UNIVERSITY OF HULL

An Evaluation Framework and Selection Tool for Education Apps Usability, With a Case Study from Health Education Apps

being a Thesis submitted for the Degree of

Doctor of Philosophy

in the University of Hull

by

Tareq Aljaber B.Sc., M.Sc.

May 2019

Abstract

Mobile apps for health education are commonly utilised to support different users. The development of these apps is increasing rapidly. A critical evaluation framework is needed to ensure the usability and reliability of Mobile Health Education Applications (MHEAs) to save time and effort for stakeholders. This project aims to assist the evaluation of MHEAs through development of an evaluation framework, which includes suitable metrics, an efficient hybrid utilizing Heuristic Evaluation (HE) and Usability Evaluation (UE). This framework determines the usefulness and usability of MHEAs, in order to improve the software engineering to create more effective ways to evaluate such software. In this framework, the Medical Apps Selection Tool (MAST) has been developed in which performance helps to select suitable MHEAs, assisting stakeholders to choose MHEAs that meet their requirements.

The thesis employs two methods to make the evaluation framework capable of performing qualitative and quantitative data analysis. The first is a qualitative method, involving interviews based on proposed selected hybrid metrics from HE and UE, with three kinds of stakeholders: Patients, Health Professionals and Software Developers, to identify specific relevant selected hybrid metrics to measure usability in MHEAs. These metrics are deployed to measure usability in different MHEAs based on ranking these apps within the evaluation framework. These metrics were converted into an evaluation questionnaire, which has been applied to several MHEAs. The second method is the translation of the outcomes form the first method to measure the usability of MHEAs and determine what stakeholders require from using MHEAs. For this purpose, it categorises stakeholders with different needs from MHEAs; this reflected in the MAST, based on matching different stakeholders with different MHEAs.

The findings of the study indicate that the evaluation framework is able to evaluate MHEAs and record usability problems. Furthermore, this evaluation framework leads to selection of the most appropriate apps by developing the MAST for stakeholders. The framework is expected to be applicable to other domains and platforms.

Keywords: Metrics, Medical Apps Selection Tool (MAST), Heuristic Evaluation (HE), Usability Evaluation (UE), Evaluation Framework, Mobile Health Education Applications (MHEAs), Stakeholders.

TO MY GREAT MUM: MAJIDAH SABHA

TO MY BELOVED WIFE: ZAINAH SAMARAH

TO MY CHILDREN: ZAID, ZAYAN AND ZIYAD (ALJABER)

TO THE SOUL OF MY FATHER: TALAL ALJABER

TO MY BROTHER AND SISTERS: SAMER, NOUR AND HADIL (ALJABER)

إلى أمي العظيمة: ماجدة صبحا

إلى زوجتي الحبيبة: زينة سمارة

لأبنائي و قرة عيني: زيد، زيان وزياد (الجابر)

إلى روح أبي الطاهرة: طلال الجابر

إلى أخي وأخواتي: سامر ، نور وهديل (الجابر)

Acknowledgements

I am deeply indebted to my supervisor, Dr Neil Andrew Gordon, whose commitment; persistent encouragement and invaluable advice have sustained me throughout the various stages of this research. He has read countless drafts of my thesis and has provided thoughtful insight and critiques.

I would like to thank my colleagues at University of Hull School of Engineering and Computer Science, Department of Computer Science, for their support and encouragement. I am also very grateful to Dr Chandra Kambhampati and Dr Bing Wang, who have inspired and guided me through the process of this research and monitored my academic progress. Moreover, I would like to thank all the secretaries' staff for the computer department especially Lynn Morrell and Helen El-Sharkawy. I would like also to thank the Skills team in the library, especially Jacqui Bartram , Lee G Fallin and Paul Chin. I would also like to acknowledge all the participants in this research, who were willing to spend their time and to share their views and priceless experience with me. Indeed, their cooperation and sincere comments are worth more than a 'thank you'.

A special thank you must go to my wonderful mother, Majidah Sabha, my amazing wife Zainah, my lovely children and to the soul of my father, for their moral support and patience throughout this challenging but invaluable journey. Their support, encouragement and belief in me have made this possible. Thanks to my brother Samer and my sisters Nour and Hadeel, for believing in my abilities and for encouraging me to start and continue this doctoral journey.

Special thanks go as well to Dr Mike Brayshaw, Dr Ali Adawi, Dr Sohag Kabir, Dr Robert Costello, Dr Koko kalambay, Dr Septavera Sharvia and Dr Janine Hatter for their support and comments. Unfortunately, it is not possible to list everyone who has helped me in some way with this thesis. My sincere thanks are extended to all parties and individuals who have assisted me, especially at the University of Hull for providing excellent resources and to the IT infrastructure unit for excellent support.

Finally, many thanks should also be extended to my best friends, Dr Amadou Gning and Othman Ozizi, for their huge support.

Related Publications

Some of the materials in this thesis have been published in several national and international conference proceedings, as below. In addition, they have been utilized, recommended, or cited, in fourteen citations in total. For example, Aljaber, et al. (2015) and Gordon, et al. (2016) in the publications list reflect part of chapters 1, 2 and 3. Aljaber and Gordon (2016) reflects part of chapter 4. Furthermore, Aljaber and Gordon (2017a) and Aljaber and Gordon (2017b) reflect part of chapters 4 and 5. Moreover, I have been involved in organizing and being a part of the committee for the 6th Departmental Poster Conference at the University of Hull and 7th PhD Experience Conference at the University of Hull. Finally, I was Co-Chair for the session entitled: Smart Design and User Experience II in the 8th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, AHFE 2017 International Conference. Los Angeles, California, USA.

Aljaber, T., Gordon, N., Kambhampati, C. & Brayshaw, M., 2015. *An Evaluation Framework for Mobile Health Education Software.* London, UK, Proceedings of Science and Information Conference (SAI), pp. 786-790.

Gordon , N., Brayshaw , M. & Aljaber , T., 2016. *Heuristic evaluation for serious immersive games and M-instruction*, In International Conference on Learning and Collaboration Technologies . Springer, Cham., pp. 310-319.

Aljaber, T. & Gordon, N., 2016. *Evaluation of Mobile Health Education Applications for Health Professionals and Patients.* Funchal, Madeira, Portugal, Proceedings of the 8th International conference on e-Health (EH 2016), pp. 107-114.

Aljaber, T. & Gordon, N., 2017a. *A guidance and evaluation approach for mHealth education applications.* Vancouver,Canada, In International Conference on Learning and Collaboration Technologies . Springer, Cham, pp. 330-340.

Aljaber, T. & Gordon, N., 2017b. *A Hybrid Evaluation Approach and Guidance for mHealth Education Applications.* los angeles,California, USA, In International Conference on Applied Human Factors and Ergonomics. Springer, Cham, pp. 282-290.

List of Abbreviations

The following Table describes the various acronyms and abbreviations used throughout this thesis.

