level of Education

Figure 4.4 Frequency of the sample according to Education classes



(c) – Experience

Figure 4.5 Frequency of the sample according to Experience classes



The Figures (4.5, 4.6 and 4.7) above display the results of respondents forall the classes of the demographic independent variables (Age, Education, and Experience). The average age was between 28 and 35 years old, and the qualification of the vast majority of the participants was undergraduate level; most of them had been working at the Recruitment department in the Ministry of Labour and using the system for more than one year, with a percentage of 63.6%. It should be noted that, due to some classes having a small number of respondents which was not adequate to represent the insight of the class to which they belong, it was merged with the closest one. For instance, only 4 participants were aged between 20 and 27 years, so we merged this class into the nearest Age Class, between 28 and 35 years, so the total of participants for this class became 67 participants. In the same way, in the Experience dimension, the class of 1 to 6 monthsexperience in using the system included just 3 participants, thus we moved them to another class, which was 7 to 12 months.

4.8.1.2 Descriptive Statistics for Questions in the form of Likert Scale

As previously mentioned, the form of the Likert scale has been utilised for the responses to identify the final opinions on the statements of all dimensions of the study, therefore, we need to calculate the weighted mean of the responses to the statements on each dimensions which reflects its importance, and the responses of each statement take a weight as follows:

Table 4.3 Scoring worded statement in a Likert-Scale questionnaire

Response	Weight
Strongly Disagree	1
Disagree	2
Neutral	3
Agree	4
Strongly Agree	5

The above procedure in (Table 4.3) determines to which class the response of each statement belongs. According to the value of the weighted mean, each dimension has been explored and measured overall by the following (Table 4.4), which shows the criterion of Likert Scale Ranges (Statistical Range).

Table 4.4 The criterion of weighted mean in Likert-Scale

Response	Weight Mean
Strongly Disagree	From 1.00 to less than 1.80
Disagree	From 1.80 to less than 2.60
Neutral	From 2.60 to less than 3.40
Agree	From 3.40 to less than 4.20
Strongly Agree	From 4.20 to less than 5.00

The classification of the responses in the following (Table 4.5) was calculated by using the statistical ranges and given a rank priority according to overall response in the mean.

Table 4.5 Overall response in the mean

Dimension	Statement	Weighted Mean	Std. Deviation	Overall Response (in mean)	Priority
Effectiveness	1- It is easy to follow the instruction to complete tasks	3.78	0.89	Agree	1
	2- The system quickly completes the tasks I want to do.	3.30	1.01	Neutral	3
	3- I could not effectively complete my tasks using this system	2.60	0.81	Neutral	2
Total		3.23	0.93	Neutral	
Efficiency	4- The system requires many steps to complete a task	3.98	0.97	Agree	4
	5- It is relatively easy to move from one step of a task to another.	3.59	0.96	Agree	2
	6- It takes a long time to complete a task using the system	2.55	1.03	Disagree	3
	7- The integration ability of the system is a crucial factor for its efficiency.	4.22	0.95	Strongly Agree	1
Total		3.58	1.26	Agree	
Easy to Learn (Learnability)	8- It was easy to learn how to use this system.	4.22	0.69	Strongly Agree	1
	9- Learning how to use the new functions of this system is hard.	2.79	1.03	Neutral	3
	10- The system uses a consistent navigational system to enable me to understand where to find information.	2.94	1.01	Neutral	4
	11- As long as I am following the system's procedure, I can get the necessary information and knowledge to complete my task.	3.36	0.87	Neutral	2
Total		3.33	1.05	Neutral	
Engaging (Satisfaction)	12- The data interface is not always consistent with the instruction menu	2.84	1.02	Neutral	6
	13- The interface of this system is well designed.	3.43	0.97	Agree	2
	14- The features of the system are very limited.	2.67	1.00	Neutral	3
	15- It is obvious that the end-user's needs have not been fully taken into consideration when the system was developed.	3.67	1.00	Agree	1
	16- I believe using this system makes me more productive in my work	2.96	1.02	Neutral	5
	17- Overall, I am satisfied with using this system for issuing work visas.	3.30	0.91	Neutral	4

Total		3.25	1.05	Neutral	
Error Tolerant	18- The system does not give error messages that clearly guide me on how to fix problems before I finish a task.	2.76	1.18	Neutral	2
	19- The system has helped me to overcome any problems I have had after I finish a task.	2.56	0.92	Disagree	4
	20- The system has at some time stopped unpredictably and it is not easy to restart it again.	2.60	1.12	Neutral	1
	21- When I make a mistake before I finish a task while using the system, I can recover it easily and quickly.	2.88	1.06	Neutral	3
Total		3.02	1.08	Neutral	
Helpfulness	22- The system does not support basic standard actions, such as save, copy, cut and paste.	3.35	1.13	Neutral	2
	23- There is enough information on the system to help me complete the tasks (such as online help, and other documentation)	2.44	1.18	Disagree	4
	24- The system enables me to provide good quality service to the citizen	3.57	1.02	Agree	1
	25- I think I would need Technical Support regularly to be able to use this system.	3.48	1.03	Agree	3
Total		3.21	1.20	Neutral	
Control	26- I quickly became skilful at operating this system.	4.06	0.71	Agree	1
	27- The data entry of the applicant into the system to issue work visas is hard and complicated	2.19	1.01	Disagree	4
	28- The amendment ability of the system affects my ability to complete a task.	3.07	1.02	Neutral	2
	29- I would not like to perform the task using this system	2.78	1.11	Neutral	3
Total		3.02	1.05	Neutral	

a. Effectiveness

This attribute was measured by three questions within the questionnaire. The overall responses are 'Natural' with a mean of 3.23. However, from the results of two positive questions it was clear that a significant number of responses indicate that the users (53.7% from Q5 and43.8% from Q8) 'Agree' about the effectiveness of the system and they had generally a positive feeling regarding its usage, so, the system helped them to complete their task quickly and they could easily follow its instructions.

b. Efficiency

The overall result from respondents on the system efficiency (measurement of the speed with which the users were-able to complete their task successfully) is "Agree" with a mean of 3.58. However, more than half of the users (70 or 57%) confirmed that they needed to operate many steps to complete a task, which would affect the speed of ending the task. In addition, a large number of respondents (57 or 47%) 'Strongly Agree' with the importance of the integration ability of the system for operating the task faster.

c. Easy to learn (Learnability)

The overall response to the statements on this attribute is ‘Neutral’ with a mean of 3.33. It is apparent that a significant number of respondents stated ‘Strongly Agree’ to the statement that it is easy to learn the system functionality (74 or 61.2%) with a high mean of 4.22.

d. Engaging (Satisfaction)

This attribute measures users’ impressions of the system and the degree to which that interface makes them satisfied with using it, and almost half of respondents confirmed that they were satisfied. There is also however a significant result for the negative statement which says that “It is obvious that the end-user’s needs have not been fully taken into consideration when the system was developed”, half of respondentssupported this statement. Regarding whether the system makes its users more productive in their work, 35 respondents ‘Agree’ while 30 respond with ‘Disagree’.

e. Error Tolerant

Error tolerance refers to the degree to which the system design helps users to prevent errors, and recover when they occur. More than 40% of users ‘disagree’ that the system would not provide help to overcome any problems after finishing the task. However, the total responses for this dimension are ‘Neutral’ with a mean of 3.02.

f. Helpfulness

The overall response for the system helpfulness attribute is ’Agree’. There is a great difference between the responses to the statement that the system helps the users to provide good quality service to the citizen through finishing their task successfully;a majority of responses (68 or 56.2%) ‘Agree’ while only 15(12.4%) respondents ‘Disagree’. In addition, 51 participants (42.1%) responded with ‘Disagree’ to the statement that the system has enough information for doing the task.

g. Control

In terms of the control creation, or the measurement of the feeling by the users that they control the system consistently, the overall response from the majority of the respondents was 'Neutral' with mean 3.02.

4.6.1.3 ANOVA analysis of variance

a. The Effect of Age

To test whether there is a significant difference between the respondents' opinions about all attributes of the study due to their being in different Age classes, the *F*- test (ANOVA) was employed, and the following table was obtained.

Table 4.6 F-test (ANOVA)

Dimensions	Age classes	N	Mean	Std. Deviation	*F*	p	Conclusion
1- Effectiveness	20 - 27	0	0.00	0.00	0.601	0.663	Not Significant
	28 - 35	67	3.43	0.71			
	36 - 42	32	3.56	0.63			
	43 – 50	22	3.55	0.71			
	51 and More	0	0.00	0.00			
2- Efficiency	20 - 27	0	0.00	0.00	2.399	0.050	Significant
	28 - 35	67	3.18	0.64			
	36 - 42	32	3.46	0.39			
	43 – 50	22	3.54	0.40			
	51 and More	0	0.00	0.00			
3- Easy to Learn (Learnability)	20 - 27	0	0.00	0.00	3.002	0.021	Significant
	28 - 35	67	3.35	0.59			
	36 - 42	32	3.69	0.40			
	43 – 50	22	3.43	0.55			
	51 and More	0	0.00	0.00			
4- Engaging (Satisfaction)	20 - 27	0	0.00	0.00	2.257	0.067	Not Significant
	28 - 35	67	2.93	0.63			
	36 - 42	32	2.98	0.38			
	43 – 50	22	3.11	0.51			
	51 and More	0	0.00	0.00			
5- Error	20 - 27	0	0.00	0.00	4.350	0.003	Significant

Dimensions	Age classes	N	Mean	Std. Deviation	*F*	p	Conclusion
Tolerant	28 - 35	67	2.85	0.69			
	36 - 42	32	2.70	0.55			
	43 – 50	22	2.49	0.54			
	51 and More	0	0.00	0.00			
6- Help fullness	20 - 27	0	0.00	0.00	1.679	0.160	Not Significant
	28 - 35	67	2.72	0.67			
	36 - 42	32	2.96	0.64			
	43 – 50	22	2.90	0.75			
	51 and More	0	0.00	0.00			
7- Control	20 - 27	0	0.00	0.00	1.358	0.253	Not Significant
	28 - 35	67	3.49	0.49			
	36 - 42	32	3.52	0.53			
	43 – 50	22	3.73	0.49			
	51 and More	0	0.00	0.00			

The results from the (Table 4.6) above show that the p-value is more than 0.05 for the dimensions of Effectiveness, Engaging, Helpfulness, and Control, which means that there is no significant difference between the individuals' responses related to their different ages.

However, for the other dimensions, (Efficiency, Easy to learn, and Error tolerant), the p-value is less 0.05, which means that there are significant differences between the individuals' responses due to their different ages.

The following (Table 4.7) highlights age classes. It shows only the significant differences on each dimension between the opinions of respondents in the Age class (I) and the corresponding respondents in Age class (J), while non-significant results have been omitted and not presented in the following table.

Table 4.7 Age classes which caused the significant differences

Dimensions	Ages classes (I)	Ages classes (J)	The difference Mean	The difference Std. Error.	p	Conclusion
2-Efficiency	36-42	28-35	0.28	0.12	0.019	Significant
	43-50	28-35	0.36	0.14	0.012	Significant
3- Easy to Learn (Learnability)	36-42	28-35	0.34	0.11	0.003	Significant
5- Error Tolerant	28-35	43-50	0.36	0.16	0.029	Significant

b. The Effect of Education

The following results in (Table 4.8) show the difference between participants' responses about all dimensions due to different Education classes.

Table 4.8 *F*-test (ANOVA)

Dimensions	Education classes	N	Mean	Std. Deviation	*F*	p	Conclusion
1- Effectiveness	1- High School	8	3.33	0.53	0.473	0.702	Not Significant
	2- Diploma	10	3.67	0.38			
	3- Under Graduate	103	3.48	0.70			
	4- Post Graduate	0	0.00	0.00			
2- Efficiency	1- High School	8	3.25	0.38	0.419	0.740	Not Significant
	2- Diploma	10	3.50	0.47			
	3- Under Graduate	103	3.31	0.58			
	4- Post Graduate	0	0.00	0.00			
3- Easy to Learn (Learnability)	1- High School	8	3.25	0.42	0.721	0.546	Not Significant
	2- Diploma	10	3.28	0.30			
	3- Under Graduate	103	3.46	0.54			
	4- Post Graduate	0	0.00	0.00			
4- Engaging (Satisfaction)	1- High School	8	2.88	0.69	0.133	0.940	Not Significant
	2- Diploma	10	2.93	0.64			
	3- Under Graduate	103	2.99	0.55			
	4- Post Graduate	0	0.00	0.00			

Dimensions	Education classes	N	Mean	Std. Deviation	*F*	p	Conclusion
5- Error Tolerant	1- High School	8	2.78	0.70	1.836	0.144	Not Significant
	2- Diploma	10	3.13	0.66			
	3- Under Graduate	103	2.67	0.66			
	4- Post Graduate	0	0.00	0.00			
6- Help fullness	1- High School	8	2.84	1.35	0.842	0.474	Not Significant
	2- Diploma	10	2.68	0.53			
	3- Under Graduate	103	2.78	0.64			
	4- Post Graduate	0	0.00	0.00			
7- Control	1- High School	8	3.66	0.57	0.565	0.639	Not Significant
	2- Diploma	10	3.70	0.40			
	3- Under Graduate	103	3.52	0.50			
	4- Post Graduate	0	0.00	0.00			

We notice from the above table that the p-value (the level of significance) in all dimensions is more than >0.05 , which means that there is no significant difference between the individuals' responses due to their different levels ofeducation. This would be related to the fact that the ability to operate the system depends on the training which is provided to the employees as users.

c. The Effect of Experience

To test whether there is a significant difference between respondents in regards to the factors related to different Experience classes, we used the F-test (ANOVA) and obtained the following (Table 4.9).

Table 4.9 *F*-test (ANOVA)

Dimensions	Experience classes	N	Mean	Std. Deviation	*F*	*p*	Conclusion
1- Effectiveness	**1- Less than 1 Month**	10	3.60	0.26	1.714	0.152	Not Significant
	2- 1-6	0	0.00	0.00			
	3- 7 -12	23	3.53	0.74			
	4- 13- 24	11	3.82	0.43			
	5- More than 24 Month	77	3.45	0.69			
2- Efficiency	**1- Less than 1 Month**	10	3.25	0.37	2.932	0.024	Significant
	2- 1-6	0	0.00	0.00			
	3- 7 -12	23	3.15	0.69			
	4- 13- 24	11	3.80	0.33			
	5- More than 24 Month	77	3.30	0.54			
3- Easy to Learn (Learnability)	**1- Less than 1 Month**	10	3.55	0.42	1.566	0.188	Not Significant
	2- 1-6	0	0.00	0.00			
	3- 7 -12	23	3.35	0.68			
	4- 13- 24	11	3.52	0.38			
	5- More than 24 Month	77	3.46	0.53			
4- Engaging (Satisfaction)	**1- Less than 1 Month**	10	3.10	0.41	7.584	0.000	Significant
	2- 1-6	3	0.00	0.00			
	3- 7 -12	23	2.83	0.37			
	4- 13- 24	11	3.62	0.48			
	5- More than 24 Month	77	2.94	0.55			
5- Error Tolerant	**1- Less than 1 Month**	10	2.85	0.49	7.832	0.000	Significant
	2- 1-6	0	0.00	0.00			
	3- 7 -12	23	2.98	0.62			
	4- 13- 24	11	3.05	0.71			
	5- More than 24 Month	77	2.63	0.60			
6- Help fullness	**1- Less than 1 Month**	10	2.93	0.33	5.863	0.000	Significant
	2- 1-6	0	0.00	0.00			
	3- 7 -12	23	3.03	0.75			
	4- 13- 24	11	3.20	0.38			

Dimensions	Experience classes	N	Mean	Std. Deviation	*F*	*p*	Conclusion
	5- More than 24 Month	77	2.72	0.68			
7- Control	**1- Less than 1 Month**	10	3.25	0.53	4.592	0.002	Significant
	2- 1-6	0	0.00	0.00			
	3- 7 -12	23	3.24	0.36			
	4- 13- 24	11	3.84	0.28			
	5- More than 24 Month	77	3.61	0.51			

The above Table (4.9) shows that the p-value is> 0.05 for dimensions (1) and (3) which means there is no significant difference between the individuals' responses according to their different Experience classes. The result shows that there are differences between the mean of factors (2-4-5-6 and 7), the p-value is less than 0.05, which means that there are significant differences between the individuals' responses due to their different Experience classes.

As in the analysis of the previous demographic dimension, we employed the Scheffe test in order to find out which classes caused the significant differences according to Experience. The following Table (4.10) presents the conclusions for the classes which have a significant difference only.

Table 4.10 Experience classes which caused the significant differences

Dimensions	Experience classes		The difference		p	Conclusion
	(I)	(J)	Mean	Std. Error.		
2-Efficiency	13-24	Less than 1 Month	0.55	0.23	0.021	Significant
	13-24	7 -12	0.65*	0.20	0.002	Significant
	13-24	More than 24 Month	0.50*	0.17	0.005	Significant
4- Engaging (Satisfaction)	13-24	Less than 1 Month	0.52*	0.22	0.020	Significant
	13-24	7 -12	0.79*	0.19	.000***	Significant
	13-24	More than 24 Month	0.68*	0.16	.000***	Significant
5- Error Tolerant	7 -12	More than 24 Month	0.35*	0.15	0.023	Significant
	13-24	More than 24 Month	0.42*	0.19	0.033	Significant
6- Help fullness	13-24	More than 24 Month	0.49*	0.21	0.022	Significant
7- Control	13-24	Less than 1 Month	0.59*	0.21	0.005**	Significant
	13-24	7 -12	0.60*	0.18	0.001**	Significant
	More than 24 Month	7 -12	0.38*	0.12	0.002**	Significant

4.6.1.4 The multiple linear regression output

A multiple linear regression analysis was applied to predict the influence of independent variables Age, Experience, and Education Level, on the dependent variables of the study. As shown in the following tables, the results of the regression indicated that each independent variable has a significant effect on a particular dependent variable. R^2 for the initial model and the change R^2 (denoted as $\triangle R^2$) are reported below each table (Field,2009)

Summary of Multiple Regression Analysis for Variables Predicting the Influence of Independent Variable (N = 121) - Tables (4.12 to 4.18)

1-

Table 4.11 ***Dependent Variable:*** **Satisfaction**

Independent Variable	B	SE B	β	t	Sig
(Constant)	2.788	.248		11.220	.000
Age	-.039	.048	-.080	-.823	.412
Experience	.192	.064	.271	3.002	.003**
Education	.009	.030	.028	.292	.771

Note R^2 = .085; $\triangle R^2$= .062

2-

Table 4.12 ***Dependent Variable:*** **Effectiveness**

Independent Variable	B	SE B	β	t	Sig
(Constant)	3.502	.394		8.898	.000
Age	-.051	.077	-.065	-.656	.513
Experience	-.175	.103	-.156	-1.706	.091
Education	-.041	.053	-.078	-.786	.434

Note R^2 = .035 ; $\triangle R^2$= .010

3-

Table 4.13 ***Dependent Variable:*****Efficiency**

Independent Variable	B	SE B	β	t	Sig
(Constant)	2.486	.410		6.067	.000
Age	.199	.080	.244	2.482	.014
Experience	.076	.107	.064	.714	.476
Education	-.102	.055	-.182	-1.856	.066

Note R^2 = .059; $\triangle R^2$= .035

4-

Table 4.14 ***Dependent Variable:*****Learnability**

Independent Variable	B	SE B	*β*	t	Sig
(Constant)	3.189	.287		11.096	.000
Age	.051	.056	.088	.910	.365
Experience	-.059	.075	-.070	-.784	.435
Education	.092	.038	.230	2.385	.019

Note R^2 = .085; ΔR^2=.062

5-

Table 4.15 ***Dependent Variable:*** **Error Tolerant**

Independent Variable	B	SE B	*β*	t	Sig
(Constant)	.211	.108		1.947	.054
Age	.900	.021	.982	42.441	.000***
Experience	.035	.028	.026	1.238	.218
Education	-.009	.014	-.015	-.652	.515

Note R^2 = .948; ΔR^2= .947

6-

Table 4.16 ***Dependent Variable:*** **Help fullness**

Independent Variable	B	SE B	*β*	t	Sig
(Constant)	.507	.093		5.434	.000
Age	-.071	.018	-.111	-3.878	.000***
Experience	.876	.024	.941	35.947	.000***
Education	-.001	.012	-.003	-.119	.905

Note R^2 = .921; ΔR^2= .919

7-

Table 4.17 ***Dependent Variable:*** **Control**

Independent Variable	B	SE B	*β*	t	Sig
(Constant)	3.212	.286		11.220	.000
Age	.051	.057	.088	.895	.372
Experience	.087	.038	.222	2.273	.025**
Education	-.059	.075	-.070	-.782	.436

Note R^2 = 0.081; ΔR^2= 0.058

The analysis results of multiple regressions of all dependent variables from Age, Experience, and Education, as shown in the tables above, hasrevealed significant outcomes (negative and positive effect). In (Table 4.11) Experience was a statistically highly significant predictive at 5% of the Satisfaction factor with the expected positive sign when the variable weight was statistically controlled t = 3.00, and Sig= .003. In the multiple regression of the factor of Effectiveness, we noted that no independent variables were statistically significant, but the variable of Experience comes close at Sig=0.091. Age was significantly predictive of the Efficiency factor when the variable weight was statistically controlled. In (Table 4.14) Education was significantly predictive of Easy to learn (Learnability) when it was statistically controlled at t=2.39.

Furthermore, Table (4.15) reveals a significant result that Age was uniquely predictive of the Error tolerant factor, when B value is positive B=0.98. In addition, in the factor of Helpfulness (Table 4.16), the variables of Age and Experience were significantly associated with this factor, (R^2 = .921; Adjusted R^2= .919).

4.6.2 Qualitative Results (Free Textbox and Interviews)

Because the analysis of the questionnaires clarified differences between the responses, qualitative analysis was performed. This section presents the outputs of analysing qualitative data that provide significant information regarding the users' perspectives and their experience of utilising the system.

The collected data from the content analysis of the free textbox within the questionnaire was transcribed as a clear data, and then it was interpreted and coded into five inductive classifications as shown in the following (Table 4.18). The source of developing and using an inductive code was through directly examining the data. These classifications are used either to confirm or explicate the quantitative results.

Table 4.18 Qualitative outcomes

Inductive categories	Question No.	Participant Responses
The speed of ending the task	6	There are too many pages that need to be opened in the system in order to complete the task.
	6	The system requires many stages to complete the task.
	8	The system does not have enough features that would help to do the task faster.
	9	The switching in use between keyboard and mouse to operate the task wastes time and affect finishing it quickly.
	17	The user needs to open many pages to gain some related information for processing the task.
	28	The productivity has been reduced due to many stages being required to finish the task.
Improve features of the system	12	The system needs to support some features, to help in processing the task.
	12	These basic standard features should be there to enhance the user performance in doing the task.
	16	The amendment ability of the system would be very helpful for the user to use this system easily.
	20	The limited features of the system prevent me providing good services to the citizens.
	14	Increase the activation of integration of the system with other systems to obtain more related information to finish the task.
	14	Integrate the system with phone text message SMS, so it can inform the applicant that the task is finished.
System Interface design	4	It would not be able to gain some information that we need, due to the mess of the system interface.
	12	The system interface does not provide tools which might help the users to do their task effectively.
	14	The system interface design would not help you to use the features of it properly.
	29	I feel bored and depressed when I use this system which has many pages and enormous dispersed columns.
System support for ending the task	10	I cannot get moreinformationthat isnecessarytoinquire and finish the task.
	22	Some issues related to the operating task need to be solved automatically by the system.
	21	There is no help section in the system.
	23	The procedure of sorting out such frequent problems would take a long time.

As with the previous process of analysing qualitative data, the interview data were interpreted and transcribed as raw data, and the transcript was used as the main source for content analysis. But afterward, analysis of these results from the interview questions was categorized based on the dimensions of the study, besides the key needs of users, and the relationships between the data were listed and described in the Table below (4.19). They are used to explain the results in more detail for identifying some usability problems and the areas of frustration with the system.

Table 4.19Explanation of qualitative outcomes

Dimension	Essential needs	The Description needs
Efficiency	Operational speed	- The consideration of reducing the number of pages and unnecessary steps for doing the task. - Working on a smooth transition through the task that moves as directly as possible. - Working on auto-transition after finishing the task to operate a new one. - Working to speed up and unify the decision on visa issuance.
Effectiveness	Reliability	- Working to track the most frequent errors in the interface and tackle them.
Error Tolerant	Self-correction	- Seeking to improve auto-selection to avoid the wrong data entry. - Working on the ability of errors corrections after finishing the task.
Help fullness	Dynamic support	- Working on providing the essential information for doing the task in the proper place clearly. - Working on providing error messages, and the possibility of resolving the problem.
Control	Give more authorization	- Working on giving and increasing the user's ability to obtain more access to information related to the task.
Engaging (Satisfaction)	Involvement of Users	- Consideration of what features of the system are most required, and include them in the process of the system development.
Easy to Learn (Learnability)	None	None

4.7 General findings and discussion

This section presents the discussion of the overall results, which were obtained from the participants' (employees') point of view. It also discusses the major issues that came across in analysing data on the usability evaluation on the targeted VI system. Moreover, it consists of some outcomes that would confirm or prove to the contrary some significant findings that were mentioned before in the literature review.

This study had the key aim of evaluating to what extent the targeted government system in Saudi Arabia, as a developing country, has conformed to usability principles. The main objectives considered to achieve this aim were to identify the system problems and issues, and also to find the users' requirements of the VI system in order to complete their tasks effectively.

The major finding from this research can be seen generally as the achievement of identifying the current state of the target VI system's usability, and particularly by discovering some major usability problems and users' requirements. In addition, the applying of query evaluation technique through employing a mixed method of quantitative and qualitative approaches assisted inachieving the aims of the study. Therefore, the results obtained from conducting this research on usability evaluation supported the studies by Bakhshi-Raiez et al. (2012) and Suliman et al. (2009) which took the view that Usability Evaluation for the system through utilizing mixed methods (quantitative and qualitative) for the assessment of its interface should be implemented in order to identify the problems of whether users of the system meet their goal successfully, and to figure out the missing functionality (Hix et al., 2004), and the issues that should be included within the improvement process to assist in finishing the task more effectively.

As shown in the methodology section of this thesis, a number of different statistics tests were applied in an attempt to analyse the gathered study data. The questionnaire survey achieved a response rate of 86%.Fowler (2008) identifies the response level as being acceptable if it falls between 5% at the lower score end and 95% at the higher end, so in respect to this statement, the response rate to the study questionnaire is within the acceptable range.

This study analysed the demographic differences of Age, Education Level, and Experience (System Usage) as independent variables for this research study through the applying of ANOVA one way F test. The result was varying as it showed some significant differences with some factors. For instance, there were significant differences in the factors of Efficiency and Easy to learn when considering Age, as many studies have found that age is an important factor that would have an effect on several aspects such as user behaviours and cognitive aspects (Al-Shafi and Weerakkody, 2010; Dwivedi and Lal, 2007a; Bigne, et al., 2005).

As regards the usability dimensions, in the data gathered from the respondents on the statements about the Effectiveness dimension it appeared that half of these answers were to ‘Agree’ that the system helped the users to finish their tasks successfully, and generally they felt good using it, though some results obtained from the qualitative data (the free textbox and interviews) stated that the system needs to complete these tasks with minimum effort. In other words, the users (employees) require the system to have some improvement in some aspects, especially when they have to use the system under pressure when there is an increase in the number of the applicants requesting the issue of a visa. For example, when the users lost their input data before completing the task for any reason during using the system, they were required to re-enter this data in order to finish the task. Though, according to the data that was categorized among the codes for improving the features of the system, this issue could be resolved by improving the system by giving it some basic features like copy, save, cut and paste, or by installing the feature of saving by default, so that the system would have the ability to save the entered data automatically, besides giving the users the choice to delete it. As a consequence, the users might avoid wasting time and effort by repeating the data entry several times.

This related point of saving time and effort as the benefits of Usability has been pointed out previously in the study by Nielsen (1993).Moreover, regarding the auto-correction issue, many of the respondents indicated in the free textbox that the system needs to support solving some problems directly, for example, one of the interviewees detailed this issue in relation to non-compliant data entry, the system should correct it automatically, rather than carrying on through further steps until finishing the task with errors, which resulted in additional work and time to correct these mistakes. Or in another proposed solution it even should give a suggestion to the users to prevent errors occurring during processing a task. At present the error messages appeared late, and in some cases did not come at all.

In relation to the previous point and the study by Ohnemus (1996), there are various benefits that can be gathered when the system is usable and easy to recover from the errors. However, it was found from the responses on the statement number (22) which inquired whether the system helped the users to solve any problem after finishing a task, that approximately more than half of the participants claimed the system made it difficult to recover errors after they finished their task, and it was a bit restrictive and users were not able to correct the errors by themselves, so that the organization might be required to call the professional maintenance authority more frequently to fix the problems which resulted. Accordingly this issue should be taken into account and the system become more flexible to enable the users to correct errors. This can be considered as significant in terms of cost savings. When the percentage of employees (users) who can recover from or correct problems in a system is high, the company may be required to call those engineering services less frequently, which leads to cost savings. Many studies have pointed out several advantages from evaluation of usability; one of them was reduced costs (Donahue, 2001; Griffith,2002; Black, 2002, Juristo et al., 2004).

Furthermore, in relation to the previous issue, one of the major outcomes from analysing the data of the study was obtained from the effect of the Experience factor on the usability of the system, in particular the Error Tolerant attribute. The results that were obtained from applying this measure were significant, which coincides with McLellan et al. (2012) Recent research has reported finding that differences in user retains could be based on the extent of a user's prior experience with the computer system.

The data confirms that the users with more experience had more ability to avoid committing errors. Karat (2000) stated that the users with less experience of using the systems were found to have substantial difficulty with error correction. In comparison to the users who used the system more extensively were found to have better overall performance. Therefore, it has become very important to consider the quantity of user experience (UX) with a specific system (Sturm et al., 2002). The interview results also indicate that the users who have more experience could develop their scope of work and navigate new options, however they would request access to more features and new functions to do their task. From this point it could be argued that, although the user who has more experience of using the system has more ability to use the system easily, and lead to increase his satisfaction. (Bokhari,2005), still he could claim more Usability problems due to having deeper knowledge of the system obtained from more frequent use. Similarly in the study conducted by Suliman et al., (2009) on Usability evaluation to identify users' expectations, it was found that the users with the most experience were more deeply concerned with the functionality of the system because they were able to find more issues.

In addition, a signal of misguiding from the users' opinions about the speed of doing the task was revealed in the statements number (4 and 6). It seems that there is a difficulty in interpreting the responses, which stated that although they agreed that the system does not take a long time to finish a task, it requires many steps to complete it. This could be illustrated by the qualitative technique through some comments in the empty boxes within the questionnaire and other interview responses. It has been indicated that the speed of completing the task by using the system was relatively good; however the system still forced the users to open several pages to operate this task, which would consume additional time for doing it. Furthermore, this issue would particularly have a negative impact on the employees working face to face who need to finish their task directly in front of the citizen. In relation to this issue, it could show the advantage of applying a survey by employing mixed methods of quantitative and qualitative approaches. Stone et al. (2005, p. 483) provided an evaluation tip which stated that *"if you use a questionnaire, then always interview as well, in order to probe the answers to the questions. So, after the participants fills in the questionnaire, use it as the basis for an interview so that you have an opportunity to explore the reasons why the participants chose the answers".*

Additionally, as shown in the responses to the question number (7) about the integration ability of the system, the vast majority of the employees have confirmed the importance of this feature of the system, and that it should be enhanced to be more integrated with associated systems such as the banks and immigration systems. Moreover the users (employees) should be granted a proper access to much reliable information, which in turn could help them to speed up finishing the task accurately with a low error rate.

Another major problem that faced the users was revealed from their responses in the dimension of Helpfulness. Thus, half of the participants confirmed that they needed technical support while they were operating the task, and the system had a lack of providing sufficient information on the system such as an online help facility and documentation information as guidelines for helping the user to finish the task and find the solution when they committed any errors that might disrupt finishing it. In relation to this concern Microsoft Corporation (2000) reported that it has improved and presented interface guidelines for its Windows computing platform (computer architecture, operating system, programming languages and related user interface), to verify that they have a consistent look and feel. IBM is another organization that has improved similar guidelines for helping the delivery teams as users of the software and systems to improve their process of driving business innovation(IBM). Besides, a number of Usability experts such as Nielsen (1993) have provided comprehensive guidelines for the designers, who believe that the commitment to following these guidelines is very important to produce usable products (ibid). Zhang et al.,(2005,p.1) stated that *"Illogical overall organization of data/information in the system, lack of task support, misfit between the nature of the task and the support provided, difficulty of navigation, and inconsistent mental models of system operation are among the major problem or difficulties users experience"*.

Furthermore, the concept of "guidance" is an applicable example of a usability property. The guidance might have a positive effect on the attribute learnabilitybut on the other hand might affect the dimension of efficiency negatively due to the user spending time on navigating the information displayed on the system guidance rather carrying on performing the task thoroughly (Scapin and Bastien, 1997; Ravden and Johnson, 1989).

The responses to the statement number (15) which was seeking to identify the employees' (users of the system) views concerning whether their needs have been fully taken into account within the system development lifecycle, revealed that the developers of the system had paid more attention to the aspects that related to the processing of a task appropriately, rather than to considering and involving the users' opinions within the improvement of the system quality, and the functions that related to the system itself which would meet their needs too. This could lead to the system failing to reach the main goals of implementing it, besides it makes users feel frustrated and annoyed. This was also pointed out in the study by Bostrom and Heinen (1977 cited in Zhang et al., 2005). The study by Boehm and Hansen, (2001) confirmed that the users' considerations must be addressed early in the development of a system in order to achieve good implementation of a usable system.

In analyzing the answers about the user satisfaction with using the system and the design of its interface, it can be clearly seen that more than half of the participants agreed to use it as a system for issuing visas, and that it has a well-designed interface. On the other hand, other users claimed that the UI of the system was tiring, and that it had quite limited capabilities for doing the task, and a messy system interface due to several dispersed columns that in some parts of processing the task did not support auto-transition and self-correction, which would lead to disrupting the completion of the task and errors occurring. This issue is of concern to the literature (Daouk and Leveson, 2001), and a similar result was observed in a study by Manresa-Yee et al., (2010).

From the results in connection with the statement number (27) which seeks to identify whether the data entry about the applicant into the system is hard and complicated, it can be very clearly observed that the majority of respondents said they 'Disagree'd or ' Strongly disagree'd with this statement. This finding would be consistent with the above explanation regarding the issue of user satisfaction with the system in terms of ease of use. However, some responses in the qualitative results explained that the interface had some weak points such as the messy columns and the lack of related information.

Based on applying and analysing the multiple regressions for study, other interesting findings were revealed. For instance, the independent variables of Age and Experience had a highly significant influence on the Helpfulness factor to determine and measure the current practice of the usability of the VI system. Moreover, the results related to the dimension of Error tolerance highlighted the positive impact of the experience.

Lastly, although applying a query technique through conducted a questionnaire and interview, could help in evaluating the VI system usability, and achieved a sufficient insight about the state of it, yet some usability problem did not reveal due to some issues were unobvious, subtle or difficult to determine the reason for occurring them. Similarly to this issue Spencer (2004) argued that in most cases people usually provide things that do not end up problems and will fail to spot things that are problems.

4.8 Chapter Summary

This chapter four has presented the first study which was concerned with the summative usability evaluation of the current system in a government organization in Saudi Arabia by utilizing a query evaluation approach. As a stage of HCD for defining the additional user requirements, this is a formal usability query technique utilizing a mixed method of quantitative and qualitative, using a number of known usability instruments, such as questionnaire, free textbox, and interviews, to measure the usability from the user's perspective, which could help to assess the system usability through which the users’ requirements have been identified. At the beginning of this chapter a brief background of related work has been provided. Then the consolidated model of the study has been introduced by a detailed explanation of the methodology.

The major findings of the study could answer to RQ1 and RQ2, so it indicated that the general views of the participants about the target VI system was as a simple one in terms of how easy it was to learn to use it; however, it has revealed several usability problems varying between major and minor, with some missing functionality, and barriers towards the usage of the system were identified from the participants’ responses.

As regards the additional dimensions of Helpfulness and Control, they generated formative information that helped to understand the most important issue in evaluating the system usability besides the users' requirements. Furthermore, it was clear that the Efficiency factor provided a better way to measure the current practice of VI system usability work from resources data in relation to achieved aims. Furthermore, we recognized that the usability of the system could be improved through measures such as formative guidance to help the users to operate the task.

In addition, several problems that came across from the participants’ responses could be taken into account and traced back in order to improve the system usability. For instance, the lack of organising the information into clear guidance for processing a task, and after that the users were sometimes confused by the huge mass of information that appeared in several pages.

Moreover, based on the analysis from one-way ANOVAs of the independent variables, such as Age, Qualification, and Experience, it found that the factor of experience presented the most significant results that would be more beneficial to the research. Furthermore, the results of the multiple linear regression analysis were powerful for helping to achieve the aim of the study through clarifying the influence of the independent variables on the dependent dimensions. In addition, the mixed method based on the user's perspective which was applied in this study, through questionnaires, free text boxes and holding interviews, was a helpful technique in terms of collecting data for evaluating the usability of the system.

The major findings that stem from the study were that:

- In the general view, according to the participants' feedback, the existing VI system was easy to learn in respect to how they could learn to use this system quickly in their daily business for completing their tasks, though they were not satisfied in regards to how the system consumed much time to complete the intended task, besides other issues that were associated with some parts of the system.
- Applying the query evaluation technique to measuring the system usability it was helpful, quite insightful and gave us the opportunity to figure out different usability problems.
- A very interesting different feedback was generated by interpreting qualitative data through applying content analysis.
- The qualitative outcomes have explained some ambiguousfindings which came with the quantitative results.
- The qualitative outcomes through conducting the free textbox and interview tools have revealed different usability problems.
- The problems that emerged with the usability evaluation will be addressed and taken into account for developing a suggested prototype user interface design.

And for the purpose of reporting the various problems identified easily, we divided them into two categories, major and minor, based on the previously analysed data as in the following (Table 4.20).

Table 4.20 Usability Problems

	Problems	Requirements No.
Major	Multiple pages to finish a task.	RQMT 1
	Lack of organising the information on the system UI.	RQMT 2
	A chaotic distribution of the columns on the system UI.	RQMT 3
	No proper alert message that an error has occurred.	RQMT 4
	Some missing help features for operating a task.	RQMT 5
	No presence of guidance to support users to operate a task..	RQMT 6
	No technical support section provided on the system interface.	RQMT 7
	The users' inability to correct errors after finishing a task.	RQMT 8
Minor	The need for an Auto-transition between the columns during operating a task.	RQMT 9
	Frequently switching between using the mouse and keyboard to finish a task.	RQMT 10

Lastly, the significant outcomes which we came across when analysing the data of the study will be addressed in the next study for improving the new interface design and implementation of the prototype process based on the interface design requirements. After that the compression usability test technique will be applied in order to validate and confirm the provide design solution.

Chapter 5: The usability testing approach for developing a prototype user interface of an internal system in a governmental organization

5.1 Introduction

Through the previous study reported in chapter four, which aimed to conduct a query usability evaluation of the current VI system in a governmental organization, the main findings highlighted some problems and additional requirements associated with several parts of the VI system usability. Following a user centred design approach, and from these problems, it has been generated a number of recommendations for improving the user interface (UI) of the VI system, and to be directed at trying to solve the errors and cover the drawbacks of its usability. Tullis and Albert (2008) stated that the majority of usability experts claimed that discovering usability problems and then providing specific design recommendations is the most crucial part of their concerns. As far as these issues and problems are part of an iterative process in which designs are being evaluated and improved, they would afford enormous value to achieve a proper design of product and provide a foundation to the usability domain.

Therefore, after these recommendations have been summarized, as shown in the ten requirements in previous chapter, the next stage would be starting the process of transforming this type of issues into a suggested prototype design. A prototyping is considered as a crucial principle of UE domain (Fulten Suri, 2000).
Therefore, in this study, we developed a potential prototype UI design based on the outcomes of the former study through using a Medium-fidelity prototype as a first step, following by High-fidelity with specific functionalities, in order to improve users' interaction (Juristo,2009). After the challenges and drawbacks of the existing system had been addressed in the suggested system (mock-up UI design), we had to prove it via an experimental test to ensure if the improved prototype fulfilled the users' needs on the existing system or not. Karat, (1997,p.693) stated

that *“ The identification of usability problems in a prototype user interface (UI) is not the end goal of any evaluation, the end goal is a redesigned system that meetsthe usability objectives set for the system such that users are able to achieve their goals and are satisfied with the product.”*. In addition, Bailey (1993) stated that offering additional experimental evidence directed to iterative design relay on usability testing would lead to achieving measurable improvements in the usability of an application. For this purpose, the study conducted a comparison usability test between the interface designs of two evaluated systems in order to highlight the differences and record the improvement. Furthermore, it employed mixed methods, quantitative and qualitative, through employing some specific standard quality criteria and different statistical approaches, such as measuring the time to complete task, calculating successful completion rates, and responses to a post study questionnaire that represents a quantitative technique. While some other techniques represent qualitative data, such as the user’s subjective impressions of each option of UI design, and note taking through observing the participants while conducting the usability test.

Initially, in this study, the demographics of Age and Educational level, were not considered as main "variable factors", but rather were being collected only as basic information within the test procedure. However, the Experience factor will be only involved in respect to making sure that the participants in the test have a sufficient ability to use both systems. This was referring to some factors; firstly, on the grounds that the main objectives of this study were concerned with developing a simulation design of the target system and employing the controlled comparison usability test between the two systems through recruiting a limited sample, and measuring their performing represented tasks, in order to verify the improvement. Thus, the influence of these demographics would become ineffective. According to Sauro (2013),*"product and domain experience generally have much more impact on usability metrics than demographics in a usability test"*

5.2 Aim and Objectives

The main aims of conducting this study are to develop a prototype design (P.D) for the purpose of conducting a comparison usability test between its characteristics and abilities and the current VI system. Furthermore, the actual users will validate the proposed design in order to determine whether it fulfils their quantitative and qualitative usability requirements that were provided earlier in the previous study of usability evaluation of the VI system, and whether they are able to use it more easily and effectively.

In order to achieve the aim of the research, the following objectives had to be attained:

- To consider precisely the usability evaluation outcomes and recommendations of the previous study for developing a prototype UI design of the existing VI system through utilising medium and high fidelity prototypes.
- To apply a comparison usability test via employing quantitative and qualitative tools to collect data from participants during or after the usability test.
- To determine specific and clear tasks that represent the main tasks of the actual VI system to be used in the usability test.
- To determine the usability measurements to evaluate subjective and objective aspects of both systems' usability. Subjective values are quantitative measures based on task time, completion rate, error rate, and participants' responses on the post study questionnaire. Objective measured values are based on qualitative measures of observing participant performance, and note taking, besides his response on open-end questions within the post study questionnaire.
- To identify statistical evidence in order to support the answers for the research questions.

One of the basic elements in usability testing is to develop research questions or test objectives rather than hypotheses (Rubin and Chrinell,2008), with one or more participants to achieve a successful experiment (Durbin,2004).
The achieved results of conducting this study will enable the research to answer to the RQ3 and RQ4, which were reported in chapter one as follow:

RQ3: How should the current VI system be re-deigned and produce a prototype design?
RQ4: What are the effects of the proposed usability design solutions on the UX?

5.3 The study instruments, techniques and procedures for collecting data

Since the main objectives of this study are associated with the comparison usability test method, during which users perform specific tasks using different designs, the actual VI system and a high-fidelity prototype, a user-based testing method is employed as the criterion for considering several techniques and tools for gathering data. such as survey (post-study questionnaire), free textbox and users' comments, taking notes, calculating time to complete a task, and percentage rate of successful task completion, and error rate.

5.3.1 Developing the Post-Study Questionnaire Design

The developing of questionnaire design in this study and collecting its potential statements were based on the literature from different studies, besides the context of this research study. The various questionnaires applied were Lewis's (1993 and 2002) post-study system usability questionnaire (PSSUQ), and Lund's (2001), measuring usability with the USE questionnaire. Generally speaking, the user satisfaction according to Bhattacherjee (2001) is measured by the extent to which previous exceptions are confirmed. So essentially, in this study, the post-study system usability questionnaire (PSSUQ) was used as a foundation for developing the questionnaire which aimed to assess users' perceived satisfaction with systems and applications (Sauro and Lewis,2012). The origin of this standard usability questionnaire was an internal IBM project called system usability metrics (SUMS), and it affords users an opportunity to provide their views on the system they recently used or tested. Furthermore, it is built to conduct an overall evaluation of a system at the end of a usability test, through three factors, System Usefulness, Information Quality, and Interface Quality, while the USE questionnaire was used to evaluate user attitude towards the system, and it stands for four dimensions, Usefulness, Ease of Use, Ease of Learning, and Satisfaction (Lund,2001)[11]. Similarly, there are several studies that reported on other dimensions, however, the three dimensions (System Usefulness, Information Quality, which represents the system ease of use, and finally Interface Quality which represents the user satisfaction) are the most effective way to distinguish between interfaces (Lund,2011; Lewis,1995).