Abbreviation	Meaning
MHEAs	Mobile Health Education Apps
HE	Heuristic Evaluation
UE	Usability Evaluation
HP	Health Professionals
P	Patients
SD	Software Developers
P/GP	Patients / General Public
MAST	Medical Apps Selection Tool
HCI	Human Computer Interaction
HPIW	Health Professionals Individual Weighting
PIW	Patients Individual Weighting
TRSM	Total of Requirements of Specific Metric
TRAM	Total of Requirements of All Metrics
SDIW	Software Developer Individual Weighting
UX	User Experience
UCI	User-Computer Interaction
CHI	Computer-Human Interaction
MMI	Man-Machine Interaction

Table of Contents

Table of Appendices

List of Figures

List of Tables

Chapter 1: Introduction

Many health professionals and patients utilize Mobile Health Education Applications (MHEAs), in order to aid their personal health and job performance. Therefore, the effectiveness of these apps is critical for improving health education, quality of life and health care systems. This chapter introduces the area of research number starting with an explanation of the research field, which is about MHEAs, the increasing of these apps and the users of these apps. In addition, the rational for researching MHEAs is provided, as well as a description of the research problems and the research methods (Qualitative and Quantitative) for this study. In addition, it describes why it is important to construct an evaluation framework and why a questionnaire is used. The chapter also introduces the research questions and objectives, which in summary are: How can we construct an evaluation framework for MHEAs? Can it provide a suitable measurement to rank MHEA? How can it be used to identify issues in MHEAs? How can it help in developing an appropriate tool to select suitable apps for different user groups? How can the framework be used to characterize the stakeholders and their requirements for MHEAs? The Objectives are: Developing the framework for MHEAs by utilizing a mix of Heuristic Evaluation (HE) and Usability Evaluation (UE) metrics, Utilising the framework to rank MHEAs, Using the framework to identify issues in MHEAs, Using the framework to develop a Medical Apps Selection Tool, Utilising the framework to characterise the profile of different stakeholders in order to identify their different requirements. Finally, this chapter describes the research flow and the organization of the whole thesis.

1.1 Field of Research and Problem Context

Computer devices such as personal computers and Laptops are still not economical devices for everybody compared to mobile phone technology. According to Chaudhuri (2012), computer and internet elevation schemes usually fail despite active support, but mobile infiltration in even the poorest countries is progressing quickly. Mobile phones now play a major role in many aspects of human life. Improvements in mobile technology have allowed a wide range of applications to be developed, which people can use to support various aspects of their life (Aljaber, et al., 2015; Harrison , et al., 2013). Everyday life in general is becoming increasingly complex and in this context, there is a need for assisting patients to improve their health education, which can improve their health condition. By one way of supporting them in this respect is with the Mobile Health Education Applications (MHEAs).

"*M-health was defined as wireless telemedicine involving the use of mobile telecommunications and multimedia technologies and their integration with mobile healthcare delivery systems*" (Istepanian & Lacal, 2003).

There is an increasing number of mobile applications, such as Machine-to-Machine communications, Electronic-Health (e-Health), and Mobile-Health (m-Health) (Adibi, 2012). The development of Mobile apps has been growing rapidly in the last seven years with Mobile Health and wellness apps reaching 130000 by mid-September 2012. (Strickland, 2012). In addition, the number of mobile health apps more than doubled between 2015 and 2016, increasing by 57% to 259,000 apps (research2guidance, 2016). Moreover, the number was approaching 325,000 by the middle of 2017 (research2guidance, 2017).

The increase in mobile technology has led to explosive growth in the numbers of m-Health applications and users. Interest from both the health aspects and software development communities has occurred because of the technology revolution. In 2009, the inaugural m-Health Summit, a partnership between the National Institutes of Health, the Foundation for the National Institutes of Health, and the m-Health Alliance, involved 800 people. A year later, the number of people attending the same conference reached 2,400 (Aljaber, et al., 2015; Zhenwei Qiang, et al., 2011). This demonstrates large increase in demand over a short period. Learning and training on health matters are increasingly important issues; particularly as more people are living longer and because of the scale and complexity of support for health and well-being (Aljaber, et al., 2015; Shareef, 2006).

This type of health education includes several areas, such as learning how to receive appropriate treatment, how to manage life in an acceptable way and how to employ and manage health requirements without barriers (Glanz, *et al*., 2008). These requirements from the patient's side are paralleled by those from health professionals, who seek to update their medical knowledge and are looking for specific information, as well as other health professional's requirements.

Mobile devices however, have various restrictions, such as processor capacity, battery life, and storage resource limitations, which affect the quality of service and user experience (Silva, et al., 2013; Ventola, 2014).

Moreover, different healthcare scenarios/contexts demand different approaches to managing risk in both the development and use of medical applications. Mobile medical apps can pose the same risks of failure as other medical devices, due to mechanical failure, faulty design, poor manufacturing quality, and user error, among other safety issues (South, 2016). The aim of e-Health/m-Health is to reduce the time, the risk and costs for stakeholders. (Pagliari, 2007).

The above-mentioned requirements for the use of mobile phone apps for various purposes, such as health education, call attention to the lack of an effective evaluation framework to measure and evaluate these M-health applications / MHEAs, in order to certify that they meet the requirements, such as usability, and meet the needs of patients, health professionals and other stakeholders.

This PhD project represents a combination of evaluation methods (which are explored in chapter 3), in order to build a framework that can evaluate existing MHEAs. The framework contains different types of evaluation methods and a tool in order to help and guide different suitable MHEAs that meet different stakeholder requirements. This will in turn to increase the desire for using health education mobile apps by health professionals, patients and other stakeholders.

This project purposes to develop a novel evaluation technique and evaluation framework for health education mobile apps, as well as helping health professionals, patients and other stakeholders to find their most suitable MHEAs to meet their requirements.

The need to use several types of MHEAs dealing with health care or health education is increasing day-by-day. These requirements are from both health professionals and patients. These wide requirements encourage the creation of a vast amount of available commercial apps, which apps vary in quality, usability, error prevention and provision of appropriate information (East, 2014). Successful MHEAs should guarantee to provide users with what they expect from the apps, without having the barriers of time wasting, usability issues, flexibility issues, or errors. Therefore, the purpose of having an evaluation system is to make sure that the outcome from MHEAs will help health professionals, patients, and other stakeholders, help to improve MHEAs in the future, determine the suitability of MHEAs and assess wither it is worth using them (Taylor-powell, *et al*., 1996).

1.2 Motivation and Scope

It is clear that finding the appropriate health education is getting more complex with the passing of time. This complexity presents many challenges for the different stakeholders, making it difficult to find suitable technology/apps in order to help them to improving their health education. Effective technology/apps could save their time and effort and provide them with efficient and effective health education.

One of the existing solutions for health education is Health Education Solution (HES), which is a form of Online Medical Education and Certification that supports health education for patients and health professionals, but the drawback of HES is that not everybody can afford to pay for HES and have Online Medical Education and Certification (Solutions, 2016).

Another existing solution for health education is the LINDSAY Virtual Human project, which is modern software that permits faculty and students to model and manipulate three-dimensional anatomy presentations and images, while including embedded quizzes, links, and text-based content (Tworek, et al., 2013) . This type of technology is valuable but it is limited to computer using only. In addition, only helps students, it could be too complicated for patients, who are beginners with health information.

Finally, literature on measuring usability is limited (Hornbask & Law, 2007). In addition, it is a lot less within the range of mobile apps (Coursaris & Kim, 2007) .

However, different stakeholders depend on different health apps, and particular, different mobile health apps. Some of these apps are helpful to meet the requirements of different stakeholders. This thesis will highlight an evaluation framework including a medical selection tool, which will help in ranking different mobile health apps in order to help different stakeholders to find the most suitable mobile health apps to meet their requirements.

1.2.1 The importance of Mobile Health Education Apps (MHEAs) as a Research Area

Research done by Ziefle (2008) showed that one of the main development areas in computing is mobile devices. According to Hernandez and Woolley (2009)., mobile phones have become essential and common among users Humans depend on a mobile phone device in many aspects of their daily life, such as commerce, transport and health education. Mobile phones could be used for variety of purposes other than the conventional aspects, such as merely using a mobile phone as a flexible replacement for a fixed desktop computer by (Traxler, 2009). Several types of information could be accessed from anywhere and anytime, to anybody, by mobile phone among the wireless technology (Traxler, 2009).

There are two types of mobile phone devices. One of them is the classical type, which contains two parts, the display screen and the keyboard. The other type is the touch screen type, which can be ether a full touch screen controller or half touch screen controller and half keyboard bottoms developed by IBM controller. The first touch screen mobile phone was the Simon Personal Communicator in 1992. However, the first smart mobile phone used in the market, in 2000, was the Ericsson R380 (Lobo, *et al*., 2011). Seven years later, Apple released the IPhone; this was the first smartphone mainly controlled by its touch screen (Lobo, *et al*., 2011). After one year, only the Android operating system was in the market for touch screen mobile phones (Lobo, *et al*., 2011).

According to Leonardi, *et al*. (2010) a touch screen setup is more flexible and easy to use by novice users, rather than the keyboard and mouse setup. Which is highlighting the Potential of increase the depending on mobile phone.

Recent estimates showed that one person in every five persons in the world own a smartphone (Heggestuen, 2013). In addition, according to Hussain, *et al.* (2015) the latest estimates indicated that in 2017 the number of mobile devices per capita would increase to 1.4 billion, as reported by the Cisco Global Mobile Data Traffic Forecast 2013 Update. (Hussain, *et al*., 2015). Recent estimates indicated that about 429 million mobile devices added in 2016, and that smart devices represent 46 % of the total mobile devices. The total number of smartphones will be over 50 % of global devices by 2021 (Cisco , 2017). This reflects the vast growth in mobile devices, which increases the potential for having more applications installed and used in mobile phone devices. Finally, the number of mobile health apps approached 325,000 apps in 2017. (research2guidance, 2017). This brings to the fore the lack of a technique to rank and distinguish the useful apps from the less useful ones.

1.2.2 The Need to Construct an Evaluation framework

In the last decade, several researchers have developed usability evaluations for systems in general, but the majority were for web apps in general, not for mobile apps or MHEAs (Xie, et al., 2018; Bobian, et al., 2016; Boudreaux, et al., 2014; Alva, et al., 2003; Ivory & Hearst, 2001). Also Kumar and Mohite (2018) claim that there is still a lot of research work required in order to establish a substantial amount of knowledge in the field of usability in design of mobile learning apps. Moreover, they summarise their finding in their research that there is no focused source for reporting on usability research conducted for mobile learning apps. Furthermore, they recommended future work to develop a particular source for reporting on the usability of mobile learning apps, besides developing a guideline that can be used to design usable mobile learning apps.

Nowadays, usability has become a significant area for smartphones, as it is essential to avoid difficulty in employing mobile apps (Baharuddin, *et al.*, 2013). One of the main factors in mobile applications, success is achieving high-level user satisfaction. Therefore, the usability testing of mobile learning apps is required to confirm that mobile learning apps are effective, practical and easy to use (Kumar & Mohite, 2016; Ali, 2013). According to Coursaris and Kim (2011), the lack of empirical research on the impact of the environment on mobile usability and the significance of user characteristics is remarkable. Furthermore, they mention in their paper that there is a lack of qualitative study on usability among such mobile studies (Coursaris & Kim, 2011). This shows the need for constructing a qualitative study to improve, rank and measure the usability for mobile apps in general and MHEAs in particular.

Moreover, according to Coursaris and Kim (2011). There is as yet, remarkably no usability evaluation framework in the context of a mobile computing environment. This confirms the needs for our research, which is building an evaluation framework for MHEAs.

Furthermore, according to Smith (2013) smart mobile phone device ownership marks a momentous growth of 10 percent between 2012 and 2013. However, the use of mHealth applications began to increase in 2013. Over 36 percent of mHealth application publishers were released in the market between 2013 and 2014 (research2guidance, 2014). According to Dubey and Mridu (2012), Usability inquiry contains different methods such as surveys, logging, or interviews. Moreover, Usability Evaluation Methods are among the top approaches in term of usability to evaluate different aspects such as systems, applications and prototypes (Bernhaupt, 2009) .

1.2.3 The Role of a Questionnaire as a Part of Our Framework

According to Hussain, *et al*. (2015), One of the methods most regularly used to evaluate the usability of a Mobile application is the questionnaire. Questionnaires are made out of a number of questions, which require a set of participants to answer these questions after having experienced the application actuality studied.

The questionnaire is an appropriate way of gathering useful comparable data on different user's needs. Of course, the questionnaire is only able to produce valid and meaningful outcomes if the questionnaire questions are clear, and asked consistently across all respondents. This means we need to be careful when we design the questionnaire questions (Fox, et al., 2009).

Recent research indicates that Medscape is one of the top mobile health application downloads (Husain & Einerson, 2015), with over 100k applications. This suggests that Medscape is a very good example to utilize in the framework. By having this questionnaire flowing, the framework could rank several MHEAs. This will help software developers to determine the usability of these apps and they can use this information to achieve the best design of mobile health applications.

1.3 Research Hypothesis and Research Questions

The research hypothesis presented in this thesis considers that there are significant differences in the needs of distinct stakeholders for - and the effective usefulness of - Mobile Health Education Applications.

The research presented in this thesis will address five research questions. Research question 1 leads on to the subsequent 4 research questions:

Q1.) How can we construct an effective and efficient hybrid evaluation framework for medical health apps utilizing an amalgam of Heuristic Evaluation (HE) and more general Usability Evaluation (UE) metrics?

Q2. **How effectively can such an evaluation framework provide a suitable measurement in order to assess and rank Mobile Health Education Apps?**

Q3. How can the proposed evaluation framework be used to identify issues and to aid in improving the design of Mobile Health Education Apps by supporting software developers?

Q4. How far and in what ways can this evaluation framework be used to enable the specification and development of an appropriate tool to select suitable apps for different user groups? For this research, the different groups are Health Professionals, Patients and Other Stakeholders, including software developers/engineering.

Q5. How effectively can the evaluation framework be used to characterize the different stakeholders and their different requirements for MHEAs?

The research hypothesis will be tested by answering the research questions. These questions were addressed by building An Evaluation Framework for Mobile Health Education Apps (MHEAs) and Medical Apps Selection Tool (MAST) tool. In addition, the evaluation framework was applied to some existing MHEAs. There were three parts in this evaluation framework.

The first part was 15 interviews with 5 patients, 5 professionals and 5 developers, which helped to determine more than 80% of the specific requirements for these groups when from using MHEAs. As according to Nielson (2000), dealing with five testing users is sufficient to find around 80% of the result. (Nielson, 2000).

These 15 interviews were recorded by using mobile software called "Voice Record". The recording were then converted into 15 word documents. After that, an analysis of the data in these word documents was applied, in addition to coding it using the "NVIVO" software version 10, which is one of the top twelve software's for qualitative research (Predictive Analytics Today, 2015).

The second part of the evaluation framework was the questionnaires, which were designed as an initial version then amended based on the mapping of the efficient hybrid selection of the UE and HE. In addition, it was ensured that this mapping met the requirements of Professionals, Patients and Developers. By matching it with the outcomes of the interviews with Professionals, Patients and Developers In addition, the questionnaires helped to in designing the last part of the evaluation framework, which was the Medical Apps Selection Tool (MAST) in order to help several kinds of stakeholders to select their most suitable MHEAs in order to meet their specific requirements.

1.4 Research Objectives

To investigate the validity of the above research hypothesis, this thesis systematically researches, develops and evaluates the proposed methodologies in five major steps, which form the key objectives of the thesis:

Objective 1: Developing an effective and efficient hybrid evaluation framework for Mobile Health Education Applications by utilizing a mix of Heuristic Evaluation (HE) and Usability Evaluation (UE) metrics.