Therefore, the post-study questionnaire in this study was constructed with three common factors of the PSSUQ with a spirit of the USE questionnaire. Furthermore, it was slightly edited and improved to include the entire aspects of the VI system interfaces, and finally included 26 statements. These statements represent the three subscales of the study for measuring the users' satisfaction, as follows:

a. The statements from 1 to 11 were designed to measure the System Quality (SysQual) factor.
b. The statements from 12 to 19 were designed to measure the Information Quality (InfoQual) factor.
c. The statements from 20 to 26 were designed to assess the Interface Quality (IntfQual)factor.

In addition, due to this research aiming to gather users' opinions, attitudes, and behaviours, it is considered as a pathway within the Human Computer Interaction (HCI) field. Furthermore, Taylor and Heath (1996) indicated that the Likert scale has become one of the leading methods of measuring social and political attitudes, because it would maintain the respondent naturally in direct involvement, and shows a high degree of validity and reliability (Coolican,2004). Additionally, in the final section of the questionnaire, another two questions have been added which include like/dislike statements with clarification, that give the participants a chance to make judgments on specific issues (Rogers et al.,2011). In order to produce the questionnaire well with checks and balances as an unbiased evaluation, some negative types of questions were included within the survey. So, the questionnaire has statements that alternate between negative and positive

[11]Measuring Usability with the USE Questionnaire, available at USE Questionnaire: Usefulness, Satisfaction, and Ease of use.

wording. Love (2005) reported that it is vital to mix-up the view of the positive and negative statements in the Likert scale within the questionnaire, in order to reduce the risk of "constant error" which is caused by the acquiescence effect (Love,2005), and to minimize extreme biases (Sauro and Lewis,2011). Therefore, some reorganization in the order of the statements has been done to obtain the final questionnaire (Appendix G).

5.3.1.1 Pilot study for the post-study questionnaire

Rubin and Chisnell (2008) stated that a pilot study is useful for obtaining a sense of whether the statements within the questionnaire are formulating the right information. Thus, a pilot study was employed to develop the quality of the questionnaire through assessing its reliability and validity. Furthermore, it was carried out in order to verify that it was in context understandable by the participants, easy for them to answer, and supportive for collecting and analysing data which would help to achieve the aims of the study. It was conducted with 12 participants from among the actual users of the VI system. As we mentioned before, in the task scenario the post study questionnaires were distributed to the participants after they finished the tasks of testing the two systems, the current VI system and the suggested prototype. Furthermore, they were asked to provide their direct feedback if any confusing questions appeared (Babbie,1990;Lazar et al.,2010). Thus, we used the Cronbach alpha coefficient statistic to evaluate the reliability of the scale for the questions considered.

5.3.1.2 Reliability

Reliability is the determination of whether a questionnaire measure develops consistently across time (Babbie,1990; Ozok,2007). We applied the measuring of reliability using alpha coefficient to indicate the level of reliability of the overall questionnaire besides its three factors, System Quality, Information Quality, and Interface Quality. Based on the participants’ responses after they finished tasks using the suggested high-fidelity prototype, and the current system, we obtained the following results.

Table 5.1 Cronbach Alpha Reliability of Post Study Questionnaire

	Cronbach's Alpha	No. of statements
Sys Qual	.799	1-12 (12)
Info Qual	.750	13-22 (10)
Intf Qual	.857	23-28 (6)
Overall	**.859**	**28**

As shown in the above (Table 5.1) the values of Cronbach's Alpha is obviously more than >0.8 in the Interface quality construct, while in the system quality and information quality were less than <0.8, namely 0.78 and 0.75.

However, this indicates the sufficient reliability of this scale to be a useful usability measurement. As the procedure of testing the internal consistency of each statement is applied, there are some statements that have the calculated Alpha more than Alpha for all statements of the construct, for example:

Table 5.2 Some Items-Total Statistics

Item No.	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q9-I detected much inconsistency in this system	.285	.814
Q14- Learning how to use new functions on this system is hard.	.199	.792

So, we followed the application of this procedure, then we removed these two statements (items). The following (Table 5.3) reports the results obtained after we re-calculated the Cronbach's Alpha.

Table 5.3 Cronbach Alpha and the reliability of Post Study Questionnaire

	Cronbach's Alpha	No. of statements
Sys Qual	.818	11 (1-11)
Info Qual	.790	9 (12-20)
Intf Qual	.857	6 (21-26)
Overall	**.861**	**26**

In addition, according to the participants' feedback, some statements have been amended. For example, the statement number (Q5 in the questionnaire) which stated "The system did not save me time when I used it", has been revised to be more clear; it became "The system consumed much time when I attempted to do the required task". Furthermore, some participants have suggested a correction of grammatical mistakes within the Arabic version of the questionnaire, knowing that this adjustment did not affect the meaning of the question itself when it was translated back to English language.

5.3.2 Planning for the usability test

The need of having a clear test plan that addresses the how, when, where, who, why, and what of a usability test, is crucial as a successful foundation for the entire test (Rubin, 1994). There should be a general procedure for conducting it that includes understanding the purpose of the test, preparing test objectives, selecting a target sample, conducting usability testing techniques, identifying a tasks list, arranging the test environment, gathering information, and analysing it and reporting the results (Dumas and Redish, 1993; Rubin, 1994). This sub-section presents an important step in developing a usability test plan.

5.3.2.1 Purpose of usability test

As we mentioned previously, the main objective of this study is to determine whether the users, "participants", can accomplish specific tasks faster, more effectively and more accurately though applying an empirical test, and by using an improved prototype UI design of the VI system with specific characteristics and abilities.

5.3.2.2. Testing laboratory sittings

The usability tests in this study took place in an informal laboratory rather than a formal usability test site, so, an office of the duty manager at the work place was used. The arrangement for having this facility has been made with this duty manager as he was considered to be a gatekeeper for the study. According to the U.S Department of Health and Human Services (no date), there is no need for conducting testing in a formal usability lab, and it can be in any of the following settings:

- A fixed laboratory having two or three connected rooms outfitted with audio-visual equipment;
- A conference room, or the user's home or work space, with portable recording equipment;
- A conference room, or the user's home or work space, with no recording equipment as long as someone is observing the user and taking notes;
- Remotely, with the user in a different location.

The decision of determining this location was taken by considering several aspects: There were difficulties of accessibility and in installing the actual VI system in another site. This office would represent a real working atmosphere, which would help the participants or users to perform tasks more naturally, and for employing methods for collecting qualitative data such as note taking data, beside managing the system being tested and giving hints, we needed to sit next to the participant. In addition, there are technical limitations and the economic issue that an equipped lab needs a high cost spending. However, this office was structured and fully equipped with all requirements, tools and insulations system for conducting the test, such as, the UI designs, the prototype and the actual VI system are set up to be used at the same time.

5.3.3 The experiment sample

As it described before the specifications of participants of this research in chapter 2, the selection of appropriate participants, who were representative of the actual users of the VI system, was a vital part of the testing process (Jonathan,2005). The number of these representative users depends on the test selected and the experimental design (Carvalho,1996). Spyridakis (1992) stated that a minimum of 10 to 12 participants must be used in an appropriate experimental design. Moreover, this study has applied a mixed method, and usability testing as a qualitative method is traditionally accepted by experts as sufficiently based on a small sample size of 5 to 20 participants. In respect to the quantitative method of testing usability Nielsen (2006) recommended 20 participants or more, because it would give significant numbers and meaningful results in the statistical analysis. Therefore, the number of selected participants that was chosen for the usability test is 32 users for more accurate results. All of them were employees at the governmental organization in recruitment administrations; they have access for using the actual VI system. Also all of them were male, because there is no female section provided in this service of issuing visas yet. In addition, after the participants filled out the background sheet (Appendix B), we found that they were aged between 28 and 38 years, most of them have experience of using the VI system for more than 2 years after it was implemented, but approximately five participants have less than 2 years' experience. The responsibilities of the chosen participants will be to attempt to perform a set of representative task scenarios that are provided to them in an appropriate manner, efficiently and timely, while the evaluator has a responsibility to keep them feeling as comfortable as possible during and after the test (Nielsen 1994). This in turn would provide formative feedback about the usability of the re-designed UI of the actual VI interactive system.

5.4 Prototyping (design solution)

As we mentioned earlier, the developing of prototypes can be either paper-based or computer-based, however, the participants who are involved in usability testing mostly prefer to interact with computer-based prototypes (Leavitt and Shneiderman, 2006). In addition, Yee et al. (2010) stated that the research devoted to developing UI is rapidly increasing, to get more advantages from technological development. Therefore, in this project we attempted to create a prototype, using the most appropriate technology with required fidelity for the potential UI design. Furthermore, Preece et al. (2002) propose that HCD should involve both low and high fidelity prototyping, although most researchers point out that most of the low fidelity prototype methods would provide equivalent results with high fidelity prototypes. Thus, the approach of developing the prototype UI in this study consists of a medium-fidelity prototype and high-fidelity prototype with iterative technique.

5.4.1 Medium-fidelity

The medium-fidelity prototype (as an early design outlining) has been developed relying on the requirements and recommendations that obtained from the previous evaluation study of the VI system, besides transferring the strengths of some existing areas and good UX which were already provided by the current (legacy) system for exploring and drawing the basic concepts of the final prototype UI design by using Microsoft PowerPoint software as a type of better method of medium-fidelity prototype. Essentially, it is designed on the computer using the Microsoft PowerPoint application because it presents a lightweight method for prototyping UI for the system, which can be useful for testing users and figuring out their requirements and then to gather their views about the system easily. Thus, the medium prototype design has been sent to the participants through their emails, we attempted to receive their feedback in a fast and easy way, as this could save us a lot of time. In addition, due to the prototyping being electronically designed it helped us to have more flexibility for modifying and updating the design than other tools such as paper and white board. Furthermore, working on the computer is considered to be faster than other tools for developing low to medium-fidelity prototypes (Clark et al.,2011). In addition, Rogers et al. (2007) stated that when the developer or designer attempt to sketch an interface design, they might need to apply such dialog boxes and approximate different icons to final design as would help the participants to understand the design properly. Boothe et al. (2012) pointed out that a medium fidelity prototype would increase users' perception of particular system usability. Righetti (2005) stated that the main advantage of using the PowerPoint to develop a prototype is its simplicity of use and that everyone can be able to develop his own, for this reason it has become the most commonly used prototyping tool. Therefore, Microsoft PowerPoint would be an appropriate tool in this stage for achieving the initial requirements. This prototype UI design contains four main pages that present the story of the process. The first page is a kick-off step for enabling users to login to the system. The remaining pages consist of five sub-pages; each presents a particular piece of information such as payment details, immigration data, completeness and the previously issued visas. The second page is about calling the details of the applicant. The third page is about the current visa application details, while in the fourth page, the final summary of the application will be shown before issuing the visa. Also, the prototype comprised a menu bar for additional functions like copy, paste and cut. The following figures (5.1 to 5.4) show the screenshots of the preliminary solution design of the medium fidelity prototype using MS power point before conducting the pilot test.

In the screenshot (5.2), the research could cover the RQMT 1 by aggregation several screens that the users need to visit during performing the actual task using the existing VI system. In addition, the same screen the RQMT2 and RQMT3 have been met, thus the non-used columns have removed, and the information re-organized in positioned in aparticualr place where it is easy to be seen by end users. In regards to the RQTM 4 and 5 , the screenshot (5.3) shows that the suggested prototype could solved these two requirements by adding a suggested alarm massage and a tools bar for adding several features like copy, past and save. In screen. In screenshot (5.4) the apperance of a guidance that the task has completed, the research could tackle the RQTM 6. However, in regards to RQTM 7,8,9,and 10, due to there are related to functional processes, they will be considered within the high-fidelity prototype design.

Figure 5.1 A screenshot of the first screen of the preliminary medium fidelity prototype

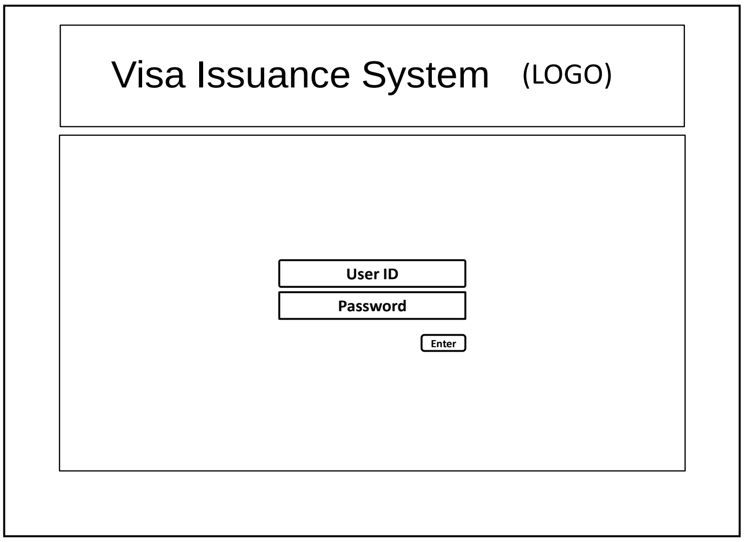


Figure 5.2 A screenshot of the main screen searching the applicant details

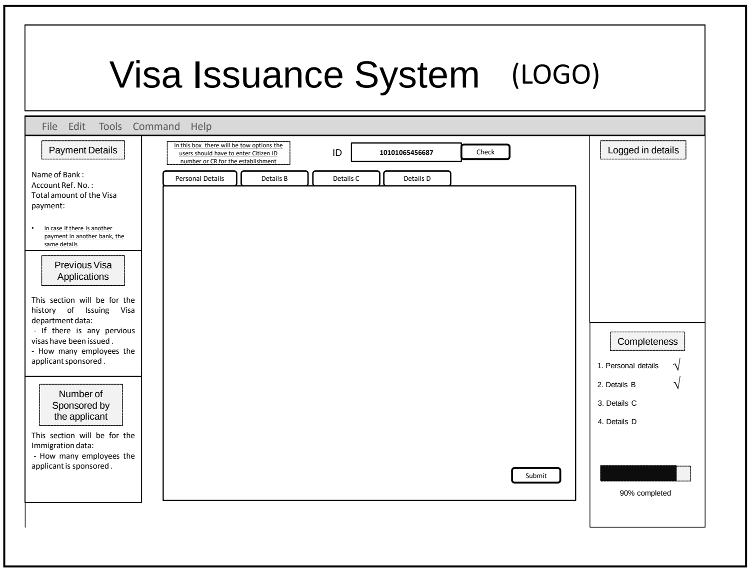


Figure 5.3 A screenshot of the main screen of processing issue a Visa

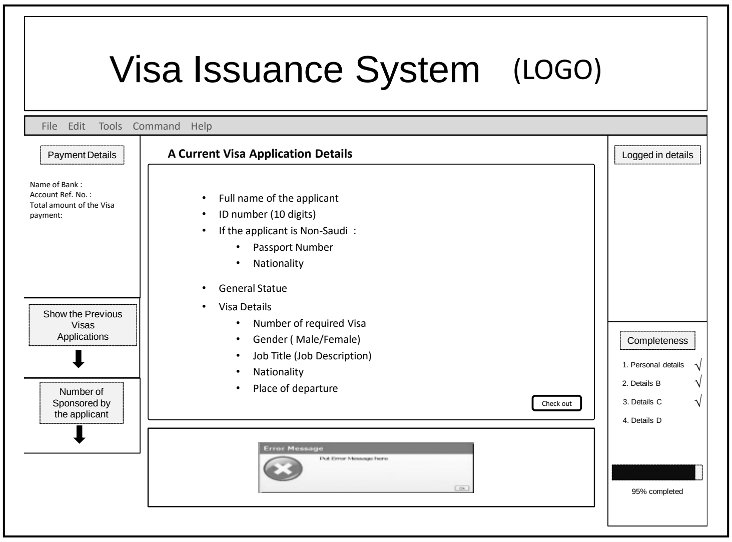
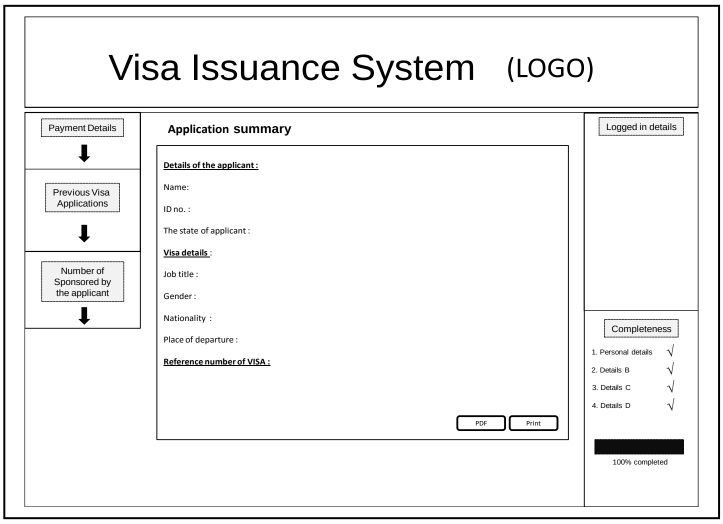


Figure 5.4 A screenshot of the last screen of submitting a Visa



5.4.1.1 Pilot study for Medium-Fidelity prototype

After the medium-fidelity prototype was created as a preliminary concept of a suggested UI prototype of the VI system as a design solution, a pilot study was employed as an informal assessment of UI design. A qualitative discussion was applied for obtaining the views of the participants on the developed UI prototype, and to find out its effects on their perceptions of usability as well as to spot any errors or issues with it. The medium prototype design was pre-tested with five participants from real users (employees) of the VI system who have been chosen from one department, they had participated previously in usability evaluation of the VI systemin the previous study. After getting access to their contact details, they received via email the file attachment of a medium-fidelity prototype design of UI. Due to the fact that the users have not done the same test before, or in other words, they are not familiar with such a type of assessment design, each user was informed in advance of the purpose of conducting this feasibility study, and provided with a short guidance and dialogue structure about the prototype, and how the process goes as a scenario of the real system. In addition, since all the participants were non-conversant with the English language, so an Arabic version of the medium prototype was developed to cover this issue, and we simply translated the titles and words. This in turn, would help them to provide related information that might help to achieve an acceptable level of usability in the UI design, and then to enhance the effectiveness of its various tools which were related to the task. However, we have selected the English version of the medium prototype to present in the screenshots, because if we chose the Arabic version, we might need to insert some induction boxes and arrows, which in turn would lead to a confusion about the prototype design. For obtaining the feedback, the Skype application and phone calls were used to receive the responses from each participant individually, after they had all had adequate time to go through it.

5.4.1.1.1 Findings of the pilot test

After the medium fidelity prototype was tested, the major results indicated a consensus of all five participants that the concept of reducing the number of pages for operating the task would be a significant improvement in the UI design of the VI system. They referred to the fact that fitting of these pages into one page would be very helpful in picking up the related information more quickly to complete their task. They emphasized in particular the payment details, which would help them to make an early decision whether to carry on finishing the Issue Visa process or to halt it in the case that no payment has appeared. Additionally, most of them could express their rational understanding of what the logical following steps of the design would be. However, some issues and problems were found with the prototype. For example, four users reported that " *the additional functions such as edit, tools, and help in the menu bar, could enable them to have more control on operating the task, but they noticed that the appearance of the toolbar on the interface design was a bit confusing with other information in one page* ". It would be better if it was removed and its features kept in a place where it will not interfere with displaying other important information such as the sub screen information. In addition, they also mentioned that they would have a better ability to learn new functions when they used this UI design. One participant mentioned especially, that “*the help option would be a useful function if any problems occurred during operating the task* ", and he compared it with the current VI system that does not support this option. However, other outcomes and comments led to iterate on the design of the medium-fidelity prototype. Three participants have mentioned some missing details that need to be added to the applicant details, for example, some required information in the current visa application page, such as “nationality, number of passport and place of departure”. In addition, two participants suggested decreasing the number of subpages, or adjusting the places of some sub-pages, in order to avoid the dispersion of their concentration when they attempt to use the system. Furthermore, they reported the importance of some key parts being shown in some subpage. For instance, in the subpage of viewing the previous visa applications, the users expressed their request to be able to choose between getting all the information related to the previous visa applications or having it filtered by particular options in order to obtain specific information. In addition, two participants recommended adding another screen after the main one for login, for choosing the type of visa and then starting the search for applicant’s details via his ID number. They claimed that this page would help them to navigate the information more freely in another screen after it opened.

Through this pilot study, the importance of rational issues for conducting field usability testing by using prototype software has been identified, and it is obvious from the various outcomes that the participants had a good perception of the medium fidelity. Based on the findings and participants' suggestions, the medium-fidelity prototype design was modified and then it was converted to the next step of developing a high-fidelity prototype. The following Figures (5.5 to 5.8) present the interfaces of the initial design solution based on the users additional requirements and recommendations, though utilizing the medium fidelity prototype, created by the researcher

Figure 5.5 A screenshot of the first screen of the medium fidelity

Figure 5.6A screenshot of the main screen searching the applicant details

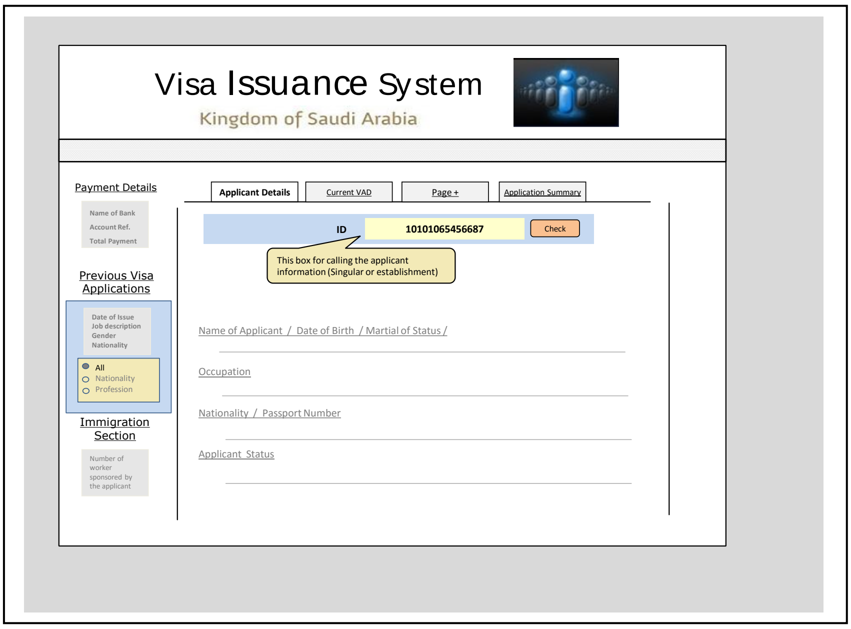


Figure 5.7 A screenshot of the main screen of processing issue a Visa

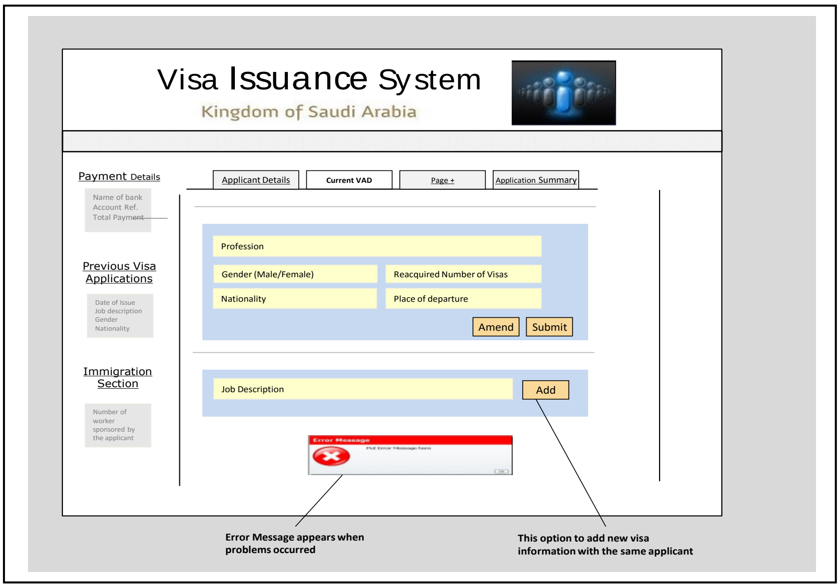
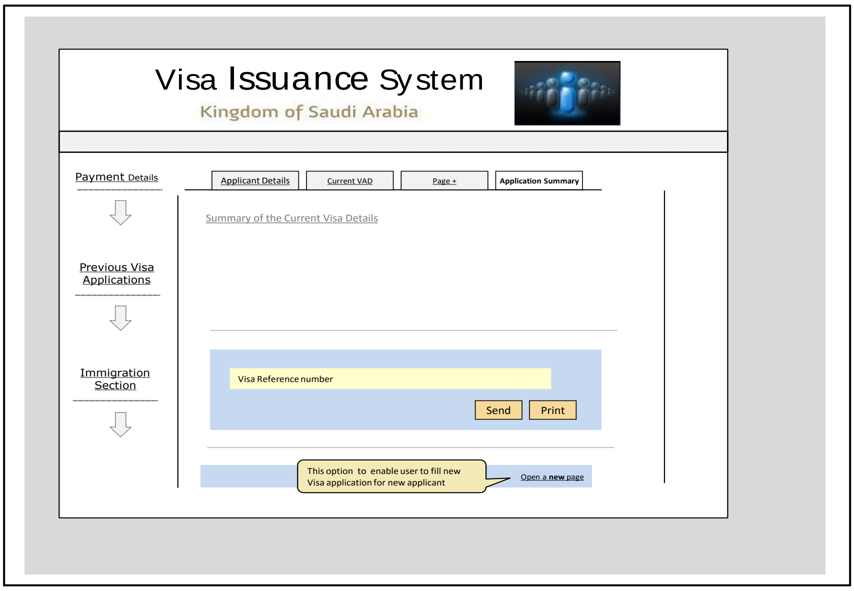


Figure 5.8 A screenshot of the last screen of submitting a Visa



5.4.2 High-fidelity prototype

After the participants' feedback in regards to the medium-fidelity prototype design had been gathered for determining the scope of prototype simulation with high-level operations, the requirements and modifications based on the outcomes were addressed and coded into the developing of a high fidelity prototype via utilizing MS Access Software. This tool has been employed for several reasons. For instance, as we were attempting to improve a prototype UI of a current system with a similar base of conceptual design to the existing system, and with selections of functionalities for enabling the participants to perform the test tasks realistically, the need of developing a version of database application “dummy data” became important. Therefore, MS Access database development tools would provide to the participants a feeling of interaction and performing the actual main tasks. Thus it might convince them to provide more precise information for it. For instance, searching for data using features such as, search by ID number of applicants, entering real data and getting a response as final output, like the issue number of a new visa for the applicant, and error messages in case problems occurred or enquiries or reports rely on the data entry. In addition, the environment of MS Access would help for creation of useful database solutions faster, which require significantly less code than alternatives. Thus, it is a great platform for prototyping (Chung,2013). In addition, we should bear in mind when improving a prototype design that it would not represent the entire protocol and functionality of the existing system, however, it would produce adequate functionality that could address the main test objectives (Rubin and Chisnell,2008).

5.4.2.1 Pilot Study for the High-fidelity prototype and main test

The main goals for conducting the pilot study in this stage were generally to assess the high-fidelity prototype design that has been developed and outlined based on the preliminary results of medium-fidelity prototype through involving the typical functions of the existing VI system within its features. At that point, in this research step, we applied this pilot study for the purpose of assessing the prototype design and to make an iteration of changes in design before implementing the latest version for further testing.

Initially, the data for this sub-study was collected through two stages:

The first stage was conducted after the outcomes and issues from the medium fidelity stage were addressed in the developing of a suggested high fidelity prototype. So, five participants from among actual users of the current VI system were recruited to be involved with this pilot test. These participants had participated previously in the evaluation of the medium fidelity prototype, so they were conversant with the foundation for conducting this project as a whole, and this might help them to provide proper views on the high-fidelity prototype. Each of these five participants before starting the evaluation individually had a small amount of training in using the design. A technical problem was encountered before we started to run this pilot study. Due to the high fidelity prototype being developed using MS Access version 2010; it needed to be installed on the PC with the same version. However, all PCs at the work place were set up limitedly to be not capable of installing any other additional software; this was because of an internal regulation from the developer of the actual VI system, and additionally to the lack of equipment. Therefore, a high specifications laptop (XPS Dell Intl Core i7) with a 15.6”-inch screen was used to solve this problem. After the system design had been installed on the laptop at the workplace and they were ready for the test, they received a short demo about the design besides a short training in using it, in order to identify its dialogue structure, functions and how it works. In addition, each participant had an opportunity to go through it and have some experience of using it. Then they have been required to present their useful feedback orally, which was recorded and noted down. Another stage was related to the participants' comments and suggestions that were collected generally on the high fidelity prototype while they performed test tasks for piloting the usability test.

Some participants have reported that there was some missing information e.g. some nationalities did not appear in the drag-down menu while the participants performed the task. However, they were informed that behind this prototype, there was dummy data only. Moreover, after the common usage nationalities were established, these few only were used for performing the tasks. Then, these comments and suggestions were considered and incorporated into the prototype design modifications to be ready for conducting the next step of testing participants' tasks in a scenario based usability testing.

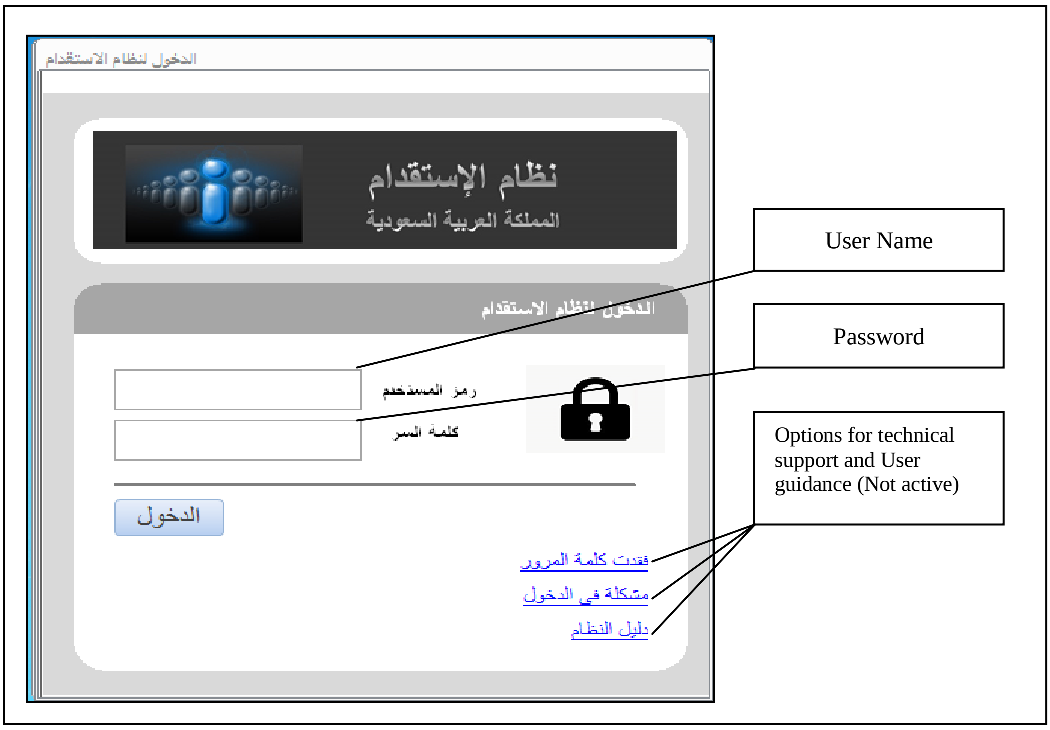
5.4.2.1.1 Findings of the pilot study

According to the participants' feedback, a number of problems and errors were discovered while they were using the improved high fidelity prototype design in the usability test. The total number of identified problems was 10, and we suggested classifying them into major and minor categories, in terms of their expected influence generally on the task performance. The following Table (5.4) presents the identified problems.

Table 5.4 Identified Problems of High Fidelity Prototype

Prob. No.	Description	Major	Minor
1	An error occurred regarding the automatic transition between the columns with three participants while performing a task. We noticed that the system gave an error message when they wanted to correct some data that already been entered.	√	
2	Missing information about the options of the type of visa, one of the participants said “it should be in a separate page that includes the field of entering the data of visa”.	√	
2	Missing information and options (like a participant has pointed out that “the system needs to include a column regarding the justification of an applicant’s request within the issuing visa process, as the purpose of the visa”).	√	
3	Other missing information included a field about the Consulate, in case the nationality does not match the place of despatch.	√	
4	Some of the participant accounts to access the prototype design did not work due to being wrongly coded, so all accounts needed to be checked be before conducting the actual test.	√	
5	Incorrect coding encountered with the P.D, so it accepts to issue a visa without payment shown in bank details.	√	
6	A title of error message appeared in a different language, and the participants could not understand it.	√	
7	In the sub-screen of bank details, a column for reference number needs to be added for obtaining the information more accurately.		√
8	Some functions titles have spelling mistakes, and others did not comply with the current terminology in the actual VI system.		√
9	Some active fields that need to be filled by users have the same colour which might be confusing to the users, so they need to be in a different colour to be easy to fill in during the tasks.		√
10	Some titles were difficult to read due to their size needing to be adjusted.		√

Figure 5.9 Screenshots of the main user interface of high fidelity prototype



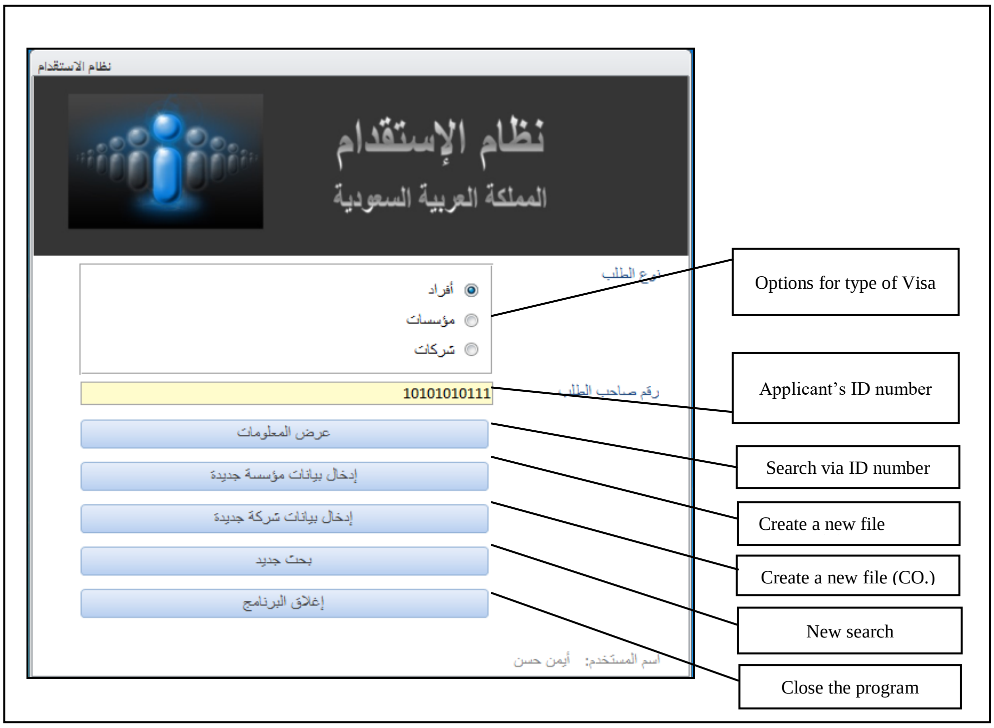


Figure 5.10 A Screenshot of the sub-screen for creating a new file for an establishment of applicant

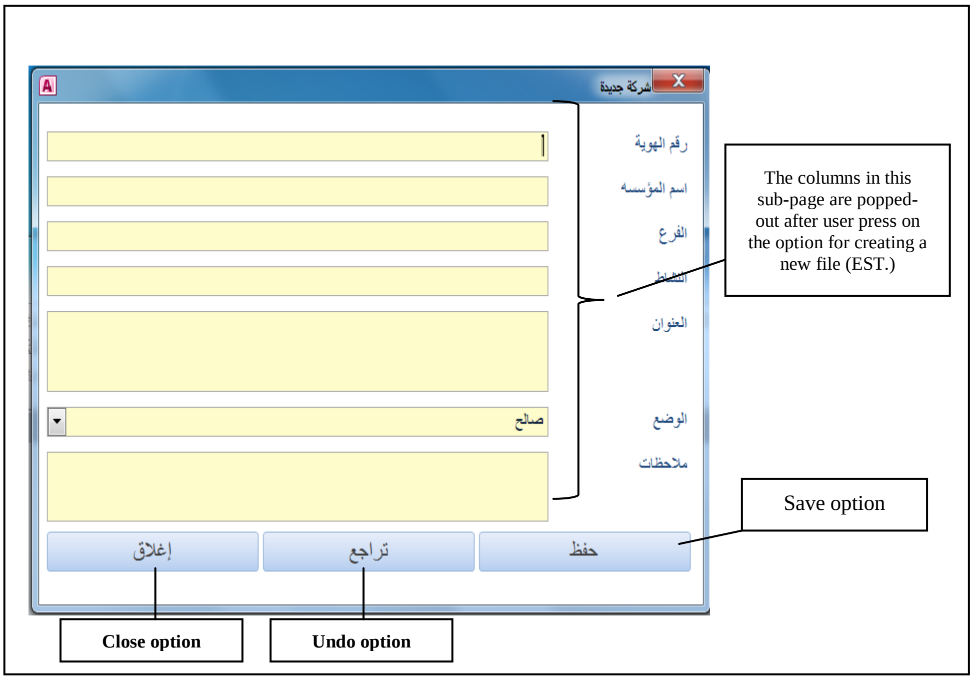


Figure 5.11 A Screenshot of the main screen for displaying an applicant details

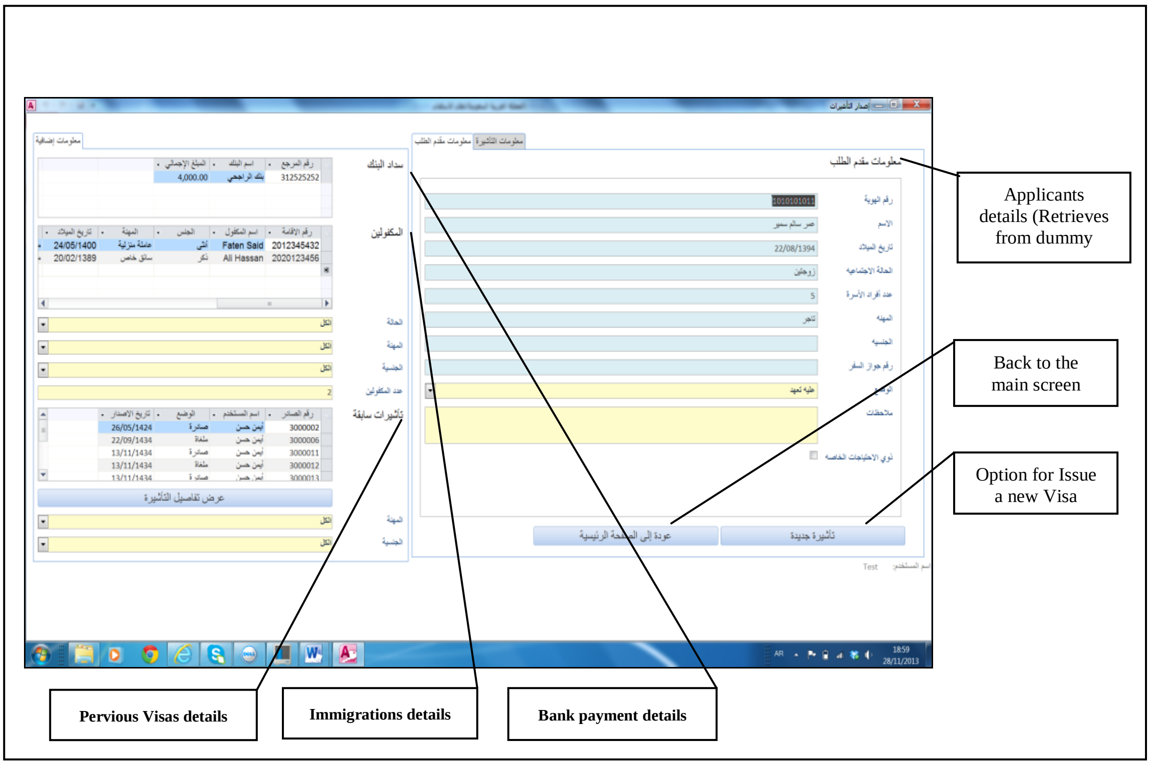
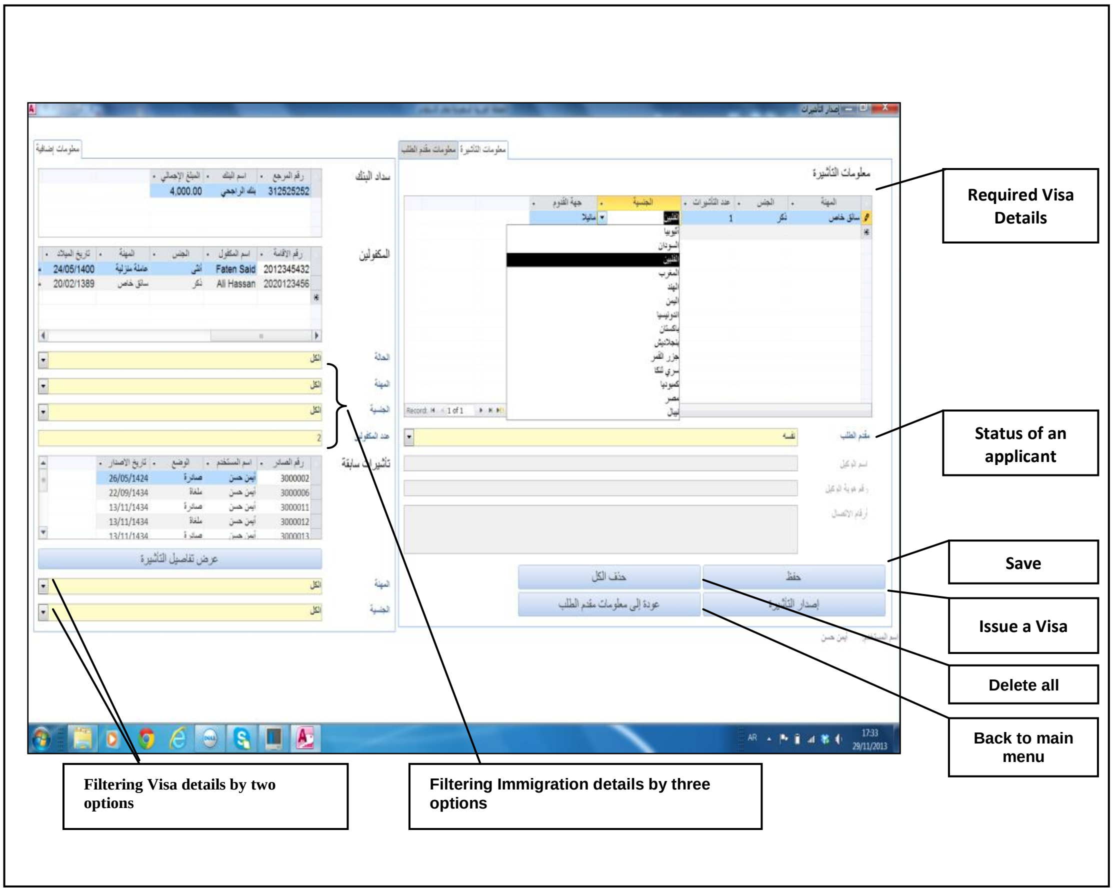


Figure 5.12 A Screenshot of the main screen for issuing a Visa



5.5 Target test tasks

The basic rule for test tasks is that they should be designed to represent real tasks, which users initially perform and complete with the existing interactive VI system every day. In addition, the tasks should cover most crucial issues of the UI design, and be sufficient to be completed within the time limits of the user test (Nielsen,1994; Kuniavsky,2003). Furthermore, most laboratory research collects task performance time, screen displays, response questions such as regarding what is being carried out well or poorly, and what parts of the system need attention. Likewise, due to the variety of functionality provided in the existing VI system, besides the limited time that was allocated for each participant to complete the entire usability test, the targeted tasks in this project are derived from the main actual tasks of the VI system that might enable participants to achieve a clear conception of the targeted system's properties. The following table presents the tasks that participants were requested to perform using both UI designs, DPD and C.S.