This objective aims to answer the first research question. To do this, we proposed the framework, developed in three stages of design. These designs developed during the research process. The first stage involved proposal of a design for the effective and efficient hybrid evaluation framework; for example in terms of efficiency, the framework reduced the number of factors from a large set of general factors set to five hybrid selected metrics. Moreover, in terms of effectiveness the framework provided positive outcomes, which proved it worked appropriately. The second stage was an updating of the first stage of design in order to specify the hybrid selected metrics for MHEAs. The third stage of design was a further updating of the second stage in order to develop the MAST tool.

Objective 2: Utilising the framework to develop a ranking measurements system so we can assess and rank Mobile Health Education Applications.

This objective aims to answer the second research question. To do so, the research applied interviews, which identified the priorities for the final hybrid selected metrics, in order to map and distribute them into the questionnaire questions. For instance, analysing the interviews provided a suitable metrics to map and use them in the evaluation questionnaire questions in order to rank and assess MHEAs. The evaluation questionnaire was the second part of the evaluation framework. It was distributed to three types of stakeholders health professionals, software developers, patients/ general public (HP, SD, P/GP) who already used several types of MHEAs, for developing their health education. The results from the questionnaire identified different rankings for different MHEAs. Moreover, it identified several MHEAs matching with the needs of different types of stakeholders, which aimed to direct different types of stakeholders to different MHEAs.

Objective 3: Using the framework to identify issues in Mobile Health Education Applications identified by the rankings, to aid in improving the design of Mobile Health Education Applications.

This objective aims to answer the third research question. To do so, the outcomes from the questionnaire were used to distinguish the positive and negative issues in the MHEAs. This will help to improve design and software engineering. Furthermore, the outcomes from the questionnaire described how the evaluation framework identified different stakeholders (HP, SD, P/GP) opinions on different MHEAs which will aid in improving the design of MHEAs.

Objective 4: Using the framework to develop a Medical Apps Selection Tool to select suitable Mobile Health Education Applications for the different stakeholder types.

This objective aims to answer the fourth research question. To do so, the data from the evaluation questionnaire were utilised to develop the MAST tool. For example, the matching section in Chapter 5 was used to create the profile data for the MAST, in order to help matching between different stakeholders (HP, SD, P/GP) for the MAST tool with different suitable MHEAs packages.

Objective 5: Utilising the framework to characterise the profile of different stakeholders in order to identify their different requirements.

This objective aims to answer the fifth research question. To do so, the outcomes from the evaluation questionnaire highlighted how the evaluation framework specified the characteristics of different stakeholders (HP, SD, P/GP) requirements. Moreover, the outcomes from the evaluation questionnaire enabled a comparison between different stakeholder’s requirements within the same MHEAs, to address their characteristics. Furthermore, the outcomes from the evaluation questionnaire delivered matching between different stakeholder’s requirements and different MHEAs.

1.5 Thesis Structure

The rest of the thesis is arranged as follows:

Chapter Two: Background Study and Literature Review

Chapter 2 defines the context of the research by discussing the generic dependability on MHEAs. In addition, it gives an overview of Heuristic Evaluation (HE) and Usability Evaluation (UE). A critical review of the literature highlighting the key principles, strengths and limitations of different usability method is provided to show the background and motivation for the methodologies developed in this thesis.

Chapter Three: The Structure of the Framework

Chapter 3 clarifies the strategies used to design the framework. In addition, this chapter explores the three stages the design went through. It also discusses one of the existing MHEAs. Finally, it discusses the design of the software and considers how to use it in the framework for designing the MAST.

Chapter Four: Metrics Selection Design and Qualitative Results

This chapter investigates the proposed selected hybrid metrics of HE and UE, so they are compatible with the architecture to map them to the questionnaire. Then, the chapter explores the differences and similarities between these metrics and between different stakeholders. The chapter also states the priority of measuring the metrics between different stakeholders, and finally measuring the weighting in between, the selected hybrid metrics and the different stakeholders.

Chapter Five: Designing the Questionnaire and Quantitative Results

This chapter investigates the quantitative analysis methods used in the evaluation framework and measures the usability possibility of several MHEAs, then expands them using the questionnaire to allow the evaluation framework to perform quantitative analysis. Three main types of MHEAs have been used, including Medscape, Epocrates and Other types of MHEAs (WebMD, UpToDate). The type of questionnaire used is a self-administered online questionnaire, which was used with three types of stakeholders Health Professionals (HP), Software Developers (SD) and Patients/General Public (P/GP) to collect and analyse different types of data from users with different backgrounds.

Chapter Six: Developing Medical Apps Selection Tool (MAST)

Having developed the methodologies to perform qualitative and quantitative analysis for the selected hybrid metrics under uncertainty chapter 6 illustrates how these metrics can be used. The thesis demonstrates this by applying the outcomes from Chapter 5 in the MAST tool, which will help stakeholders to select the most suitable MHEAs depending on user type and user requirements from MHEAs. As the tool contains various types of MHEAs, it can be used by different types of users.

Chapter Seven: Discussion

Chapter 7 provides an overview of the evaluation framework and guidance tool and discusses the finding of the research, mapped against the research questions and the objectives. In addition, the chapter discusses the approaches used to identify the efficiency of the framework, the tool, and improvements to the framework. Moreover, it provides an evaluation of the techniques used for the Interviews, Questionnaires and MAST. In addition, the chapter identifies some limitations of the research and offers recommendations for future research in MHEAs evaluation.

Chapter Eight: Conclusion

The final chapter contains a summary of all the chapters in this thesis and the outcomes, which are the key contributions of this thesis. Chapter 8 illustrates all the methods employed in this thesis, and evaluates, and interprets the research findings. The chapter ends with an explanation of the significance of the research and an overall conclusion.

1.6 Conclusion

This chapter has explained the rationale for conducting this research; this research is important because it addresses a growing area and because of the impact in enhancing MHEAs in terms of the user interface and usability. From the explanation of the research problem above, it seems sensible to explore further the usability evaluation area, principally that related to MHEAs. The following chapter therefore will examine previous research on usability evaluation for MHEAs and identify the potential area in which this study can contribute.

Chapter 2: Background Study and Literature Review

Chapter One provided the introduction to the thesis in addition outlined the objectives of the research. Chapter Two provides an overview of the literature on mobile applications and their evaluation, especially Heuristic Evaluation and usability evaluation. It begins with an introduction to Heuristic Evaluation and usability evaluation, then describes some of the usability metrics, including some definitions and explores some previous studies, for instance, evaluation models, previous studies of mobile devices and mobile HCI. In addition, it clarifies the limitation of these studies. An overview, of HCI and mobile apps. Then, it discusses usability issues in Mobile devices, such as memory capacity, limited connectivity and battery life. After consideration of, evaluation of mobile apps, the chapter concludes.

2.1 Introduction to Heuristic Evaluation and Usability Evaluation

In our daily life, measurement has become an important concern. For example. We calculate distance, and measure speed, in order to anticipate when we will reach a destination. This reflects that the measurement is essential for understanding the world. If there is no measurement, there is no method to control usability requirements or to evaluate after developing a product which it meets the requirements of its users (Jokela, *et al.*, 2006).

Measurement is utilized broadly in information systems, for the purpose of controlling software projects, which could verify whither the projects are successful. Measurement is required to evaluate the standing of projects, processes, products and resources (Todhunter, et al., 2017; Zuse, 1997). According to Norman and Shari (1998), measurement is defined as:

"The process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to describe them according to clearly defined rules".

The process of obtaining evidence about the characteristics of units referenced from this definition of the measurement in real measurement. A characteristic is a feature or aspects of a unit, while a unit is an object or an event in the real world (Norman & Shari, 1998) .

Measurement in Human Computer Interaction (HCI) research is based on principles of human performance from human factors and psychology. Several basics for measuring interfaces, including error rate, time taken, and completion rate, they are still in use today. In this thesis, the word 'measurement' is defined as the process for evaluating quality characteristics. In addition, the word 'metric' is used in evaluating the usability of an interface. Measurements are utilized in this research to develop an evaluation framework for Mobile Health Education Apps (MHEAs), which contains metrics for the purpose of assessing usability.

The main concern of usability is to detect how easy it is to use interfaces. Some years ago, information systems begin to be a part of our daily life. A key concern was to create system interface that were easy to use and easy to learn. In other words, usability is the most important issue for a system interface. Usability is defined by the International Standards Organisation (ISO 9241-11) (ISO 9241-11, 2018) as *"the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"*.

Many software developers utilize usability methods that they are familiar with, but not all methods are appropriate to all applications. The ISO 9241-11 Guidance on Usability (Effectiveness, Efficiency and Satisfaction) is the most cited when measuring usability (ISO, 1998). However, this guidance is very general and difficult to relate to specific domains (Bertoa, et al., 2006; Seffah, et al., 2006). However, numerous new ISO standards have been published since ISO 9241-11, such as ISO 9241-220, which defines usability in simple terms and concepts.

ISO/IEC 25022 and 25023 give examples of measures to use in usability evaluation, and ISO/IEC 25066, which was published in 2008, defines what should be integrated in usability evaluation reports for usability surveys, tests and inspections. ISO/IEC 25066 recognised that the extent of the variation needed in practice was too vast to support the development of a single suggested format (Theofanos & Quesenbery, 2005; Bevan, et al., 2016).

In the past, many usability metrics was purposely generated to be used in related to desktop applications. However, these metrics were not necessary applicable to mobile apps (Zhang & Adipat, 2005). Furthermore, the demand for mobile apps is growing enormously. According to Ankeny (2010), it was expected that the number of mobile app downloads would will approach to almost 50 billion in 2012. However, recent statistics indicated that the number of mobile app downloads worldwide in 2017 was 178.1 billion. Moreover, this number is expected to approach 258.2 billion in 2022. (Statista, 2018).However, according to Seffah, *et al.* (2006) , there is a lack of usability measurement for mobile apps; although, according to Brodkin (2008), mobile phone devices will be the primary Internet devices by 2020.

Within the context of this thesis, the term usability is represented as Heuristic Evaluations (HE) and Usability Evaluation (UE) to rank mobile applications, and in particular Mobile Health Education Applications (MHEAs).

2.1.1 Usability Methods and Techniques

There are several types of usability evaluation methods, some for websites or apps and others for devices. Usability evaluation methods include: cognitive modelling methods, inspection methods, inquiry methods, prototyping methods, testing methods and others. These main categories of methods include specific methods, such as Heuristic Evaluation, focus groups, questionnaires/surveys, usability inspection, Heuristic Evaluation, one to one interviews and some other methods and techniques. Figure 2.1 represents these methods, which part from the wider collection of usability methods available to utilize in this research. This section presents the three major methods utilized in this research, in order to develop a hybrid usability evaluation. They are: inspection, inquiry and testing methods.

Usability Inspection Methods: This is a formal usability inspection, which follow the software inspection methodology, then adjusts it to usability evaluation. (Nielsen & Mack, 1994). Usability inspection methods include various methods such as Heuristic Evaluation, cognitive walkthrough and other methods. This research utilized Heuristic Evaluations as a part of the Hybrid Usability Evaluation. Heuristic Evaluations is explored in depth in following section.

Usability Inquiry Methods: Usability Inquiry Methods are led through verbal inquiry, such as interviews, questionnaires, focus groups, or observations of the use of the apps. By the usability assessor in actual work situations, in addition to collecting users opinions in order to understand their preferences and needs from the system (Chen, 2018). This research utilized questionnaires and interviews as a part of the Hybrid Usability Evaluation. They are explored in depth in chapters 4 and 5.

Usability Testing Methods: Usability testing is the procedure by which products are tested by those who are going to use them. (Ghasemifard, *et al*., 2015). Usability testing methods include various methods such as metrics and other methods. This research utilized metrics as a part of the Hybrid Usability Evaluation. Metrics are explored in depth in section 2.1.3.

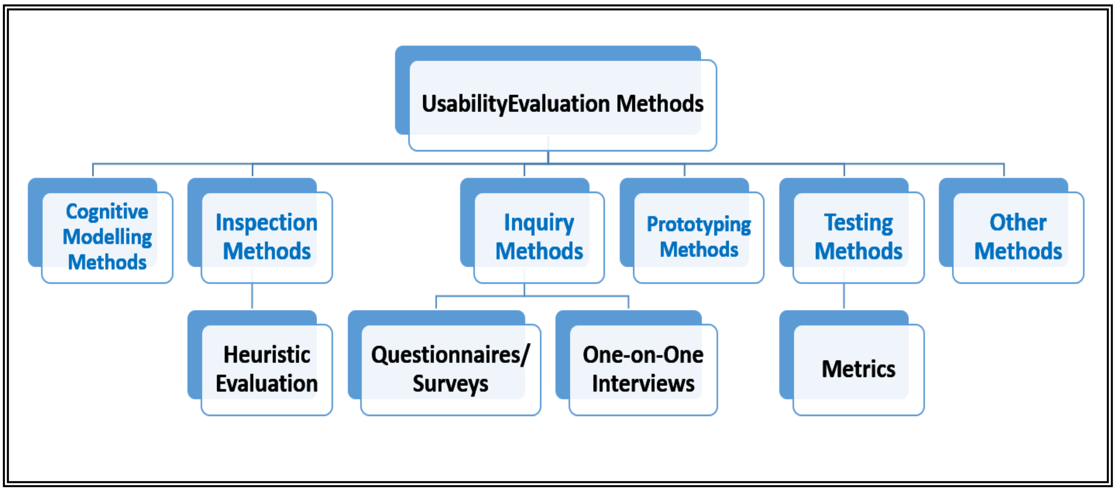


Figure 2.1: Overview Usability Evaluation Methods

This section highlights the specific usability methods, which were used in this research, including questionnaires/surveys, metrics, Heuristic Evaluation, one to one interviews and some other methods and techniques.

In regard to the questionnaire, this is explored in details in Chapter Five. Moreover, regarding metrics, these are explored in section 2.1.3. In addition, regarding the one to one interviews, this is explored in chapter four. Heuristic Evaluation is explored in section 2.1.2. Figure 2.2 illustrates the usability evaluation methods utilized for this research, which are one to one interviews, Heuristic Evaluation, questionnaires and metrics. The one to one interviews have been explained in chapter 4; Heuristic Evaluation has been addressed through selecting different types of evaluators. The Questionnaire has been explained in chapter 5; and finally the metrics have been addressed in chapters 3 and 4.