Table 5.5 Test Tasks (Experiment 1)

Task No.	Task description
1	Log in to the system and display an applicant's data by entering his ID number into a system.
2	Search for the details of previous visas of the applicant.
3	Search for the number of employees who are sponsored by the applicant.
4	Data entry to create a new file for a corporation by entering its data, which includes name, commercial registration number (CR no.), and the nature of corporation's activities..
5	Data entry for issuing a new visa with different requirements.
6	Amendment to the details of Issued Visa.
7	Proceed to re-print Issued Visa.

All seven tasks shown in (Table 5.5) are identical for all participants who were involved in this study. Each task has been chosen as an indicator to measure a particular aspect of the test.

Task 1 is aimed at ascertaining if participants could use the UI of a system successfully, to display required details. The applicant data will be retrieved from the dummy data, which was created for the research purposes.

Task 2 and 3 are selected to test how easy it is for participants to use search functionality, which in turn, enables them to assess the information quality through using the navigation system for finding some details.

Task 4 and 5 are designed as the key tasks for determining the system usefulness through various measurements such as task times and completion rate. This would help to decide the improvement progress of a suggested prototype design.

Task 6 is selected to ascertain whether the participants can use control functionality of a system, like the amendment ability of modifying the issued details.

Task 7 is designed to determine if participants could reach efficiently a targeted option of processing and printing an issued visa.

5.5.1 Test tasks procedure and scenario

The common view of the usability testing area is that a task procedure explains how features of a system contribute to the user's experience, using a suggested UI design as an initial implementation of a system's features (Rosson and Carroll, 2002). A usability test is intended to determine the extent to which an interface facilitates a user's ability to complete particular tasks using both DPD and C.S.

Before holding the actual usability test, all participants were informed of the aim of conducting this project and the structure of the test, how the plan goes, and how the importance of their participation would help to achieve the success of the test. This might allow them to feel comfortable and create a less pressured atmosphere (Hsieh and Huang,2008). Moreover, they were asked to complete a general pre-test demographic information and background sheet, also each of them was notified of his order number in the test and which group he belonged to (A or B), which would help in getting back to the same participant if the need arose.

In addition, the task scenarios generally attempted to cover all aspects that related to usability features and main functions of the actual VI system. Thus, a short demonstration about the new suggested prototype with its all-basic functionalities was presented clearly and in an easy way. Then they had a short training to execute all scenarios on both systems designs and to ensure they understood what to do. Rubin and Chisnell (2008) stated that providingtraining to participants would increase their skill level in performing the tasks. Especially in our case due to the participants having already experience in the existing VI system, so for making the test to have a moderation form, they should have also some experience of the suggested design. Besides, as all of them have already an access to use it, they were also given a test user ID for logging in to the DPD. As mentioned before the test was held in an informal lab at the office of the duty manager in the workplace. The test ran while the practitioner was sitting next to the participant who was able to ask any question in each session, but not while performing the tasks.

As one of the chief objectives in this project was to compare two UI designs, so we attended to apply within-subject design by testing 32 participants. They were given a guideline of the test and then were requested to carry out the series of seven tasks that have been mentioned before, using both systems, DPD and C.S. This test design mainly intends for each participant to perform all targeted tasks using the different UI design. Nielsen (1999) stated that in within-subject testing, users can not be novices when they start to use the other system, because they already have learned how to use the first system and picked up some skills. However, Rubin and Chisnell (2008), and Sauro and Leweis (2012) pointed out that biased results would be gained from the usability test when one UI design is tested before the other, because the users might learn to perform the tasks from using the first UI design or maybe this effect would be reversed and the users would have difficulty in adapting to the second UI design. Therefore, in order to tackle this issue, and to minimize carryover effects, a way of switching the order of using the two designs by participants to perform all tasks was applied. This is recognized as a counterbalancing technique (Lewis,2006;Rubin and Chisnell,2008).

For implementing this solution, at the beginning, we divided the selected participants into two groups (A and B), each group had the same number of 16 participants. Principally, the usability test was conducted in four days, each day had 8 participants who were tested in two sessions on both systems alternately. So on days 1and2, participants in group (A) were asked in the first session to use the DPD, and then in the second session to use the C.S. While on days 3 and 4, the usability test was started by asking the participants in group (B) to use the C.S and then the DPD in the second session, as shown in (Table 5.6).

Table 5.6 Within-Subjects method and counterbalancing technique

	Session 1							**Session 2**						
	Prototype Design (DPD)							Current VI System (C.S)						
	Group A							Group A						
1	T1	T2	T3	T4	T5	T6	T7	T1	T2	T3	T4	T5	T6	T7
2	T1	T2	T3	T4	T5	T6	T7	T1	T2	T3	T4	T5	T6	T7
	Current VI system (C.S)							Prototype design (DPD)						
	Group B							Group B						
3	T1	T2	T3	T4	T5	T6	T7	T1	T2	T3	T4	T5	T6	T7
4	T1	T2	T3	T4	T5	T6	T7	T1	T2	T3	T4	T5	T6	T7

This technique would enable the usability measurements to achieve an unbiased assessment of both systems (Bakhshi-Raiz et al.,2012). Consequently, each version, the prototype design DPD and current system C.S, were tested with all participants equivalently. In addition, each participant among his group (A and B) was invited in the session to perform the usability test individually. Bastien (2010) stated that most of the usability tests were employed with an individual test participant in a session. According to Kuniavsky (2003,p.53) employing *"one-on-one tests can quickly reveal an immense amount of information about how people use a prototype"*. Likewise, each participant was requested to carry on and finish all tasks, even if he was impeded by any difficulties or other external factors. In addition, as we mentioned before the tests were conducted in a quiet office of a duty manager in the work place, which would help the users to feel that the tests were realistic (Nielson and Loranger, 2006).

Figure 5.13 A participant engaging in an informal usability test

In addition, the participants were encouraged to use the systems and perform the tasks with think-aloud protocol, so that we could observe and take notes in order to identify the UI design challenges and frustrations that caused participants' errors or difficulties. Kyan (1991, p66) stated that "

Furthermore, during the test, the time that it took to finish each test task was calculated; even a task that was incorrectly done. Additionally, the actual task time was only considered when the participant continued to perform a task until he finished it. In addition, the rates of completion of a task, and the number and types of errors committed, were also recorded and codified. So, once the participants gave the agreed sign of starting the usability tests, their interaction with the targeted system was observed and notes were taken. However, the time on task was only measured from when the participant started the targeted task, which means that each task time was counted; however, this did not include the time of recording subjective evaluation during the test. The picture below (Figure 5.10) shows a stopwatch in the device (Samsung Galaxy S4) which was used in the test to calculate the time that participants consumed to perform the tasks.

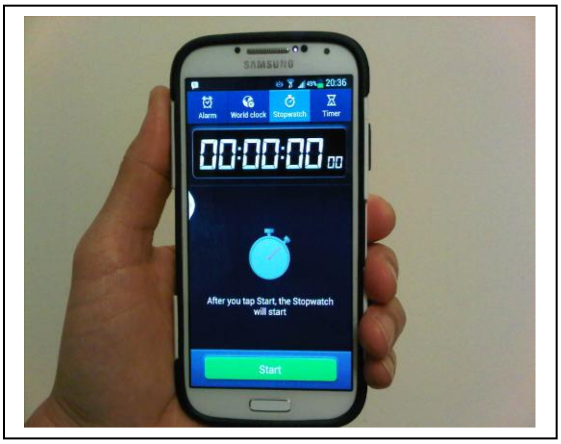


Figure 5.14 A stopwatch device used for calculating the time task

After the participants succeeded in completing all test tasks of all sessions using both designs, P.D and C.S, they were requested to provide data through responses to a survey approach (post-study questionnaire) which represents a quantitative method. This study questionnaire aims to gather participants' feedback and to assess which aspects they like and dislike on both systems. Therefore, all participants in each group (A and B) had to fill out the questionnaire two times, one about the current UI design and another about the prototype UI design. Furthermore, their responses and notes in the empty boxes under each question within the questionnaire were considered as qualitative data.

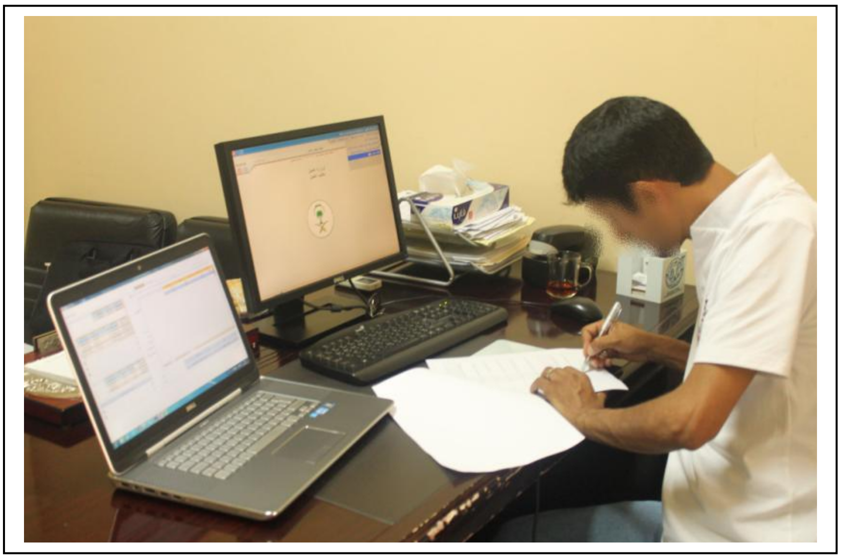


Figure 5.15 A participant filling out the post-study questionnaire after finished all test tasks

5.5.2 Pilot study of Usability within-subjects test

The main purposes of this pilot test are to learn how to perform a simple usability test with the suggested prototype in order to verify that it is well designed, and to check whether its script is clear to the participants, which would help to figure out if there are any errors could be encountered, then to sort them out and incorporate these solutions into redesigning the final experiment before running the study with a larger number of participants. Neilson(1994) pointed out that all usability tests should have a trial of the test procedure, and it can be for any case, either on a few pilot subjects, or for a large test once the initial pilot test reveals severe deficiencies in the test plan.

Therefore, 12 participants were invited to be involved in this sub test officially via making a phone call or sending emails to their personal email addresses. All of them were asked to be involved with the assessing of task procedure within this research individually at an informal lab in the work place that was prepared to conduct the usability test. All participants are employees in the governmental organization in the Recruitment Department; they have an access to work on the existing VI system as they use it in their job duties. Those twelve participants will be invited to take part in the final test, and that is because the number of employees who have access in this branch of the government organization for using the original system is limited (36). Therefore, eliminating those participants would affect the sample size in the final test. Teijlingen and Hundley (2001) stated that including pilot study participants in the main test would cause either positive or negative changes in their behaviour due to the fact they have already been involved previously in the test, and they may respond differently. However, in some research it is not possible to exclude the participants who already took part in the pilot study, because this might result in having too small a sample in the main study.

After all participants were recruited, we divided them into two groups (A and B), each included six participants. Ingroup (A) the participants were asked to start the test by using P.D in the first session, followed by using the C.S in the second session. While group Bstarted the test reversely, starting with C.S and ending with P.D. Furthermore, all participants in groups A and B have been given a brief explanation about the aim of conducting this usability test, including a short training on using the high fidelity prototype by presenting a short demo of it, to show them how the system works in general. So, each participant carried out all these series tasks while time on task was recorded, including calculating the length of the time on each task whether completed successfully or not. Furthermore, the notes were taken down in a paper sheet.

5.5.2.1 Reliability of usability test and the Procedure

For analysing the gathered data from 12 participants who performed all tasks in two sessions using both designs P.D and C.S, and to ensure the test reliability, we applied the one sample *t* - test statistic for comparing the means of tasks in both designs (Odeh and Adwan,2009; Sauro and Lewis,2010). We could obtain the average time in seconds for performing each task using both systems, P.D and C.S. The following (Table 5.7) shows the statistical results for the pilot study of the calculated task time.

Table 5.7 One Sample *t* - test

	Developed Prototype Design								Current System						
TASK	N	Mean	Std. Dve	t	Mean Diff.	Sig. 2-tailed	95% confidence interval of the Difference		Mean	Std. Dvi.	t	Mean Diff.	Sig. 2-tailed	95% confidence interval of the Difference	
							Lower bound	Upper bound						Lower bound	Upper bound
T1	12	50.42	7.597	22.99	50.42	.000	45.59	55.24	56.91	6.626	29.76	56.92	.000	52.71	61.13
T2	12	55.17	3.762	50.80	55.17	.000	52.78	57.55	98.25	41.34	8.233	98.25	.000	71.98	124.5
T3	12	54.50	9.060	20.84	54.50	.000	48.74	60.26	66.67	13.75	16.78	66.67	.000	57.92	75.40
T4	12	178.4	15.75	18.72	178.3	.000	168.4	188.4	186.8	34.57	18.72	164.9	.000	164.9	208.8
T5	12	251.6	17.72	49.18	251.6	.000	240.32	262.8	352.4	22.32	54.70	352.4	.000	338.2	366.6
T6	12	206.9	15.33	4.676	206.9	.001	109.5	304.3	164.6	18.06	31.57	164.6	.000	153.1	176.0
T7	12	62.92	16.53	13.19	62.92	.000	52.42	73.41	67.58	10.73	21.82	67.58	.000	60.76	74.40

All participants were encouraged to complete the targeted tasks; even if any challenges or errors might happen. Actually, all of them showed a good ability of using the suggested DPD after they had a short training as mentioned before, however, some problems and issues occurred and were reported and noticed. Some of these problems were related to the procedure of employing the usability test, others about the DPD.

For instance, one of the earliest issues encountered during the test was about the confusion over when the participants should start performing a task and when they should end it, in order to know the exact start of the timing or stop it. To cope with this issue, an agreement was proposed regarding the start and end task time, so when the participant attempted to start a task, he verbally said “start”, and when he finished it he said “done”.

Another issue has been taken into account for the final test, which was the contravention that occurred during running the test, which led some participants several times to abandon performance of a task. For example, the room door was knocked on several times which caused confusion in running the test, due to some staff in the same administration not being aware about the test, and some applicants having queries about their applications. Consequently, to avoid this occurring again in the final test, an announcement was circulated among the employees about the time of holding the final test and the fact that the office would be occupied. In addition the gatekeeper in the work place would handle any queries and issues that were encountered by the citizens. In addition, some of the participants received phone calls while performing tasks, which interrupted them and made it impossible to run the test well. Therefore, later on all the participants were asked politely to switch off their mobiles before starting the test.

Some participants have pointed out that the font size of some titles like “Visa Details” should be adjusted to be more clear. Furthermore, other participants have suggested that the active columns that were used for performing the tasks should have a different colour. They claimed that it would be easier for them to differentiate them from other columns for displaying information.

One expected issue was reported within the usability test, due to the fact that the suggested P.D did not cover all the protocol of the actual C.S., some participants as they thought aloud, raised some points that were not related to our usability test but rather related to the work regulations. For example, three participants asked about the amendment time limit for permitting to change issued visas. Also another two participants claimed that the P.D does not provide all the possibilities of the choice of profession and nationality required for a visa. However, this issue was explained to them in detail and was presented to the participants who are selected to be involved in the final test.

In addition, according to the results obtained for calculating time on task, we decided to use the time for each task as benchmarks for timings that established the average time for performing the tasks. Generally, the results indicated that each test session was taking approximately between 35 and 50 minutes, including around 20 minutes for the tasks test.

5.6 Results analysis and interpretation of findings

This section presents the outcomes and analyses the gathered data and fundamentally comprises transforming quantitative and qualitative data into useful results. This data was based on five main types of data, task time data, task completion rate data, and post study questionnaire as quantitative measurements, besides observation notes and comments, and a number of participants' subjective views that included positive and negative statements, these measurements were collected as a qualitative data.

5.6.1 Quantitative Data

The gathered data and information on the quantitative measures have been analysed with regards to the implementation approach of a comparison usability test within this project. Furthermore, they are commonly analysed by descriptive statistics in order to figure out users' performance measures (IAR,2011).

5.6.1.1Task times

For assessing and analysing the outcomes of the task times, we applied the method of calculating the total duration of time that participants spent on a test task for each session of using both the current system and the prototype design.

This includes all tasks that participants successfully completed, partly completed, or failed to complete.

In addition, due to the target sample in this study being 32 participants, and >25 (Sauro, et al.,2012), is considered a large sample, so we applied and calculated the common statistics that describe the task timings which consist of sample median, mean and standard deviation.

Table 5.8 Sample Statistics for each task

Sample Statistics	Task1		Task2		Task3		Task4		Task5		Task6		Task7	
	P.D	C.S	P.D	C.S	P.D.	C.S	P.D.	C.S	P.D.	C.S	P.D.	C.S	P.D.	C.S
Geometric Mean	53.7	53.14	52.95	79.97	51.79	69.31	186.19	238.78	237.81	359.09	74.25	178.05	57.97	69.13
Median	50.00	52.00	53.50	98.00	52.00	69.00	184.00	239.00	237.50	358.00	72.00	178.00	59.00	69.50
Std. Deviation	11.28	9.78	5.08	7.89	7.11	12.88	23.33	11.58	11.96	22.41	14.22	13.60	4.98	4.65
Minimum	42.00	38.00	45.00	84.00	41.00	70.31	155.00	200.00	218.00	320.00	57.00	140.00	45.00	59.00
Lower quartile (LQ)	47	48	48	92	47	59	172	232	229	341	65	172	54	65
Maximum	82.00	79.00	64.00	112.00	69.00	119.00	258.00	260.00	275.00	400.00	110.00	205.00	65.00	77.00
le(UQ)	62	58	57	105	57	76	200	247	244	372	82	188	62	73
Range	40.00	41.00	19.00	28.00	28.00	67.00	103.00	60.00	57.00	80.00	53.00	65.00	20.00	18.00
Inter quartile range	15	10	9	13	10	17	28	15	15	31	17	16	8	8
Sample Size (n)	32	32	32	32	32	32	32	32	32	32	32	32	32	32

Note. P.D= Prototype Design; C.S= Current System

Lower quartile (LQ) = $= \frac{1}{4}(n+1)th\ value$

Upper quartile (LQ) = $= \frac{3}{4}(n+1)th\ value$

Inter quartile range = Upper quartile (LQ) - Lower quartile (LQ)

The results of task times in the table above (5.8) were obtained from the examination of all tasks on both systems, P.D and C.S, in favour of measuring the system design's efficiency. Noticeably, the participants had more difficulty in completing some tasks than others and this was reflected in the average task times. Noticeably, in tasks 4 and 5 as the key tasks of the usability test, which represented the process of issuing visas and creating a new account number for an applicant's corporation, it obvious that the participants needed more time to finish these two tasks using C.S than using P.D, and they had significantly different completion time rates. So, the participants needed to complete Task 5 on P.D (M = 238.1 sec., SD = 11.96) compared to the subjects using C.S (M = 1319.91 sec., SD = 381.50). Additionally, from comparing the geometric mean of most tasks, as Nielsen (2001) recommended, it is obvious that an improvement of specific percentage could be observed, i.e. in task2 the participants could finish this task with C.S in 79.97sec. While using P.D, they took 52.59 sec. So, the improvement would be calculated as follows:

(79.97-52.59)= 27.02 sec. /79.97= 0.338 ≌ 34 %

Following this procedure;

Task1:(improvement -1.05%),Task3:(improvement 25.28%), Task4:(improvement 22.02%), Task5:(improvement 33.93%), Task6:(improvement 58.29%) , and Task7:(improvement 16.14%).

As we mentioned before the sample for this study is (>25), thus, the computing of sample median is a better measure for estimating the population median, and it should be applied as the best average task (Sauro and Lewis,2012). Love (2005) stated that "*one of the advantages that the median has over the mean as measure of central tendency is that is unaffected by extreme scores in one direction*". Initially, it is deemed as the centre point of time to complete each task that represents the time required by all participants to complete it, or it is the time that is exactly in the middle position when all the completion times are recorded in scaling order. Figure (5.16) shows the average time spent on each task for both the prototype design and the current VI system.

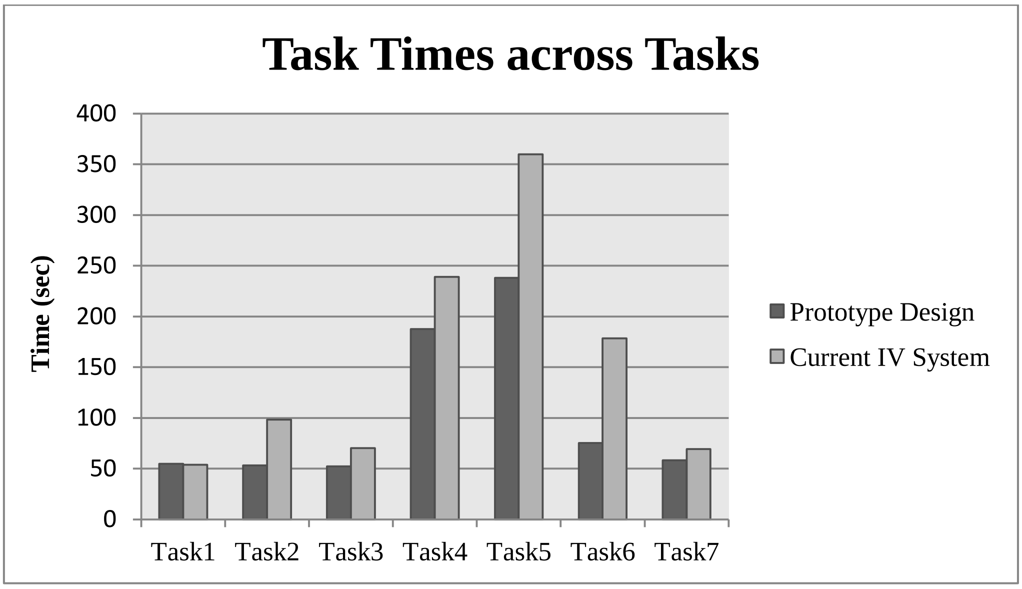


Figure 5.16 Comparison of sample median times spent on each of the 7 tasks for both design

a. Confidence interval around the Median

As regards to the task time considered as a certain type of data, it "*tends to be skewed and the median tends to be a better estimate of the middle values than the mean"* (Sauro and Lewis,2012,p.33). Consequently, when the sample median has been provided as an estimate of the average, the confidence intervals should be included within the analysis of results of task times. In particular, for a large sample such as in this study with 32 participants, it does make sense to calculate a confidence interval (CI) around the median, because it is such an observed interval estimate of a population parameter, and is utilised to measure the reliability of an estimate. In other words, the confidence interval has the same units of measurement as the variables from which they are calculated (Lewis,2006).

There are a number of methods with all confidence interval formulas to calculate them, though in this study the method of binomial distribution was applied to estimate the confidence interval because it should be appropriate for large samples. The following equation constructs a confidence interval with the most common median (0.5) around any percentile such as (0.05, 0.97, or 0.25).

$$np \pm z_{(1-\frac{\alpha}{2})}\sqrt{np(1-\rho)}$$

n is the sample size
ρ is the percentile expressed as a proportion (0.5 for the median)
$z_{(1-\frac{\alpha}{2})}$ is the critical value form the normal distribution (1.96 for a 95% confidence level)
$\sqrt{np(1-\rho)}$ is the standard error

Table 5.9 Descriptive statistics for overall task times results

System	N	Median	Std.Deviation	Std. error	95% Confidence Interval		Min	Max
					Lower bound	Upper bound		
P.D	32	720.5	36.58	12.67	707.82	733.17	644	759
C.S	32	878.5	168.64	58.43	820.06	936.93	692	1141

Note. P.D= Prototype Design; C.S= Current System

The Table above (5.9) shows that overall on average the participants spent 878.5 Sec. (Range: 820.06 – 936.93) to finish all seven tasks using the current VI system, which takes more time than the prototype design where the participants spent 720.5 Sec. (Range: 707.82–733.17). This main result from all data analyses indicates that task completion times differed statistically, so DPD shows a significant improvement over the C.S.

Because the paired *t*–test is not very sensitive to non-normal data, so the deviation from normality has to be pretty dramatic to make the paired *t*–test inappropriate (Mackdonalds,2014). In addition, in order to obtain the magnitude of difference (MacFarland,1998), between the two system designs and to examine if participants had a more positive performance with the developed prototype, Wilcoxon's matched-pairs signed ranks test was used for ranking the data to provide two rank totals, one for each system. Furthermore, it is a non-normally distributed paired samples t-test (McDonald, 2014; Hole, 2011). The data in (Table 5.10) shows the calculated sum of the ranks of the positive and negative differences of the overall pairing of the task time needed for the participants to finish performing all tasks of the work research.

Table 5.10 Ranks

		N	Mean Rank	Sum of Ranks
DPD - CS	Negative Ranks	0[a]	.00	.00
	Positive Ranks	32[b]	16.50	528.00
	Ties	0[c]		
	Total	32		

a.DPD < CS
b.DPD > CS
c.DPD = CS

Table 5.11 Test Statistics[a]

	DPD - CS
Z	-4.937[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

A Wilcoxon signed-rank test (Table 5.11) shows that a difference exists between the users' overall time of finishing the test tasks, and it is a statistically highly significant change using developed prototype designDPD in individuals compared with the existing system, with less time (Z = -4.937, p = 0.000).

5.6.1.2 Task Completion Rate

As we mentioned before this measurement of task completion rate or success rates is considered a vital measure in the usability evaluation (Nielsen, 2001). Initially, the procedure of collecting and recording data is based on the simpler binary measure, in which a task success is coded as the number (1) while a task failure or uncompleted is coded (0). Therefore, the participants were asked to finish all the tasks, and after that, only the completion rate is reported as a percentage.

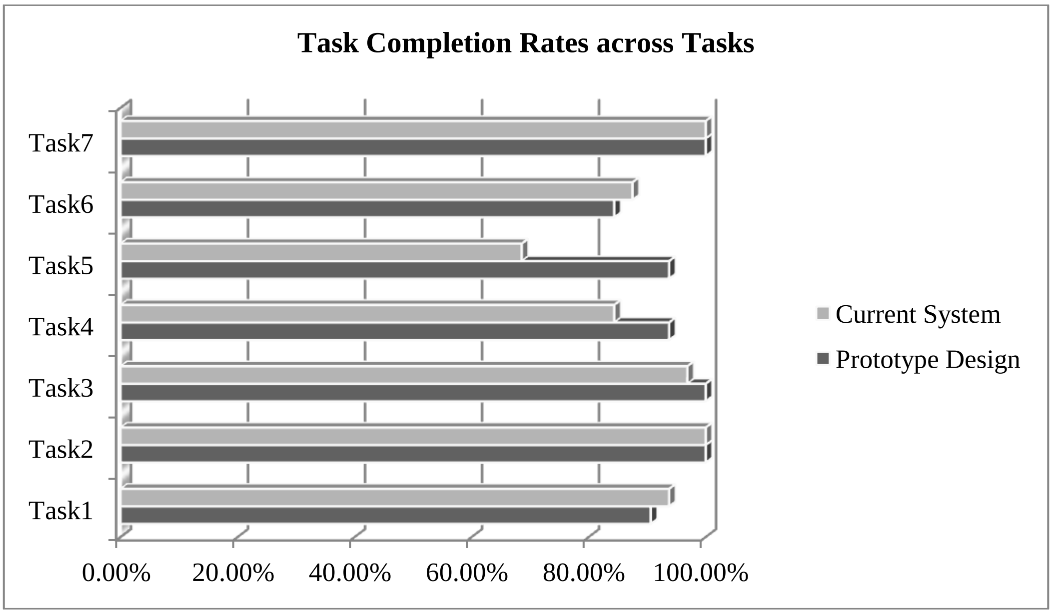


Figure 5.17 Comparison of percentage of participants who successfully completed each of 7 tasks

Figure (5.17) shows the percentage of completeness and accuracy of targeted tasks that participants could finish successfully by using both systems. As mentioned previously, the Tasks 4 and 5 were considered as key tasks in this usability test, and it is obvious that the participants had difficulty in performing this task using C.S with 78.13% completion rate. In tasks 2 and 7, the participants were able to complete all the targeted tasks fully 100%. This would be referring to the difficulty of performing these tasks of assessing the efficiency, which were ranging between easy to medium. One of the noticeable results we came across was related to Task 6 and the outcomes of task times in the same task. So, we noticed that although the participants could finish the task 6 using DPD at times faster than when they used C.S, the rate of completing the task successfully using DPD was lower than the completion rate using C.S in the same task. From analysing the results of observation techniques, we found that some errors having occurred with some participants during performing the task due to a software bug which causes it to generate incorrect results. For example, in task 5, when the participants wanted to enter data and move from a column to another without selecting any information from the given option, the cursor was jammed and the participant needed to go back again to start processing the test task .

According to Sauro and Lewis (2005) it is essential to include confidence intervals around the difference (the effect size), when reporting the completion rates, in order to find out if there is sufficient evidence that more participants could finish the test tasks using the DPD over the CS. However, because the test was conducted with a large sample a within-subjects comparison technique was applied in the test design, and each participant was tested twice (matched pairs); we needed to determine if there is a significant difference between completion rates using both system designs, or any dichotomous variables. Therefore, the McNemar test was used, which initially counts the number of participants who had discordant pairs of swapping reciprocally between pass (Successfully completed) and Fail (Not successfully completed). In addition, the McNemar's test is on a 2x2 contingency table; however, in order to present the outcomes of the concordant and discordant participants from seven tasks clearly, we compromised the results in one (Table 5.12) which would help to identify whether they performed the tasks better or not, using the improved prototype. Furthermore, the McNemar would help to test the difference between paired proportions.

Table 5.12 Concordant and Discordant Responses for all tasks tests

			Current VI System Pass								Current VI System Fail							Total
			T1	T2	T3	T4	T5	T6	T7		T1	T2	T3	T4	T5	T6	T7	
						(a)								(b)				(m)
Prototype design Pass	**T1**		27	-	-	-	-	-	-		2	-	-	-	-	-	-	29
	T2		-	32	-	-	-	-	-		-	0	-	-	-	-	-	32
	T3		-	-	31	-	-	-	-		-	-	1	-	-	-	-	32
	T4		-	-	-	25	-	-	-		-	-	-	5	-	-	-	30
	T5		-	-	-	-	18	-	-		-	-	-	-	10	-	-	28
	T6		-	-	-	-	-	22	-		-	-	-	-	-	4	-	26
	T7		-	-	-	-	-	-	32		-	-	-	-	-	-	0	32
						(c)								(d)				(n)
Prototype design Fail	**T1**		3	-	-	-	-	-	-		0	-	-	-	-	-	-	3
	T2		-	0	-	-	-	-	-		-	0	-	-	-	-	-	0
	T3		-	-	0	-	-	-	-		-	-	0	-	-	-	-	0
	T4		-	-	-	2	-	-	-		-	-	-	0	-	-	-	2
	T5		-	-	-	-	2	-	-		-	-	-	-	2	-	-	4
	T6		-	-	-	-	-	5	-		-	-	-	-	-	1	-	6
	T7		-	-	-	-	-	-	0		-	-	-	-	-	-	0	0
Total		(r)	30	32	31	27	20	27	32	(s)	2	0	1	5	12	5	0	32(N)

The concordant pairs were presented in two squares, (a) where the participants completed successfully the seven tasks using both designs, and square (d) the participants oppositely fail in performing tasks using both P.D and C.S., in our test only one participant has failed in Task6 using both systems.

The discordant pairs were displayed in square (b) where the participants could pass the tasks using the DPD but failed in using the C.S., on the contrary, in square (c) the participants had switched from passing the tasks using the C.S to failing in P.D. Furthermore, (N) in the table shows the total number of participants in each task. N= cells (a+b+c=d).

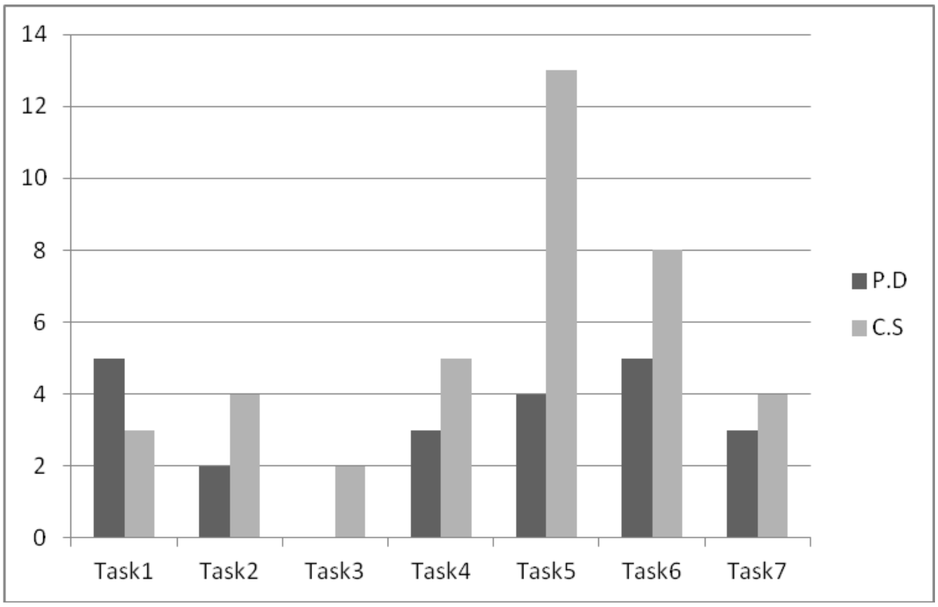
5.6.1.3 Error Rate

The gathered error rate data after the participants completed the targeted tasks was aggregated from each task as number of errors occurred in the failure or uncompleted tasks completion, besides the others which were observed and recorded such as any unintended action or omission by a user while performing the task (Sauro and Lewis,2010).

Table 5.13 Number of Errors in each task

	P.D		C.S	
Task	**No. Error**	**Problem description**	**No. Error**	**Problem description**
T1	5	-Incorrect login to the system. -Wrong clicks by getting confused which to use between the mouse and tab on keyboard.	3	-Incorrect login to the system -Incorrect choice screen for displaying applicant details
T2	2	-Wrong clicks	4	-Wrong clicks on the mouse instead of tab -Wrong clicks by choosing wrong buttons.
T3	0	-	2	-Chose incorrect option for displaying the required details.
T4	3	-Entering data into the wrong field of another dedicated case. -Wrong clicks	5	-Lost data after choosing wrong option for correcting the details. -Wrong clicks
T5	4	-The inability to make any amendment on data entry of applicant until a full filling of fields -Entering incorrect data into specific fields.	13	-The System displayed the required information late. -The wrong choice between several opened screens for finishing the tasks. -The wrong choices of required details of visa. -Wrong clicks.
T6	5	-System Miscoding resulted in that the data of some previously issued visa did not accept the change on its details.	8	-The system did not allow for the requested amendment of an issued visa: " System regulations" -Wrong clicks.
T7	3	-Wrong clicks.	4	-Wrong clicks -Wrong entry data of issued visa number to give an order to reprint it.

Figure 5.18 Total number of Errors for each task



According to the results of the error rate, from analysing each task individually, we are going to report the main errors. In Task 1, the error of wrong login to the DPD has occurred with three participants, which caused some delay in finishing the task. This more likely happened, because the participants were not familiar with the user account that has been given to them for accessing the DPD system, and they tried incorrectly to type their account for using C.S instead. On the other hand, there was only one participant who failed to login to the C.S correctly.

As mentioned before Task 2 and 3 were designed to evaluate generally the information quality by assessing the search functionally. It obvious that the participants found that it was easy to use both system designs to search for some related information. However, two participants committed an error in task two using C.S by wrong clicking using the mouse rather than keyboard. In task 3 there were 2 participants who chose a wrong option for calling up the details of previous visas. One participant claimed that he was sometimes getting a bit confused with the mass of useless columns on the UI of C.S.

In Task 4 and 5 as they were selected as main tasks for the usability test, it is clear that the participants were struggling to perform them using the C.S design, in particular Task 5 when it is compared to their performance using the P.D in the same task. The reported errors in this task using the C.S was related to the fact that six participants carried on with the task, but at the end they could not finish it due to there being no payments shown earlier. Another four participants made a wrong choice of required visa details.

The C.S in Task 6 did not accept modification of visas in three cases, as due to the regulation the visas had expired, thus, the participants needed to call up another visa to conduct the task.

In the Task 7, three participants had errors by entering the wrong number of the issued visa using the C.S, while all the participants could finish this task efficiently and reached the option of giving an order to print issued visas using the P.D.

5.6.1.4 Post study questionnaire results

As mentioned before, in our study, the post study questionnaire has been used as a tool for measuring the participant’s satisfaction. For statistical analysing of the gathered data from the post study questionnaire as continuous data , and as the participants completed it after using both systems following the within-subject approach, we applied Paired-sample *t*-test (Lazar et al, 2010 ; Sauro and Lewis, 2012)

Table 5.14 Paired Sample *t* test

Items	Paired Differences					t	df	Sig. (2-tailed)
	Mean diff	Std. Devi	Std. Error Mean	95% Confidence Intervals				
				Lower Limit	Upper Limit			
Pair 1- Overall, I am satisfied with using this system for issuing work visas.	1.53	1.22	.21527	1.09221	1.97029	7.113	31	.000***
Pair 2- I could effectively complete the tasks and scenarios using the system	0.88	0.87	.15391	.56109	1.18891	5.685	31	.000***
Pair 3- This system has the functions and capabilities I need to perform my work.	1.47	0.98	.17381	1.11426	1.82324	8.450	31	.000***
Pair 4- I needed to learn a lot of things before I could get going with this system	0.50	1.44	.25400	-.01804	1.01804	1.969	31	.058
Pair 5- The system consumed much time to do the required task.	1.13	1.86	.32919	.45361	1.79639	3.417	31	.002**
Pair 6- The system required only a few steps to accomplish the tasks.	2.56	.84003	.14850	2.25964	2.86536	17.256	31	.000***
Pair 7- It was not easy to learn how to use this system.	0.91	.81752	.14452	.61150	1.20100	6.271	31	.000***
Pair 8- Data entry of new applicant information was easy and clear.	0.28	.81258	.14364	-.01172	.57422	1.958	31	.059
Pair 9- Whenever I made a mistake while using the system, I could recover it easily and quickly.	0.56	.94826	.16763	.22062	.90438	3.356	31	.002**
Pair 10- I felt in control when I was using the system.	1.00	1.13592	.20080	.59046	1.40954	4.980	31	.000***
Pair 11- When I used the system, it was easy to move from one step of a task to another	0.69	1.30600	.23087	.21664	1.15836	2.978	31	.006**
Pair 12- It would be difficult to be skilful at using and controlling this system in a high degree of performance.	-0.061	1.77687	.31411	-.70313	.57813	-.199	31	.844
Pair 13- I could easily find the required information to complete tasks.	1.38	1.03954	.18377	1.00021	1.74979	7.482	31	.000***
Pair 14- I think I would need technical support to be able to use this system.	0.53	1.75948	.31103	-.10311	1.16561	1.708	31	.098
Pair 15- Navigating information through the menus and toolbars of system was not easy to do.	-0.13	1.896	.355	-.809	.559	-.373	31	.712
Pair 16- The system has not helped me to overcome any problems I have had after I finish a task, such as amend the input data.	0.50	1.31982	.23331	.02415	.97585	2.143	31	.040*
Pair 17- The organisation of information on the system user interface was clear.	1.63	1.12880	.19955	1.21802	2.03198	8.143	31	.000***
Pair 18- The information provided for the system was easy to understand.	1.10	1.01176	.18172	.72566	1.46789	6.036	30	.000***
Pair 19- I understood the function of each buttonin the systemeasily.	0.69	1.11984	.19796	.28376	1.09124	3.473	31	.002**
Pair 20- The system provided a capability (such as UNDO feature) which enables me to quickly reverse mistaken actions.	1.66	1.28539	.22723	1.19282	2.11968	7.289	31	.000***
Pair 21- The system features (like default saving data and copy) helped me to perform tasks successfully.	2.34	1.09572	.19370	1.94870	2.73880	12.100	31	.000***
Pair 22-The interface of this system was pleasant to use.	0.59	1.34066	.23700	.11039	1.07711	2.505	31	.018*
Pair 23- I would be happy to use the system UI again.	1.44	1.24272	.21968	.98945	1.88555	6.543	31	.000***
Pair 24- It is obvious that user needs have been taken into consideration when the system was developed.	2.09	1.02735	.18161	1.72335	2.46415	11.529	31	.000***

Level of significant *P<0.05,**P<0.01,***P<0001(two-tailed tests)

The results in the Table (5.14) from utilising the paired t-test for analysing the data, based on the sampling distribution of mean differences, indicated that the vast majority of the pair's questions (19 out of 26) have a significant statistical difference between the participants' responses on the questionnaire on both systems, so p-value < 0.05 level of significance. Moreover, there are three pairs that have a low significance while only two pairs (12 and 15) are not significant. Since we found that 79.17% of the paired questions were statistically significant this would indicate that the users' satisfaction has been improved through using the DPD.

In addition, to assist this outcome, and in aiming to have more clarification on the post study questionnaire's outcomes, the total average values of results for each of the sub-scales within the post study questionnaire will be calculated, which can be correlated to some extent rather than if they were calculated independently. As we mentioned before the average will be counted for the questions as follow:

a. System Quality (SysQual): Average the responses to statements 1 through 9
b. Information Quality (InfoQual): Average the response to statements 10 through 18
c. Interface Quality (IntfQual): Average the response to statements 19 through 24

Before conducting the analysis, we ought to present the criterion of weighted mean for each response on the Likert scale within the post study questionnaire, so based on this, number one can specify the overall response of each statement (in mean). The following Table (5.15) represents this criterion of weighted mean for each sub-scale in using a 5- Likert scale within the post study questionnaire.

Table 5.15 The Criterion of weighted mean

Weight Mean	Response
From 1.00 to less than 1.80	Strongly Disagree
From 1.80 to less than 2.60	Disagree
From 2.60 to less than 3.40	Neutral
From 3.40 to less than 4.20	Agree
From 4.20 to less than 5.00	Strongly Agree

Table 5.16 Descriptive Statistics

	P.D				C.S			
	Mean	**Std. Devi.**	**Min**	**Max**	**Mean**	**Std. Devi.**	**Min**	**Max**
SysQual	4.00	0.34	3.45	4.73	2.94	0.34	2.27	2.27
InfoQual	3.59	0.54	2.88	4.75	2.89	0.44	2.13	2.13
IntfQual	4.14	0.53	3.20	5.00	2.39	0.64	1.20	2.40

As a consequence of this calculation, the results (means and standard deviations) from analysing the data are reported in Table (5.16) . It is obvious that the mean of the sub variables in P.D was between 3.6 and 4.1., which means the participants were satisfied and “Agree” d in total while in the C.S between 2.9 and 2.3. This reflects that the participants’ satisfaction was “Neutral”.

a.Wilcoxon matched-pairs signed-ranks test

In order to determine whether there are differences between the two systems, C.S and P.D, in terms of the participants' satisfaction, Wilcoxon matched-pairs signed-ranks test was applied as a non-parametric test since the sample size of (32) participant was considered as a small size, besides the utilised within-subject design (Cairns and Cox,2008). The data in the following Table (5.17) showed the sum of the ranks of the positive and negative differences of each pair of the test tasks for each sub-scale of the post study questionnaire, and for the total.

Table 5.17 Ranks

		N	Mean Rank	Sum of Ranks
SysQualCS - SysQualPD	Negative Ranks	32[a]	16.50	528.00
	Positive Ranks	0[b]	.00	.00
	Ties	0[c]		
	Total	32		
InfoQualCS - InfoQualPD	Negative Ranks	28[d]	16.04	449.00
	Positive Ranks	2[e]	8.00	16.00
	Ties	2[f]		
	Total	32		
IntfQualCS - IntfQualPD	Negative Ranks	31[g]	16.00	496.00
	Positive Ranks	0[h]	.00	.00
	Ties	1[i]		
	Total	32		
Total CS – TotalDPD	**Negative Ranks**	**32[j]**	**16.50**	**528.00**
	Positive Ranks	**0[k]**	**.00**	**.00**
	Ties	**0[l]**		
	Total	**32**		

a. SysQual(CS) < SysQual(PD)
b. SysQual(CS) > SysQual(PD)
c. SysQual(CS) = SysQual(PD)
d. InfoQual(CS) < InfoQual(PD)
e. InfoQual(CS) > InfoQual(PD)
f. InfoQual(CS) = InfoQual(PD)
g. IntfQual(CS) < IntfQual(PD)
h. IntfQual(CS) > IntfQual(PD)
i. IntfQual(CS) = IntfQual(PD)
j. Total(CS) < Total(PD)
k. Total(CS) > Total(PD)
l. Total(CS) = Total(PD)

Table 5.18 Test Statistics[a]

	SysQualCS - SysQualPD	**InfoQualCS - InfoQualPD**	**IntfQualCS - IntfQualPD**	**Total CS-DPD**
Z	-4.939[b]	-4.461[b]	-4.867[b]	-4.937[b]
Asymp. Sig. (2-tailed)	.000***	.000***	.000***	.000***

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

5.6.2 Qualitative Data

In addition to quantitative data analysis, the study employed qualitative methods for interpreting the data. Zimmerman and Muraski (1995) stated that the data could be simply reported as what has been found in a study by a researcher, or can even be interpreted in a way that shows whether this analysed data supports a problem statement.

This sub section presents the gathered qualitative data which included the observational and note taking techniques employed while the participants were attending the sessions of using both systems. Besides the comments that have been collected from the participants' responses during and after the test in the empty fields, besides the part of explanation in the opened questions section within the post study questionnaire. For summarising the enormous amount of qualitative data obtained, we employed content analysis technique to compress this data into specific categories based on the participants' comments and feedback and through the observations and questionnaire methods (Lazar et al, 2010; Stemler,2001). Due to most of this data being provided within the post study questionnaire, and since each question was reflected as measuring one of the three sub-scales, we suggested using the three sub-scales of the study (system quality, information quality, and interface quality) as key themes within the qualitative data analysis. The following (Table 5.19) shows the number of positive and negative statements that we gathered from the comments collected in the observations technique and the participants' subjective views.

Table 5.19 Usability positive and negative statements

Content Classification	Parti. No.	P.D		C.S	
		+	-	+	-
Sys. Qual. (System usefulness)	1	-The system has limited steps to finish the tasks. - The basic features, like copy and paste, are very helpful in terms of moving the data entry from one column to other.	- The system needs to enhance the automation movement from one step to another, so I would be able to perform the tasks more consistently.	- I liked how the system was integrated with other systems, like the immigration system, which enables me to get the related information for doing the tasks easily.	-The system forced me to open many screens to finish a task.
	2	None	None	None	- It consumed a lot of time to issue the visa, due to the many stages that I had to follow.
	5	- The default saving of data feature (auto save) in the system saves my time and effort, so I am not worried to lose my data entry.	None	None	- The system does not support the feature of saving data, so it would be possible for the data entry to get lost.
	7	- The system was easy to use, and I was able to learn to use it quickly.	- Some options and icons like the technical support on the first screen were not activated, and I supposed if they were, I would feel more control of using it.	None	None
	8	- The system was easy to use, and included most of what I need to be able to finish the tasks in a short time.	None	None	None
	15	-Filtering the previous issued visas with different options (like job and nationality) and find out their details without needing to check each one individually was supportive for making a quick decision.	None	None	- The lack of providing a proper way for searching on applicant details in the system would lead the users to get different decisions regarding the issue visa for the applicant.
	17	-I liked the Undo feature, it makes the correction of mistakes easier.	None	None	None
	22	None	None	None	- The system does not support the automation movement between the

					columns, so I wasted some time in choosing between the tab and mouse while I was performing the task.
	23	- I liked the feature of categorizing the options for searching for some previous details i.e. searching about the previous workers by their nationalities, it saved my time.	None	None	None
	27	None	None	None	- The system forced me to open many screens to finish a task that consumed much time to end it. This would delay serving the applicant.
	29	- The system opens a new screen directly for processing a new applicant after I finished the previous one.	None	None	- I need to call a new page for starting to process a new applicant, this might consume time with heavy workload.
	30	- Entering the applicant data in the system for doing the task was easy and straightforward.	None	None	None
Info. Qual.	1	- Adding a column for the delegation information was a remarkable improvement that enables me to do the tasks more accurately.	None	None	- The system does not have any option of a direct technical support.
	2	- Showing the bank account details was very helpful for finishing the tasks.	None	- The system UI has a clear displaying of information, and I could pick it up easily.	None
	5	- Picking up early information related to completing the tasks would help me to take the decision faster. i.e. provided information about the amount of money in the bank account for an applicant, and the previous issued visa details.	- The system needs to be improved in terms of completing the missing information and options in some tasks i.e. some occupations were not included in the options.	None	None
	16	None	None	None	- Some required information for completing the task, like the bank details, it just would come in the last stage of the issuing of the visa, which considered at the end as wasting time.
	19	- There was no need to open many screens for doing the task. So, all the required information was available in one screen, which in turn meant I could quickly finish my tasks with efficiency.	None	-The system ability for integration with other systems, was very helpful for gathering related information, however, it needs to have more improvement.	None
	22	- I liked using this system; because after I used it, I could efficiently finish the task in a short time.	None	None	None
	27	None	None	None	None
	29	- All the required information appeared earlier in the task process, so I can take the decision to issue a visa for the applicant.	None	None	None
Intf. Qual.	1	-The main screen (interface) has all the required steps and information to finish my tasks.	None	None	-The system has many steps to finish the task, which would lead to wasting my time.
	5	None	None	- I like the user interface of the system; it was pleasant and easy to understand.	None
	7	- I liked much the limitation of several screens into one screen for doing the tasks.	None	None	None
	11	- There was no need to open many screens for doing the task. So all the required information was available in one screen which in turn meant I	None	None	None

		could speed up finishing my tasks efficiently.			
	16	- The user interface of the system was very clear and simple; because I could understand how to use it quickly.			- The system has many unnecessary columns and icons that would let the users to get more confused.
	19	- I liked the reduced number of required columns that need to be filled in for doing the tasks which enhanced the clarity of UI.	None	None	None
	22	None	None	None	- The system has a lot of distracting information and many steps for doing the task that would make the user feel bored to use it.
	26	- I likedthesimplicity ofthe system, and my sense ofthe ability tolearnitquickly.	None	None	None
	32	None	None	None	None

5.7 Discussion

In this study, the overall outcomes from the usability testing indicated that the improvement of a developed prototype design (DPD) with some functional usability features has been proved through applying the experimental approach with typical users of the VI system. It has shown the low, moderate, and significant differences between performing the tasks using DPD or C.S by the participants. Accordingly, considering the usability requirements in the form of functionality in line with the nature of the VI system afforded useful information for providing improved features that led to enhancing the system usability, and then to satisfying its users (Juristo,2009). Although this issue in the study would support the view of Juristo on the one hand, on the other hand, there was a slight difference regarding his suggestion that usability features with specific functions should be dealt with at the requirements phase. Since the original system in our study is already implemented, however, we identified several requirements for improving the system ease of use that have been provided by the typical users. In the sense that employing the iteration design method through a usability test for improving the system design can be at any stage of the system development lifecycle as required by the user until fulfilled to their satisfaction. In addition, affording such a true iterative design would enable the comprehensive overhaul and rethink of a system design (Rubin and Chisnell, 2008).

At the beginning of the study were investigated the research questions by applying different analysis, paired t-test was enrolled for testing the gathered data on tasks time and successfully accomplished tasks rates were indicated. The first paired t-test in Table (5.7) tested whether the participant could finish the tasks by using DPD. rather than using C.S. The test outcomes indicated that there were significant differences, and that the participants could finish all the tasks faster using DPD than C.S. except Task 1,which was insignificant, due to the participants having incorrect login to the DPD system, which caused some delay in finishing the task. The second calculated paired *t*-test shown in (Table 5.14) indicated that there were significant differences between most of the participants' responses on the statements within the post-study questionnaire about the two systems. It has been noticed that some participants had failed to complete some tasks, though their responses on the post study questionnaire were still positive, and expressed their satisfaction and improved UX. As we stated previously, it would reflect their motivation for being involved with the improvement of the system, which they use on a daily basis. This issue came across in the study by Travis (2008), but he argued that the possible reason behind this issue could be related to giving the participants an amount of money as encouragement to participate in the test. It also may be for another reason as Wiklund et al (1997) stated that the participants may have sympathy with the practitioner who conducts the test, and they do not want to hurt his feelings. Regarding the measure of successful completion and error rates, we used the McNemar test for counting the participants who had discordant pairs, so we could identify the number of participants who had discordant pairs, changing between successful and unsuccessful completion on the two system designs.

Conducting the test in an informal laboratory at the work place on the basis of many points that were mentioned previously, was helpful for obtaining useful results. However, we should bear in mind several issues and problems that would impact on running the usability test negatively. For example, one of the major issues we came across was that since the study was implemented in the governmental organization which provides such services for the citizens, by the busy employees as they were the participants in the study, it is obvious that due to the anticipation of several interventions that might occur during the test, it was a challenge to conduct it in a proper and quiet atmosphere. To mitigate the effects of this issue, having such a good access, chiefly with a gatekeeper of the workplace, and then generally with the whole sample, could obtain for the usability test an appropriate arrangement. Nielsen (2009) pointed out the importance of a gatekeeper role as an anchor person in the experimental design for ensuring a good usability, besides the need for the presence of someone with whom participants have a trusted relationship for enabling access to accurate research data (Herrod,2012). Furthermore, as we mentioned previously, at the beginning of setting up the test, we faced some major problems, like the installation of the prototype design on the workplace PCs, besides the misunderstanding of the test procedure by some participants, especially as they were part of a large sample for usability testing. Therefore, our case follows the view of (Nielson, 1994), which recommended conducting a pilot test study that represented all subjects that will be included in the actual test for achieving a clear understanding of all its aspects, besides overcoming any issue that might be encountered.

Following the above challenges, the experimental usability test in the study, as a concept was brand new for most of the participants, in as far they had not been involved in any test before, and this could have a likely negative impact on either the test plan or running the test itself. What is more, due to the study employing within-subject technique, the participants had to use both systems, the C.S they had already had experience of using, while the DPD was a new system that has been developed for the purpose of the study. In light of this, in order to conduct the comparison usability test between the two systems in a balanced way, we had to ensure that each system can be used by the participants almost equally. Accordingly, in our case as we intended to implement a simulated system with some functional features; the participants needed to have a small amount of training to undertake the evaluation test. As stated by Souro (2009), after providing training materials to users before conducting an actual usability test, it might have a significant impact commonly on the test, but most likely not on perceived ease of use. Although all the participants had a short training before using the improved system, it was noticed that most of the participants who were already involved in the pilot study of the test procedure, could perform the tasks in the actual test slightly faster than others regardless of the errors, this might be due to the fact they had already trained and then had more experience using the P.D. This was clearly shown with the first six participants in the group (A) who started the test using the DPD.

In respect to the literature about the orientation to reduce a type of fidelity prototype due to the constraints by different aspects like the industrial design process, time and budget limitations, most of the studies claim that this reduction would provide similar outcomes to a higher fidelity prototype (Virzi et al, 1996; Sefelin et al.,2003; Walker et al, 2002). In contrast, Sauer et al (2010) stated that overall many studies provide a general recommendation that the benefit from reducing fidelity prototypes was not sufficient to obtain a more accurate analysis of this pattern. Hence, the combination of developing medium and then high fidelity prototypes of the VI system and implementing them in consecutive stages in this study, would help to achieve firstly a quick and easy access to designing the recommendations, getting feedback from the participants, and conducting the modifications (Preece et al,2002). Then to use a more functional design that is applicable to the typical tasks of the VI system that led to having a significant influence to assist the implementing of the comparison usability test. In addition, since this study was concerned with re-engineering the existing system in its later stages, so the fidelity affected a prototype's usefulness, and medium and high fidelity were considered for the prototype design. Galitz (2007) presented three reasons for the lesser usefulness of low fidelity prototypes in later stages of design, namely that they are; (1) limited in functionality, and in general do not consist of a system's features;(2) they have less interaction, and they would prevent use of some tools such as a mouse and keyboard; and (3) they have limited features.

Another key issue was identified via observing the participant’s behavior, that informing the participants about their substantial role and the value of their views to the project could increase their motivation, which in turn augmented their interaction with the test. So, we noticed that most of the participants tried to show their knowledge by providing numerous comments whether writing or verbally as a contribution to the system development. According to Hinderer and Arbor (1998) utilizing effective participant recruiting to make the participants likely to use the system is considered a substantial method for collecting reliable data when conducting the usability test for high-technology products and services. However, we kept in mind, to avoid convincing the participants via any words or action to agree with the improvement of the VI system, and we let them freely express what they factually have seen. Furthermore, some of these provided comments were not related to the aims of study, but rather related to the regulations of the work. For example, they suggested that the system needs to extend the allowed time limits either for amendment of the details or cancellation of the issued visas. Accordingly, it consumed a bit of time in the test session, which could impact on conducting the next sessions. This point corresponded with what Lazar et al. (2010,p.271) stated that "*the more that users talk, the more their task or time performance data may be influenced*". Therefore, we provided to them a short demo with relevant materials about the test and its aims, so they would have a clear vision of the entire test and what they needed to do precisely to avoid wasting time.

Aaccording to the targeted tasks which were designed for the purpose of conducting the usability test, we noticed that the most significant difference was in the tasks 4 and 5, because these two tasks represented the daily frequent tasks of the original VI system, and which, based on the pilot test study, took the participants the longest time compared to other tasks. Hence, we found that most of the participants had difficulty in finishing these two tasks with the C.S compared to their performance with DPD. However, according to the obtained results, the rest of the tasks the participants performed showed only small differences. We would claim that this was due to these tasks being around easy to medium in terms of the difficulty, while the tasks 4 and 5 were a bit complicated and harder to finish. Furthermore, due to the VI system being implemented only for the purpose of providing such particular services to the citizens, by issuing the visa to them, we do not find a variety of major tasks to be tested within the study, however, the tasks 4 and 5, as far as we expected they supported the study aspects. In addition, another issue that was related to the test tasks as well as an obstacle to the experiment design was confounding variables which affect unintentionally the experiment in general, and on the measured values of the dependent variables (tasks time, successful completion rate, and error rate) in particular. So, the variation between cases of targeted test tasks and their length had an influence on how long it takes the participants to finish the tasks. To cope with this issue and to eliminate the impact of it, we considered the moderation of each case to be rather complicated, when we created the dummy data, besides that, to be appropriate to help in achieving the aims of the study. Similarly, Cairns and Cox (2008), in their research found that in testing users with different interfaces for text message entry, the length of the message was obviously a confounding variable;it had an effect on how the people could enter in the different interfaces, and how long it took, regardless of the interfaces.

Due to the difficulties and high cost for accessing the whole population sample, we could not obtain exact accurate data. Consequently, the need for achieving a method to identify how good our estimates are became very important. Thus, a confidence interval would help since it is a way that provides a range of values that could possibly have a specified chance of containing the unknown population parameter (Sauro and Lewis,2012).

Considering this study as a typical usability test that aimed to specify participants' efficiency, effectiveness, or satisfaction with good UX that the targeted system should achieve, we utilized these specifications as measurements for testing the usability through employing different tools and techniques (Juristo,2009). According to the study by Frokjaer et al, (2000), the experimental usability test of a system for complex tasks should involve the three measures of efficiency, effectiveness, and users' satisfaction through UX .

As we mentioned earlier in the section of study measurements the examining of the efficiency measure was based on the row time (task time) which is the most commonly reported usability metric (Sauro and Lewies, 2009). It refers to measures of how long the participants take to finish the targeted tasks using both systems, DPD and C.S,Table (5.16) summarizes the results. A study conducted by

Hornbaek (2005) reviewed the current practice in how usability is measured, and found that more than half of studies measured time as task completion time. However, the rest of the studies were measuring the time in a different way i.e. the time taken for parts of a task, the time consumed in the help function, or the time spent in different parts of the user interface design. In our case study, from measuring the task time, it was obvious that the participants had taken significantly less time using DPD to complete the task 4 and 5, as major tasks for using the VI system. But in other tasks there were no big differences in the calculation of task time, and the participants could finish them rather similarly. This outcome coincides with our expectation regarding the improved design, since we addressed most of the users' requirements that were generated from the previous study, in the development process for designing a suggested prototype. In addition, this study was conducted in a governmental organization that was established for providing such services to the citizens and other residents through establishing the issuing visas system, and they attempt to serve a large number of applicant in a specified time, which they need to be faster. Therefore, we noticed that the majority of participants, as typical users of the VI system, emphasised most of the aspects that related to improving the system efficiency. This comes in accordance with what Nielson (1993,p.34) stated in his previous study "*the user efficiency is often seen as the most important attribute of usability*".

In this study, the effectiveness was measured through enrolling two criteria, the completion rate and the number of errors that were made by the participants whether they were committed directly or recorded by the observation technique. According to Molich et al (1999) the effectiveness measure in the usability test is dependent on several aspects, i.e. the method, the selected tasks, and the person who has a responsibility for conducting the test. So, from using these two measures for testing the factor of effectiveness, and then analysing the obtained results, we noted that the measure of completion rate presented evidently the differences between two systems, because it might be the test was involved with functional tasks. Therefore, finishing the tasks of the systems would be influenced by the effectiveness of each task functionality. This issue is corresponded with Len et al (1997). However, the number error rate as measure of effectiveness on another hand, did not give much illustration about the state of the effectiveness except the number of errors. This goes back to the quality of available recording of the test against the massive amount of provided information.

Initially, this measure as Bevan (1995) described consists of comfort and acceptability of use. Thus, in our study, as we described previously, to identify this view and measure users' subjective satisfaction, a post study questionnaire has been developed that included three sub scales, system quality, information quality, and interface quality. Overall results indicated that the participants were satisfied with the suggested prototype design with its improvements and had a good experience of using it.. In addition, the overall average responses of the majority of participants to using DPD were "Agree", while they were as "Neutral" about the C.S. Robertson et al (2002) also measured average satisfaction in their study. However, there were two statements that were insignificant, and both were related to sub measures of information quality. Moreover, in our study through applying content analysis to the qualitative data, and categorizing the data into three themes by using the three constructs of the questionnaire, we could obtain other supporting results that related to the users' satisfaction (Zimmerman and Muraski,1995). Corbin and Strauss (2008) stated that the content analysis would provide description that *"embodies well-constructed themes/categories, development of context, and explanations of process*".

One of the important results obtained from the correlation between considering the three principal subscales as measuring the user satisfaction and experience, we found that the association of information quality with other sub factors was concordant with the DeLone and McLean (1992) and Dista and MacGregor (1995) studies. They analysed many empirical research studies which were concerned with the measures of IS user satisfaction and experience, they identified the information quality as one of the major factors that affect user satisfaction and would lead to IS success within the organization. As this review presented, then the improving user satisfaction and experience through the quality of the information from the information system should be performed for further research.

However, in regards to the factor of easy of learning, it has not been considered in this study due to the results of the usability evaluation in this first study showing that it did not have a significant influence. Besides that, the participants have already an adequate experience of using the implemented VI system, which makes this factor useless in terms of comparison with the suggested prototype design.

5.7 Chapter Summery

This chapter presented the second study which included the first experiment in the research. It was principally aiming to develop a simulated design of the existing VI system based on the problems and recommendations that were generated from the previous research on usability evaluation of the current VI system, followed by the attempt to validate this developed prototype through conducting an experimental approach for a comparison usability test of two systems with actual users. It has been clearly shown from the useful information collected by analysing both quantitative and qualitative data, that the additional usability features of the improved prototype design were verified by the participants, prospective users of the actual VI system. One of the most obvious improvements that the participants were concerned about was the system efficiency aspect through measuring the task time. Thus, most of the average times of using P.D for each task were slightly different than using the C.S. This might relate to their needs to speed up the servicing of a large number of citizen and other residents in a short time. In addition, another evidence of improvement was related to the factor of Effectiveness through measuring the task completion rate, so in particular, the task 5, as the core of the tasks, we found noticeable differences between the numbers of participants who were not able to complete the task successfully. Additionally, due to this study being concerned about the usability of the VI system as an internal interactive system of the government administration from its employees' viewpoint as users for this system, it was obvious that conducting usability testing in this direction needs to take into account other aspects which might vary in their entirety from usability evaluation for web design from the commercial perspective of different users.

Fundamentally, the study involved the combination of medium and high fidelity for developing the suggested prototype design. At the beginning, the medium fidelity prototype by using the Microsoft PowerPoint helped to address the recommendations easily into an initial design, and apply the modifications later after getting the feedback from the participants. After that, the suggested design was formed into a high fidelity prototype in order to add some functions that enabled the participants to perform the actual tasks. For this purpose, and to create a small set of dummy data, Microsoft Access was utilised. After the suggested prototype has been developed and implemented, and the experimental test has been conducted, the gathered data about the comparison of the two system designs was mostly analysed by using SPSS and a web-calculator. It was obvious that including the statistical method for quantitative data gave a good indicator for employing a comparison usability test.

The access for conducting the usability test in an informal lab in the work place, and working as a team with the sample size of 32 participants who have experience of using the current VI system of between 3 and 5 years, could help to gather reliable results, and give clear insight into usability issues of both designs DPD and C.S. This is what Rubin and Chisnell (2008) stated: *"The testing process for an exploratory test is usually quite informal and almost a collaboration between participants and test moderator, with much interaction between the two"*

From observing the participants' performance during the test, it was apparent that the majority of them have shown an aspiration to contribute to this project through their extensive views and suggestions about the proposed P.D and how it could be improved. In addition, most of them expressed their admiration of the prototype design verbally. For example, one of the participants said, "*I am so impressed to see this improvement; this is precisely what I need ...*", and another participant said " *I'm looking forward to using this system in reality, because it represents my views ...*". On the other hand, we noticed that a few participants mentioned some negative views, like one participant stated " *I believe, I need more time to get used to using this system ...*", and another participant said " *I think, the system interface needs to have more improvement in terms of its design ...*"

Since the government administrations endeavour to enhance their internal software systems to be more capable of integration with other systems, and to become more effective as a reliance on electronic services, the continuation of carrying out more experimental studies with implementing the iteration design approach needs to be considered to obtain more usable interactive systems. Therefore, the usability recommendations and certain features provided by the HCD approach in this research, with impact on the system design, would be an essential ingredient for the iterative design process. This iteration would be necessary due to having identified that other UI aspects and specific functionalities can be taken into account into the system UI design and then to further usability evaluation. For example, the capability of enabling the system to make recommendations to the user through displaying the checklists of the required documents that the applicant has to provide for obtaining the visa. This might help to increase the efficiency of the system by speeding up the decision-making by users. Furthermore, linking the system with the feature of archiving these required documents for finishing the task, would enhance the system quality as a source of information, which in turn would increase the users' satisfaction in using it. Consequently, in the next stage in this direction of conducting an iterative design approach a more usable UI design will be proposed, such as an appropriate guidance to those involved in the improvement of the VI system for developing its ease of use.

The analysis of data in this experiment of usability testing can be described as converting raw quantitative and qualitative data into comprehensive results that can be utilized to produce recommendations for developing the usability of a product. In addition, through the final results , the study could meet the two research questions (RQs 3 and 4).

Furthermore, based on the issues raised and recommendations provided, they will be able to enhance the target system until reaching the accepted level of usability that is required by those concerned, whether by the employees or the governmental administration. In particular, the participants' views that are associated with the rules and regulations regarding the system work procedure, or even changing them, on the effectiveness of making decision to issue visas to the number of applicants.

Besides increasing the number of applicants, the information on work regulation could cause slowness in the process of finishing the task based on the slow pace of decision-making, and it might lead to making different decisions about a certain case.

In consequence, as more of the issues were identified and corrected through the usability testing, a more controlled experiment on the UI design of prototype, and for further research on the usability testing to improve the internal system, it is recommended to conduct more experimental projects constantly through applying an HCD approach in the government administration.

Last but not least, reaping the benefits of this study and presenting it as guidance for applying it more broadly to inherit a more usable software system, may require increasing capabilities, and assigning different roles and responsibilities to team work.

Finally yet importantly, most of the participants showed their full co-operative manners and welcomed participating in this project. This issue is considered a vital factor and fulcrum, which would help this research study to achieve its main goals and then make a success of the project as a whole.

Chapter6: Designing the User Interface of a Government System Using Iterative Prototyping Method.

6.1 Introduction

Based on the outcomes of the second study (the experiment one), which indicated that the improvement of the suggested prototype has been verified through conducting a comparison usability test with the current VI system, it was questioned that the developed design was significantly different and the current VI system was enhanced through the new release design because the participants had performed the main tasks using it with less time, better task completion rate and reduced occurrence of errors. However, the employees (as participants in the study were still providing suggestions and recommendations that needed to be added to the prototype design, in order to enhance their work performance, to cope with the increasing workload and to meet some diverse demands which had newly arisen. Such as, to increase the number of applicants for visas, and for information about the regulations for granting of visas to those applicants (Al-Motaryi, 2012)[12]. Accordingly some participants pointed out that "*besides increasing the number of applicants, the information on work regulation could cause slowness in the process of finishing the task based on the slow pace of decision-making, and it might also lead to making different decisions about a certain case*". In addition, other participants stated that some part of the decision relies on the submission of support documents by applicants, and they asked "*if the documentation can be computerised and fitted within the system design with an accessible functionality, which might help for feeling more control over using the system by mitigating the frequent checking on related documents, and to keep them from being lost*". Furthermore, the participants' reviews were mainly focused on the aspects related to the efficiency and effectiveness as the major factors that had been tested earlier, and reported in the second study.

[12] *A columnist at " Al Eqtisadiah newspaper* 'The Economy' which is a local Saudi daily newspaper, published by Saudi Research and Publishing Company. Accessed online at http://www.aleqt.com/2012/10/07/article_699370.html

Therefore, in light of the identification of general capabilities needs, recommendations, and challenges, in this study an experimental test with iterative design approach was employed as a cyclic process of prototyping and as a proper method to ensure the incorporation of the participants' input into a system design (Stone et al.,2005). Nielson (1993) stated that *"redesigning user interfaces on the basis of user testing can substantially improve usability"*. According to Rubin and Chisnell (2008) applying an iterative design and testing method would lead a project to make steady and quick progress. Furthermore, Hix and Rex Hartson (1993, p.252) stated that "*working from concrete to abstract is the way humans naturally investigate, understand, and assimilate new concepts and solve problems*". This means that as far the prototype is the key for supporting the formative evaluation and iterative design process, it often begins as concrete, and then it moves on to be more abstract with requirements specifications. In addition, since the main task of the target system is issuing visas to the applicants, which was considered to be an individual decision making process by employees in the administration who were participants in the study. Therefore, comprehensive help concepts for working functions were suggested as additional support features for the prototype design. For instance, automated assistant decision-making (AADM) technology has been considered to speed up the decision making process by involving a computerised system. Turban (1995) provided a definition of AADM as "*an interactive, flexible, and adaptable computer-based information system, especially developed for supporting the solution of a non-structured management problem for improved decision making. It utilizes data, provides an easy to use interface, and allows for the decision maker's own insights".* Philmayo (2011) stated that the construction of AADM tends to follow an iterative design or prototyping technique. Thus, the work regulations and policy on issuing visas have to be transferred into code, and then the system with automation function is able to study each entered case of an applicant rapidly and in its entirety. After that it provides to the employees, as participants in our study, an accurate suggestion of the current status of the applicant's requirements via displaying a check list of possible allowed visas that can be issued. In addition it even shows the rationales in case of rejection of the application for non-entitlement to obtain a visa. This would increase the design efficiency through supporting the employees as users of the system to make a decision faster, and provide them with an attractive and effective solution to some of the issues which in turn would be considered as an improvement of the system usability. In addition, another possible dynamic support concerns an archive feature which was considered based on the participants' suggestion regarding their needs for the target VI system. Therefore, it is involved in the design for helping them to preserve the applied documents, and to rapidly refer to these documents as necessary. This would also enable them to have more control over using the system by shortening the time for checking the documents which would lead them to be more productive and satisfied. In consequence, for this stage of the development of the target system and the further stages, it is necessary that new versions and future releases of the system perform to be at least within the predetermined standard or much better (Rubin and Chisnell,2008).

6.2 Aim and objectives

As it mentioned before in chapter one regarding the third study in this thesis, the main aim of conducting the study is to assess and validate the new version of the developed prototype design (DPD) from the employees' point of view as target users of the study. The design which was developed relies on the outcomes picked from the former experiment in the second study, as a stage of the VI system development cycle.

In order to fulfil the aim of the research, the following objectives are performed:

- To redesign the suggested prototype through adding a new set of features (i.e. AADM technology and the Archiving function AF).
- To apply an experimental approach to evaluating the efficiency of the participants' performance in completing the test tasks.
- To determine the representative tasks for conducting the usability test.
- To determine the measurements of the study such as calculating task time, task completion rate, error rate, and subjects' satisfaction, in order to compare the new version of the developed prototype with the previous version in some predetermined usability standard which is generated from the experimental one.

6.3 The iterative design and prototyping for proposed UI

The purpose of this section is to propose the suggested prototype based on the results of the preliminary experiment in study two. So, an extended version was developed with rapid prototyping with the notion of additional features that have been proposed as functional specifications of the user interface design. An iterative approach was utilised during the development process for introducing such a refinement in the design of the VI system that should allow improving its usability, and then meeting its users' requirements. Firstly, we attempted to develop the suggested prototype with a user interface (UI) functionally, via involving the technology of automated assistant decision making AADM, as a key feature that enables the system to have an ability to inspect a set of circumstances when the participants started the test and entered data.

6.3.1 Automated Assistant Decision Making (AADM)

Increasing complexity of environment and systems, besides the natural human capacities becoming gradually insufficient for data sizes, processing skills, and decision speeds, are considered as burdens and challenges which in some cases could be severe usability problems that need to be addressed effectively by the systems' developer (Dahm,2010 cited in Maybury,2012). In addition, the improvement of the user's performance has become important for obtaining the possible advantages which can be provided by other technology advances. Therefore, the automation for developing the system can afford increased efficiency, effectiveness, safety, besides other benefits such as increased performance, and reduce the cost by relying on fewer users for getting the work done . Australian Government (2007) described the term automated system in their Practice Guide as a computer system that automates major components (or all) of an administrative decision-making process. The vital distinctive aspect of an automated system is its capability to examine a set of circumstances (data that has been entered by the user) by applying enacted rules (agency policy or procedures) to decide dynamically what additional information is needed, or what choices or list of information to show to the user, or what conclusion is to be proposed. Shneiderman and Plaisant (2010) pointed out that the degree of dependence on using the function of automation increases because most procedures become more standardized and the pressure for productivity rises. Furthermore, they stated that "*With routine tasks, automation is desirable, since it reduces the potential for errors and the users' workload*" (Shneiderman and Plaisant, 2010, p.91).

In addition, the second part of this feature of AAMD, relates to assisting or supporting the users in making the decision and finishing the tasks effectively. Druzdzel and Flynt (1999) stated that the concept of decision support systems (DSS) is very wide, and it can be defined differently according to the researcher's point of view. Accordingly, other types or names in some cases can be used as synonyms for DSS with various forms and ways which refer to particular domain knowledge (ibid). However, Finaly (1994) could define a DSS comprehensively as "*a computer-based system that aids the process of decision making*". Consequently, this feature would help the system user and support the decision making process by collating and presenting the relevant information for their consideration, besides the recommended state of each particular case.

6.3.1.1 Degree of automation in decision-making system

Automated systems have different degrees of automation based on the intervention of the system in the process of making the decision to finish the tasks. So, these systems might be only partially automated, by automating just parts of the administrative decision-making process, and notifying the users of the points where they need a human decision to be involved, or entirely by providing a final decision regarding a particular case.

For our purpose, in this study, the main regulations and policy which are relevant to the VI system for particular cases which are considered in performing the main test tasks, have been computerised and coded using MS Access software that has been selected for several reasons. Such as, its flexibility for creating a system design that works with a database, which would meet the research's need for building dummy data for the purposes of the experiment. Across that, the design of this new version of the prototype was initially based on the basic layouts of the original one, and the new feature has been included in the design through an active button. In regards to adding another additional feature of Archiving, as a function that has been requested by the participants to preserve the applied relevant documents from loss, the prototype designed was used to include this new function in the VI system, and we intended to test its importance for adopting it within the system. Thus, we simply added another option to the main menu for processing the archiving feature on the prototype UI design, which displays data on a form. For facilitating the participants in navigating this function with the proposed artefact, we created a dummy data of particular applicant documents, such as a national ID card, a proof of income letter, an authorization letter, etc. Therefore, the participant while performing the tasks related to the Archiving function would be able to search for any previous documents which are linked to specific cases through selecting two main searching options; issued visa number or type of document.

6.3.2 Pilot study of the prototype design

Prior to the developed prototype being used for the actual experiment, a pilot test was employed using a qualitative approach that aimed to review and gather the participants' feedback regarding the new version of the prototype. Therefore, six male employees were selected from the main sample that has been chosen for conducting the final test. They were invited to participate in this sub study via emails and phone calls, knowing that their contact details have been collected previously in the second study. Additionally, all of them have a good knowledge generally of what they need to do regarding the design by providing their feedback about it. A created folder of the release version with the sheet of main tips for using it has been shared via their accounts in the Dropbox application. In addition, the instructions for using this system prototype were explained to them along with providing all the relevant information, such as the user ID and password for accessing the system, and different dummy numbers such as citizen's ID and the previous visa number for enabling them to perform the functions entirely.

6.3.2.1 Findings of pilot study

At first, three of the participants have reported that they faced a problem of not being able to use the prototype design. One of them could not use it at all, and two participants only after they had successfully logged in. After we investigated the issues, we found that the first participant had tried to use a different version of Access 2007, so he could not open the system from the beginning and he received an error message (case 1) as shown in (Figure 6.1).

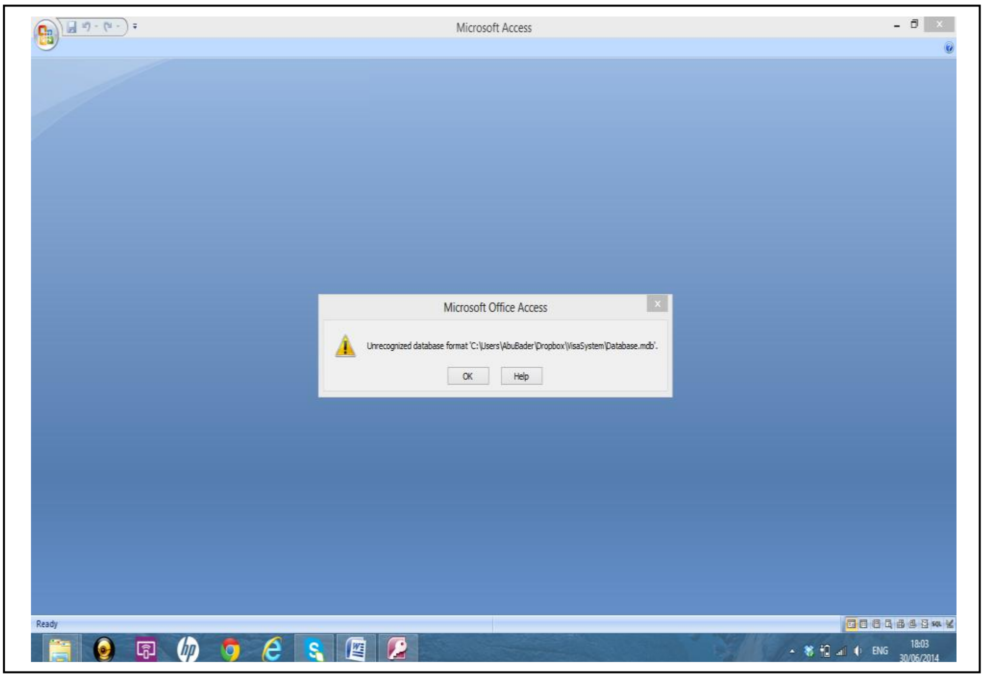


Figure 6.1 A screenshot of the error message for case 1

Similarly, another issue was encountered with another two participants, as they could open the system and log in, but they could not go into any further stages of using the system. The problem in this case 2 (Figure 6.2), was using an incompatible 32-bit version of Office 2010, not 64-bit for which the design was developed.

Therefore, all of them were informed that they needed to utilise Microsoft Access 2010 or the latest 64-bit version to be able to use the suggested system design fully. In addition,another issue related to the error messages was revealed after one participant committed a mistake while performing Task1, when he tried to enter information that does not exist, so, an error message appeared in the English language rather than Arabic. After we investigated the problem, we found that the system should set display language as default input for the target language (i.e. Arabic), in order to receive the messages in the same language format (case 3 in (Figure 6.3).

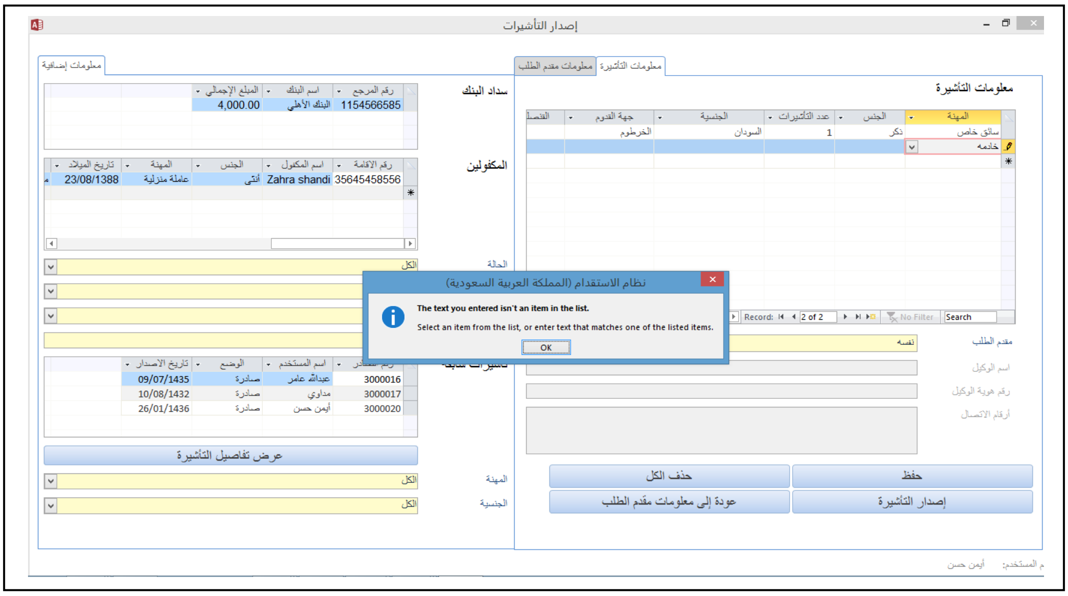


Figure 6.2A screenshot of the error message for case 3

After these issues were solved, feedback and some comments were obtained from the selected participants on the suggested prototype design. In regards to the feature of the AADM system, all participants admired the way that the concept of involving this attribute within the system would speed up the decision which needs to be taken to complete the main task of issuing a visa to the applicant. It would also eliminate several issues like providing different decisions upon a specific case. However, one participant argued that it would be difficult to rely on the computer as a machine to make a final decision for different cases with several regulations, and also he mentioned that "*This might lead me to get confused in making the decision whether to follow my own views or consider the computer presentation*". It was explained to him clearly that in such system feature based Information Communication Technology (ICT), the automated system would just guide the employees (users) and assist them throughout the decision making process by studying and presenting the relevant information for their consideration. At the end the employee will be in charge for taking the final decision.

Three participants pointed out that the interface of this feature would be better if the order of displaying the presented relevant information were re-arranged as follows: main details of the applicants, status of the case, and the unrealized conditions. One participant said, *"Reshaping the presented information in this order form would be better for enabling me to read and pick out the relevant information pretty quickly"*. Also, a missing option had been figured out in the main design of the immigration section, it was about the transfer of sponsorship from the applicant. In addition, it has been suggested that the screen which opens to include automatically the presented information that assists decision making, would be better as a sub-screen rather than a whole main screen for using the system. They claimed that managing the information in a sub-screen as assistant message would be functional for the employee (user) to link the information with the one in the main screen as fast as possible to have a quick decision. Furthermore, two participants proposed that a small box needed to be added at the end of displaying the relevant information, so it can be ticked when the decision has been made based on the system assistance, or remain blank if it does not rely on that. They stated that, "*this is useful as a record when it is needed to return to the issued visa and check-up the source of decision that has been taken toward the visa*".

Although most of the participants indicated the importance of this function being implemented within the VI system, a few comments were gathered from the pilot respondents regarding AF. For instance, the major comment was about the need for adding an option to search the previous documents by the type of document, besides the issued visa number; hence, it would become easy to reach the required document.

Therefore, based on the previous illustration of participants' feedback, some modification and editing was adopted to the prototype simulation design with operational appliance to be expanded with a higher level of functionality. The following figures display screenshots of the final prototype design.