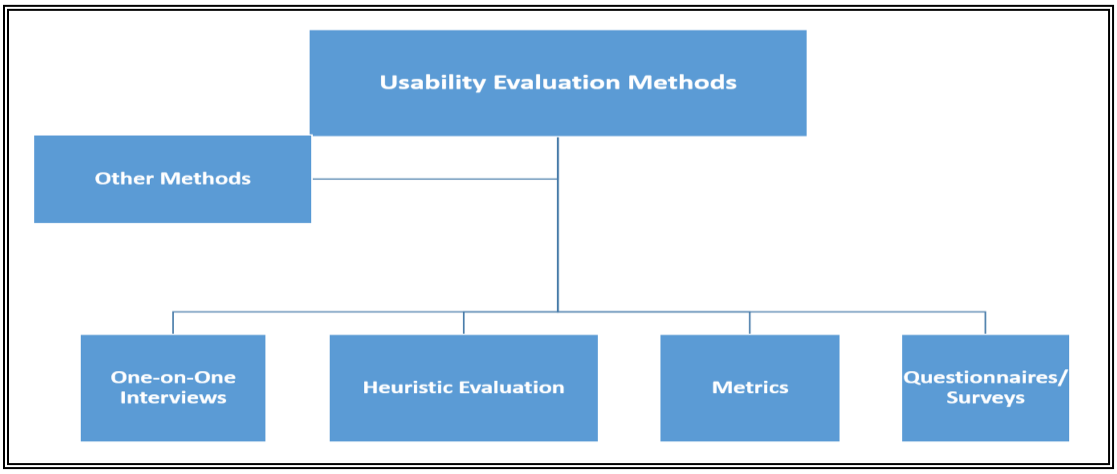


Figure 2.2: Focused Usability Evaluation Methods for this Research

Hybrid Usability Approaches

Literature proposes a few hybrid usability evaluation approaches, some of them for websites and others for mobile apps. However, there is a lack of integrating a true hybrid usability evaluation approach in order to evaluate mobile health apps, identify usability issues in such apps, and provide an effective ranking of them. This is necessary in orders to meet the specific requirements of different stakeholders from different mobile health apps. Below we present some of the existing hybrid usability approaches.

Research was done by Sivaji, *et al*. (2013) in regard to developing a hybrid usability methodology. They integrated Heuristic Evaluation with laboratory testing for this purpose. However, they employed the two approaches separately, then combined them afterwards in order to obtain the hybrid usability methodology. Also they only applied their methods to website interfaces. Therefore, our approach is a more integrated approach, since our approach is to develop a true hybrid through collecting different measurements and metrics chosen from the Nielsen Heuristic Evaluation and other general usability evaluation approaches; we then generate a set of proposed hybrid usability metrics. In addition we selected those following semi-structured interviews, in order to obtain the final set of hybrid selected usability metrics. This final set was then used within our evaluation framework to rank and measure the usability of different mobile health apps.

Research done by Tselios, *et al*. (2008) concerned the effective combination of hybrid usability methods in evaluating educational applications of ICT. They proposed using four categories of usability methods: usability inspection methods, user testing methods, exploratory methods and analytic methods of evaluation. They performed a mapping between the categories of usability methods with the classes of learning systems. However, they applied the full set of 10 general principles for Nielsen Heuristic Evaluation, without determining whether it is necessary to use all of them. They applied their research to three cases, all of which were for website and desktop applications (Tselios, *et al*., 2008). In contrast to their approach, in our approach we filtered our proposed metrics through semi-structured interviews, in order to determine the suitable metrics to measure usability issues. Moreover, our approach is for mobile apps, in particulars heath apps, which is shown in the application later.

Some related research was done by Brown, *et al*. (2016), regarding interface design recommendations for computerised clinical audit and feedback. However, they considered hybrid inspection usability methods only, by using a combination of Heuristic Evaluation and cognitive walkthrough methods, to develop a single protocol to identify usability issues. However, their approach was to use a hybrid mix of methods within the usability methods category along with inspection usability methods. In contrast, our approach is a hybrid of three usability categories: usability inspection methods, usability inquiry methods and usability testing methods.

Finally, some research done by Kronbauer, *et al*. (2012), focused on smartphone applications usability evaluation. They developed a hybrid model to evaluate smart phones. Their evaluation method involved integrating, in the same experiment, metrics data, contextual data and users' feelings about the used applications. They developed three units within their model, mapping unit, traceability unit and assessment unit, in for hybrid evaluation. However, it was not a true hybrid as they did not utilize both heuristics evaluation and usability evaluation. Also they did not determine specific metrics to use for measuring usability and finally, they did not focus on mobile health apps.

The Hybrid Usability Evaluation in this Research

The evaluation in this research is a hybrid usability evaluation, since we employed both Heuristic Evaluation and usability testing. Moreover, we included both of experts and novices for providing the feedback. Also, this research explored different profiles for them, which have been combined and in addition, we executed two or three tests at once. This is again a hybrid approach, which is explored in later chapters. Here, the Heuristic Evaluation is reflected by utilizing the expert part as (SD) who are experts in software (as experts should be expert in interfaces since interface design should be by software engineers). It also includes HP who are expert in mobile health apps. Furthermore, the usability testing was reflected by P who are novices in software and novices in mobile health apps, while SD are novices in mobile health apps, and HP novices in software. Figure 2.3 presents the hybrid usability evaluation from Heuristic Evaluation and usability testing. Then Figure 2.4 presents this hybrid approach which is a combination of usability testing by users and expert's user testing (HE) with a single set of metrics (5 metrics) which we distil from Jakob Nielsen's Heuristic Evaluation and other usability metrics.

Why Developing the Hybrid Usability Evaluation

This research focused on the design of Hybrid Usability Evaluation for three reasons. Firstly, to the best of our knowledge, there is no such research developed to evaluate mobile health education apps using a Hybrid Usability Evaluation approach. Secondly, there is no need to use all different types of usability methods to evaluate mobile health education apps. Thirdly, developing a Hybrid Usability Evaluation would provide a variety of, evaluation and ranking contributions through having different feedback from different evaluators and different evaluation methods. This would help in determining the specific procedure for developing mobile health education apps in order to meet the exact requirements of different users. According to Nielsen (1995), different people face different usability problems. Hence, it is possible to improve the effectiveness of the method significantly, by involving multiple evaluators.