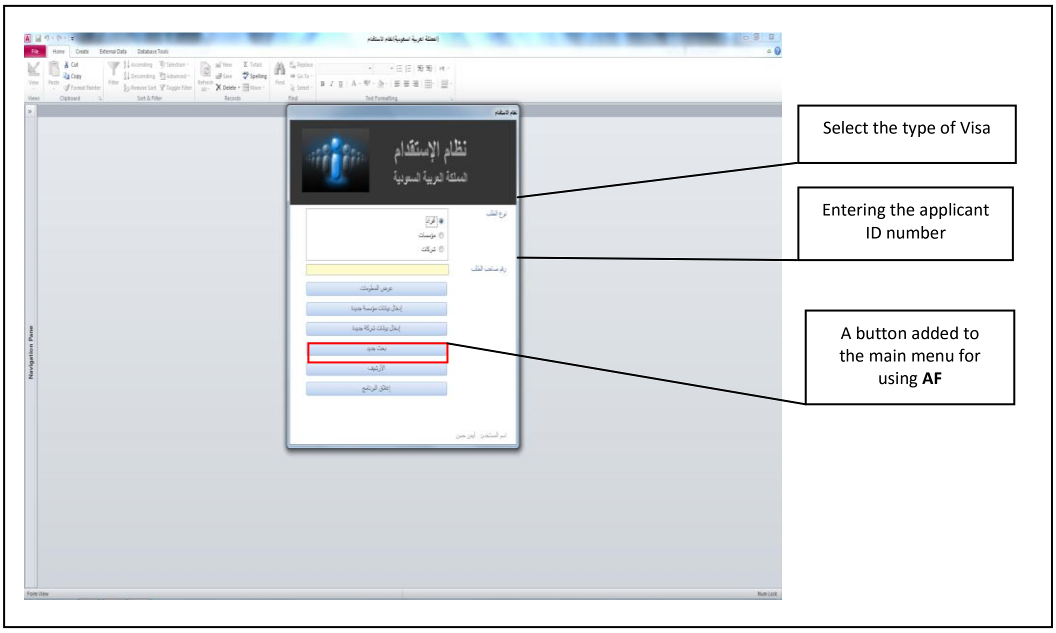


Figure 6.3 A screenshot of user interface of main menu showing additional function of the developed design

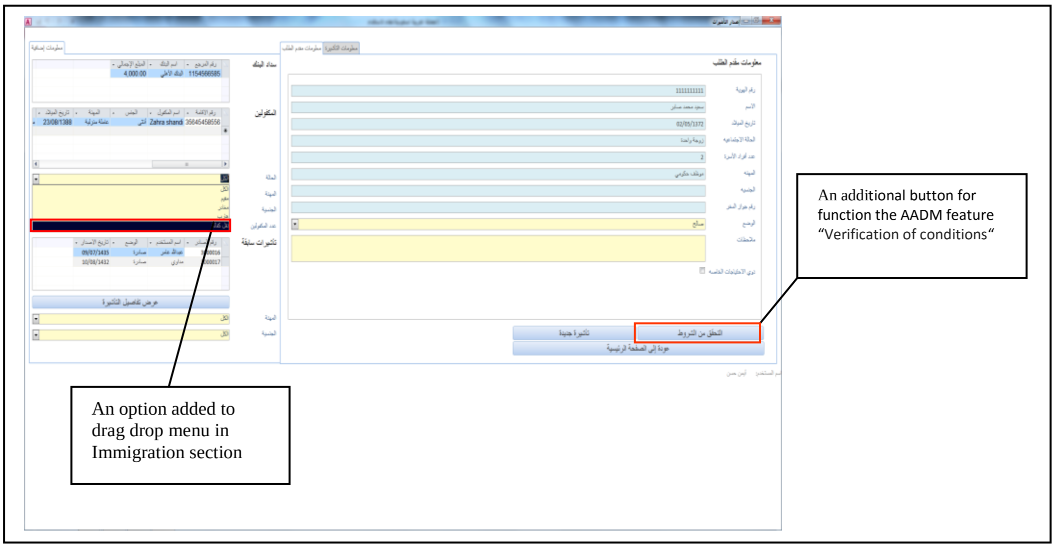
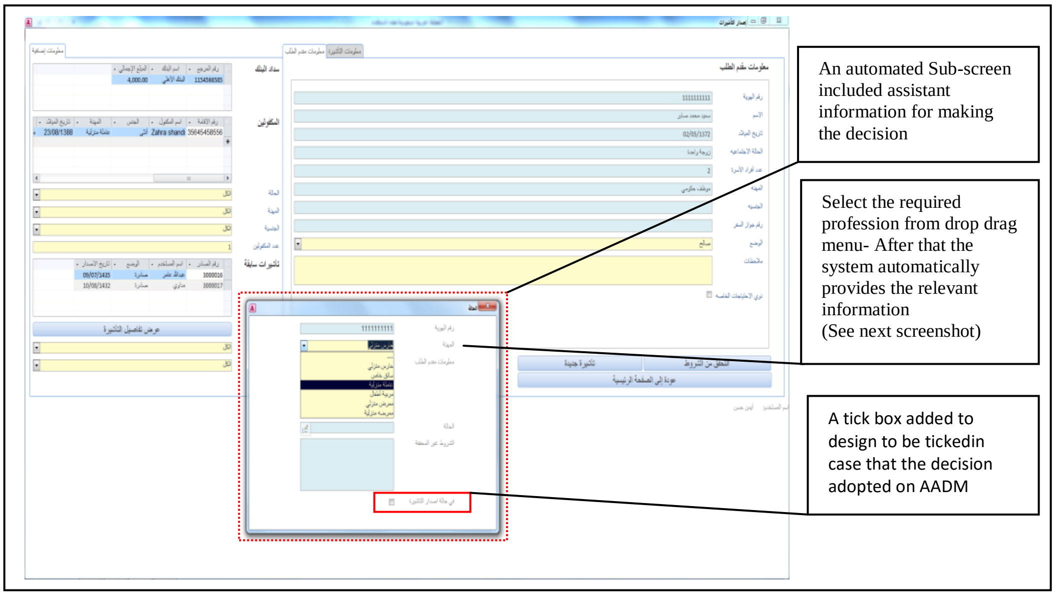


Figure 6.4A screenshot of the user interface showing an additional feature in the processing page, with modified option



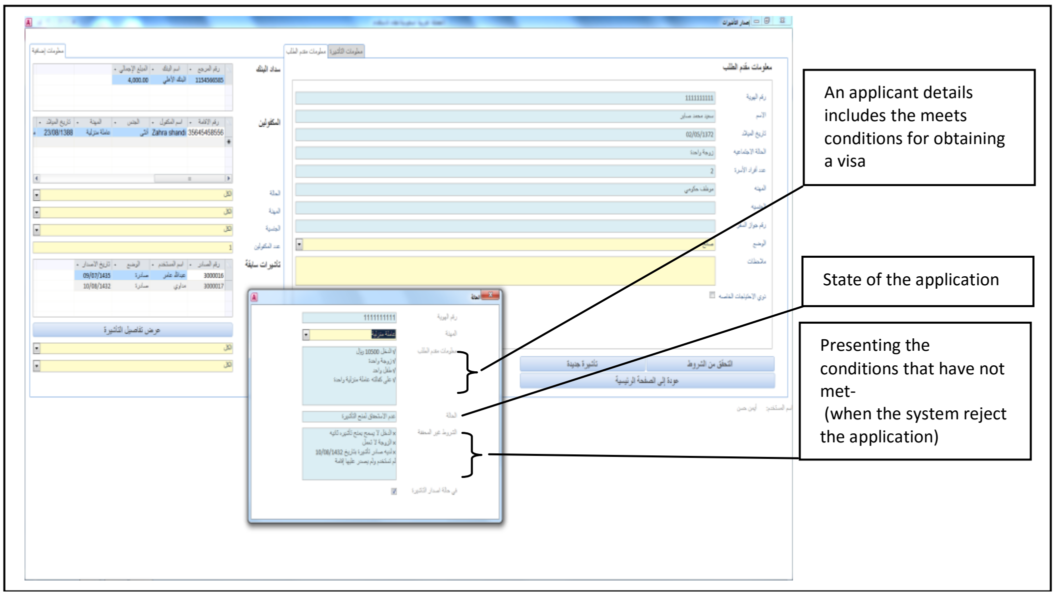


Figure 6.5 Screenshots of the user interface illustrating functionality of AADM feature with modification

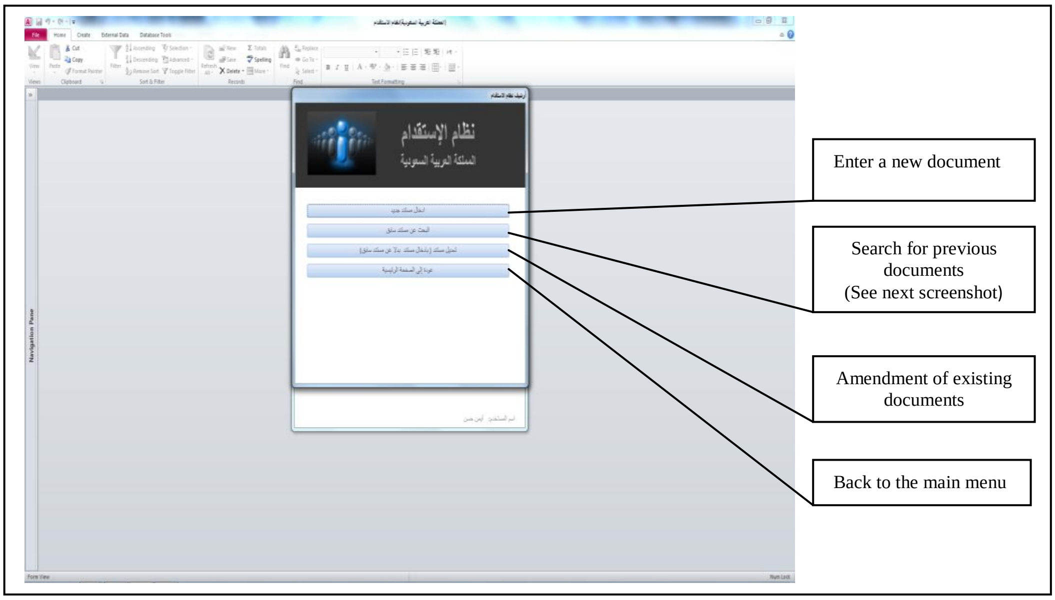
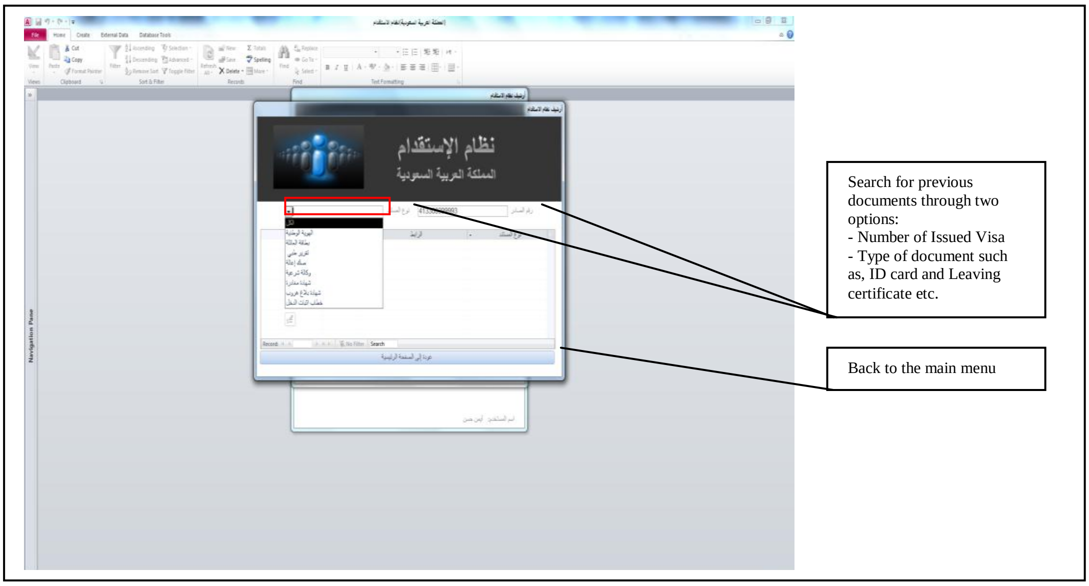
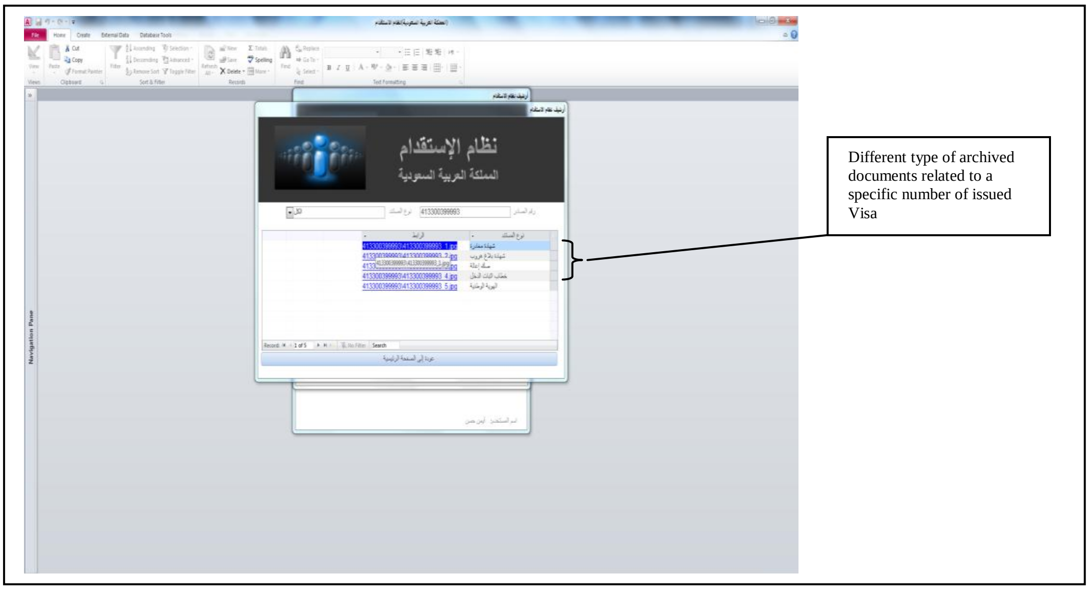
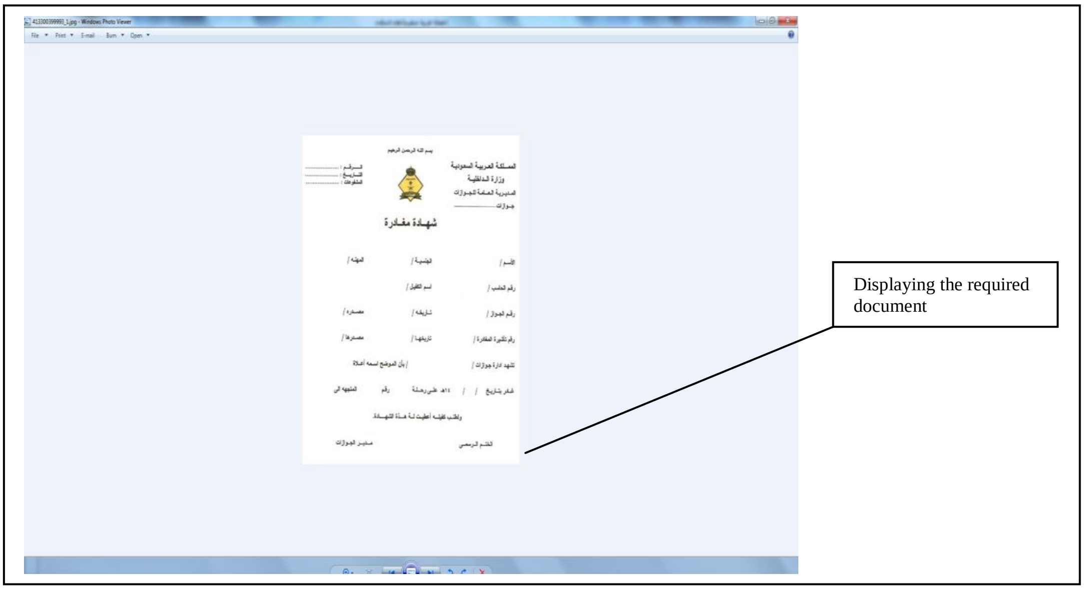


Figure 6.6A screenshot of the user interface showing functionality of Archiving as a new function in the developed design

Figure 6.7 Screenshots of the user interface illustrating functionality of AF, with the modification of search options







6.4 Pilot study for the test setup and procedure

Following the pilot study of a high fidelity prototype, a pilot usability test was carried out to ensure that the test setting and plan were accurate and on the right track, and then to refine the test process especially the test tasks. Reynolds (2014) stated that *"Run a pilot test or two beforehand to help you shape tasks to ensure you get the information you are looking for in the follow-up test"*.

Initially, due to the major goal of this study being to assess and validate the suggested prototype design which was re-developed based on employees' views in experiment one, it has been considered to follow the same setting for usability testing which was implemented in the second study. So, the test experiment was determined to be held in the office of the assistant manager at the work place of the government organization, as a simple single room setup is the most common standard for testing (Ruben and Chisnell,2008). The room or informal test laboratory was provided with testing equipment, such as a PC with the suggested system installed (prototype design Version A and B), also other tools like a Dell laptop with camera for recording the test, paper and pen for recording the comments and notes, besides providing all the required documents (Figure 6.8).

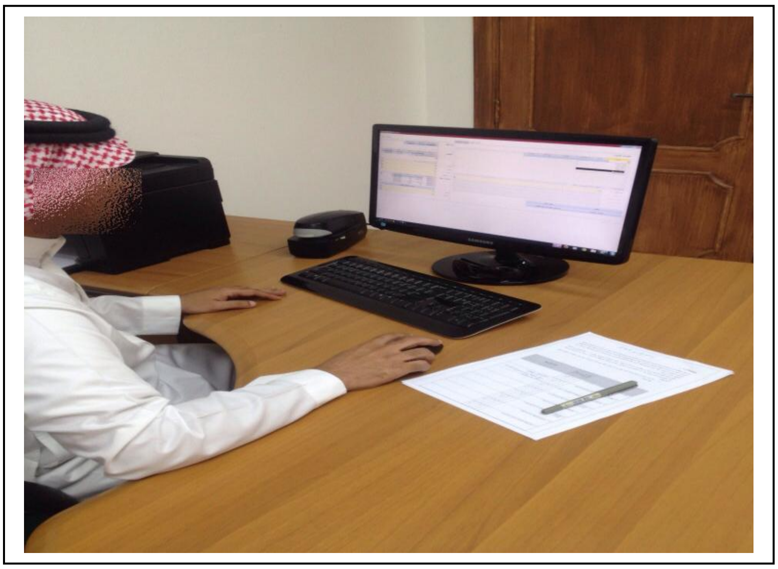


Figure 6.8A participant performing a task in an informal usability test lab

Firstly, four participants (employees from the same department in the government organization) were recruited as pilot users to participate in this sub-test and to perform individually the selected tasks using the UI of the developed prototype of both versions. All of them will be participants in the final test due to the limited number of available resources to implement the test, which would affect the prescribed sample for study if we exclude them from participation in it. At the beginning, each participant was informed of the purpose of the pilot test by an information sheet, and they were given a short demo of the new functions and features which have been developed in the suggested system design. Afterward, the instructionsfor using this system and the selected test tasks list were handed out to the participants to read and they were told to start performing them while verbalising their interaction with the system, and provide any comments during the test. The time was calculated during which the participant was performing the test tasks, meanwhile, he was observed and notes were taken and written down as we were located next to him as the test moderator. After each participant finished the test session, he was asked to fill out the post study questionnaire which has been built and used in the first experiment.

6.4.1 Pilot usability test findings

Since we were following the same test setup as in the first experiment, a few issues were noticed and reported after the three participants were encouraged to complete all the targeted tasks and finish the session. For instance, one of the issues encountered was about the test tasks; after we gave the list of tasks sheet to the participants, two of them took time to read through it, besides all of them still asked for more explanation. This issue was resolved by deciding that the list of selected tasks should be read to the participants in the final test, each task ata time (Hix and Harston, 1993), rather than handing them out only, so it would help to consolidate the test procedure and tasks in their mind by the double actions of giving and reading them. However, the test tasks were clear as the participants stated, as they didnot comment upon these test tasks themselves. Another issue was regarding the confusion in calculating the actual time taken for the test tasks. Therefore, a signal of starting and ending the tasks should be posted by the participant in order to start and end the counting of the time. Furthermore, since the pilot participants were asked to verbalise their interaction while performing the test tasks (think aloud), this was distracting most of the participants which caused some delays when we counted the task time. So, in the actual test, they will not be encouraged to think aloud while the task time is being calculated. The timer of the

Samsung Galaxy S4 was usedas it had fulfilled the purpose in the first experiment. Furthermore, in regards to the test length, according to the calculation the whole test time would take between one hour and one and a half hours. In addition, other minor issues occurred and have been taken into account, such as asking the participants politely to keep their mobiles switched off during the test, besides sorting out other sources of disturbance from the applicants'enquiries, which was done in the previous test.

6.5 The study techniques, internments, and procedures for the Data collection

As it presented earlier in chapter three, this study combined method of qualitative and quantitative through employing different techniques and instruments for collecting data to achieve best results. Rogers et al., (2011) stated that for collecting data in usability testing a mixture of quantitative and qualitative methods is often used.

6.5.1 Test Tasks for Experiment (2)

For examining each feature that has been considered on the developed prototype design, there should be at least one task that practices it (Kuniavsky, 2003). Therefore, the representative tasks for the current experiment included five tasks; three fixed set tasks (1, 2 and 3) were considered from experiment one for conducting the comparison test with the previous version of the prototype design in order to identify the improvement on the new release of the prototype design. Initially, the tasks (1 and 3) were designed as the key tasks for determining the system usefulness through different measurements such as task times and completion rate. This in turn would help to determine the improvement through involving the new features in the prototype design, like the feature of AADM. Task 2 was chosen as an indicator of whether the participants could use some functions and finish the task with better control overusing the system after it had been developed. For example, the amendment ability of modifying the issued details was improved by adding the missing information andan option to drag-drop a menu. Tasks 4 and 5 were designed for the purpose of investigating the effect of the additional feature of the Archiving Function through gathering the participants' views on such a new concept in the system design. The following (Table 6.1) shows the targeted tasks for this experiment.

Table 6.1 Test Tasks

Task No.	Task description
1	Data entry for issuing a new visa with different requirements.
2	Amendment of the details of Issued Visa.
3	Data entry to create a new file for a corporation by entering its data, which includes: name, commercial registration number (CR no.), and the nature of corporation's activities.
4	Navigating the Archive function from the main menu.
5	Search for the applied documentation of previous visas.

6.5.2 The test plan and procedure

The design of the main experiment in this study followed a similar usability testing approach setting to the preliminary experiment in the second study. Some changes were decided based on the findings of the pilot usability test that was discussed previously. Accordingly, the usability test was conducted in the same quiet room as an informal laboratory (the office ofthe duty manager at the work place in the governmental administration). It has been coordinated with him as a gatekeeper for implementing this experimental test, which in turn he arranged with the sample for the test and passed on the invitations for the usability test. Additionally, all the participants have prior awareness officially that the researcher was conducting this project, and of the confidential treatment of their provided information.

After the new version of the developed prototype design was ready for conducting the final experiment, and before running the test sessions, all participants were welcomed and thanked for their acceptance of being involved in this usability test. Then they have been informed by a brief introduction about the purpose of conducting this further stage of research. Afterwards, all the participants have been given an ID for logging into the suggested prototype system, and the instructionsfor performing the test (Appendix F) were provided to them by reading it directly. Initially, the scenarios of tasks was read to each participant, which enabled us to interact with all participants and to make sure that they gained a good understanding of what they should do, and then to control the progress of the usability test as whole. Besides that a copy of the list of tasks was handed out as a reference in time of need (Rubin and Chisnell,2008). In addition, the signs of starting and ending the test tasks were explained to each participant; they were told to say "Task number". (T1 to T5), I'm ready," to start counting the time, and "Ok, I finished it" when he stopped work on the task, in order to facilitate recording of the time taken on the tasks. Additionally, each participant was notified that there were additional features that had been developed into a new version of prototype design. Thus, a short demo of the design functions was delivered, and in addition a brief training session for using this new release of the prototype design was carried out, with provision of additional explanation when needed.

Initially, within-subject test design with counterbalancing technique was employed in this experiment for enabling comparison of the two versions on specific criteria, and in order to avoid the bias towards one version rather than another and to limit the effects of learning (Rubin and Chisnell, 2008). Thus, the sample of 26 participants was divided into two groups (A and B), each participant in group A started using the previous version (V1) of the prototype to perform all five test tasks and then version2 , while the participants in group B used the new release of the prototype at the beginning. In addition, the whole test session for the individual participant was divided into two parts, in each part he tested one version. Following this, each participant was invited individually to take part in accomplishing all five test tasks using the new version of the developed prototype, and three tasks with the previous version. Bastien (2010) stated that "*most of the usability sessions are run with a single test participant*". Principally, the experiment as a whole was conducted over five days; each day has one session with 5 participants, except the last day will be with 6 participants. So, two days and a half were scheduled for conducting the usability test with each participant for each group.

Table 6.2 Within-Subjects method and counterbalancing technique for the test

Day	Session							
	Prototype Design (Version A)			Prototype Design (Version B)				
	Group A							
1	T1	T2	T3	T1	T2	T3	T4	T5
2	T1	T2	T3	T1	T2	T3	T4	T5
	Prototype Design (Version B)					Prototype Design (Version A)		
	Group B							
3	T1	T2	T3	T4	T5	T1	T2	T3
4	T1	T2	T3	T4	T5	T1	T2	T3
5	Prototype Design (Version A)			Prototype Design (Version B)				
	Group A (3 participants)							
	T1	T2	T3	T1	T2	T3	T4	T5
	Prototype Design (Version B)					Prototype Design (Version A)		
	Group B (3 participants)							
	T1	T2	T3	T4	T5	T1	T2	T3

6.5.2.1 Think aloud protocol, Observations and note taking

In addition, each participant was encouraged to perform the five tasks as quickly and accurately as possible (Cairns and Cox, 2008). Also, he was asked to interact with the system verbally, or in other words to think out loud while he was communicating with the design and to express his views while performing the targeted tasks using the suggested prototype design being tested (Hegarty and Wusteman, 2011; Nielson,2012). This procedure of think aloud was considered for helping to evaluate the functionality, strengths and weaknesses of the system, besides its usability (George, 2005). So, as think aloud protocol was applied in this experiment to collect useful participant's feedback as qualitative data, each participant was informed that he has been observed directly while running the experiment (Hertzum and Jacobsen, 2003), and his reactions and comments are being recorded and noted down for later analysis, such as when he mentioned his feelings, faced challenges, or got confused during or after performing the tasks .

6.5.2.2 Post study questionnaire

Once the participants had finished performing all test tasks using both systems, they were asked to fill out the post-study questionnaire that had been used previously in the first experiment, as a tool to gather subjective data regarding the participants' opinions and attitudes towards both versions, the old one (V1) and the iterated version (V2), and then for measuring their satisfaction aspect and if their experience got improved.

In regards to the reliability and validity, these have been calculated already by using Cronbach’s alpha coefficient, which was "0.861"for the overall questionnaire and for each construct as follows: System Quality 11 statements, "0.818", Information Quality 9 statements, "0.79", and Interface Quality 6 statements, "0.857".

As mentioned before the questionnaire has a part for collecting qualitative data by adding an empty box after each statement within the questionnaire. Furthermore, an open ended question was added as another section within the questionnaire, aiming to collect a participant's preference towards the prototype design he has just tested, either what he likes or he does not like (Appendix G).

6.5.3 The measurements of study

In this experiment, the four most common quantitative measurements have been selected as types of dependent variables as follows (Lazar et al.,2010):

- Task time (how long each task took to be finished using both versions (A and B), and this includes all tasks regardless of whether the participants had successfully completed, partly completed, or failed to complete them).
- Task successful completion rate (how many participants could correctly finish each task).
- Error rate (numbers of errors occurring during performing the test).
- Subjective measure (user satisfaction).

Therefore, the task times, successful completion rate, and the error rate, were calculated, and after that the participants were asked to complete a post study questionnaire to measure their satisfaction with the system usability.

6.6 Results and data analysis

In this section the results of the quantitative method, that included the main four types of data, task time, completion success rates, error rates, and satisfaction ratings via post study questionnaire, are analysed. In addition the qualitative data collected via the tools of observation (pathways participants took), note taking (comments, recommendations and problems experienced) and the qualitative part within the questionnaire, will be presented for the current usability test (answers to open-ended questions, recommendations and problems experienced).

6.6.1 Quantitative results

6.6.1.1 Task time

As we mentioned earlier, in this measurement the time was calculated during which an individual participant was performing the test tasks, whether he could finish them successfully, partly finished them or failed to complete them.

The following Table (6.3) displays the data collected from 26 participants via utilizing this measurement of calculating the time. It includes the averages of task time in seconds for all test tasks which have been carried out by the participants in both groups A and B after they swapped between the two versions (1 and 2) of theDPD for starting the test session. The purpose of this description statistics is to identify whether there is a difference or not in the speed of the participants' performance when they started using one version rather than the other.

Table 6.3 The averages of task time across test tasks for each group

13 Participants Group A	Sessions started by using Prototype Design (PD) Version (1)					
	T1		T2		T3	
	DPD/V1	DPD/V2	DPD/V1	DPD/V2	DPD/V1	DPD/V2
Average time in Sec.	247.85	213.92	70.38	61.46	178.15	159.15
13 Participants Group B	Sessions started by using Prototype Design (PD) Version (B)					
	DPD/V2	DPD/V1	DPD/V2	DPD/V1	DPD/V2	DPD/V1
Average time in Sec.	224.23	236.08	62.08	65.15	167.31	170.92

The Table (6.3) above presents the calculation of average task time results that all participants in each group (A and B) needed to complete the entire test tasks after they swapped between the two versions (1 and 2). These recorded results comparing the averages of task time for each task using the same version, show that the participants' performance has improved by consuming less time for completing the test tasks.

From the outcomes, it was obvious that the participants' performance was boosted after they finished the tasks using the first version (1 or 2) before the other. For instance, comparing the average time for finishing Task 1, the participants in group (B), who started the test using the V2, could perform and finish the task using V1 with an average time of 236 seconds, better than group (A) who started the test using V1, who spent an average time of about 247 seconds for completing Task1 using it. Similarly, in the same Task1, the participants in group (A) spent less time to finish this task using V2 than group (B) with a time-lag of 11 seconds.

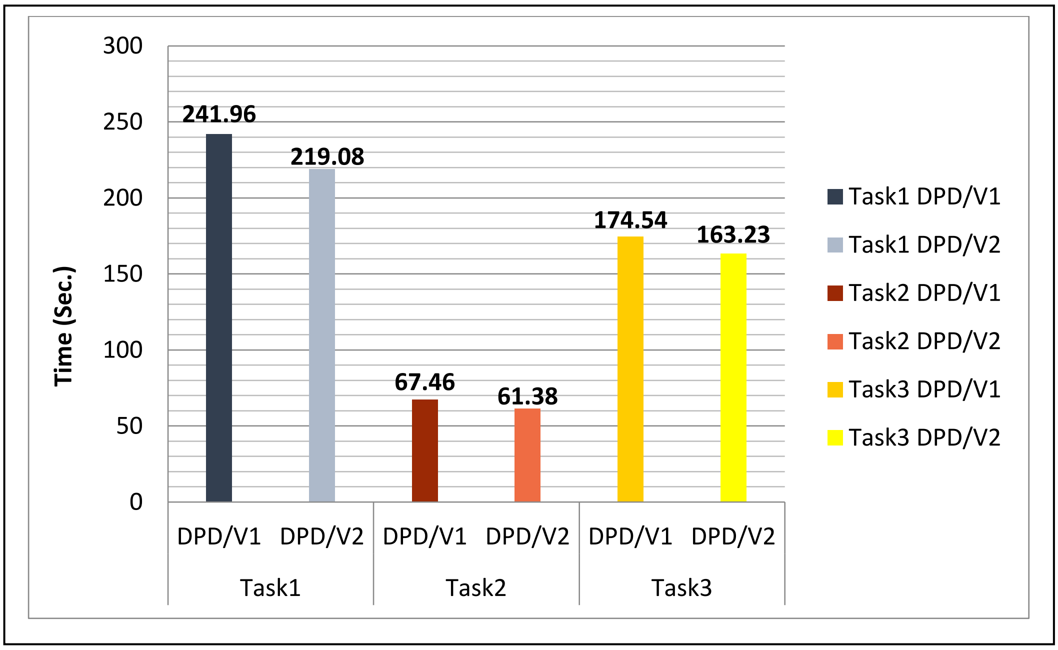


Figure 6.9 Task Times across Tasks for all samples

Figure (6.9) shows the total averages of task time in seconds spent on the performance of test tasks using both versions (1 and 2). The results show that there was a marked improvement in the participants'performance as they could finish Task1, which was considered a key task inthe study, associated with taking the decision on issuing a visa to the applicant, faster with 23seconds difference using V2/PD rather than V1/PD. The improvement is thus 9.5%. On the other hand, in regards to the Tasks 2 and 3, there were slight differences in the task time on finishing the two tasks, using both versions ofDPD. So, the participants improved a little when they performed the tasks using V2/DPD, for example, they could finish Task 2, when they used V1, in an average time of 67 seconds, and when they used V2 they spent an evenly matched time with an average time of 61 seconds. Likewise, the participants completed Task3 using V1 in 1minute and 75seconds, while they took less time when they performed the same Task using V2 with 1minute and 63seconds. This was a relative improvement on the speed of performance.

For comparing the results obtained to determine whether there was a significant difference in the time on task when the participants performed the test task while using both versions of DPD reciprocally, the paired-samples *t*-test analysis was conducted for this measurement (Newsom, 2008). Table (6.4) below summarises the descriptive statistics data of paired-sample *t*-test of each task. The results showed that the means (Average time) of using DPD/V1 to perform all the test tasks were higher than when they used DPD/V2 to complete these selected tasks. This indicated that the participants could finish the task using the new release version ofDPD efficiently, with less time. A question research related to this measurement is needed to determine if these outcomes mean that we can expect that there would be differences in mean DPD/VB for participants as users and if it would help their performance to speed up finishing the tasks.

Table 6.4 Descriptive Statistics for task time

		Mean	N	Std. Deviation	Std. Error Mean
Pair1(Task1)	PD/V1	241.962	26	12.101	2.373
	PD/V2	219.077	26	20.576	4.035
Pair2 (Task2)	PD/V1	67.731	26	7.302	1.432
	PD/V2	61.769	26	5.922	1.161
Pair3 (Task3)	PD/V1	174.539	26	7.415	1.454
	PD/V2	163.231	26	7.825	1.534

The following Table (6.5) shows the results of the paired *t*-test, which comprises the key outputs, such as, the test statistic, the degree of freedom and the P-value.

Table 6.5 Paired Samples Test for task time

		Paired Differences					T	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	T1(DPD/V1 -DPD/V2)	22.88	22.94	4.499	13.617	32.152	5.086	25	.000***
Pair 2	T2(DPD/V1-DPD/V2)	5.96	6.34	1.243	3.401	8.522	4.794	25	.000***
Pair 3	T3(DPD/V1-DPD/V2)	11.30	9.91	1.944	7.303	15.312	5.815	25	.000***

*** *P*<0.001 at 95%

As stated previously, a paired *t*- test was conducted to compare the time in seconds that a participant needed to finish the test tasks using DPD/V1 and DPD/V2. From the results obtained, it was obvious that there was a highly significant difference on performing task1 using DPD/V2, and the participants could improve their speed of performance in taking a decision, through consuming less time (t = 5.086, df=26, P<0.001).

In order to obtain the magnitude of difference (MacFarland, 1998), between the two prototype design versions and to examine if participants had a more positive performance towards the new version of the prototype, Wilcoxon matched-pairs signed- ranks test, was used for ranking the data to provide two rank totals, one for each version. Furthermore, it is a non-parametric version of a paired samples t-test (Hole, 2011). The data in Table (6.6) shows the calculated sum of the ranks of the positive and negative differences of each pair of the test tasks.

Table 6.6 Ranks

		N	Mean Rank	Sum of Ranks
TASK1DPD/V2 -DPD/V1	Negative Ranks	25[a]	13.00	325.00
	Positive Ranks	1[b]	26.00	26.00
	Ties	0[c]		
	Total	26		
TASK2DPD/V2 -DPD/V1	Negative Ranks	21[d]	14.52	305.00
	Positive Ranks	5[e]	9.20	46.00
	Ties	0[f]		
	Total	26		
TASK3DPD/V2 –DPD/V1	Negative Ranks	22[g]	15.02	330.50
	Positive Ranks	4[h]	5.13	20.50
	Ties	0[i]		
	Total	26		

a. TASK1PDV2 < TASK1PDV1
b. TASK1PDV2 > TASK1PDV1
c. TASK1PDV2 = TASK1PDV1
d. TASK2PDV2 < TASK2PDV1
e. TASK2PDV2 > TASK2PDV1
f. TASK2PDV2 = TASK2PDV1
g. TASK3PDV2 < TASK3PDV1
h. TASK3PDV2 > TASK3PDV1
i. TASK3PDV2 = TASK3PDV1

Table 6.7 Test Statistics[a]

	TASK1PDV2 - TASK1PDV1	TASK2PDV2 - TASK2PDV1	TASK3PDV2 - TASK3PDV1
Z	-3.798[b]	-3.295[b]	-3.941[b]
Asymp. Sig. (2-tailed)	.000***	.001***	.000***

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

b. Setting up the study questions

According to Sheskin (2003) the Wilcoxon matched-pairs signed-ranks test is conducted in a usability testing situation using a design with two independent samples. Thus, in this study, for investigating the difference between the two versions in regards to Task 1 and 3, measuring the system usefulness through utilising the measure of task on time.

The results of the analysis data indicated that there was a significant difference in each task.

Task 1: z = -3.798, T (N = 26) = .000 and Task3: z = -3.295, T (N = 26) = .000, p<.05

Task 2: z = -3.798, T (N = 26) = .000, p < .05

This could be interpreted as the participants' performance having improved in terms of completing the task using DPD/V2 faster than using DPD/V1. In addition, the results of analysing Task 3 indicated that the participants could finish the task using DPD/V2 with better control than using DPD/V1.

6.6.1.2 Task completion rate

The measuring of task successful completion rate is considered one of the crucial criteria for evaluating the usability of a product (Nielsen, 2001; Sauro and Lewis,2012), and comparing it (Rummel,2014). Therefore, it has been applied to this experiment in order to calculate whether the test tasks have been completed or not. The utilising of this measure for gathering data relied on a binomial distribution, "a binary coding", so that a successful task completion was coded with the number (1), whilst a task failure or uncompleted task was coded as (0) (Sauro and Lewis,2012). Accordingly, after the participants had finished the selected test tasks, the data was recorded and coded to be reported as a percentage. The following (Figure 6.11) shows the total percentage of successfully completed test tasks after the participants had used both versions of the prototype. The result indicates that in Task1 as a key task for the experiment, 73% of the participants could finish it successfully using the DPD/V1, while after they used DPD/V2, their performance improved by 2.25%.The percentage was 88% of the participants who could complete the task successfully. In Task2, the percentage of successful task completion was the same using both versions (1 and 2) with 85%. Evidently, in overall terms the successful task completion rates were generally positive and quite high across all the test tasks except Task2, and this would be related a number of errors occurring in this task. The percentage of uncompleted tasks included the participants who committed errors or faced some problems, and did not know what was going on, and then preferred to abandon the task and start to perform another one.

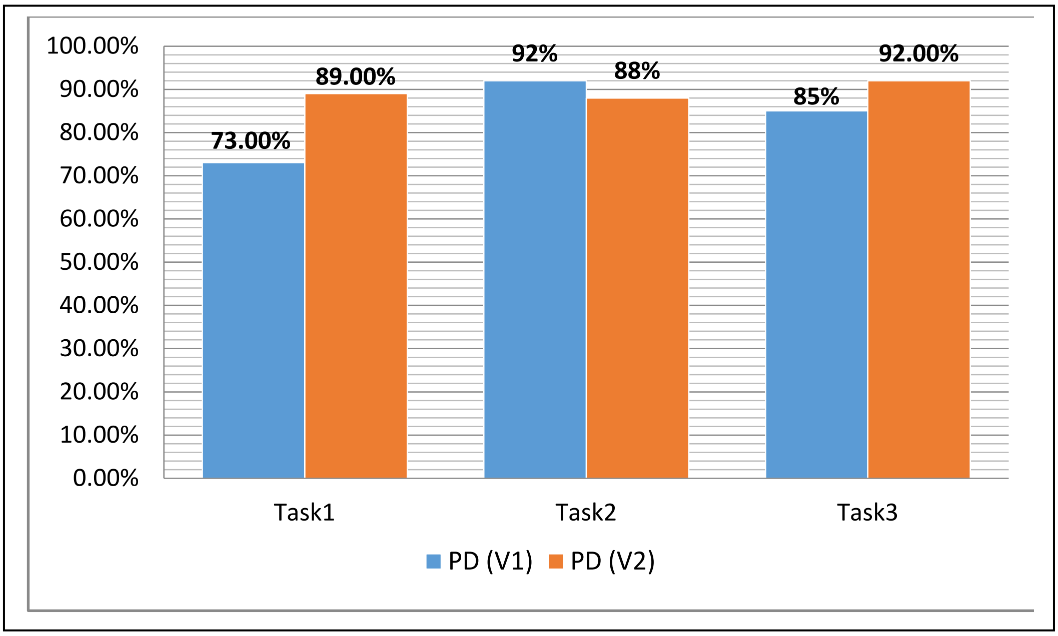


Figure 6.10 Task Completion rate across tasks

In addition, since within-subject technique was adopted for the experiment, and each participant had to use both versions of the prototype design, so the McNemar Exact Test was used for analysing the gathered data on completion rates (Sauro and Lewis,2012). In addition, it was decided to determine if there was sufficient evidence for a significant difference between the two DPD versions in the participants' performance in completing the test tasks through generating *p*-values by examining if the proportion of discordant pairs is greater than 0.5 for all participants. Essentially, it uses a 2X2 table that presents the number of concordant and discordant responses (McNemar, 1969, cited in Sauro and Lewis, 2012). The following Tables (6.8, 6.9and 6.10) display the total number of concordant and discordant responses in all test tasks.

Table 6.8 Concordant and Discordant Responses for Task 2

	DPD/V2 Pass	DPD/V2 Fail	Total
DPD/V1 Pass	21 (a)	3 (b)	23 (m)
DPD/V1 Fail	2 (c)	0 (d)	3 (n)
Total	23 (r)	3 (s)	26 (N)

Table 6.9 Concordant and Discordant Responses for Task 1

	DPD/V2 Pass	DPD/V2 Fail	Total
DPD/V1 Pass	15 (a)	3 (b)	18 (m)
DPD/V1 Fail	6 (c)	2 (d)	8 (n)
Total	21 (r)	5 (s)	26 (N)

Table 6.10 Concordant and Discordant Responses for Task 3

	DPD/VB Pass	DPD/VB Fail	Total
DPD/VA Pass	19(a)	2 (b)	21 (m)
PD/VB Fail	4 (c)	1 (d)	5 (n)
Total	23 (r)	3 (s)	26 (N)

From Table (6.9), we found that under the concordant pairs, there were 15 participants who could finish the task using both designs (cell a), and only two participants who failed to complete the task successfully using both DPD/V1 and DPD/VB (cell d). While in the discordant pairs, there were three pairs (33.33%) who failed on finishing the task using DPD/VB but finished it successfully using DPD/V1 (cell b), and there were 6 pairs (66.67%) who failed on DPD/V1 and passed on DPD/V2 (cell c). Similarly, in Task2, which was presented in Table (6.8), as concordant pairs, there were 20 participants who could finish the task successfully (passed) using both designs (cell a). Furthermore, there was not any participant who failed in completing Task 2 using both versions of the prototype design (cell d). On the other hand, within the discordant pairs, there were three participants who failed to perform Task 2 using both versions of the design (1 and 2) (cell b and c). In Task3, within the concordant pairs, there were 19 participants who completed the task on both designs (cell a), and one participant who failed on both designs. As for the discordant pairs, there were two participants who passed onDPD/V1 but failed on design DPD/V2 (cell b), moreover, four participants failed on DPD/V1 and passed on DPD/V2 (cell c). In total, we noticed that the majority of participants who performed the test tasks differently performed better when they used DPD/V2, such as, in Task1 (6 out of 9) and in Task 3 (4 out of 6). Furthermore, as mentioned before, McNemar's exact test was employed for analyzing the gathered data from the measurement of task completion rate, through calculating the one-tailed *p*-value with the exactprobabilities. However, according to Agresti and Coull (1998), in most user based research, generating mid-*p* value within the data analysis would work better, and it would lead to more accurate decisions over the long run. Furthermore, since the total number of participants tested in this experiment was considered to be a small sample size (26<30), calculating the exact probabilities tend to be conservative by having a limited number of possible values instead of taking on any number of values (Sauro and Lewis, 2012). Therefore, to solve this problem we need to simulate a continuous result through using a middle point between the exact probabilities, so mid-*p* value was calculated and then extracted with two-tailed value (Sauro and Lewis, 2012; Fagerland et al., 2013). The tools used for calculating Mcnemar's test, mid-*p*-values, and the sign test, were an online calculator[13], in addition to Microsoft Excel via the function BINODIST.

Table 6.11 Calculating of Mcnemar Test for Task Completion Rate

	***p*-value**		**One-tailed Mid-*p*-value**	**2-tailed Mid-*p*-value**	**Proportions**		**Difference (Unsigned)**
	(DPD-V1)	**(DPD-V2)**			**(DPD/VA)**	**(DPD/VB)**	
Task1	0.002	0.018	0.011	0.02*	18/26=0.692	21/26=0.807	0.115 (+)
Task2	0.031	0.156	0.109	0.219	24/26=0.923	23/26=0.885	0.038 (-)
Task3	0.016	0.094	0.063	0.125	21/26=0.808	23/26=0.885	0.077 (+)

P<0.05*

The results in Table (6.11) show that there was a significant difference only at the 5% level, in the key test Task 1, between the two versions of the prototype (1 and 2) in regards to the task completion rate. Thus, the participants could perform better on version 2 than on 1, with a positive ratio improved by 11.5% based on the difference of proportions between the two versions. On the other hand, there were no significant differences apparent between the DPD/V1 and V2 as P>0.05, besides a decline in the performance of completing Task 2 by using DPD/V2.

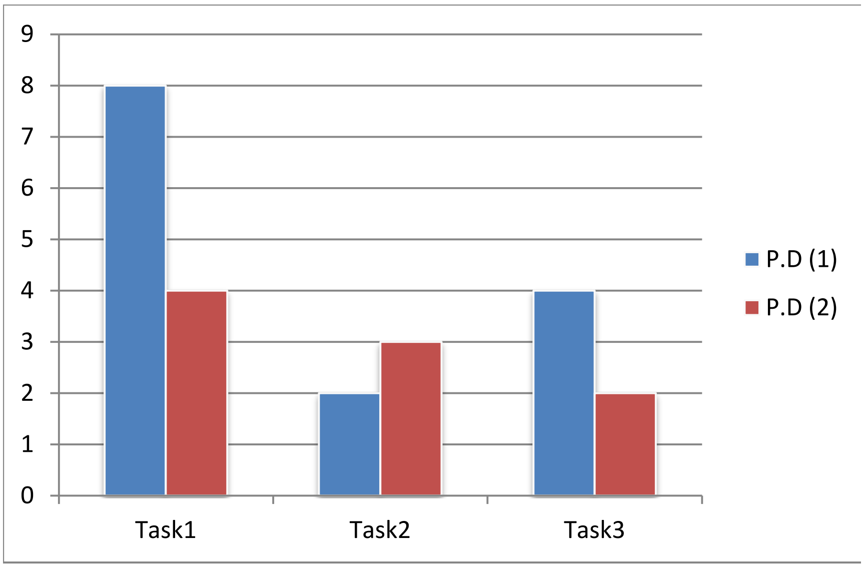
[13] The calculator was used from the following link, http://vassarstats.net/propcorr.html

However, there was a slight improvement in the participants' performance on finishing Task3.

6.6.1.3 Error Rate

In this second experiment, as stated before, the data collected through employing the error rate measurement included the total number of errors which occurred during performing the test tasks by all participants. These errors were observed and recorded when each participant committed them directly in the course of processing the tasks using UI design or pursuing the test scenario, such as giving an incorrect order by pressing the wrong button within the UI of prototype design for processing a task, and ending the task at a different point of the target scenario. Additionally, the data gathered from applying this measurement of error rate was only analyzed simply using quantitative and qualitative methods, since few errors were identified across the test tasks. So, as it was analyzed quantitatively, Figure (6.11) presents the total number of recorded errors across the test tasks. It can be seen that the majority of errors in the experiment overall occurred in Task1 using both versions, 12 out of 23 errors. More than half of these 12 errors were committed when the participants finished it usingDPD/V1; in particular, the participants in group B who started the test session using DPD/V2. Additionally, 5 out of 8 errors were about incorrect starting of the task, either by attempting to click the wrong button or navigating in the wrong part of the UI, where the new feature of AADM has been added to the new version of the prototype design. The analyzing of this issue qualitatively will be reported in the qualitative results section. However, in Task 2, which was concerned with the amendment ability of the system, it is obvious that a few errors occurred in performing the task in general. Furthermore, there was a slight difference between the two versions by one error more occurring when the participants performed the task using DPD/V2. Most of the errors, 4 out of 5, that happened in the task were about choosing the wrong button for processing the task. For instance, three participants clicked mistakenly on the same information field in order to amend the requested visa details, when they supposed to press on the dedicated button of amendment. The other two participants chose a wrong part and button for searching and calling up the previous visa details. In the last task, Task 3, similarly few errors occurred, and the participants could complete the task usingDPD/V2 with less error. Two out of the total five errors in this task occurred when the participants attempted to create a new account for the applicant by choosing the wrong option.

Figure 6.11 Total number of Errors for each



6.6.1.4 Subjective satisfaction results (post study questionnaire)

For analyzing the 26 participants' responses to the post study questionnaire, and to discover the outcomes, the total average value of results for each of the sub-scales within the post study questionnaire was calculated, which can be correlated to some extent, rather than each statement being calculated independently. As we mentioned previously, the average was calculated for the questions as follow:

a. System Quality (SysQual): Average the responses to statements 1 through 11
b. Information Quality (InfoQual): Average the responses to statements 12 through19
c. Interface Quality (IntfQual): Average the responses to statements 20 through 26.

Before conducting the analysis, we ought to present the criterion of weighted mean for each response in the five point Likert scale within the post study questionnaire. The following (Table 6.12) represents this criterion of weighted mean for each sub-scale within the post study questionnaire.

Table 6.12 The Criterion of weighted mean

Weight Mean	Response
From 1.00 to less than 1.80	Strongly Disagree
From 1.80 to less than 2.60	Disagree
From 2.60 to less than 3.40	Neutral
From 3.40 to less than 4.20	Agree
From 4.20 to less than 5.00	Strongly Agree

Table 6.13 Descriptive Statistics of Post Study Questionnaire

	PD/V1				PD/V2			
	Mean	Std. Devi.	Min	Max	Mean	Std. Devi.	Min	Max
SysQual	3.13	.307	2.36	3.73	3.77	.272	3.36	4.45
InfoQual	3.05	.199	2.78	3.56	3.12	.222	2.67	3.67
IntfQual	3.39	.356	2.83	4.00	4.17	.366	3.00	4.17
Total	3.17	.185	2.69	3.58	3.61	.169	3.35	3.88

The descriptive statistics data in (Table 6.13) included all the three constructs of the post study questionnaire in each version, besides the calculation total. The results indicated the mean of all the three constructs of theDPD/V2 and the total was higher than the means inDPD/V1, particularly in the SysQual and IntfQual sub-scales. Thus, based on the measurement of the scales, which are presented in (Table 6.12),this would reflect generally the increase of the users' satisfaction towards the iterative prototype design.

As the ultimate object of this experimental study was the comparison between the two versions (1 and 2) of the developed prototype design, to figure out whether there are any differences between them, the paired-samples *t* test was utilized for analyzing the collected data. Lazar et al. (2010,p76) stated that "*since the data points contributed by the same participant are related, a paired-samples t test should be used*".

Table 6.14 Paired *t*- test for the sub-scales of post study questionnaire

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	SYSQUAL/V1 - SYSQUAL/V2	-.63287	.33276	.06526	-.76727	-.49846	-9.698	25	.000***
Pair 2	INFOQUAL/V1 - INFOQUAL/V2	-.06838	.27591	.05411	-.17982	.04307	-1.264	25	.218
Pair 3	INTFQUAL/V1 - INTFQUAL/V2	-.14744	.28801	.05648	-.26376	-.03111	-2.610	25	.015**
Pair 4	Total (V1)-(V2)	- .44675	.20499	.04020	- .52954	- .36395	- 11.113	25	.000***

Level of significant *P<0.05,**P<0.01,***P<0001

Table (6.14) presents the results of analysing the data after utilising the paired t-test, based on the sampling distribution of mean differences for each sub-scale in both versions of prototype design and for the total of the questionnaire. The results indicated that the pairs of SysQual and IntfQual showed significant statistical difference between the participants' responses on the questionnaire on both versions; *p*-value < 0.05 level of significant. However, the pair2 of InfoQual was the only one of the sub-scales that was not significant. This would indicate that the participants' overall satisfaction has increased across the developing prototype design. In addition, the pair 4 represented the difference in the overall user's satisfaction with the two versions of the prototype design. The result showed that there is a significance difference, *p*=.000 < 0.05,

a. Wilcoxon matched-pairs signed-ranks test

In order to determine whether there were differences between the two versions of the prototype design in terms of the participants' satisfaction, Wilcoxon matched-pairs signed-ranks test was applied as a non-parametric test used since the sample size of 26 participants was considered to be small, besides the within-subject designutilised (Cairns and Cox,2008). The data in (Table 6.15) shows the sum of the ranks of the positive and negative differences of each pair of the test tasks for each sub-scale of the post study questionnaire, and for the total.

Table 6.15 Ranks

		N	Mean Rank	Sum of Ranks
SysQual (V2) - SysQual (V1)	Negative Ranks	0[a]	.00	.00
	Positive Ranks	25[b]	13.00	325.00
	Ties	1[c]		
	Total	26		
InfoQual (V2) - InfoQual (V1)	Negative Ranks	7[d]	12.86	90.00
	Positive Ranks	15[e]	10.87	163.00
	Ties	4[f]		
	Total	26		
IntfQual(V2) - IntfQual (V1)	Negative Ranks	3[g]	7.17	21.50
	Positive Ranks	13[h]	8.81	114.50
	Ties	10[i]		
	Total	26		
Total(V2)- (V1)	**Negative Ranks**	**0[g]**	**.00**	**.00**
	Positive Ranks	**25[k]**	**13.00**	**325.00**
	Ties	**1[l]**		
	Total	**26**		

a. SYSQUAL(V2) < SYSQUAL(V1)
b. SYSQUAL(V2) > SYSQUAL(V1)
c. SYSQUAL(V2) = SYSQUAL(V1)
d. INFOQUAL(V2) < INFOQUAL(V1)
e. INFOQUAL(V2) > INFOQUAL(V1)
f. INFOQUAL(V2) = INFOQUAL(V1)
g. INTFQUAL(V2) < INTFQUAL(V1)
h. INTFQUAL(V2) > INTFQUAL(V1)
i. INTFQUAL(V2) = INTFQUAL(V1)
g. Total V2 < TV1
k. Total V2 > TV1
l. Total V2 = TVA

Table 6 .16 Test Statistics

	SYSQUAL(V2) – SYSQUAL(V1)	INFOQUAL(V2) – INFOQUAL(V1)	INTFQUAL(V2) – INTFQUAL(V1)	Total (V2) - (V1)
Z	-4.383[b]	-1.195[b]	-2.436[b]	**-4.378[b]**
Asymp. Sig. (2-tailed)	.000	.232	.015	**.000**

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

6.6.2 Setting up the study questions

For investigating the questions in relation to the measure of satisfaction, the Wilcoxon matched-pairs signed-ranks test was employed, since the usability testing situation uses a design with two independent samples (Sheskin,2003). Therefore, to examine the differences between the two versions through the participants' responses to the post study questionnaire, the questions are spelled out below:

The results of the analysis using Wilcoxon matched-pairs signed-ranks testdata indicated that there was a significant difference in the overall participants' satisfaction toward the two versions ofDPD.

SYSQUAL: z = -4.383, T (N = 26) = .000, p < .0001

INFOQUAL: z = -1.195, T (N = 26) =.232, p > .05

INTFQUAL: z = -2.436, T (N = 26) = . 015, p < .05

TOTAL : z = -4.378, T (N = 26) = . 000, p < .0001

6.6.3 Qualitative results and analysing

As mentioned before, the qualitative data in this experiment was collected via the observation and note taking methods while the participants were performing verbally the test tasks using both versions of the prototype design. It was also collected through the comments that have been given by the participants regarding some statements within the questionnaire after they were encouraged to provide their views in depth, besides in response to the open-ended questions section within the post study questionnaire. In this section analysis and interpretation of the qualitative data was employed as supportive to the quantitative data for achieving better illustration of the participants' point of view on the proposed prototype design, and explanation of some ambiguous quantitative results.

In the overall qualitative results, after the participants have assessed the two versions (1 and 2), they were fairly similar regarding some of the positive design aspects, which have been validated in experiment one. For example, providing all the required information for initiating the tasks in one screen, and the system having only limited steps to complete the test tasks, were mentioned as great improvements for the suggested prototype design by the vast majority of participants. In addition, among the comparable responses were some about some features which have not undergone the process of development for producing the new version of the prototype design. For instance, in relation to the system feature of auto-save, approximately more than half of the participants (71%) were impressed with this feature, as they stated this mostly within the questionnaire in the section where they were asked to list something they like and explain why, saying that the auto save or the default saving of data feature in the system would protect the data entry from being lost, thus saving time and effort. However, there were some disparate qualitative results, including positive and negative comments; this was especially related to the new additional features considered for the developed prototype design, and the other aspects which have been improved in the new version of the prototype design. To indicate these further dissimilar thoughts of the participants on both versions of the prototype, we summarized the frequent recorded positive and negative comments, and then reported them in a form that would support the questions and hypothesis of the study (Zimmerman and Muraski, 1995).

Through the observation technique, we found that most of the errors were committed when the participants finished the test using DPD/V2 and started the other session of the test using DPD/V1, their engagement with the new features of the prototype in the first session and its effects carried over to the next session. They attempted to processTask1 with the same functions as in the previous test using DPD/V2. One of those participants stated that "*Although I've been informed about the test procedure clearly, I started this second session with thoughts of using the same system design which was unlike the intended one*".

6.6.3.1 Analyzing the results of AADM feature

The obtained outcomes in regards to the feature of AADM were that most of the participants showed their positive attitude and agreement to adding this feature within the system, as it would be such a remarkable improvement that would grant them a lot of advantages for their performance as users. The participants' feedback was reported based on the technique or tool that has been used for obtaining this result. The following table provides some insight into the participants' reactions via collecting notes from the observation technique and written comments on the post study questionnaire, and they are quoted and displayed as follows:

Table 6.17 Positive and Negative qualitative responses on AADM

a. Observation Technique
(+) - P5*"It would help to unify the decisions that need to be taken for issuing the visas in some complicated cases..."* - P17 *"This feature is definitely going to save my time and effort..."* - P18 *"I like this feature and how it shows me how the applicants' conditions meet the requirements..."* - P21 *"This feature is great, it enables me to grant a quick check on the required information to issue the visas to the applicants as soon as possible..."* (-) P17 *Actually, I think relying on the computer for taking the decision would not cover some aspects of the visa applicants, like increasing their income..."* P11 *"... , but sometimes, I feel I would not trust the computer's decisions..."*
b. Post-Study Questionnaire
(+) - P1 *" The function of verification of the conditions as an informational base would provide a great assistance to the employees for completing their tasks effectively, and then matching the presented information with supplied documents, which would eliminate manipulation and fraud"* - P7 *"This system has the functions and capabilities I need to perform my work effectively".* - P17 "*I believe that,as long as the system keeps assisting me for making a quick and accurate decision, this would increase my productivity and provide better services to the public".* - P18 *"This feature would help relatively the employees to identify the required information visibly on the screen rather than to be obtained as a hardcopy among many other paper documents"* - P23*"I like this feature, and I think it would help a lot to cope with the increasing number of applicants, because I'll be able to finish my work tasks faster"* *(-)* P5 "*This improvement of adding AADM would not succeed with the system, because there are enormous regulations and they are changed occasionally by the administrationin accordancewith the requirements and circumstances, which might make the employee a bit worriedabout taking a decision based on the system assistant"* P25 *"I believe that the complexity of some cases for issuing visas, would make the implementation of this feature difficult"*

6.6.3.2 Analyzing the results of Tasks 4 and 5 in relation to the Archiving Function

The results obtained from performing Tasks 4 and 5, using both UI versions, were analyzed only qualitatively, since they were designed as new test tasks for the purpose of evaluating the additional feature of Archiving Function (AF) which was only proposed in the new version of the prototype design, and was also not considered to be within the comparison usability test between the two versions through conducting various quantitative measurements that have been applied previously.

According to the participants' feedback in regards to this feature, either by the verbal comments during performing the tasks or by writing them as explanation in the post study questionnaire, there was almost a consensus on the effectiveness of integration of this feature within the process of the system development. For example, one of the participants (P1) has commented in the section of the open ended question within the post study questionnaire (Section b1), that said, "Please list the particular aspect(s) of prototype UI that you liked, and why?", that "*The archiving option with its several functions, it would help the applicants to reduce the number of required documents that need to be carried and submitted for the application. And this in turn would assist a lot the employees to accommodate and accept the volume of applicants, in particular those who are reluctant to review the administration repeatedly, such as the individual pursuers".* From observing another participant (P5) while performing Task 4, he said "*this feature certainly would adress an important part of our needs of the system to provide faster and better services to the applicant*". Furthermore, the participant (P22) mentioned that there is no need for many papers if we have this feature. From other participants' comments in regard to this feature, we could summarize their responses as being that fundamentally considering the Archiving feature within developing the system should be taken into account from the earliest time, as it corresponds with their aspiration. So, it would grant them better control channels upon the tasks they intend to do, such as a quick check up on the required documents, and complete them efficiently, while it would meet one of the administration's goals of implementing a paperless environment. On the other hand, on both versions, similarly some participants reported and provided some recommendations for improvement and comments about problems they faced and aspects that they felt needed to be adjusted. For example, some error messages still need to be re-written as they were unclear to the participants.

6.7 Discussion

The rationale for conducting this experiment was to test the improvement of the DPD, after it has been through the iterative design process, and to confirm it through enrolling several measures, tools and techniques to achieve a unique strategy for improving system usability. As Nielsen (1993) stated, "*Redesigning user interfaces on the basis of user testing can substantially improve usability*"**,** as has been already corroborated empirically in a study by Juristo (2009), as the usability and software design are associated. The quantitative and qualitative results obtained in this experiment study highlighted the importance of issues related to verifying of this relationship.

a. Sample and Procedure of the experiment

In regards to the test sample, it was mentioned earlier that 26 participants were recruited to take part in this usability test. All of them were involved previously in the former experiment, so they already had a proper knowledge about the purpose of the experiment in general, and had gained experience of using the old version of the proposed (DPD) system. Consequently, each participant had to have a short training in using the new version of (DPD) together with presenting a brief demo regarding the additional features that have been added to it, i.e. the feature of AADM. Nielsen (1993) in his study mentioned that at the stage of developing a UI when it is transmitted to another generation where the users will gain experience of how to interact with the old techniques, it has become necessary to provide them with some training in using the new ones; otherwise they will be completely confused. Yet, in the same study, he provided conversely a view about the involving of experienced users or participants in the usability test, as he stated that *"it is normally quite easy to conduct user testing with novice users in order to measure learnability and initial performance, error rate, and subjective satisfaction. Usability measures for expert user are unfortunately harder to come by, as are measures of the ability of experienced users to return to the system after a period of absence."*

In addition, in regards to the sample size, and how many test users should participate in a usability study, there is controversy revolving among the HCI community regarding this issue. For example, Nielsen (2012) and many other researchers (Lazar et al.,2010; Virzi,1992) indicated that five users are sufficient for conducting a usability test, and they attributed that it will identify about 80% of usability problems. On the other hand, Nielsen (2012) in the same study, clarified that some tests should recruit more than 5 users based on the exceptions in certain cases, such as that a study that approaches quantitative methods would require at least 20 users. Furthermore, other research studies have revealed that choosing only five users in the test is not enough to identify a majority of usability problems (Lindgard and Chattaichart, 2007;Spool and Schroeder,2001). In spite of the previous issues having been in a broad debate in the HCI community, in our study, we could provide some considerations for choosing this sample size and characteristic as follows:

- The proper accessibility to the sample was found with the employees who work in the west branch of the government organisation, compared to the other employees in other branches, besides the provision of appropriate facilities to implement the test setting of an informal lab in the work place. Lazar et al. (2010) pointed out that the test location can be determined by the availability of the places for the test or of participants. Additionally, this factor of having an opportune access to the sample, besides well coordinated arrangements for conducting the experiment, has a significant impact on completing the test in five days sufficiently.
- The probability of emergence of some negative implications in case of choosing another sample of employees in another branch, like increased cost (such as travelling and living expenses where another branch and sample are), and the need for extra time (re-presenting the test, and providing the training from the beginning), which would in turn affect the conduct of research properly.
- For enabling the statistical analysis of significant differences between the two versions in the study we invited 26 participants, and also for obtaining meaningful outcomes (Rubin and Chisnell, 2008).
- Eliminating the participants who participated before in our case would have a negative impact on the sample size, which in turn could lead to having too small a sample to conduct the experiment (Teijlingen and Hundley, 2001).

- Conducting the test with experienced participants would help to save time in terms of shortening the definition of the test's purpose and procedure to the participants, as they will quickly come to grips with it. Moreover it would give the experiment a balanced quality, since a participant should have an experience with both versions ofDPD, and otherwise the test would be biased. As consequence, considering the experienced participants to be involved in our usability study of improvement of an internal software system became important at this stage of iterative design.

Our assumptions and conclusion towards this issue were congruent with a study by Faulkner (2003) which stated that deciding the number of user "participants" to test the usability of interface, can be determined by several considerations, and variables over which the practitioner have diverse levels of control, such as which types of users are available or accessible to the practitioner, how important is the mission of a system, or having any possible effects on any specific usability problem. Thus, those variables can have a deep impact on the number of test users needed to gather accurate and valid results in a specific time.

From the previous test, the participants showed their preference for non-use of recording devices, such as a camera, even if they have knowledge about the privacy and confidentially of the test. However, some of them agreed only to the taking of a snapshot for research purposes but without their faces appearing. According to Mackay (1995) people mostly feel uncomfortable being recorded, which could force them to change their usual behaviour; which they do not like to do.

Furthermore, in regards to the test tasks, there were five tasks that have been designed to fulfil the aims of the study. Three of them were selected from the primary experiment, due to the study using the iterative design, so the essential tasks should be re-tested in order to assess and confirm the improvement. Besides that, the three tasks (1,2 and 3) were mainly representing the typical tasks of the target system, which was basically implemented for issuing visas to the citizens. Task1 was determined to bes a key task amongst them, because it was associated directly with operating the new feature of AADM, and for measuring the efficiency,which determines the state of the system usability. Nielsen (1994) and Kuniavsky (2003) pointed out that the chosen tasks to be performed during the test should include the representation of the most important parts of the UI design , and the participants should be able to finish these tasks within the time limits of the test.

Another important issue was discovered through the findings of the pilot study of the test procedure, which was related to the providing of the task scenario and the list of tasks to the participants before starting the test session. It should be delivered and explained to the participants both by handing them out, and being read to them as well, since the participants raised some queries about the given information, and this wasted some of the scheduled task time. This finding was in line with the view of Hix and Hartson (1993) that the evaluator can either hand the participants the copy of the written list of test tasks, or he can read it out loud to them. However, in our experiment, we preferred to utilise the two actions of apprising the subjects of the test plan and the tasks list, besides giving it to the participants to refer to in case they need to do so.

b. Iterative design

Considering this experimental study as a complementarystep to the previous stages of the research,which has applied the human-centered design (HCD) approach, it tackled the issues and resolved some major problems based on the user experience, and then redesigned the suggested prototype design, and issues that came across in the previous experiment, and what would arise recently, in order to develop a successful prototyping design. So, the sequences of these stages of design in our experiment study were considered as central to the iterative design process, as the study by Kies et al. summarized (1998). Following this approach of iteration design over prolonged periods, as Stringer et al. (2005) pointed out, it would help to understand the users properly, determine what a system's best use is, recognize an appropriate use condition, and how the technology would fit within the design. Therefore, in our case, the iterative prototyping design was considered with additional features i.e. AADM and AF, which were suited to the users' needs, and compatible with the nature of their daily work, in order to assist them to complete the tasks effectively. This would coincide with what Rubin and Chisnell (2008,p.6) stated, that through integrating a usability testing, as evaluation methods, during an iterative design process, it would generate products and services that are " *useful and usable, and possibly even delightful"*. In addition, another study by Matera et al. (2006) mentioned that the core of iterative design approach is that it can be certain about the effectiveness of improving particular design decisions, and evaluate them through the use of application prototypes.

c. Additional feature of AADM

It was evident from the results of answering the study's questions, that we achieved one of the main aims of the research, to improve the usability of the current VI system based on users' point of view, and by involving a set of new features of computerised system for supporting making a decision, and enhancing the control of performing the actual tasks through introduction of an Electronic Archiving System. This integration of AADM as ICT into the system development would be a vital step towards achieving many advantages for the government organization as it attempts to improve its software system. The Australian Government (2007) indicated that ICTs are shifting the way government operates. In addition, Juristo et al. (2007) emphasized the importance of considering the features that enhance the software system's usability at the earliest stages of the development lifecycle. Furthermore, this outcome supported what Macleod (1994) mentioned, that "*Workproductivity and efficiency are higher when using IT systems with good usability,and there are fewer 'user errors'; less training of staff is required to enable effectiveand efficient use of the system, users are more satisfied, and there may be lower staff turnover"*

In regards to considering the AADM feature in the design, the study derived some benefits from the quantitative and qualitative results. For instance, it helped to increase the speed of completing the test tasks.. According to the results of calculating the task time in Task 1 which was the major task of the study designed for measuring the system efficiency, the participants' performance improved as they could finish it using the new version of DPD with less time, by 23 seconds difference. In addition, in finishing Tasks 2 and 3, the participants could save 6 and 11 seconds respectively. Other benefits and positive views were drawn from the qualitative results after the participants presented their insights regarding this feature of AADM, through providing the comments and notes. Some of these thoughts were quoted and displayed in Table16. One of the participants' feedback illustrated that the actual activation of this technological feature within the system as a functional assistant would alleviate the sensation of responsibility towards giving the correct decision, due to the system providing the user with required information, which was expected to be more reliable and accurate. Another participant provided such an interesting potential advantage, as he stated that an applied Automated Assistant system would help to quickly check through the applicant's condition, at the same time it would enable making sure that the documents supplied corresponded to the related information that was presented from the system. Furthermore, P18 in Table16 presented a view that applying this computer support would help to reduce reliance on paper, so it complies with Goldsmith and Egger (2004) who described the AADM as creating less paperwork.. Additionally, some studies have presented other views of advantages with respect to AADM in government administration. For example, the study by Citron (2008) believes that applying the AADM system is "cost effective", since it uses a computerised system in the decision making process, thus it reduces the human role, relying on the level to which the decision making is automated, besides, there is an economic impactofreducing the use ofpaper as stated before (ibid). Another advantage was reported by William (2005), that relying for making the decision on the automation of computer support, ensures consistency, as the computer system could interpret rules and provide supporting information for decision-making with logic and accuracy, on the facts of what the applicant's state showed. This agreed with the statement of the Australian Government[14](2007,p.22) that the AADM system has a significant impact on "*the accuracy, transparency and accountability of the administrative decision-making process*". Additionally, a positive attitude to utilising AADM was also found from analysing the participants' comments to be a significant factor in their confidence to provide better service in their work. On the other hand, a few participants in this experiment and a number of other published studies have proposed negative

[14]Australian Government, *Automated Assistance in Administrative Decision-Making*, Best Practice Guide, February 2007.

opinions and argument regarding the utilising of AADM in the government administration. For example, one of the participants (P5) has provided his own perception as he stated that "*This improvement of adding AADM might not succeed with the system, because there are enormous regulations and they are changed occasionally by the administration in accordance to the requirements and circumstances, which might affect either the system giving an accurate decision or the employee to be worried over taking a decision based on the system assistant*". The Australian government (2007) mentioned a view in its Best Practice Guide, it may correspond with an aspect of this attitude; if the delivering of the ongoing rules and policies of the system which would lead to a complicated and changing legislative environment, did not have an appropriate capacity of budget and the right personnel for updating the rules upon which a system works, this may result in a risk of making inaccurate and/or unlawful decisions. Another argument was provided by the study of Citron (2008,p.1253), which explained some reasons for challenging AADM, the main reason was assigned to the belief that automation system would expose the *"procedural safeguards at the foundation of administrative law"* to risks of fragmentation by several issues. For instance, the process of translating a set of rules as human language into computer code would lead in some way to significant change in the substance of the rules, or what is called misinterpreting rules (Australian Administrative Review Council, 2003). This might be because as Grimmelmann (2005) claimed, the artificial software languages, which are comprehensible to a computer, have a limited vocabulary when compared to the human languages. Furthermore, another issue was pointed out by Citron (2008), regarding the human discretion, depending on the facts, it becomes restricted or impossible with computerised systems.

It should be noted however that it was obvious from the obtained findings from this study, in respect to iterating the UI design and improving the system usability, which led to reshaping the design and providing it with a package of functions as guidance, that it represented clear and understandable information, and afforded a suitable means of entering data and instructions easily and efficiently to the users (Macleod,1994) for developing the usability of an internal VI software system. This supported the view that was mentioned in the published study of

Juristo,(2009), which intended to enable developers to spot functional usability requirements without relying on a usability expert.

d. The measures of the test

1. Task time

The results obtained from analysing the measure of time on task showed that there were generally remarkable differences in the participants' performance in finishing the test tasks faster when they used the new release version. Initially, the finding from the calculation of task time results, which is displayed in table 3 as the average task time for each target task when all the participants in both test groups (A and B) performed them across all the sessions of the test tasksusing both versions (1 and 2), was about their performance and how it improved after they swapped between the versions. For example, comparing the average time for finishing Task 1, the participants in group (B), who ended the test session by using the (V1), performed and finished Task 1 using (V1) with an average time of 236 seconds better than group (A) who started the test using (V1) and spent an average time of about 247 seconds for completing the Task1 using the same version of (1). Similarly, in the same Task1, the participants in group (A) spent less time to finish this task using (V2) than group (B) with a time-lag of 11 seconds. This issue would describe the need of employing the counterbalancing technique within the test to minimize the effects of learning on the test results. Therefore, this technique was implemented, running the test sessions with each participant who attempted to use both versions alternately. Tullis and Albert (2008) stated that the participants generally learn to use the product as well as their experience with it grows, thus, the way to solve this issue is to control for order effect by implementing a counterbalancing technique. What is more, after applying this technique within the experiment, it could also help to avoid obtaining biased results, due to apparently the participants' performance has been boosted after they finished the tasks using the first version (1 or 2) before the other. This might relate to the aim of the experiment for comparison between the old version of prototype design and the new release version which has been developed through involving new features and functions, subsequently; both versions (1 and 2) had the same design instruction layouts, which would increase the likelihood of the occurrence of the learning influence.

2. Error rates

Through analysing the results obtained from applying this measurement, it has been found that the overall rate of errors that were committed across the entire test tasks using the two versions of DPD was 23, the errors over the new release ofDPD were fewer than the ones using the old version, even if the participants have experience of using this version rather the new one. This would possibly refer to the fact that most of the needs, problems, and drawbacks had been tackled and improved in the new version of DPD. Nielsen (1993) pointed out that the system should have a low error rate, which means that its users should commit as few errors as possible during performing tasks using it, and even then they will be able to recover from them easily. However, only in Task2, the participants committed more errors when they used the new version. Furthermore, it should be noted that most of the tasks' errors occurred generally in Task1 which was designed as a key task for the test. In detail, through the observation technique, we found that most of these errors in Task1 were committed after the participants finished the test using DPD/V2 and started the other session of the test using DPD/V1. So when they attempted processing Task1using DPD/V1 with the same functions as in the previous test session using DPD/V2. This would relate to their engagement with the new features of the prototype in the first session and its effects being retained in the next session. One of those participants mentioned that "*Although I've been informed about the test procedure clearly,I started this second session with thoughts of using the same system design which was unlike the intended one*". This issue as it occurred in this experiment frequently should be taken into account especially when the study aims to conduct a comparison test between two versions where they are quite similar and have unified layout design, but the new version has been improved with new functions over the old one. There was also a noticeable issue that has been discovered through analysing of the observation results, which was that related to the number of errors committed in Task 2 as stated above. We noticed that the participants committed more errors when they performed Task 2 using DPD/V2 than DPD/V1, however, they consumed less average time to complete the same task using the new version.

3. Satisfaction survey

As mentioned before, the subjects' satisfaction rating was collected through administrating a post-study questionnaire. It was handed out to each participant individually at the end of the test session to give scores on a 5-point Likert-scale to indicate his preferences, for getting an insight into the subjective impressions towards the usability of a system (Wallach and Scholz, 2012), besides, the participant was encouraged to provide any comments or suggestions for improvement on the version he had just tested. This measurement of satisfaction in this experiment study, showed how much pleasanter the new version of the prototype design was to use, and the participants were satisfied because they could interact with the new additional features easily, and perform the represented tasks faster and more accurately. The helpful dynamic of presented features for developing the P.D. may have allowed this improvement of performance to improve the participants' satisfaction and generated significant differences in their responses. In regards to the constructs of the questionnaire, the analysis of the results showed that there was a significant difference in SysQual T (N = 26) = .000, $p < .0001$ and IntfQual T (N = 26) = . 015, $p < .05$ between the two versions (1 and 2), whilst in the InfoQual there was no significant difference. These results indicated that applying the iterative design for proposing the new version ofDPD with AADM and AF, was satisfied the participants because it provided the usefulness and effectiveness for them to take a quick decision and then complete the tasks faster. On the other hand, in respect to the information quality, it would have a slight improvement through adding some missing information which has been reported by the participants, like the adding of an option for transfer of sponsorship in the immigration section, and adjusting display of some error messages. This minor amendment might lead to no significant difference in the participants' responses about the information quality, but in the entire participants' satisfaction the Wilcoxon test analyses showed that there was a significant difference in the participants' responses which would explain the significant improvement in their attitude and performance with the new release of prototype design. Furthermore, an important observation was recorded and it might be related to this concern for information quality. So, even after the adjusting of the error messages, yet some participants were unsatisfied and still providing comments about revising or rewriting the message context or redesigning the message format as whole, and this would be referring to the multiplicity of linguistic terms.

Finally, a few participants revealed their resistance to this proposed improvement, as it appeared from the observation notes that they had not absorbed the concept of prototype designs, as they claimed that it needed to involve all the protocol of the current VI system with more professional implementation.

6.8 Chapter summary

This chapter has documented the second experiment of the study, which was designed to investigate and validate the improvement of the new release of prototype design after it had addressed the outcomes and issues from the primary experiment. Accordingly, thesuggestedprototype design was refined in line with the system employees' perspective, as the users of the actual VI system, to fulfill their needs in their daily work, which was highly associated with the decision-making process for issuing visas to the applicants. Therefore, the iterative technique was conducted including involving an Automated Assistance Decision-Making (AADM) feature, and the other additional feature of ArchivingFunction (AF). In respect of assessing and validating the improvement of adding these features and adjusting the issues that came across in the previous test, a comparison usability test was conducted between the two versions of the prototype design (1 and 2) with various tools and techniques employed for collecting the quantitative and qualitative data.

The results confirmed the most critical usability improvement strategy, which was detected to be including more assistant features within the system and validating them further. It showed that the participants' performance was generally improved when they used the new version of prototype design compared to their performance when they used the old version. This performance advantage appeared greatly in relation to involving an advance in information technology (IT) of the Automated Assistant Decision-Making (AADM) feature. The main rationale for involving this feature within the iterative prototype design was to improve the system usability, through accelerating the process of decision making which needs to be undertaken to perform and complete the main tasks in an easier, faster, and more accurate manner. Accordingly, the study could answer to the RQ6, by involving additional features ICT to the prototype design which led for having a positive effect upon users' experience. What is more, the qualitative outcomes presented the other likely advantages pointed out by the participants, such as overcoming of the problem of taking varying decisions for issuing a visa regarding one specific case, and increasing the confidence in providing accurate decisions, because the system would keep up to date with the latest outgoing instructions and regulations. Hence, the most important two aspects that can be deduced in using AADM were firstly that it motivated the participants to improve their performance and provide better service to the applicants for visas, and secondly the possibility of taking lessons from other successful examples like the Best Practice Guide provided by the Australian Government.[15] Nonetheless, a number of challenges might face the developer when the AADM feature is considered within the system functions, such as to evolve more accurate and consistent measures for evaluation and reviewof the decision taken.

In addition, the study confirmed that a concept or new approach can be proved by utilizing a prototype (Berkun,2000), as it presented an appropriate layout instruction that described the Archiving function process, with the most important selected steps of assisting finishing the main tasks effectively. This was by considering the participants'(employees’) requirements from the former experiment in respect to the targeted current system for issuing visas, in order that they could be provided within the guidance on its improvements to the system's developers.

In this experiment, it seems clear from the findings that the conducted iterative design approach has increased the participants' satisfaction, which may lead to the conclusion that the process of iterative design can be repeated as long as the participants still provide comments and suggestion. Therefore, these issues will be taken into account, and so the improving and changing of the design will be

[15] This guide prompts effective collaboration between Australian Government agencies and seeks to ensure that their use of automated systems for administrative decision-making consistently reflects administrative law values.

applied iteratively for conducting another experiment to compare the results in order to gain the most usable system (Nielsen,1993).

Finally, there was a key finding that has been drawn from the experience of the participants' expressions and their reactions after they interacted with the iterative prototype design and discovered that most of their responses, opinions, and suggestion of their needs were taken into account, which is that they showed a lot of optimism to provide a better performance in their work when this system is implemented in reality as the final system. This might lead to the need to involve the human factor (users) within the system development life cycle in many government organizations in Saudi Arabia that attempt to develop their internal software systems, which would help to provide usable systems that could satisfy the system users and improve their experience, and consequently it might improve their productivity and increase team morale as a whole (Juriso,2009). Therefore, to conclude, the outcomes presented in this study will be employed as a part of guidance for developing and boosting the system usability via considering re-designing aspects in order to gain more successful systems in governmental organizations.

Chapter 7: Conclusion

7.1 Overview

This chapter presents the conclusions of this thesis by summarising the important findings that have been drawn from conducting the research. Firstly, the chapter discusses the main findings and illustrates how the aims and objectives of this research have been achieved in addition to how the main research question has been addressed. The chapter then provides the main contributions of the thesis. Finally, the chapter presents the limitations of this research.

7.2 Introduction

Regardless of the widely acknowledged importance of usability as a key factor of the overall quality in software system development (Mazumder and Das, 2014; Ferre and Medinilla, 2007; Costabile, 2001), it is evident that there are some software systems with poorly designed and unusable features (Juristo, 2009; Maguire, 2001; Och Dag et al., 2001).This is mostly due to development schemes that have not properly addressed usability by not taking into account the users' needs, which has been conveyed by the users with reports of frustrating experiences who find the system difficult to learn and complicated to use (Maguire, 2001).

As a means of addressing this problem within these existing systems, a number of studies have been conducted to identify solutions via the evaluation of the usability and the re-design of the UI of these Legacy systems, aiming to obtain a usable design and to improve UX. In particular, this includes many organizations in the business context, such as the IBM and Microsoft corporations; however, in the government context, Buie and Murray (2012) stated that although the government systems have the largest user base of any technology, the literature shows that there is a lack of attention to UX in government systems around the world.

The main aim of this research study was to propose the appropriate UEMs to be provided as a developed methodological framework, which would comprehensively and effectively investigate the usability problems areas of existing government systems in a developing country, such as Saudi Arabia. Developing this framework by considering the human-centered design (HCD) as the main approach of research work reported in this thesis was an endeavor to put the user and the user's experience of the target system in the centre of the re-design considerations in the development process. In addition, this was an attempt to stimulate awareness of usability and its evaluation methods in this particular developing country context to enhance the failed parts of the system and to re-design it with new features, which would improve the performance of the end user while providing good services to the public in developing countries.

For the methodological design, firstly, the literature was investigated to determine the most appropriate approaches that could be used to obtain a comprehensive evaluation of the usability of an internal government system from the users' perspectives. Chapter three explains the justification behind the selection of each of the methods and discusses their pros and cons. Thus, a mixed method of quantitative and qualitative approaches was utilized in the three studies reported in this thesis by employing a number of instruments and techniques, such as a questionnaire, interviews, usability testing, thinking aloud and observation. Furthermore, each procedure, technique and instrument was tested for validation and reliability before they were used to obtain validated results.

7.2 Summary of the three studies with the discussion of findings

The aims of research were achieved by identifying the answers of the research questions and meeting the objectives of the three studies that were conducted for this purpose. The following is a summary of the main findings of the thesis, followed by an evaluation of the research questions.

First study: *Query technique-based usability evaluation of a current system in a government organization from a user perspective*

The research work reported in chapter four of the thesis aimed to evaluate the usability of the current government system of Visa Issuance to enhance it and then to improve the overall UX. The key finding obtained from this study was that implementing a query evaluation technique using a survey method for evaluating this system's usability was significantly positive and helpful, as it enabled the researchers to achieve the main objectives through providing a good insight of the level of usability and several usability problems areas. This survey technique was used for collecting quantitative and qualitative data by employing two main tools, a questionnaire and an interview, and a free text box was utilized within the questionnaire study for gathering the participants' comments as qualitative data. In addition, making the users the main focus in the system development process could help to identify several problems associated with the system usability from the participants' responses. For example, there was a lack of organisation in the information to be used as a guide for processing a task, so the users of the target system were sometimes confused by the massive information presented in several pages. Furthermore, one of the main findings that was highlighted in the first study was related to the qualitative results that supported the quantitative outcomes by providing an explanation for vague results in the quantitative analysis. In addition, it helped in identifying several usability problems that the quantitative methods could not discover. Therefore, applying a mixed method approach could benefit the research in achieving the study objectives. Moreover, from the participant experience or system usage perspective, we found that the participants who had more experience in using the VI system highlighted more problems. They were not satisfied and indicated that the system needed improvement. This result was the opposite of what has been stated in the literature because it shows that there a significant positive relation between 'system usage' and 'user satisfaction'.

In addition, one of the main findings was that by utilizing the method of the query evaluation technique, a number of major and minor usability problems were identified within the existing government system. Based on this finding, considering these problems and recommendations as additional user requirements of the system could help identify the areas in the current system that need to be re-designed for improvement. Therefore, the query technique can be used as a basis for re-shaping the existing design for further development.

Second study: ***A usability testing approach for developing a prototype user interface design of an internal system in a government organization***

In the main work reported in this study at this stage after the usability problems and recommendations of the first study had been reported and summarized, the process of transforming the additional requirements into a simulation design that would represent the users' needs began. A prototyping design was applied for this purpose because it is considered an important principle of the usability engineering (UE) process (Fulten and Suri, 2000).

Based on the first experiment that was conducted in this thesis, there were a number of positive findings. One of the major findings emerged after the comparison usability test was conducted between the two systems: the current VI system and a suggested prototype design P.D. with 32 participants as actual users. The obtained results confirmed the improvement of the prototype design P.D., which initially was developed based on the outcomes of the first study reported in chapter four. For instance, after comparing the geometric mean of most tasks, as Nielsen (2001) recommended, it is obvious that an improvement of a specific percentage could be observed, i.e. in Task2, the participants could finish this task with C.S in 79.97sec. While using P.D, they finished in 52.59 sec. Thus, the improvement would be calculated as follows:

(79.97-52.59)= 27.02 sec. /79.97= 0.338 ≅ 34 %

Following this procedure:

Task1:(improvement -1.05%), Task3:(improvement 25.28%),

Task4:(improvement 22.02%), Task5:(improvement 33.93%),

Task6:(improvement 58.29%) and Task7:(improvement 16.14%).

This shows remarkable evidence that the research work of the first study was moving in the right direction, as it supported the achievement of the main objectives of the study. On the other hand, the conformation of the proposed prototype design could even validate this first experiment, which was conducted in the second study reported in chapter five.

Furthermore, the most noticeable improvement of the prototype design P.D. was identified in regards to increasing the system efficiency, as it was one of the problems that were reported during the first study. So, through calculating the measurement of task time that the participants needed to finish each test task, the results indicated that most of the participants could finish the selected tasks using P.D. with less average time across almost all test tasks, especially the main tasks of the test, which were tasks 5 and 6; however, only in task number one did the participants consume more time using P.D than using the current system C.S. For example, the participants needed to complete Task 5 on P.D (M = 238.1 sec., SD = 11.96) compared to the subject using C.S (M = 1319.91 sec., SD = 381.50).

Another major finding is that the study involved the combination of medium and high fidelity for developing a suggested prototype design (P.D). According to Preece et al. (2002), HCD should involve a mixed level of fidelity prototyping. Thus, initially, the medium fidelity prototype was used using Microsoft PowerPoint, which helped to address the recommendations easily into an initial design and to apply the modifications later after receiving the feedback from the participants. Boothe et al. (2012) pointed out that a medium-fidelity prototype would increase users' perceptions on a particular system's usability. Righetti (2005) stated that the main advantage of using the PowerPoint to develop a prototype is its simplicity of use in which enables the researchers or participants to develop his own. After that, the suggested design was coded and formed into a high-fidelity prototype to add the required functions that enabled the participants to perform an actual task. For this purpose, and to create a small dummy database, Microsoft Access software was utilised for obtaining a functional design to conduct a compression usability test.

For obtaining proper access to conduct the usability test in an informal lab at the workplace, in addition to the good interaction from the sample of 32 participants who had between 3 and 5 years of experience in using the current VI system, this two factors of a good access to the sample besides the participants' cooperation could help to gather reliable and useful results and generate a clear insight into the usability issues of both the P.D and C.S designs. Rubin and Chisnell (2008) stated that ‘the testing process for an exploratory test is usually quite informal and almost a collaboration between participants and test moderator, with much interaction between the two’.

Conducting the usability test in an informal laboratory at a workplace in a government organisation in Saudi Arabia had several benefits returned to research. For instance, there was a low-cost compared to testing in a formal laboratory. Furthermore, since the research work was conducted at the same workplace where the actual participants worked, it was familiar to them; however, challenges were encountered before and during the test session, e.g. limited capabilities of using the PCs at the workplace to install any other software and applications, such as Microsoft Access, to present our developed prototype design. In addition, there was a rather noisy background. Nevertheless, these problems and other issues were resolved, which is reported in chapter five, to ensure that the test went well.

Third study: *An iterative design approach for a developed prototype of a governmental organization software system with comparison usability testing from the user's perspective*

According to the findings of the experiment in the second study, which was reported in chapter 5, it was indicated that the improvement of the suggested prototype has been validated through conducting a comparison usability test with the current VI system. The developed design was significantly different compared to the current VI system because the participants had performed the main tasks using it in less time, with a better task completion rate and with a reduced occurrence of errors. Yet, some suggestions and recommendations have been raised by the participants as additional requirements that need to be included within the simulation design to improve their work performance, to cope with the increasing workload and to meet recent diverse demands, such as an increase in the number of applicants for visas, which requires information about the regulations for granting visas to those applicants (Al-Motaryi, 2012). Accordingly, this study aimed to employ an iterative design as a cyclic process of prototyping and as a proper method to ensure the incorporation of the participants’ input into a system design (Stone et al., 2005). Nielson (1993) stated that 'redesigning user interfaces on the basis of user testing can substantially improve usability'. According to Rubin and Chisnell (2008), applying an iterative design and testing method would cause a project to gain steady and quick progress.

The findings of this study revealed that applying an iterative design approach for improving and modifying the developed prototype design based on the first experiment outcomes has enhanced several aspects of the system usability, such as the efficiency, as the participants could finish the tasks faster when they used the new version (V2) of the prototype design compared to the old version (V1). For example, when comparing the average times for finishing Task1, the participants could perform and finish it using (V1) with an average time of 241.96 sec, whilst they finished the same task using the (V2) faster with a lower average time of 219.08 sec. Moreover, the participants' satisfaction increased based on comparing the results of a paired t-test for analysing the post-study, and the results showed that the overall participant satisfaction increased across the developing prototype design with a significance difference, p=.000 < 0.05. These findings may lead to the conclusion that the process of iteration design can be repeated as long as the participants still provide comments and suggestions to develop the most usable system (Nielsen, 1993).

One important finding is that the employment of HCD as the main approach for conducting the work reported in this thesis and employing both formative and summative system evaluations helps to achieve the systematic discovery of useful functions. In the case of the third study of this thesis, identifying the Automated Assistance Decision-Making (AADM) feature was such a function based on an understanding of the work domain.