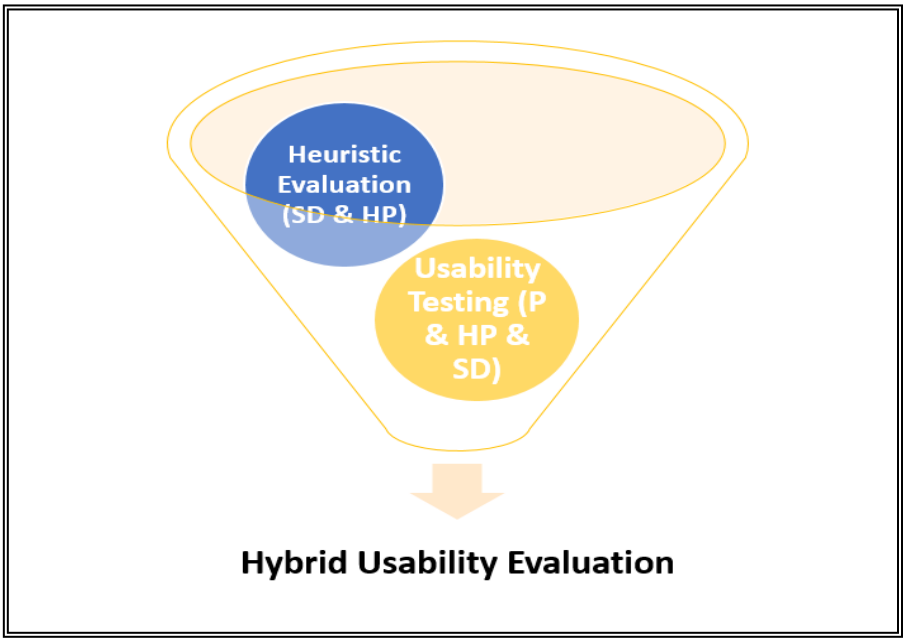


Figure 2.3: Hybrid Usability Evaluation

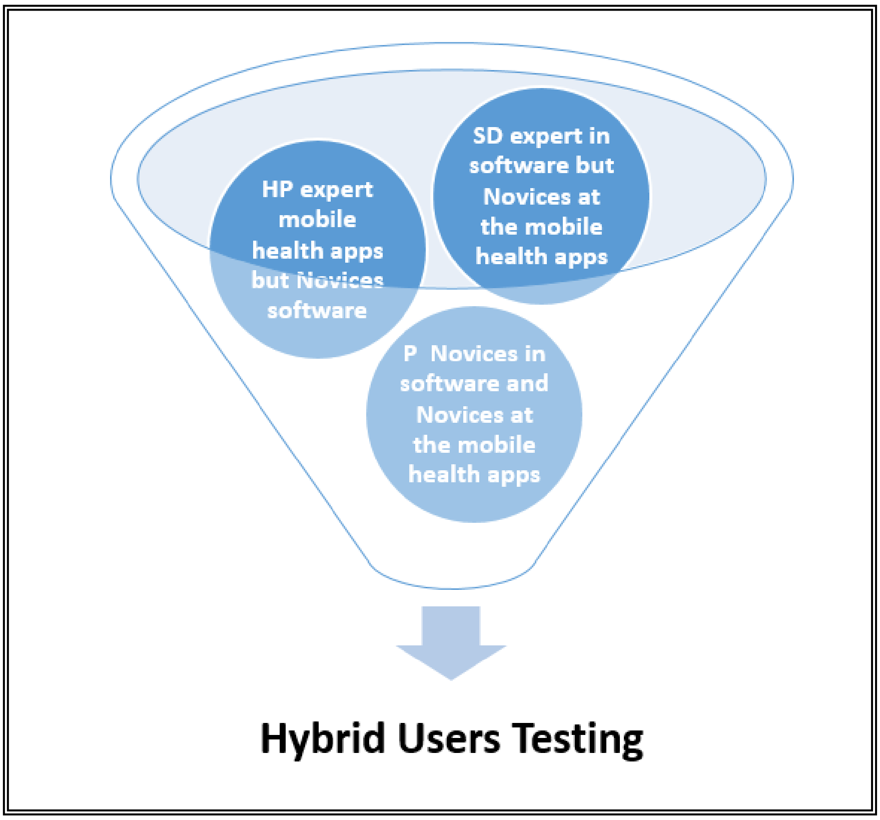


Figure 2.4: Hybrid Users Testing

2.1.2 Heuristic Evaluation

Heuristic Evaluation is defined by Nielsen and Molich (1990), as "*an informal method of usability analysis where a number of evaluators are presented with an interface design and asked to comment on it*''.

Heuristic Evaluation is an informal method of usability analysis, where several evaluators are presented with an interface design and required to comment on it. (Nielsen & Molich, 1990). According to Nielsen and Molich (1990), there are four ways to evaluate user interface: formally, automatically, empirically and heuristically. Formal evaluation involves some analysis technique, automatic employs a computerized procedure, empirical is by experiments with test users and heuristic is by simply looking at the interface and passing judgement according to one's own opinion.

They discussed four advantages of Heuristic Evaluation: it is cheap, it is intuitive and easy to motivate humans to do it, it does not need advance planning and finally it can potentially be used in the early stage of the development process.

The disadvantage of using Heuristic Evaluation is that it sometimes identifies usability issues without supplying direct suggestions to solve them.

According to Brayshaw, *et al*. (2014) although in practice evaluation in the context of educational software often hard to do; it is a vital part of the software development lifecycle. Although formal experimental empirical studies can be difficult to set up, organise and run, they discussed their own experience with a lightweight alternative method. They reflect in their paper on some issues they found, and provide details about their case studies to demonstrate the application of Heuristic Evaluation as an alternative possible solution. Brayshaw, *et al.* (2014) found that Heuristic Evaluation is an effective tool and quick to use.

This research focused on utilizing the principles of Nielsen's heuristics as a part of the hybrid evaluation, since it is one of most famous heuristics to use. However, many researchers have since utilized the principles of Nielsen's heuristics in their research. Appendix 11 illustrates some of the updates since 1995 of Nielson's heuristics.

Nielsen's heuristics

Nielsen’s heuristics “’*is a usability engineering method for finding the usability problems in a user interface design so that they can be attended to as part of an iterative design process. Heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognized usability principles (the "heuristics")* (Nielsen, 1995).

Heuristic Evaluation is done by looking at an interface and trying to come up with an opinion about what is good and bad about it, but according to Smith and Mosier (1986), most people probably make Heuristic Evaluation on the basis of their own intuition and common sense.

According to Nielsen (1989) in reality, the majority of user interface evaluations are Heuristic Evaluations. Nevertheless almost nothing is known about this kind of evaluation since it has been seen as inferior by most researchers. However, Nielsen expressed the belief that a good strategy for improving usability in most industrial situations is to study those usability methods, which are expected to see practical use.

According to Molich and Nielsen (1990), any system designed for people to use should be easy to learn and remember, effective, and pleasant to use. According to Nielsen (1995) , in general, Heuristic Evaluation is hard for a single individual to do, because one person will never be able to discover all the usability problems in an interface. Fortunately, experience from several different projects has shown that different people find different usability problems. Hence, it is possible to improve the effectiveness of the method significantly, by involving multiple evaluators.

The principles of Nielsen’s heuristics are: Visibility of system status, Match between system and the real world, User control and freedom, Consistency and standards, Error prevention, Recognition rather than recall, Flexibility and efficiency of use, Aesthetic and minimalist design, Help users recognize, diagnose, and recover from errors, Help and documentation. (Nielsen, 1995).

There are some other forms of Heuristic Evaluation, such as Gerhardt-Powals' cognitive engineering principles and Weinschenk and Barker's classification. Moreover, another important point to mention is user experience (UX).

User experience is "*a person's perceptions and responses that result from the use or anticipated use of a product, system or service*" (ISO DIS, 2008).

According to Allam, *et al*. (2013) User experience(UX) is measured for many reasons to help to improve it, to determine where a product stands relative to the competition, to help you in determining where to focus one's efforts and to inform anyone interested in improving the UX (Allam, *et al*., 2013)**.** Moreover, another important point to mention is Human-computer interaction (HCI) which will be discussed in section 2.4 in this chapter.

2.1.3 Usability Evaluation Metrics and Methods

According to Shivade and Sharma (2014), the usability metric is a very popular metric for evaluating the understand ability, learnability and attractiveness of a software. They mentioned in their paper that usability evaluation is a critical technique to estimate the function of human computer interaction (HCI).

According to Nielsen (1993), usability is defined by the following five components:

- **Learnability:** The user should be able to find the system easy to learn, which will allow the user to quickly get some work done with the system.
- **Efficiency:** The user should be able to use the system efficiently, once they learn it users should have a high level of productivity possible.
- **Memorability:** The user should be able to remember the system easily, so if the user does not use it for a long time, the user should be able to use the system without learning everything again.
- **Errors:** Users make a few errors during the use of the system, so the system should have a low error rate, so the users can easily recover from them; catastrophic errors must not occur.

- **Satisfaction:** Users should be satisfying with the system; they like it because the system pleasant to use.