For this additional feature of AADM and another additional feature of ArchivingFunction (AF), which have been added to the prototype design to fulfil the system users' needs in enhancing their performance, a major finding was revealed. Namely, the participants' performance was generally improved when they used the new version of prototype design compared to their performance when they used the old version. This performance advantage appeared particularly when considering the ICT of the Automated Assistant Decision-Making (AADM) feature.

The main rationale for including this feature within the iterative prototype design was to improve the system usability by accelerating the process of decision making, which needs to be undertaken to perform and to complete the main tasks in an easier, faster and more accurate manner. In addition, the qualitative outcomes described other possible advantages that would be gained when this feature is considered for the actual implementation. One example is to overcome the problem of examining varying decisions for issuing Visas regarding one specific case to increase confidence in providing accurate decisions because the system would keep up-to-date with the latest instructions and regulations. Hence, the most important two aspects that can be deduced in using AADM were firstly that it motivated the participants to improve their performance and to provide better service to the applicants for Visas. Secondly, the possibility of taking lessons from other successful examples, such as the Best Practice Guide, which was developed by the Australian Government (2007), was provided; however, a number of challenges may face the developer when the AADM feature is considered within the system functions, such as evolving more accurate and consistent measures for the evaluation and review of the decision taken (ibid).

As a final point, it could be said that the core findings of this thesis illustrate that there is a need for implementing an HCD approach for developing government systems to consider the users' points of view and requirements of the targeted system and to support its development process. In addition, in respect to our case in targeting the current VI system, the research work was circulated to be presented as guidance for developing the usability of government systems.

Finally, yet importantly, most of the participants gave their full co-operation and welcomed participation in this project. After they interacted with the prototype design and discovered that most of their responses, opinions and suggestions were taken into account, they showed considerable optimism in providing better performance in their work when this re-designed system is entirely implemented as the actual final system. This issue is considered a vital factor and fulcrum that would help this research study achieve its main goals.

The results demonstrate that user involvement in the development of information systems will enhance both system usage and the user's satisfaction with the system. Further, the study provides evidence that the user's satisfaction with the system will lead to better system usage.

7.3 Research questions evaluation

The evaluation of the research work conducted in this thesis confirms that the main research question presented in chapter 1, section (1.3) which motivated the thesis, was answered: How can the existing usability methods, techniques and tools be utilized, improved and integrated to enhance the design of current internal government systems? It has been answered by illustrating how the developed methodological framework demonstrates the benefit of the combination of these exiting techniques, procedures and tools in the three studies presented in chapters 4,5 and 6.

In addition, there have been six sub-research questions under investigation in this study.

7.3.1 Research question one (RQ1)

What are the usability problem areas of a current VI system that have influenced the UX?

To identify the existing usability problems in a current government system, a query evaluation technique was employed in this research, which was based on actual users' perceptions. Thus, the data were collected through the closed questions of the developed questionnaire and the open-ended questions of the semi-structured interviews. The findings indicated that a number of usability problems have been found in the target government system. These problems are summarized in Table (4-21) in chapter 4.

7.3.2 Research question two (RQ2)

What is the usability state of the current VI system based on its users' perspectives?

To investigate this question, seven dimensions were considered within the questionnaire (Effectiveness, Efficiency, Easy to learn, Error tolerance, Engaging (Satisfaction), Helpfulness and Control) for measuring the current system usability from the users' viewpoints, which provided formative information that could help to achieve a comprehensive vision regarding the usability of a current VI system. The findings indicated that the dimension of Efficiency was the critical factor that needed to be focused on to achieve the aim of the research.

7.3.3 Research question three (RQ3)

How should the current VI system be re-designed to produce a prototype design?

To answer this question, the additional user requirements that were identified from the first study were considered as a design solution and basis for developing the simulation design of the VI system. In addition, the best practices and good UX that the current (legacy) system already provides were transferred, such as the primary design of the workflow. The findings are displayed in the four main screens in Figures (5-1 to 5-4) after the prototyping method was utilized.

7.3.4 Research question four (RQ4)

What are the effects of the proposed usability design solutions on the UX?

To examine the effects of the proposed prototype design on the current VI system, the compression usability test was applied as experiment 1, followed by conducting the post-study questionnaire. In addition, the performance measurement was applied by using different performance criteria, such as task time, task completion rate and error rate, for distinguishing the differences between the current government system and the prototype design. The findings indicated that the developed prototype design DPD, as a proposed design solution, improved and addressed the usability problems found in the first study. For example, the participants could finish Task5, as the main test task, with less time using DPD, as mentioned in the previous section.

7.3.5 Research question five (RQ5)

How should the UI of prototype design of the VI system be re-designed for the actual users of the VI system?

To investigate this question, the findings of experiment 1 can be used, which revealed the improvement on the proposed design solution and the additional requirements, such as the need for enhancing the efficiency of decision making for issuing Visas. Accordingly, the iterative design was applied for re-designing the DPD by involving additional features, such as utilizing AADM as an ICT design solution. The findings were presented with some changes and developments in the previous DPD (V1), as shown in Figures (6-4) to (6-8).

7.3.6 Research question six (RQ6)

What are the effects of the proposed re-designed prototype with additional features on the UX?

To determine the effects of re-designing the proposed design solutions, another usability compression test was applied (experiment 2) with the same performance criteria used in experiment 1, which included calculating the time spent finishing test tasks, the number of successful tasks completed and the error rate. The findings revealed that the participants' performances improved in terms of time consumed finishing the three test tasks of experiment 2. Similarly, the participants completed more test tasks successfully using DPD/V2 than using DPD/V1. Therefore, the UX evolved with the improvement of the participants' interactions with each proposed redesigned prototype.

7.4 Main contributions of the research

First, the research work reported in this thesis has evaluated the usability of an existing government system in Saudi Arabia. This usability evaluation has identified a number of usability problems in the target government system. Based on these identified problems, which were considered additional requirements in this research, the proposed design solutions were developed to enhance the usability of the current government system. The final results indicated that the DPD could resolve the generated usability problems, which leads to better user interaction with the government systems and results in the improvement of the UX. According to these research outcomes, the necessity of considering the usability of government systems is illustrated, and the outcomes also identified the essential improvement of usability in the current government system. In this context, this research has developed a set of usability guidelines for developing and re-designing current government systems. Subsequently, this research has made contributions in three areas of knowledge, usability, specifically usability engineering (UE), user experience (UX) and knowledge regarding the development of an existing (Legacy) government system.

7.4.1 Contribution to knowledge regards to usability and UE

According to the usability aspects considered in the first study based on the five usability attributes (5E), Efficiency, Effectiveness, Easy to learn, Error tolerance and Engaging', presented by Quesenbery (2004), for assessing the current state of usability of the target government system, this research has extended the usability dimension by adding two further aspects, 'control' and 'helpfulness', to achieve wider usability insight and to be appropriate for the nature of the context and aims of this study. The collected data from these two aspects was analysed, and it confirmed the significant results related to the system usage. This usability extension could provide a deeper understanding regarding the usability of the current government system, which enables the involvement of proper methods for improvement.

For the selected main approach, the research has made a contribution to the consideration of a Human Centred Design (HCD) approach for involving users in the process of re-designing the current government system in a developing country context to enhance its usability. Therefore, the research provides concrete guidance for the elicitation and validation of particular techniques and tools and how these techniques and tools can be used to support the re-design of usable government systems.

In addition, the final version of the proposed prototype design, which was developed based on the perspectives of actual users and represented their needs through the involvement of ICT functions, could be used as a typical solution design for the identified problems and issues that are associated with the current VI and could be presented to system developers. This developed prototype design was empirically confirmed and validated by comparing it with the current VI system in the first experiment as well as in the second experiment with the old prototype version, which was iteratively developed and re-designed.

7.4.2 Contribution to knowledge regarding UX

This study has contributed to the field of UX by employing and improving two study questionnaires for measuring user satisfaction and understanding their preferences regarding the VI system in a Saudi Arabian government context. Thus, the obtained results from both questionnaires of the study indicated a significant improvement and success of the suggested prototype as a re-design solution for usability problems of the existing VI government system. This favourable outcome confirms that the developed methodological framework is an exemplar for integrating existing techniques and tools for enhancing system usability in a government context as well as improving the UX.

Based on achieving the aims of study by utilizing the selected UEMs, which were the most effective in dealing with the current usability problem areas of the existing government system, a methodological framework was developed and provided. This methodological framework was tested and validated by conducting the three studies for the research. The findings clarify the effectiveness of the selected UEMs for conducting the evaluation and for improving the design of the current government systems in a Saudi Arabian context, which would help to create a better information society. For instance, from the employees side, as users for the target system of government origination, would get an advantage to improve their experience for finishing the assigned tasks. Which in turn, enhance the general informational performance offered by the administrations of the specific government organization, which ultimately reflects the improved services provide to the citizens. Consequently, it could be generalized for governments in other developing countries or for research in different areas of the same fields of study.

7.4.3 knowledge regards an existing (Legacy) government system development.

According to the review of the related existing literature, this research study is considered to be one of the rare studies that was conducted in the domain of usability in current government systems by employing a combination of different techniques, procedures and tools. Furthermore, this research focuses particularly on a rapidly developing country in which the government seeks to improve its systems in various sectors to provide better services to its citizens and other residents in addition to attempting to change the previous negative image of the performance of its systems in general. Therefore, presenting the research work in this study as guidance would be a useful achievement that can contribute specifically to the development process of the current government systems in a developing country. Moreover, an important consideration in determining that this research study could be used as guidance that can be provided to government system developers is that it is easy to understand and the implementation process costs less. This aspect is considered to be a key contribution to this research, and because the obtained results indicated the improvement of real users' performances, optimism regarding this research has spread among the employees who were the participants for conducting this research. Accordingly, an administration official in this government organization pointed to the significant possibility of making a recommendation to present this study to system developers to be taken into consideration for the next development cycle.

7.5 Research limitation

As with any research, this study has a number of limitations that could affect the findings obtained and limit the research capabilities as a whole.

- Considering the query evaluation technique through using a questionnaire and an interview is an initiative method for collecting information from participants and obtaining their perspectives of a current VI system was useful in identifying a number of critical problems within the system usability; however, as with any other method, this technique has some disadvantages that could affect the quality of the collected data. For instance, due to the participants being actual end users, they provided the related information subjectively based on their experiences of using the target system, but they could not reveal the detailed technical errors that experts could.
- A noticeable limitation of this research is one of the sample characteristics in which all recruited participants for conducting this research were male due to there being no female employees in this government department who were users of the current VI system who could be involved in the study at that time. Therefore, the perception of the female gender was not included, and thus the result did not reflect a female perspective. The participation of females would have further contributed to the results of the study.
- Another limitation that could be related to the sample characteristic is that all participants were non-English language speakers. Subsequently, the questions in the study questionnaires and semi-structured interviews in addition to the important information that needed to be delivered to the participants to perform the test tasks in the two experiments reported in chapters five and six had been translated into the participants' language, which was the Arabic language. After that, the responses and gathered information were translated back into English for the purpose of analysis to achieve the research objectives, as the back translation strategy suggested by Brislin (1970) was applied in the study. Although this strategy would help in generating a more in-depth understanding of the versions encompassing the target language, the poor translation in some cases of the research may have led to losing some sense of the true meaning of the words that needed to be translated. Moreover, it consumed time, which was one constraint on this research.

- For the first study reported in chapter four, there was a limitation regarding the different intensity of workloads between the different offices in which the participants were recruited, as the system users were distributed in different areas. For example, most of the heavy workloads as well as the crowd of applicants, were concentrated in the main branches of this government organisation at the major regions, e.g. eastern and central regions, which would have a negative impact on the employees' opinions in these offices as users of the VI system. Because they would have more cases and so the systems' issues affect their performance more.
- For the experiment studies that are reported in chapters five and six that were conducted in an informal laboratory in a government work environment, the research was exposed to other challenges and limitations. Some examples include the difficulty in getting approval to access the targeted sample, finding available and proper places to hold the test and conducting a quiet test session with a noisy background due to many applicants being present. This created changes in the test settings, and there was then the difficulty of getting another approval for conducting the test sessions in a quiet office. Accordingly, this consumed time from the setup to conducting the study, which in turn impacted the research plan.
- The study focused on the target VI system, which is integrated with other systems, such as the immigration systems and bank systems. Thus, regarding the suggested P.D, which was built with high fidelity using Microsoft Access to create dummy data, it could not cover all aspects that represented the actual protocol of the system workflow, which might have led to the participants losing a sense of realism while they performed the tasks using the P.D. Therefore, in interpreting the test results, it should also be kept in mind that the prototype was just an initial prototype and was missing many features, such as those mentioned.
- Tests are almost always artificial situations. Even though we tried to conduct the test in the informal lab within the same workplace, it still only represented a mock-up of the actual situation of performing the tasks.

7.6 Chapter Summary

This chapter has discussed the main findings of the research work presented in thesis that was carried out to achieve the aim of this PhD research. A thorough discussion of the main findings has been presented. In addition, the main contribution of this research has been illustrated. Finally, the limitations and suggestions for future work were discussed.

References

Abdul Rauf, B., ShaifqurRehman, S. and Shahid, A. (2010) 'Survey Based Usability Evaluation of MS WORD', *International Conference on User Science Engineering.* , 193-198.

Abran, A., Khelifi, A., Suryn, W. and Seffah, A. (2003) 'Usability meanings and interpretations in ISO standards', *Software Quality Journal,* 11(4), pp. 325-338.

Abras, C., Maloney-Krichmar, D. and Preece, J. (2004) 'User-Centered Design', *Bainbridge, W.Encyclopedia of Human-Computer Interaction.Thousand Oaks: Sage Publications,* 37(4), pp. 445-456.

Acock, A.C. (2008) *A gentle introduction to Stata.* Stata press.

Adams, A. and Cox, A.L. (2008) 'Questionnaires, in-depth interviews and focus groups', .

Agner, L., Tavares, P. and Ferreira, S.B.L. (2011) 'Scenario and Task Based Interview to Evaluate Usability of Computer Assisted Data Collection', in *Design, User Experience, and Usability. Theory, Methods, Tools and Practice.* Springer, pp. 349-358.

Al-Baydani,M.(2011), 'Slow procedures cause congestion on the Labour Office in Al-Baha', *Al-madina16 Augast.* Available from: <http://www.al-madina.com/node/321565>**.** [22 February 2012].

Al-Edwani,S.(2010), 'The applicants are complaining from long waiting: they demand of increasing the number of employees', *Al-Madina11 Novmber.* Available from: <http://www.al-madina.com/node/273305>. [18Janurary 2013].

Al-Harthi,E.(2008), 'The applicants congestion at the recruitment office in Jeddah delay transactions', *Okaz 03Augast.* Available from: <http://www.okaz.com.sa/okaz/osf/20080803/Con20080803214227.htm>. [19 February 2012].

Al-Motaryi, M. (2012) '' The permitting for married women to recruit workers - Six thousand Saudi women recruited domestic workers '', *Al Eqtisadiah Newspaper 7October.* Available from :<http://www.aleqt.com/2012/10/07/article_699370.html.> [04 March 2014]

Al-Shafi, S. and Weerakkody, V. (2010) 'Factors affecting e-government adoption in the state of Qatar'. *In the European and Mediterranean Conference on Information Systems (EMCIS) Abu Dhabi,UAE.*

Arnold, D.M., Burns, K.E., Adhikari, N.K., Kho, M.E., Meade, M.O., Cook, D.J. and McMaster Critical Care Interest Group (2009) 'The design and interpretation of pilot trials in clinical research in critical care', *Critical Care Medicine,* 37(1 Suppl), pp. S69-74.

Atlantis Technologies (2002) Benefits of Reengineering Legacy System, *The evolution of competitive advantage, available online at: http://www.atlantistechgroup.com/Resources/Migration/Benefits.htm* (09 September 2015)

Australian Government (2007) Automated Assistance in Administrative Decision-Making (Better Practice Guide), *Commonwealth Ombudsman*, *available online at: http://www.ombudsman.gov.au/docs/better-practice-guides/aaadm_guide.pdf (10 Augst 2013)*

Aziz, N.S. and Kamaludin, A. (2014) ' Assessing Website Usability Attributes Using Partial Least Squares', *International Journal of Information and Electronics Engineering,* 4(2), pp. 137-144.

Babbie, E. (1990) *Survey Research Methods.* 2nd edn. CA: Belmont:Wadswroth Publishing.

Bailey,T. (1993) '' Discretionary Effort and the Orgainization of Work: Employee Participantion and Work Reforms since Hawthorn, Columbia University, New York, Working Paper.

Barnum, C.M. (2010) *Usability testing essentials: ready, set... test!* Elsevier.

Barnum, C.M. (2002) *Usability testing and research.* New York: Longman.

Bastien, J.C. (2010) 'Usability testing: a review of some methodological and technical aspects of the method', *International journal of medical informatics,* 79(4), pp. e18-e23.

Berkun, S. (2014) *The Art of UI Prototyping*. Available at: http://msdn.microsoft.com/en-us/library/ms993294.aspx (Accessed: 05/10 2013).

Bevan, N., Kirakowski, J. and Maissel, J. (1991) 'What is usability', *Proceeding of the Fourth International Conference on HCI,Stutgart.*

Bevan, N. (2009) 'Extending quality in use to provide a framework for usability measurement', in *Human Centered Design.* Springer, pp. 13-22.

Bevan, N. and Schoeffel, R. (2001) 'A proposed standard for consumer product usability.', *HCI.* , 557-561.

Bevan, N. (1995) 'Measuring usability as quality of use', *Software Quality Journal,* 4, pp. 115-150.

Bhattacherjee, A. (2001) 'Understanding information systems continuance: an expectation-confirmation model', *MIS quarterly,* , pp. 351-370.

Bianchi,A., Caivano,D. Marengo,V. and Visaggio,G. (2003) "Iterative Reengineering of Legacy Systems," *IEEE Transactions on Software Engineering*, vol. 29, pp. 225- 241

Bias, R. (1991) 'Interface-Walkthroughs: efficient collaborative testing', *Software, IEEE,* 8(5), pp. 94-95.

Bias, R.G. and Mayhew, D.J. (2005) *Cost-justifying usability: an update for an Internet age.* Morgan Kaufmann.

Bland, M. (2000) *An introduction to medical statistics.* Oxford University Press.

Blandford, A.E., Hyde, J.K., Green, T.R. and Connell, I. (2008) 'Scoping analytical usability evaluation methods: a case study', *Human–Computer Interaction,* 23(3), pp. 278-327.

Bokhari, R. (2005)" The relation between system usage and user satisfaction", *The Journal of Enterprise Information Management* 18 (2), pp. 211-234

Boothe, C., Strawderman, L. and Hosea, E. (2013) 'The effects of prototype medium on usability testing', *Applied Ergonomics,* 44(6), pp. 1033-1038.

Bowman, D.A., Gabbard, J.L. and Hix, D. (2002) 'A survey of usability evaluation in virtual environments: classification and comparison of methods', *Presence: Teleoperators and Virtual Environments,* 11(4), pp. 404-424.

Brinck, T., Gergle, D. and Wood, S.D. (2002) *Designing Web sites that work: Usability for the Web.* Morgan Kaufmann.

Bryman, A. and Cramer, D. (2009) *Quantitative data analysis with SPSS 14, 15 and 16: A guide for social scientists.* Routledge New York.

Buchenau, M. and Suri, J.F. (2000) 'Experience prototyping', *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques.* ACM, 424-433.

Buie, E. and Murray, D. (2012) *Usability in Government Systems.* 1st edn. Elsevier

Bygstad, B., Ghinea, G. and Brevik, E. (2008) 'Software development methods and usability: Perspectives from a survey in the software industry in Norway', *Interacting with Computers,* 20(3), pp. 375-385.

Cairns, P. and Cox, A.L. (2008) *Research methods for human-computer interaction.* Cambridge University Press New York, NY, USA.

Carr, M. and Verner, J. (1997) 'Prototyping and software development approaches', *Department of Information Systems, City University of Hong Kong, Hong Kong,* .

Amelia, A. and Carvalho, A. (2002) Usability Testing of Educational Software: Methods, Technique and evaluators.University of Minho,

Casaló, L.V., Flavián, C. and Guinalíu, M. (2010) 'Generating trust and satisfaction in e-services: the impact of usability on consumer behavior', *Journal of Relationship Marketing,* 9(4), pp. 247-263.

Catarci, T., Dongilli, P., Di Mascio, T., Franconi, E., Santucci, G. and Tessaris, S. (2004) 'An ontology based visual tool for query formulation support', *ECAI.* , 308.

Church,S. (2013) Usability and User experience, Jisc, UK, Available online at https://www.jisc.ac.uk/guides/usability-and-user-experience (Accessed 18 September 2015)

Cockton, G. (2012). *Usability Evaluation*. In Soegaard, M. & Friis, R. (Ed.), *Encyclopedia of Human-Computer Interaction*. Aarhus: The Interaction Design Foundation.

Costabile, M.F. (2001) 'Usability in the software life cycle', *Handbook of software engineering and knowledge engineering,*(1), pp. 179-192.

Crosby, P. and Spyridakis, J. (1993) 'A primer on usability testing for developers of traveler information systems', *Transpn. Res.-C,* 1(2), pp. 143-157.

Daouk, M. and Leveson, N.G. (2001) 'An approach to human-centered design', *Workshop Human Error and System Development, Linkoping, Sweden.*

Dix, A. , Finlay,J. , Abowd, G. and Beale, R. (2004) *Human-Computer Interaction*. 3rd ed. Pearson Education Limited , England.

Donahue, G.M. (2001) 'Usability and the bottom line', *IEEE Software,* 18(1), pp. 31-37.

Druzdzel, M.J. and Flynn, R.R. (1999) 'Decision support systems. Encyclopedia of library and information science. A. Kent', *Marcel Dekker, Inc.Last Login,* 10(03), pp. 2010.

Dubey, S.K., Gulati, A. and Rana, A. (2012) 'Usability Evaluation of Software Systems using Fuzzy Multi-Criteria Approach', *IJCSI International Journal of Computer Science,* 9(3), pp. 404-409.

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

Durbin, C.G.,Jr (2004) 'How to come up with a good research question: framing the hypothesis', *Respiratory care*, 49(10), pp. 1195-1198.

Downey, L., and Rosales,S. 2011, 'A brief history of user experience in government system', in E. Buie & D Murray (eds),Usability in government systems Morgan kaufmann , USA, pp. 1-15.

Ehn, P. (1989). *Word-oriented design of computer artifacts* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Faulkner, L. (2003) 'Beyond the five-user assumption: Benefits of increased sample sizes in usability testing', *Behavior Research Methods, Instruments, & Computers,* 35(3), pp. 379-383.

Fernandez, A., Insfran, E. and Abrahão, S. (2011) 'Usability evaluation methods for the web: A systematic mapping study', *Information and Software Technology,* 53, pp. 789-817.

Ferré, X., Juristo, N., Windl, H. and Constantine, L. (2001) 'Usability basics for software developers', *IEEE Software,* 18(1), pp. 22-29.

Fidgeon, F. (2014) *Usability Testing: How to statistically compare the usability of different designs*. Available at: http://www.spotless.co.uk/insights/comparing-usability (Accessed: 03/19 2014).

Finlay, P. (1994) *Introducing decision support systems.* 2nd edn. Manchester: NCC/Blackwell.

Finstad, K. (2010) 'The Usability Metric for User Experience', *Interacting with Computers, Elsevier B.V.,* 22, pp. 323-327.

Fischer, G. (1991) 'Supporting learning on demand with design environments', *Proceedings of the International Conference on the Learning Sciences.* Citeseer, 165-172.

Fischer, G. and Ostwald, J. (2002) Seeding, evolutionary growth, and reseeding: enriching participatory design with informed participation , T. Binder, J. Gregory, I. Wagner (Eds.), Proceedings of the Participatory Design Conference (PDC 2002), CPSR, Malmö University, Sweden, pp. 135–143.

Folmer, E. and Bosch, J. (2004) 'Architecting for usability: a survey', *Journal of Systems and Software*, 70(1), pp. 61-78.

Fowler Jr, F.J. (2008) *Survey research methods.* Sage publications.

Ferre,X. and Medinilla,N.(2007)"How a Human-Centred Approach Impacts Software Development", *J.Jacko(Ed.):Human-Computer Interaction, Part I,* LNCS 4550,pp 68-77.

Freiberg, M. and Baumeister, J. (2008) 'A survey on usability evaluation techniques and an analysis of their actual application', *Institute of Computer Science, University of Wurzburg, Germany,* .

Frøkjær, E., Hertzum, M. and Hornbæk, K. (2000) 'Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?', *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* ACM, 345-352.

Galitz, W.O. (2007) *The Essential Guide to User Interface Design.* Third edn. Canada: Wiley Publishing, Inc.

Gediga, G., Hamborg, K. and Düntsch, I. (2002) 'Evaluation of software systems', *Encyclopedia of computer science and technology,* 45(supplement 30), pp. 127-153.

Gediga, G., Hamborg, K. and Düntsch, I. (2002) 'Evaluation of Software Systems', *In: A. Kent and J. G. Williams, eds. Encyclopedia of Computer Science and Technology,* 44, pp. 162-196.

George, D. (2003) *SPSS for windows step by step: A simple study guide and reference, 17.0 update, 10/e.* Pearson Education India.

George, C. (2005) 'Usability testing and design of a library website: An iterative approach', *OCLC Systems & Services,* 21(3, 2005), pp. 167-180.

Gil-García, J.R. and Pardo, T.A. (2005) 'E-government success factors: Mapping practical tools to theoretical foundations', *Government Information Quarterly,* 22(2), pp. 187-216.

Glass, R.L. (2002) *Facts and fallacies of software engineering.* Addison-Wesley Professional.

Gliem, J.A. and Gliem, R.R. (2003) 'Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales', Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

Goldsmith, S. and Eggers, W.D. (2005) *Governing by network: The new shape of the public sector.* Brookings Institution Press.

González, M.P., Lorés, J. and Granollers, A. (2008) 'Enhancing usability testing through datamining techniques: A novel approach to detecting usability problem patterns for a context of use', *Information and software technology,* 50(6), pp. 547-568.

Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM, 28*(3), 300-311.

Greenbaum, J., and Kyng, M. (1991). *Design at work: Cooperative design of computersystems*. Hillsdale NJ: Erlbaum.

Grönlund, Å. and Horan, T.A. (2005) 'Introducing e-gov: history, definitions, and issues', *Communications of the Association for Information Systems,* 15(1), pp. 39.

Gulliksen, J., Harning, M.B., Palanque, P., van der Veer, G.C., Wesson, J. (2008) Engineering Interactive Systems , *EIS, Joint Working Conferences EHCI*, DSV-IS 2007, HCSE , Salamanca, Spain, 22-24 (Eds.).

Hamborg, K., Vehse, B. and Bludau, H. 'Questionnaire Based Usability Evaluation of Hospital Information Systems', *Electronic Journal of Information Systems Evaluation,* , pp. 21-30.

Han, S.H., Yun, M.H., Kwahk, J. and Hong, S.W. (2001) 'Usability of consumer electronic products', *International Journal of Industrial Ergonomics,* 28(3), pp. 143-151.

Harms, I. and Schweibenz, W. (2000) 'Testing Web Usability', *Information Management & Consulting,* 15(3), pp. 61-66.

Hartson, H.R., Andre, T.S. and Williges, R.C. (2001) 'Criteria for evaluating usability evaluation methods', *International Journal of Human-Computer Interaction,* 13(4), pp. 373-410.

Hassenzahl, M. (2008) 'User experience (UX): towards an experiential perspective on product quality', *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine.* ACM, 11-15.

Hegarty, R. and Wusteman, J. (2011) 'Evaluating EBSCO host Mobile', *Library Hi Tech,* 29(2), pp. 320-333.

Hinderer, D. (1998) 'Challenges in participant recruiting for usability testing', *Professional Communication Conference, 1998. IPCC 98. Proceedings. 1998 IEEE International.* IEEE, 417-426.

Hix, D., Gabbard, J.L., Swan, J., Livingston, M.A., Hollerer, T., Julier, S.J., Baillot, Y. and Brown, D. (2004) 'A cost-effective usability evaluation progression for novel interactive systems', *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on.* IEEE, 10 pp.

Hix, D. and Hartson, H.R. (1993) *Developing user interfaces: ensuring usability through product & process.* John Wiley & Sons, Inc.

Hoegh, R.T., Nielsen, C.M., Overgaard, M., Pedersen, M.B. and Stage, J. (2006) 'The impact of usability reports and user test observations on developers' understanding of usability data: An exploratory study', *International Journal of Human-Computer Interaction,* 21(2), pp. 173-196.

Holzinger,A. (2005)‟Usability Engineering Methods for Software Developers‟, *Communications of the ACM,* 48(1),pp 71-74

Horsky, J., Schiff, G.D., Johnston, D., Mercincavage, L., Bell, D. and Middleton, B. (2012) 'Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions', *Journal of Biomedical Informatics,* 45(6), pp. 1202-1216.

Hsieh, S. and Huang, S. (2008) 'Usability Evaluation: A Case Study', *APIEMS 2008 Proceedings of the 9th Asia Pacific Industrial Engineering & Management Systems Conference.*

Hub, M. and Čapková, V. (2010) 'Heuristic evaluation of usability of public administration portal', DEO, Narsingh, et al. APPLIED COMPUTER SCIENCE: Procedings of International Conference on Applied Computer Science (ACS). Malta: WSEAS Press.

Hub, M. and Zatloukal, M. (2008) 'Methodology of fuzzy usability evaluation of Information systems in Public administration', *WSEAS Transactions on Information Science & Applications,* 5(11), pp. 1573-1583.

ISO DIS 9241-210:2010. *Ergonomics of human system interaction -Part 210: Human-centred design for interactive systems*(formerly known as 13407). International Standardization Organization (ISO). Switzerland.

Ivory, M.Y. and Hearst, M.A. (2001) 'The state of the art in automating usability e valuation of user interfaces', *ACM Computing Surveys (CSUR),* 33(4), pp. 470-516.

Johanson,C.and Johanson,T. and Zhang,J. (2005) A user-centered framework for redesigning health care interfaces, *Journal of Biomedical Informatics* (38) 75rna.

Jordan, P.W. (1998) *An introduction to usability.* CRC Press.

Juristo, N. (2009) 'Impact of usability on software requirements and design', in *Software Engineering.* Springer, pp. 55-77.

Juristo, N., Moreno, A.M. and Sanchez-Segura, M. (2007) 'Analysing the impact of usability on software design', *Journal of Systems and Software,* 80(9), pp. 1506-1516.

Juristo, N., Moreno, A.M. and Sanchez-Segura, M. (2007) 'Analysing the impact of usability on software design', *Journal of Systems and Software,* 80(9), pp. 1506-1516.

Juristo, N., Moreno, A.M. and Sanchez-Segura, M. (2007) 'Guidelines for eliciting usability functionalities', *Software Engineering, IEEE Transactions on,* 33(11), pp. 744-758.

Kabli,F.(2013), 'Recruitment office in Jeddah', *Al-madinah 03Augast.* Available from: <http://www.al-madina.com/node/284824>. (Accessed 12March 2014).

Kamarck, E.C. (2007) 'The End of Government', *As We Know It: Making Public Policy Work.Boulder, CO,* .

Karray,F., Alemzadeh,M. , Abou Saleh, J. and Arab,M. (2008) ' Human Computer Interaction : Overview on State of the Art' *International Journal on Smart Sensing and Intelligent System*, 1(1),pp 137-159.

Karwowski, W. (2006) 'The discipline of ergonomics and human factors', *Handbook of human factors and ergonomics,* 3.

Kesseler, E. and Knapen, E.G. (2006) 'Towards human-centred design: Two case studies', *Journal of Systems and Software,* 79(3), pp. 301-313.

Kensing,F. and Blomberg,J. (1998) Participatory Design : Issues and Concerns, *Computer Supported Cooperative Work (*7)pp.167–185.

Kies, J.K., Williges, R.C. and Rosson, M.B. (1998) 'Coordinating computer-supported cooperative work: A review of research issues and strategies', *Journal of the American Society for Information Science,* 49(9), pp. 776-791.

Kirakowski, J. and Corbett, M. (1993) 'SUMI: The software usability measurement inventory', *British journal of educational technology*, 24(3), pp. 210-212.

Kiraakoswski, J. (2000) ' Questionnaires in Usability Engineering', Human Factors Research Group, Cork, Ireland, 3rd ED.

Kirwan, B. and Ainsworth, L.K. (1992) *A guide to task analysis: the task analysis working group.* CRC press.

Kuniavsky, M. (2003) *Observing the user experience: a practitioner's guide to user research.* Morgan kaufmann, USA

Kuniavsky, M. (2010) *Smart Things: Ubiquitous Computing User Experience Design* . Morgan kaufmann, USA.

Kyng,M. (1991) Designing for cooperation: cooperating in design, Communicating of the ACM, 34(12),pp. 65-73.

Landauer, T.K. (1995) *The trouble with computers: Usefulness, usability, and productivity.* Taylor & Francis.

Lauesen, S. (2005) *User interface design: a software engineering perspective.* Pearson Education.

Law, E.L., Roto, V., Hassenzahl, M., Vermeeren, A.P. and Kort, J. (2009) 'Understanding, scoping and defining user experience: a survey approach', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 719-728.

Lazar, J., Feng, J.H. and Hochheiser, H. (2010) *Research methods in human-computer interaction.* John Wiley & Sons.

Lazar, J. (2005) *Web usability: A user-center design approach.* Boston, MA: Pearson Addison Wesley.

Le Peuple, J. and Scane, R. (2003) *User interface design.* Crucial.

Leavitt, M.O. and Shneiderman, B. (eds.) (2006) *Research-Based Web Design and Usability Guidelines.* 2nd edn. Washington, DC: GSA.

Lecerof, A. and Paternò, F. (1998) 'Automatic support for usability evaluation', *Software Engineering, IEEE Transactions on*, 24(10), pp. 863-888.

Lewis, C. and Wharton, C. (1997) 'Cognitive walkthroughs', *Handbook of human-computer interaction,* 2, pp. 717-732.

Lewis, J.R. (2006) 'Usability testing', *Handbook of human factors and ergonomics,* 12, pp. e30.

Lewis, P., Thornhill, A. and Saunders, M. (2007) *Research methods for business students.* Pearson Education UK.

Lindgaard, G. and Chattratichart, J. (2007) 'Usability testing: What have we overlooked?', *Proceedings of the SIGCHI Conference on Human factors in Computing Systems.* ACM, 1415-1424.

Little, D. (2014) *User Interface Design*. Available at: http://www.slideshare.net/littleddotnet/user-interface-design-definitions-processes-and-principles-15366100 (Accessed: 11/16 2014).

Love, S. (2005) *Understanding mobile human-computer interaction.* Butterworth-Heinemann.

Lund, A. (2001) *Measuring Usability with the USE Questionnaire.STC Usability SIG Newsletter, .*

Macleod, M. (1994) 'Usability in context: Improving quality of use', *Human Factors in Organizational Design and Management–IV (Proceedings of the International Ergonomics Association 4th International Symposium on Human Factors in Organizational Design and Management, Stockholm.*

Madan, A. and Dubey, S.K. (2012) 'Usability evaluation methods: a literature review', *International Journal of Engineering Science and Technology,* 4(2), pp. 590-599.

Maguire, M. (2001) 'Context of Use within usability activates', *Int. J. Human-Computer Studies* (55), pp. 453-483.

Manresa-Yee, C., Ponsa, P., Varona, J. and Perales, F.J. (2010) 'User experience to improve the usability of a vision-based interface', *Interacting with Computers,* 22(6), pp. 594-605.

Maybury, M.T. (2012) 'Usable advanced visual interfaces in aviation', *Proceedings of the International Working Conference on Advanced Visual Interfaces.* ACM, 2-3.

Mayhew, D.J. (1999) 'The usability engineering lifecycle', *CHI'99 Extended Abstracts on Human Factors in Computing Systems.* ACM, 147-148.

Mayhew, R. and Media,D (2015) Examples of Cultural Diffrences in the Workplace, Hearst Newspapers, Houston Chronicle, Available at: http://www.chron.com/about/newsroom(Accessed: 13/09/ 2015).

Mayo, P. (2013) *Orgnization support system: DSSs , GDSSs , EISs , Groupware, Multimedia*. Available at: http://www.slideshare.net/philmayo/organization-support-systems (Accessed: 14/10/ 2013).

Mazumder, K. and Das, U. (2014) ' USABILITY GUIDELINES FOR USABLE USER INTERFACE', *IJRET: International Journal of Research in Engineering and Technology*, 03(09), pp. 70-82.

Metzker, E. and Offergeld, M. (2001) Integrating Human-Centered Design Methods into Development Process, in Little, M. and Nigay,L. (Eds.) Engineering for Human-Computer Interaction, 8th IFIP International Conference, EHCI,pp. 19-33

McDonald, J.H. (2014) *Handbook of biological statistics.* 3rd edn. Maryland: Sparky House Publishing Baltimore.

McGuffin, M.J. and Balakrishnan, R. (2005) 'Fitts' law and expanding targets: Experimental studies and designs for user interfaces', *ACM Transactions on Computer-Human Interaction (TOCHI),* 12(4), pp. 388-422

Microsoft, C. (2015) *Usability in Software Design*. Available at: https://msdn.microsoft.com/en-us/library/ms997577.aspx (Accessed: 01/19 2014).

Microsoft, C. (2013) *UI Guidelines vs. Usability Testing*. Available at: https://msdn.microsoft.com/en-us/library/ms997578.aspx (Accessed: 05/08 2013).

Moore, G.C. and Benbasat, I. (1991) 'Development of an instrument to measure the perceptions of adopting an information technology innovation', *Information systems research,* 2(3), pp. 192-222.

Moorthy, J.S., bin Ibrahim, S. and Mahrin, M.N. (2014) 'Developing Usable Software Product Using Usability Risk Assessment Model', *International Journal of Digital Information and Wireless Communications (IJDIWC),* 4(1), pp. 95-102.

Muller, M.J., and Kuhn, S. (Eds.) (1993). *Communications of the ACM* special issue on participatory design, **36**(6), June 1993.

Narasimhadevara, A., Radhakrishnan, T., Leung, B. and Jayakumar, R. (2008) 'On designing a usable interactive system to support transplant nursing', *Journal of Biomedical Informatics,* 41(1), pp. 137-151.

Nebe,K. & Zimmermann, D. (2007) Aspects of Integrating User Centered Design to Software Engineering Processes. In: Jacko, J.A. (ed.) HCI, LNCS,vol.4550,Springer,Heidelberg.

Newsom (2008) *Simple Within-Subjects Tests.* USP534 Data analysis.

Nielsen, J. (2015) *Agile User Experience Projects*. Available at: http://www.nngroup.com/articles/agile-user-experience-projects/ (Accessed: 04/10 2014).

Nielsen, J. (2015) *Usability Metrics*. Available at: http://www.nngroup.com/articles/usability-metrics/ (Accessed: 09/12 2012).

Nielsen, J. (2001) 'Success rate: the simplest usability metric', *Jakob Nielsen's Alertbox,* 18.

Nielsen, J. (1999) *Designing web usability: The practice of simplicity.* New Riders Publishing.

Nielsen, J. (1995) 'Usability inspection methods', *Conference companion on Human factors in computing systems.* ACM, 377-378.

Nielsen, J. (1994) 'Usability inspection methods', *Conference companion on Human factors in computing systems.* ACM, 413-414.

Nielsen, J. (1993) 'Iterative user-interface design', *Computer,* 26(11), pp. 32-41.

Nielsen, J. (1993) *Usability engineering.* Cambridge, MA: Academic Press Inc.

Nielsen, J. and Molich,R. (1990) 'Heuristic evaluation and user interface', *ACM,* pp. 249-256.

Och Dag, J.N., Regnell, B., Madsen, O.S. and Aurum, A. (2001) 'An industrial case study of usability engineering in market-driven packaged software development', *Elicitation and Management of User Requirements in Market-Driven Software Development,* , pp. 85.

Ogata, S. 'Effectiveness of Introducing Human-Centered Design Process', .

Ogunsola, L. (2005) 'Information and Communication Technologies and the Effects of Globalization: Twenty-First Century "Digital Slavery" for Developing Countries--Myth or Reality', *Electronic Journal of Academic and Special Librarianship,* 6(1-2), pp. 1-10.

Oulanov, A. and Pajarillo, E.J. (2002) 'CUNY Web: Usability study of the Web-based GUI version of the bibliographic database of the City University of New York (CUNY)', *The Electronic Library,* 20(6), pp. 481-487.

Ozok, A.A. (2009) 'Survey design and implementation in HCI', *Human-Computer Interaction: Development Process,* 253.

Pace, S. (2004) 'A grounded theory of the flow experiences of Web users', *International journal of human-computer studies,* 60(3), pp. 327-363.

Polson, P.G., Lewis, C., Rieman, J. and Wharton, C. (1992) 'Cognitive walkthroughs: a method for theory-based evaluation of user interfaces', *International Journal of man-machine studies,* 36(5), pp. 741-773.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T. (1994) *Human-computer interaction.* Addison-Wesley Longman Ltd.

Quesenbery, W. (2011) 'Evaluation in Government Environments', in Buie, E. and Murray, D. (eds.) *Usability in Government Systems.* 1st edn. Elsevier, pp. 317330

Quesenbery, W. (2013) *What Does Usability Mean: Looking Beyond 'Ease of Use'*. Available at: http://www.wqusability.com/articles/more-than-ease-of-use.html (Accessed: 09/13 2013).

Quesenbery, W. (2004) 'Balancing the 5Es: Usability', *Cutter IT Journal,* 17(2), pp. 4-11.

Rauf, A., Batool, S., Rehman, S. and Shahid, A. (2010) 'Survey based usability evaluation of MS Word', *User Science and Engineering (i-USEr), 2010 International Conference on.* IEEE, 193-198.

Riihiaho, S. (2000) 'Experiences with usability evaluation methods', *Licentiate thesis.Helsinki University of Technology.Laboratory of Information Processing Science, .*

Robson, C. (2002) *Real world research.* Blackwell publishers Oxford.

Rogers, Y., Sharp, H. and Preece, J. (2011) *Interaction design: beyond human-computer interaction.* 3rd edn. John Wiley & Sons.

Rogers, Y., Sharp, H., Preece, J. and Tepper, M. (2007) 'Interaction design: beyond human-computer interaction', *netWorker: The Craft of Network Computing,* 11(4), pp. 34.

Rosson, M.B. and Carroll, J.M. (2001) *Usability engineering: scenario-based development of human-computer interaction.* Elsevier.

Rubin, J. (1994) 'Handbook of Usability Testing: How to plan, design, and conduct effective tests. New York': Wiley .

Rubin, J. and Chisnell, D. (2008) *Handbook of usability testing: how to plan, design and conduct effective tests.* John Wiley & Sons.

Rudd, J. and Isensee, S. (1991) 'Twenty-two tips for a happier, healthier prototype', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* SAGE Publications, 328-331.

Rudd, J. and Isensee, S. (1991) 'Twenty-two tips for a happier, healthier prototype', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* SAGE Publications, 328-331.

Sauro , J. (2014) *What Goes into a Usability Test Plan?*. Available at: http://www.measuringu.com/blog/usability-test-plan.php (Accessed: 04/05 2014).

Sauro, J. (2014) *Are both positive and negative items necessary in questionnaires?*. Available at: http://www.measuringu.com/positive-negative.php (Accessed: 11/26 2014).

Sauro, J. and Kindlund, E. (2005) 'A method to standardize usability metrics into a single score', *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM, 401-409.

Sauro, J. and Lewis, J.R. (2011) 'When designing usability questionnaires, does it hurt to be positive?', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2215-2224.

Sauro, J. and Lewis, J.R. (2009) 'Correlations among prototypical usability metrics: evidence for the construct of usability', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 1609-1618.

Sauro, J. and Lewis, J. (2005) 'Estimating Completion Rates From Small samples using Binomial Confidence Intervals : Comparisons And Recommendations', *PROCEEDINGS of the HUMAN FACTORS AND ERGONOMICS SOCIETY 49th ANNUAL MEETING, .*

Sauro, J. (2014) *10 Essential Usability Metrics*. Available at: http://www.measuringu.com/blog/essential-metrics.php (Accessed: 04/16 2014).

Scholtz, J. (2004) 'Usability evaluation', *National Institute of Standards and Technology, .*

Schuler,D. & Namioka,A. (1993) Participatory Design: Principles and Practices, Lawrence Erlbaum Associates, Hillsdale,NJ (Eds.)

Seacord,R., Plakoshm,D. & Lewis,G. (2003) *Modernizing Legacy Systems: Software Technologies, Engineering Processes and Business Practices.*

Seffah, A., Donyaee, M., Kline, R.B. and Padda, H.K. (2006) 'Usability measurement and metrics: A consolidated model', *Software Quality Journal,* 14(2), pp. 159-178.

Serco Ltd. (2002) *Usability requirements: how to specify, test and report usability,* available online at: http://www.usabilitynet.org/prue/requirements.htm (Accessed 09/09 2015)

Shackel, B. (2009) 'Usability–Context, framework, definition, design and evaluation', *Interacting with Computers,* 21(5-6), pp. 339-346.

Shackel, B. (2009) 'Usability–Context, framework, definition, design and evaluation', *Interacting with Computers,* 21(5-6), pp. 339-346.

Shackel, B. and Richardson, S.J. (1991) *Human factors for informatics usability.* Cambridge University Press.

Sheskin, D.J. (2003) *Handbook of parametric and nonparametric statistical procedures.* 3rd edn. crc Press.

Shneiderman, B. (2010) *Designing the user interface: strategies for effective human-computer interaction.* Fifth edn. Addison-Wesley Reading, MA.

Shneiderman, B. (1998) *Designing the user interface-strategies for effective human-computer interaction.* Reading: Addison Wesley Longman.

Shneiderman, B. (1986) *Designing the user interface-strategies for effective human-computer interaction.* Pearson Education India.

Shneiderman, S.B. and Plaisant, C. (2005) 'Designing the user interface 4 th edition', *ed: Pearson Addison Wesley, USA,* .

Silverman, D. (2006) *Interpreting qualitative data: Methods for analyzing talk, text and interaction.* 3rd edn. Sage.

Smith, M.F (1993) Software Development: Prototypically topical, Br J Healthcare Comput Informat Manag. 10(6), pp. 25-27Sommerville, I. (2001) 'Software documentation', *Software Engineering,* 2, pp. 143-154.

Spencer, D. (2004) " What is usability" , *STEP TWO DESIGN PTY LTD, available at : http://www.steptwo.com.au/papers/kmc_whatisusability/ (Accessed* 14/09/2013)

Spool, J. and Schroeder, W. (2001) 'Testing web sites: Five users is nowhere near enough', *CHI'01 extended abstracts on Human factors in computing systems.* ACM, 285-286.

Spool, J. (2014)*UIE Tips: Becoming a Design-Infused Organization*. Available at: http://www.uie.com/brainsparks/ (Accessed: 06/16 2014).

Spyridakis, J.H. (1992) 'Conducting research in technical communication: The application of true experimental designs', *Technical communication,* 39(4), pp. 607-624.

Stary, C. and Eberle, P. (2008) 'Building up usability-engineering capability by improving access to automated usability evaluation', *Interacting with Computers,* 20(2), pp. 199-211.

Stone, D., Jarrett, C., Woodroffe, M. and Minocha, S. (2005) *User interface design and evaluation.* Morgan Kaufmann.

Stringer, M., Rode, J.A., Toye, E.F., Blackwell, A.F. and Simpson, A.R. (2005) 'The WebKit tangible user interface: A case study of iterative prototyping', *Pervasive Computing, IEEE,* 4(4), pp. 35-41.

Szekely, P. (1995) 'User interface prototyping: Tools and techniques', *Software Engineering and Human-Computer Interaction.* Springer, 76-92.

Tashakkori, A. and Teddlie, C. (2010) *Sage handbook of mixed methods in social & behavioral research.* Sage.

Taylor, B. and Heath, A. (1996) 'The use of double-sided items in scale construction', .

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L.P., Robson, R., Thabane, M., Giangregorio, L. and Goldsmith, C.H. (2010) 'A tutorial on pilot studies: the what, why and how', *BMC medical research methodology,* 10, pp. 1-2288-10-1.

Thomas, D.R. (2006) 'A general inductive approach for analyzing qualitative evaluation data', *American journal of evaluation,* 27(2), pp. 237-246.

Trivedi, M.C. and Khanum, M.A. (2012) 'Role of context in usability evaluations: A review', *arXiv preprint arXiv:1204.2138,* .

Troll Covey, D. (2002) 'Usage and usability assessment: Library practices and concerns', .

Tullis, T.S. and Stetson, J.N. (2004) 'A comparison of questionnaires for assessing website usability', *Usability Professional Association Conference.* , 1-12.

Tullis, T. and Albert, B. (2008) *Measuring the user experience: collecting, analyzing, and presenting usability metrics.* Elsevier inc.

Ulrich, W. (2002) Legacy System : Transformation Strategies, Prentic-Hall Englewood Cliffs, 1st Ed.

UniqueSoft LLC. (2013) *Legacy Re-Engineering Advantages* , Available at: http://www.uniquesoft.com/about-us.html (Accessed : 11/10/2015).

UsabilityNet (2014) *What is usability?.* Available at: http://www.usabilitynet.org/management/b_what.htm (Accessed: 04/19 2013).

Virzi, R.A. (1992) 'Refining the test phase of usability evaluation: How many subjects is enough?', *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 34(4), pp. 457-468.

Virzi, R.A., Sokolov, J.L. and Karis, D. (1996) 'Usability problem identification using both low-and high-fidelity prototypes', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 236-243.

Virzi, R.A., Sokolov, J.L. and Karis, D. (1996) 'Usability problem identification using both low-and high-fidelity prototypes', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 236-243.

Walker, M., Takayama, L. and Landay, J.A. (2002) 'High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* SAGE Publications, 661-665.

Wallach, D. and Scholz, S.C. (2012) 'User-centered design: why and how to put users first in software development', in *Software for People.* Springer, pp. 11-38.

Wharton, C., Rieman, J., Lewis, C. and Polson, P. (1993) 'The Cognitive Walkthrough Methods: A Practitioner's Guide', *University of Colorado at Boulder, .*

Wichansky, A.M. (2000) 'Usability testing in 2000 and beyond', *Ergonomics,* 43(7), pp. 998-1006.

Wiklund, J., Davidsson, P., Delmar, F. and Aronsson, M. (1997), "*Expected consequences of growth and their effect on growth willingness in different samples of small firms*", in Reynolds, P.D., Bygrave, W.D., Carter, N.M, Davidsson, P., Gartner, W.B., Mason, C.M. and McDougall, P.P. (Eds), *Frontiers of Entrepreneurship Research*, Babson College, Babson Park, MA

Wirtz, S., Jakobs, E. and Ziefle, M. (2009) 'Age-specific usability issues of software interfaces', *Proceedings of the IEA.*

Wixon, D. and Wilson, C. (1997) 'The usability engineering framework for product design and evaluation', *Handbook of human-computer interaction,* 2, pp. 653-668.

Wong, B. and Blandford, A. (2003) 'Field research in HCI: a case study', *Proceedings of the 4th Annual Conference of the ACM Special Interest Group on Computer-Human Interaction.* ACM, 69-74.

Xenos, M. (2001) 'Usability perspective in software quality', *Usability Engineering Workshop, The 8th Panhellenic Conference on Informatics with International Participation, Southern Cyprus.*

Zaphiris, P. and Kurniawan, S. (2007) *Human computer interaction research in web design and evaluation.* IGI Global.

Zazelenchuk, T. (2015) *Data collection for usability research*. Available at: http://www.userfocus.co.uk/articles/dataloggingtools.html (Accessed: 09/13 2013).

Zhang, P., J. Carey , D. Galletta, J. Kim , D. Te'eni and B. Wildemuth (2004)" The Role of Human Computer Interaction in Management Inofmraiton system Curricula: A Call to Action" *Communications of the AIS* (13), pp 357-379.

Zhang, P., Benbasat, I., Carey, J., Davis, F., Galletta, D.F. and Strong, D. (2002) 'AMCIS 2002 Panels and workshops I: Human-computer interaction research in the MIS discipline', *Communications of the Association for Information Systems*, 9(1), pp. 20.

Zhang, T. and Dong, H. (2009) 'Human-centred design: an emergent conceptual model', .

Ziefle, M. and Bay, S. (2008) 'Transgenerational Designs in Mobile Technology', *In: Lumsden, J. (ed.),* , pp. 122-140.

Zviran, M. and Erlich, Z. (2003) 'Measuring IS user satisfaction: review and implications', *Communications of the Association for Information Systems,* 12(1), pp. 5.

Appendices

Appendix A: Ethical Approvals

1) The first study

School of Information Systems, Computing and Mathematics
David Gilbert, Head of School, Professor of Computing
Jasna Kuljis, Head of Information Systems and Computing, Professor of Computing
Tony Rawlins, Head of Mathematical Science, Professor of Mathematics

Brunel University West London

Brunel University, Uxbridge,
Middlesex UB8 3PH, UK
Telephone: +44(0) 1895 274000
Fax: +44(0) 1895 251686
Emails:
Yongmin.Li@brunel.ac.uk
Annette.Payne@brunel.ac.uk
Lampros.Stergioulas@brunel.ac.uk
Zidong.Wang@brunel.ac.uk

Date: 22nd January 2013

STATEMENT OF ETHICS APPROVAL

Proposer: Faisal Mohammed Dahlan Baglin
Title: Query technique based Usability Evaluation of a current system in the public organization from user perspective.

The school's research ethics committee has considered the proposal recently submitted by you. Acting under delegated authority, the committee is satisfied that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that you will adhere to the terms agreed with participants and to inform the committee of any change of plans in relations to the information provided in the application form.

Yours sincerely,

Zidong Wang

Professor Zidong Wang
Chair of the Research Ethics Committee
SISCM

2) The second study

School of Information Systems, Computing and Mathematics
David Gilbert, Head of School, Professor of Computing
Jasna Kuljis, Head of Information Systems and Computing, Professor of Computing
Tony Rawlins, Head of Mathematical Science, Professor of Mathematics

Brunel UNIVERSITY WEST LONDON

Brunel University, Uxbridge,
Middlesex UB8 3PH, UK
Telephone: +44(0) 1895 274000
Fax: +44(0) 1895 251686
Emails:
Yongmin.Li@brunel.ac.uk
Annette.Payne@brunel.ac.uk
Lampros.Stergioulas@brunel.ac.uk
Zidong.Wang@brunel.ac.uk

Date: 13th August 2013

STATEMENT OF ETHICS APPROVAL

Proposer: Faisal Mohammed Dahlan Baglin

Title: A usability testing approach to evaluate a high fidelity prototype user interface of a current system in the public organization: a user perspective.

The school's research ethics committee has considered the proposal recently submitted by you. Acting under delegated authority, the committee is satisfied that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that you will adhere to the terms agreed with participants and to inform the committee of any change of plans in relations to the information provided in the application form.

Yours sincerely,

Zidong Wang

Professor Zidong Wang
Chair of the Research Ethics Committee
SISCM

3) The third study

School of Information Systems, Computing and Mathematics
David Gilbert, Head of School, Professor of Computing
Martin Shepperd, Head of Information Systems and Computing
Steven Noble, Head of Mathematical Science, Professor of Mathematics

Brunel
UNIVERSITY
LONDON

Brunel University, Uxbridge,
Middlesex UB8 3PH, UK
Telephone: +44(0) 1895 274000
Fax: +44(0) 1895 251686
Emails:
Yongmin.Li@brunel.ac.uk
Annette.Payne@brunel.ac.uk
Lampros.Stergioulas@brunel.ac.uk
Zidong.Wang@brunel.ac.uk

Date: 24/07/2014

STATEMENT OF ETHICS APPROVAL

Proposer: Faisal Mohammed Dahlan Baglin

Title: The rapid prototyping of automated assistance decision making and additional feature in governmental administration system with the iteration approach: An employee perspective.

The school's research ethics committee has considered the proposal recently submitted by you. Acting under delegated authority, the committee is satisfied that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that you will adhere to the terms agreed with participants and to inform the committee of any change of plans in relations to the information provided in the application form.

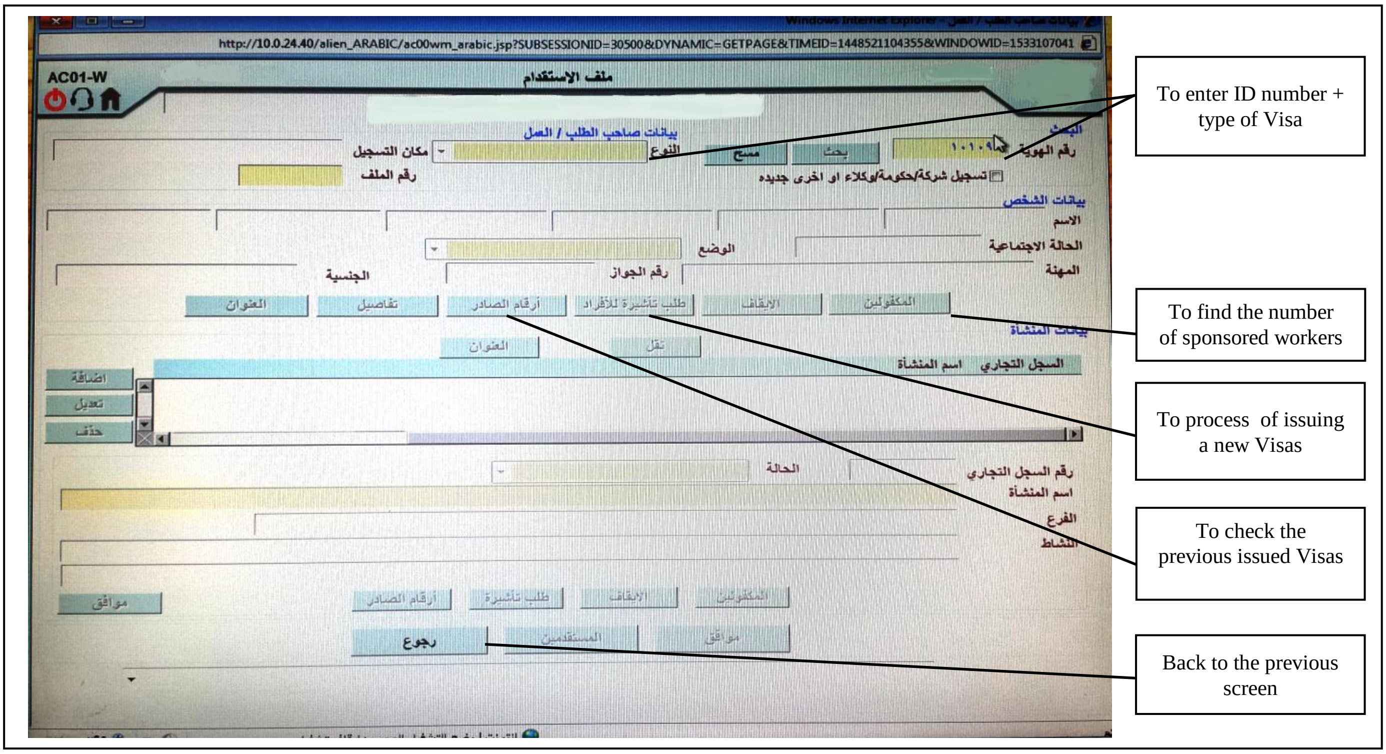
Yours sincerely,

Zidong Wang

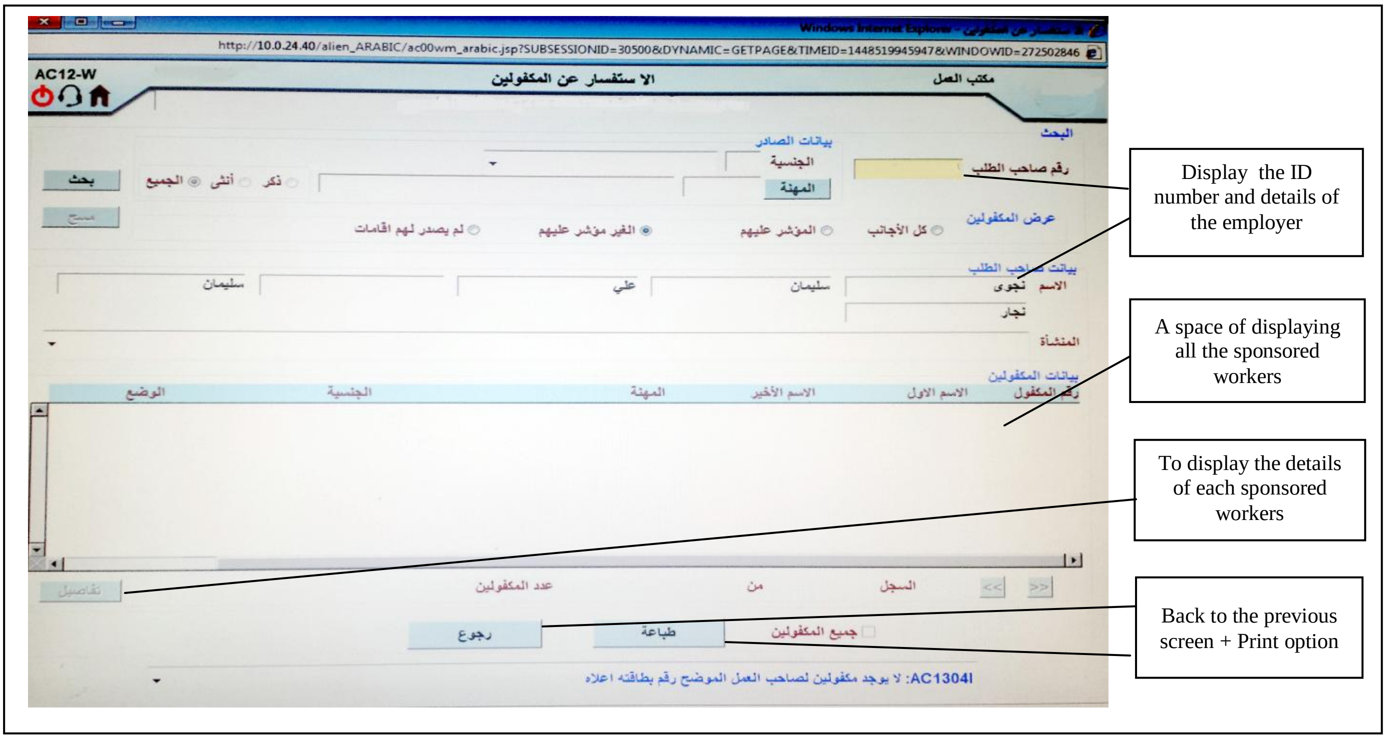
Professor Zidong Wang
Chair of the Research Ethics Committee
SISCM

Appendix B: Screenshots of the current VI system

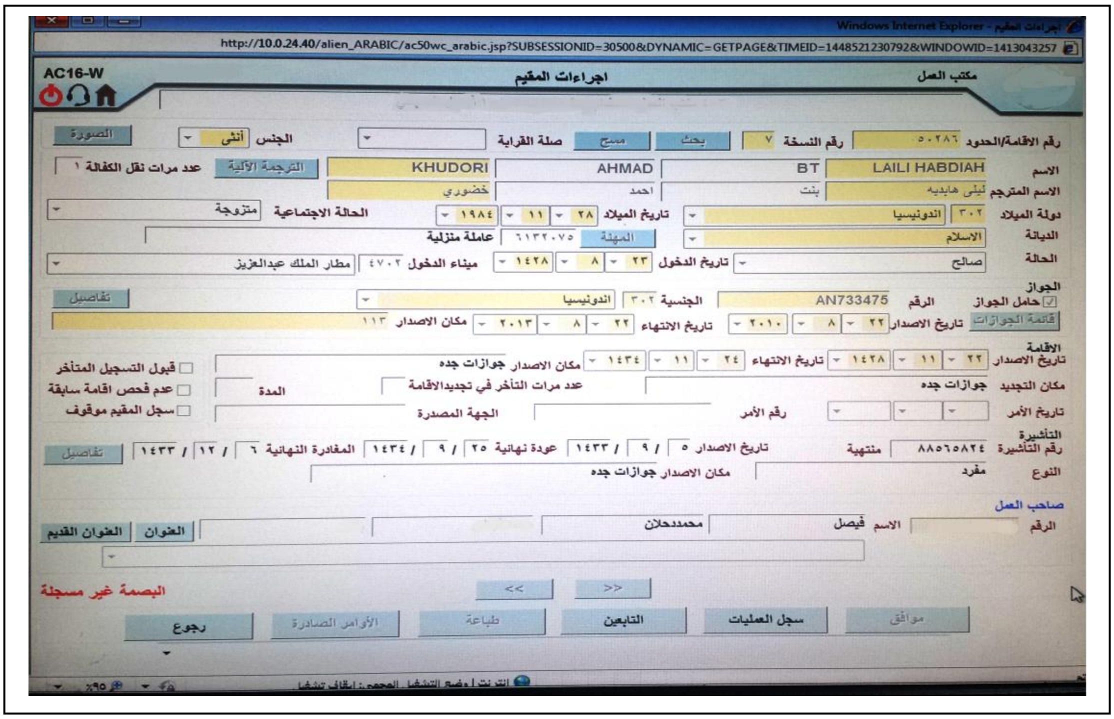
1) The main screen of the VI system



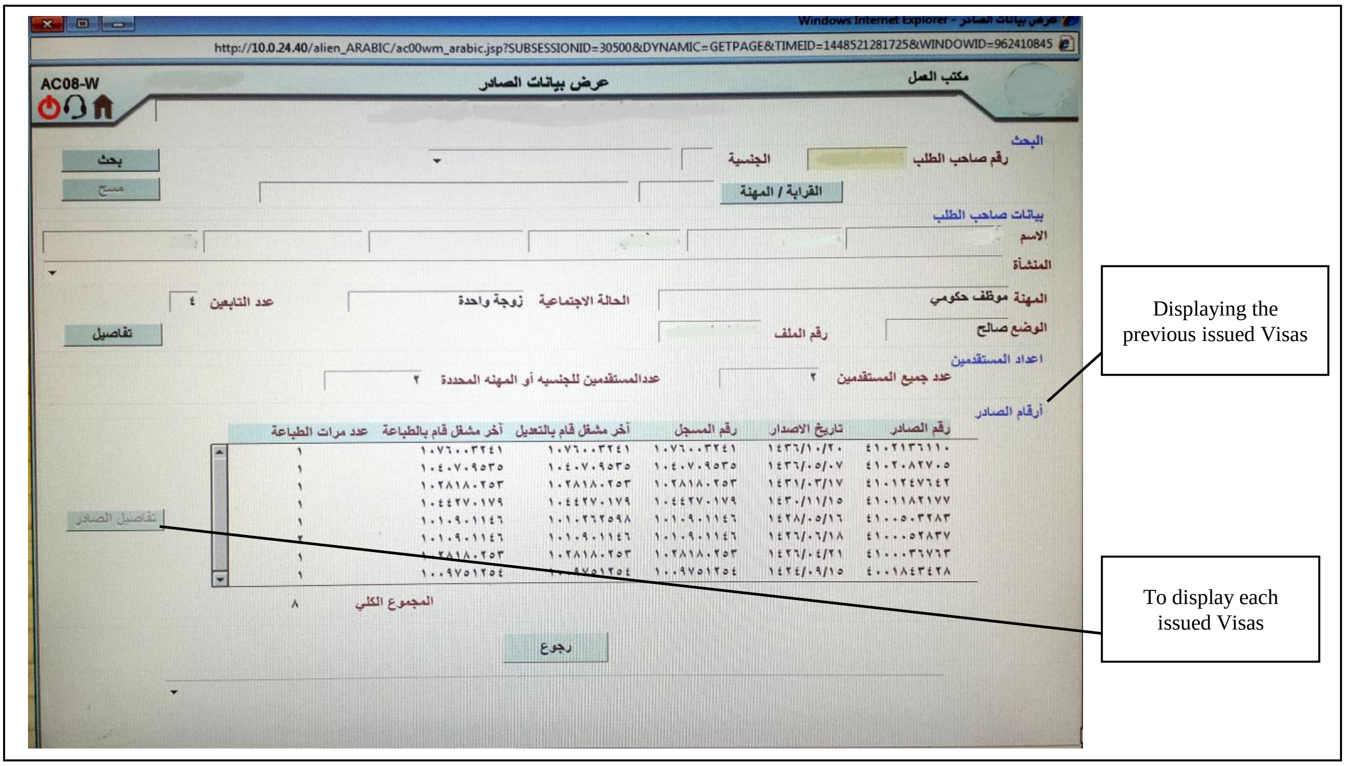
2) The screen for finding the number of sponsored workers



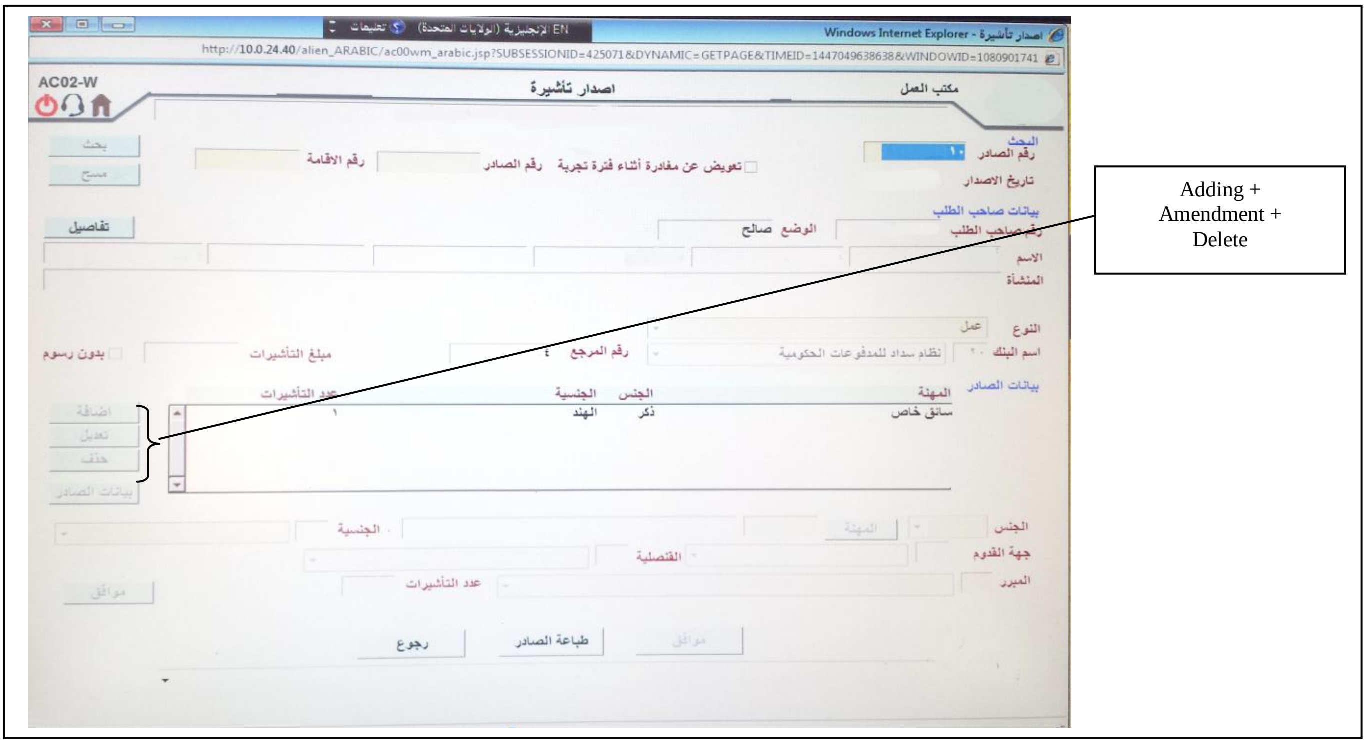
3) The screen for displaying the details of a sponsored worker



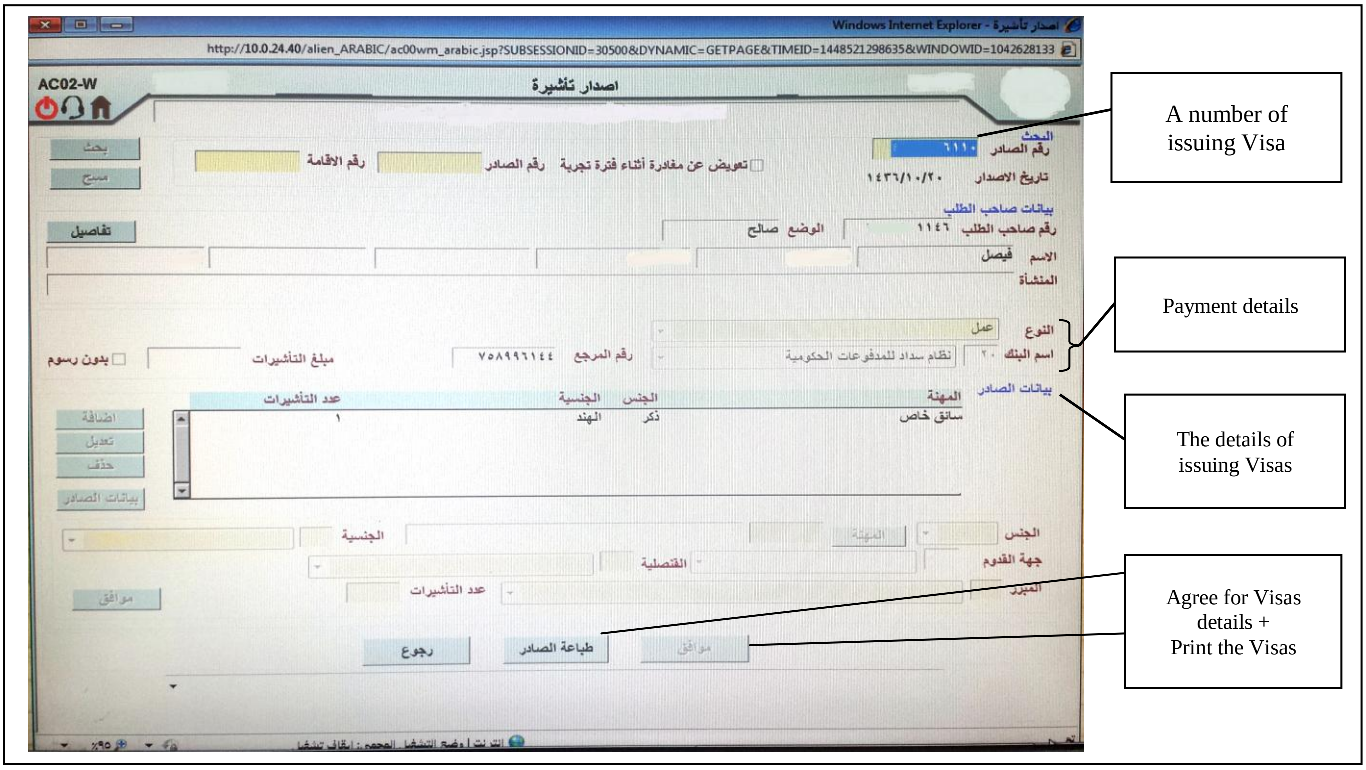
4) The screen for displaying the previous issued Visas



5) The screen of displaying the details of an issued Visa



6) The last screen of processing of issuing new visas



Appendix C : *Evaluation Usability Questionnaire*

Dear Participant,

I am a PhD research student at Brunel University London, UK. As part of my thesis, I am conducting a survey to evaluate the usability of VI system in order to identify its problems and issues that need to be solved for improving the system abilities.

The questionnaire designed for this study consists of two parts. The first part asks about the respondent's demographics. The second part presents questions that measure the perceptions about the system' ease of use included several dimensions.

Your participation in this study will allow me to identify your perspective on the system you are currently used. Also, it would be quite helpful to understand what aspects of the system particularly concern you and which aspects you are satisfied with.

The collected responses will be strictly confidential and data from this research will be reported anonymously. Your answers are valuable to us. Please take your time to answer it, and it should take around 7-11 minutes to complete.

Thank you very much for your time and support,

Researcher

If you have any queries or would you like further information please do not hesitate to contact me by email:
faisal.baglin@brunel.ac.uk
Also you can contact the school of information system and computing at Brunel University as per the following contact:
Professor Zidong Wang. Tel: +44 (0)1895 266021. Fax: +44 (0)1895 251686
Email: zidong.wang@brunel.ac.uk
Information Systems and Computing
St John's 137. Brunel University, Uxbridge
UB8 3PH. United Kingdom

Section one: Personal Information

1- What is your..?

Age

- 20-27
- 28-35
- 36-42
- 43-50
- 51 or More

Education level

- High school graduate
- Diploma or the equivalent
- Under graduate
- Post graduate

Period of Experience using the system

- Less than 1 Month
- 1 - 6
- 7 - 12
- 13 - 24
- More than 24 Month

Section two: The main dimensions

Please read each question and choose the option that most closely matches your attitude regarding the following statement:

Note: It would be helpful if you elaborate explanation upon your answers, and please include it in the space below the questions.

Effectiveness	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Please explain your answer (if applicable):
1- It is easy to follow the instruction to complete tasks.	1	2	3	4	5	-- --
2- The system quickly completes the tasks I want to do.	1	2	3	4	5	-- --
3- I could not effectively complete my tasks using this system	1	2	3	4	5	-- --

Efficiency	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Please explain your answer (if applicable):
4- The system requires many steps to complete a task.	1	2	3	4	5	-- --
5- It is relatively easy to move from one step of a task to another.	1	2	3	4	5	-- --
6- It takes a long time to complete a task using the system	1	2	3	4	5	-- --
7- The integration ability of the system is a crucial factor for its efficiency.	1	2	3	4	5	-- --
Easy to learn	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	**Please explain your answer (if applicable):**
8- It was not easy to learn how to use this system.	1	2	3	4	5	-- --
9- Learning how to use new functions on this system is hard.	1	2	3	4	5	-- --
10- The system uses a consistent navigational system to enable me to understand where to find information.	1	2	3	4	5	-- --
11- As long as I am following the system's procedure, I can get the necessary information and knowledge to complete my task.	1	2	3	4	5	-- --
Engaging (Satisfaction)	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	**Please explain your answer (if applicable):**
12- The data interface is not always consistent with instruction menu.	1	2	3	4	5	-- --
13- The interface of this system is well designed.	1	2	3	4	5	-- --
14- The features of the system are very limited.	1	2	3	4	5	-- --

15- It is obvious that end-user's needs have not been fully taken into consideration when the system was developed.	1	2	3	4	5	-- --
16- I believe using this system makes me more productive in my work.	1	2	3	4	5	-- --
17- Overall, I am satisfied with using this system for issuing work visas.	1	2	3	4	5	-- --
Error Tolerant	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	Please explain your answer (if applicable):
18- The System does not give error messages that clearly guide me on how to fix problems before I finish a task.	1	2	3	4	5	-- --
19- The system has helped me to overcome any problems I have had after I finish a task.	1	2	3	4	5	-- --
20- When I make a mistake before I finish a task while using they system, I could recover it easily and quickly.	1	2	3	4	5	-- --
21- The system has at some time stopped unpredictably and it not easy to restart it again.	1	2	3	4	5	-- --
Helpfulness	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	Please explain your answer (if applicable):
22- The system does not support basic standard actions, such as save, copy, cut and paste.	1	2	3	4	5	-- --
23- There is never enough information on the system to help me complete the tasks (such as online help, and other documentation)	1	2	3	4	5	-- --
24- The system enables me to provide good quality service to the citizen	1	2	3	4	5	-- --
25- I think I would need Technical Support regularly to be able to use this system.	1	2	3	4	5	-- --

Control	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Please explain your answer (if applicable):
26- I quickly became skilful at operating this system.	1	2	3	4	5	--- ---
27- The data entry of applicant in the system to issue work visas is hard and complicated	1	2	3	4	5	--- ---
28- The amendment ability of the system affects my ability to complete a task.	1	2	3	4	5	--- ---
29- I would not like to use this system ever day	1	2	3	4	5	--- ---

Appendix D: *The interview questions*

The following questions were used as guidance during conducting the interview

The semi-structure interview's questions

1- Efficiency
• Do you think a task takes more time than expected? If yes, why? • Do you mind to tell us, what is your suggestion to enhance the performance of the system through its interface for doing the task?
2- Effectiveness
• What are the main parts of processing the tasks that the employees have done many efforts? • Where are employees (users) likely to get stuck? • What mistakes do employees (users) make "particular in data entry"?
3- Satisfaction
• Does ease of use aspect of the VI system make an impact on the employees' performance? • Do you agree for using this system to issue visas to the applicants?

Thank you for your time,

Appendix E:*Participant Information and Consent Form*

VI System Usability Research Study

Dear Participant,

Thank you for your agreement to participate in this research project, which is officially approved by the competent authority of the Ministry of Labour. The purpose of conductingit is to assess and confirm the improvement of a suggested design solution through applying a comparison usability test between this new system and the original Visa Issuance system, based on several measurementssuch as efficiency, effectiveness, and user satisfaction.

This experiment test does not test your own abilities in anyway, but rather it tests your performance on using the particular developed system. You are in no way obligated to participate. The study is completely voluntary and you can withdraw at any time or any point in the test without incurring any liability.

Before you go through the information we would like to provide some of your personal details as follow.

Name:……………………...

Age:…………………………………...

Period of Experience with the Visa Issuance System ……………………………………………...

RISKS

Your participation in this study will not expose you to anyanticipated physical, mental, or social risks.

ADVANTAGES

Throughconfirming this improvement and identifying more issues and requirements, the system developers will be able to better accommodate user needs and make the system more usable which would improve your performance on finishing your intended tasks.

CONFIDENTIALITY

The collected information in this research study will be kept strictly confidential, and used only for the research purposes. In addition, all personal recorded information will be stored on the researcher's personal computer and hard disk, so it will be visible only to the researcher and will not be shared with anyone outside of the study. Once the study is complete, all email correspondence and your personal information will be deleted and discarded. Any data collected will be returned to you or destroyed if you choose to withdraw from the study at any time.

CONTACT

If you have any queries or you would like further information please do not hesitate to contact me by email: faisal.baglin@brunel.ac.uk

Also you can contact the School of Information Systems and Computing at Brunel University as per the following contact:

Professor Zidong Wang. Tel: +44 (0)1895 266021. Fax: +44 (0)1895 251686

Email: zidong.wang@brunel.ac.uk
Information Systems and Computing
St John's 137. Brunel University, Uxbridge
UB8 3PH. United Kingdom

YOUR DETAILS

CONSENT

Please sign your full name below if you agree to the terms of this research project.

I have read and understand the above information and I agree to participate in this study.

Participant's signature: …………………. ………….. Date: …………………..

Appendix F:*Participant Instructions (Experiment 1)*

Before starting the test session and performing the test tasks, you will be given a brief introduction and short training on how to use the suggested design solution, and then you will be provided with a user name and password for enabling you to use the developed system. Also dummy data of the applicants' ID numbers will be given to enable you to perform the typical tasks on the developed design system.

The following script is about seven typical tasksthat you will aim to perform and complete using the two systems, the current VI system and the developed prototype design. During the test session you may face some difficulties in performing a particular test task, do not worry if you feel you are unable to finish it, please just say that you would like to stop, and carry on to the next task. Keep in mind, the aim of this research is to collect information about how usable the system is. However, with regard to technical problems with the system or with the computer you are using, you will be assisted in solving the identified problems.

In the matter of time, to complete the session would take 48 minutes including working on all seven tasks, and to complete after that the post-study questionnaire. While completing these tasks, you are encouraged to "think aloud". Please say aloud what it is you are thinking as you work on the tasks in order to be able to record yourperforming on and interaction with the system for the test sessions. The collected information will be used in the data analysis to detect usability issues and additional users' requirements.

Task 1

Scenario:

After you have obtained the user name and password, login to the system, then try to review the details of an applicant whose dummy ID number you have been given by entering this number into the system.

Task 2

Scenario:

Search for the details of previous visas on the applicant's record by entering the particular numbers of issued visas which showed in the details of the applicant.

Task 3

Scenario

Search for the number of employees who are sponsored by the applicant by choosing the specific option on the user interface of the developed system design.

Task 4

Scenario

Create a new file for an applicant's corporation by entering the given dummy data, which includesname, commercial registration number (CR no.), and the nature of the corporation's activities

Task 5

Scenario

From the dummy data you have been given, create a data entry for issuing a new visa with different requirements e.g. Nationality and Occupation.

Task 6

Scenario

Change the details of the visa you have just issued during the visa issuing process.

Task 7

Scenario

Suppose that the issued visa has been lost and you would like to re-issue this visa. Find out how to process re-printing the issued visa.

Appendix G:*Participant Instructions (Experiment 2)*

Before starting the test session and perform the test tasks, will be given a brief introduction and short training to use the two suggested solution design versions, and then you will be provided by a user name and password for enabling to use these developed systems. Besides dummy data of the applicants ID numbers will be given to be able to perform the typical tasks on the developed design system.

The following script is about five typical tasksthat you will aim to perform and complete using the two versions of developed prototype design (V1 and V2). During the test session you may face some difficulties in performing a particular test task, do not worry if you feel you are unable to finish it, please just say that you would like to stop, and carry on to the next task. Keep in mind, the aim of this research is to collect information about how usable the system is. However, with regard to technical problems with the system or with the computer you are using, you will be assisted in solving the identified problems.

In the matter of time, to complete the session would take 35 minutes including working on all five tasks, and to complete after that the post-study questionnaire. While completing these tasks, you are encouraged to “think aloud”. Please say aloud what it is you are thinking as you work on the tasks in order to be able to record yourperforming on and interaction with the system for the test sessions. The collected information will be used in the data analysis to detect usability issues and confirm the system improvement.

Task 1

Scenario:
From the dummy data you have given, make data entry for issuing a new visa with different requirements e.g. Nationality and Occupation.

Task 2

Scenario:
Change the details of the issued visa you have just issued during the issuing visa process.

Task 3

Scenario
Create a new file for an applicant’s corporation by entering the given dummy data, which includes/name, commercial registration number (CR no.), and the nature of corporation‘s activates.

Task 4

Scenario
Navigate the Archive function from the main menu.

Task 5

Scenario
Search for the applied documentation of previous Visas.

Appendix H: *Post-Study Questionnaire*

Dear Participant

Thank you for agreeing to participate in this research project. The following questionnaire gives you the opportunity to provide and share any information of your thoughts towards the system you just tested. All of your answer and feedback will be very helpful to achieve a good understanding of your interaction and reaction to the system. Please take your time to answer it, and it should take around 7-10 minutes to complete it.

For obtaining as great degree results as possible, please write comments to elaborate on your response.

1) Please read the following statements carefully and think about all the tasks that you have done with the system while you response to them and select the choice that you feel it is the best reflects of your though about each statement as it relates to the user interface you tested.

SysQual	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	
1- Overall, I am satisfied with using this system for issuing work visas.	1	2	3	4	5	Your comment:
2- I could effectively complete the tasks and scenarios using the system	1	2	3	4	5	Your comment:
3- This system has the functions and capabilities I need to perform my work.	1	2	3	4	5	Your comment:
4- I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5	Your comment:
5- The system consumed much time to do the required task.	1	2	3	4	5	Your comment:
6- The system required only a few steps to accomplish the tasks.	1	2	3	4	5	Your comment:

7- It was not easy to learn how to use this system.	1	2	3	4	5	Your comment:
8- Data entry of new applicant information was easy and clear.	1	2	3	4	5	Your comment:
9- Whenever I made mistake while using the system, I could recover it easily and quickly.	1	2	3	4	5	Your comment:
10- I felt in control when I was using the system.	1	2	3	4	5	Your comment:
11- When I used the system, it was easy to move from one step of a task to another	1	2	3	4	5	Your comment:
Info Qual	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	
12- It would be difficult to be skilful at using and controlling this system in a high degree of performance.	1	2	3	4	5	Your comment:
13- I could easily find the required information to complete tasks.	1	2	3	4	5	Your comment:
14- I think I would need Technical support to be able to use this system.	1	2	3	4	5	Your comment:
15- Navigating information through the menus and toolbars of system was not easy to do.	1	2	3	4	5	Your comment:
16- The system has not helped me to overcome any problems I have had after I finish a task, such amend on the input data.	1	2	3	4	5	Your comment:
17- The organisation of information on the system user interface was clear.	1	2	3	4	5	Your comment:
18- The information provided for the system was easy to understand.	1	2	3	4	5	Your comment:

19- The system did not give error messages that clearly informed me how to fix problem.	1	2	3	4	5	Your comment:
20- I could not control the display data flexibly.	1	2	3	4	5	Your comment:
Intf Qual	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	
21- I understood the function of each buttonin the systemeasily.	1	2	3	4	5	Your comment:
22-The system provided a capability (such as UNDO feature) which enables me to quick reverse mistaken actions.	1	2	3	4	5	Your comment:
23-The system features (like default saving data and copy) helped me to perform tasks successfully.	1	2	3	4	5	Your comment:
24-The interface of this system was pleasant to use.	1	2	3	4	5	Your comment:
25-I would be happy to use the system again.	1	2	3	4	5	Your comment:
26- It is obvious that user needs have been taken into consideration when the system was developed.	1	2	3	4	5	Your comment:

2) Please list the particular aspect(s) of the system you like and explain your choices.

-

-

-

3) Please list the particular aspect(s) of the system you dislike and explain your choices.

-

-

-

Thank you for your time,