According to Shivade and Sharma (2014) softwar usabilty evaluation is an approach to evaluate the usability of every software included in a mobile application. Usability metrics incloud the folowing factors: response time, accuracy, operability, attractiveness, learnability, satisfaction, productivity, effectiveness and efficincy (Seffah, et al., 2006).

2.1.3.1 Usability Analyzer Tool: A Usability Evaluation Tool For Android Based Mobile Application

Usability Analyzer Tool is a usability evaluation tool developed in Java. It is used for two purposes. The first one to calculate the usability of applications built on android. The second is to compare usability criteria for different but similar kinds of android application (Shivade & Sharma, 2014).

According to Shivade and Sharma (2014) usability engineering deals with the requirements specification of the user. They used three techniques to evaluate software usability, which are usability inspection, usability testing and usability inquiry. These are presented in Figure 2.5.

- Usability Inspection: focuses on usability related aspects of a user interface examined by software developers, users, usability experts and other professionals. In terms of Heuristic Evaluation, Shivade and Sharma focused more on examining the user interface for the applications by the usability experts, and spent some time on different users, software developers and other professionals.
- Usability testing: focuses on a prototype model of the actual system, and tests it to determine if the system is working as per the user specification or no. In addition, to find out whether the user interface supports the users to achieving their tasks.
- Usability Inquiry: focuses on using questionnaires to determine what users like and dislike.

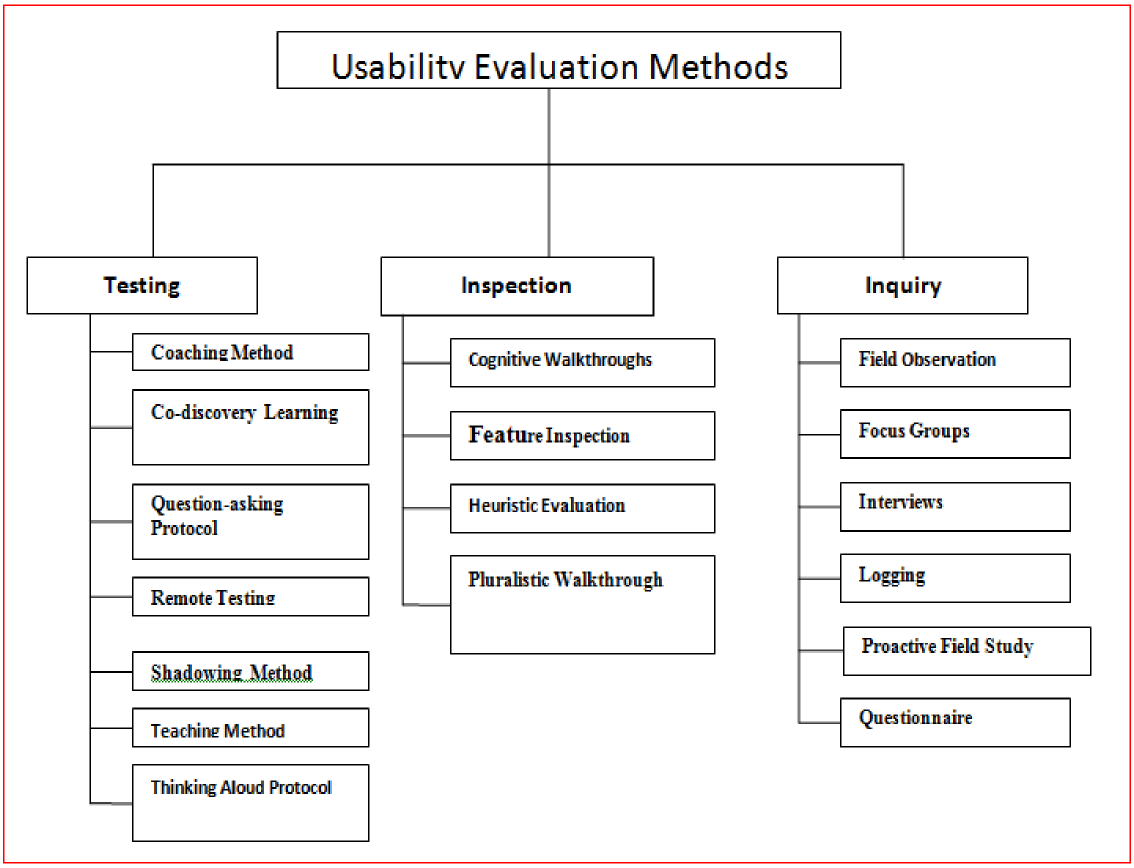


Figure 2.5: Usability Evaluation Methods

(Shivade & Sharma, 2014).

An advantage of Usability Analyzer Tool is that it enabled measurement of application usability from the real user's viewpoint. This will provide critical outcomes of application usefulness, based on the feedback taken from real users.

Shivade and Sharma (2014) considered two basic methods of usability evaluation in their tool; a questionnaire and logging. In the questionnaire method, they asked specific questions to users and received some answers from them these answers; they determined the usability of software. A questionnaire is very cheap usability evaluation method to use. The second evaluation method is Logging, which is collecting automatically all the statistics about the user and the system interaction. This method is extremely useful, as it provides accurate information on the use of the system.

Shivade and Sharma (2014) considered four android apps for usability evaluation. The evaluation consisted of four parts: usage of the apps, feedback questionnaires, usability evaluation using formulas and making usability graphs and comparison graphs.

1) Usage of apps: They described the usage of an application in this section.

2) Feedback Questionnaire: They discussed the feedback for the questions they used in their apps. They depended on efficiency, effectiveness, attractiveness, response time, productively and understand ability, as factors under usability metrics. These questions helped Shivade and Sharma to mesure the usability of any software products.
3) Usability evaluation using formulas: They discussed the usability factors they considered in their project, such as Efficincy (how users achieve each task efficiently [Efficiency = Task Effictivness / Task Time]). Learnability (how long does it take the user to learn how to use a function correctly [0 < T the shorter the bettre]. T means time taken to learn how to use a function correctly).
4) Making usability graphs and compairision graphs: They used this part to help organizations to find out usability problems. Organizations could improve product's usability by allowing the organiation to check products usability from the graphs. (Shivade & Sharma, 2014).

Overall Shivade and Sharma's (2014) usability evaluation tool was a usability result for software ranging from 0 to 1 and their calculation range is :

> 0 < usability factor < 1. So if the usability factor is near to 1, the software is more usable, and if its close to 0, the software is less usable. (Shivade & Sharma, 2014).

Shivade and Sharma's (2014) evaluation tool showed the lack of apropriate evaluation tools to evaluate software. Also they used their tools for general apps, but in our framework we are dealing with health education mobile appliaction. In addition Shivade and Sharma, unlike our proposed framework, focuss on usability only, wheres in our framework we adobt a new approach by combining usability evaluation with Heuristic Evaluation. Also they did not determine whether their metrics met the requirments of the users of the apps. This is a critical point to consider in order to detect the critical requirments to be mesured . From Shivade and Sharma's usability evaluation tool, we used some of the usability factors they used, in our proposed metrics, before filtering them through interviews, as well as some of the usability evaluation formulas.

2.1.3.2 Goal Question Metric (GQM)

Goal Question Metric (GQM) was created by Basili *et al*. (1994). GQM is a popular metric approach to measure the quality of software, which could be used to several measurement areas such as effort, schedule and process conformance. (van Solingen & Berghout, 1999; Basili, et al., 1994). Figure 2.6 presents GQM.

GQM was originally used to specify and evaluate goals for a particular project and environment. Its goal covered several areas, including improving quality, measuring progress and planning (Basili, *et al*., 1994).

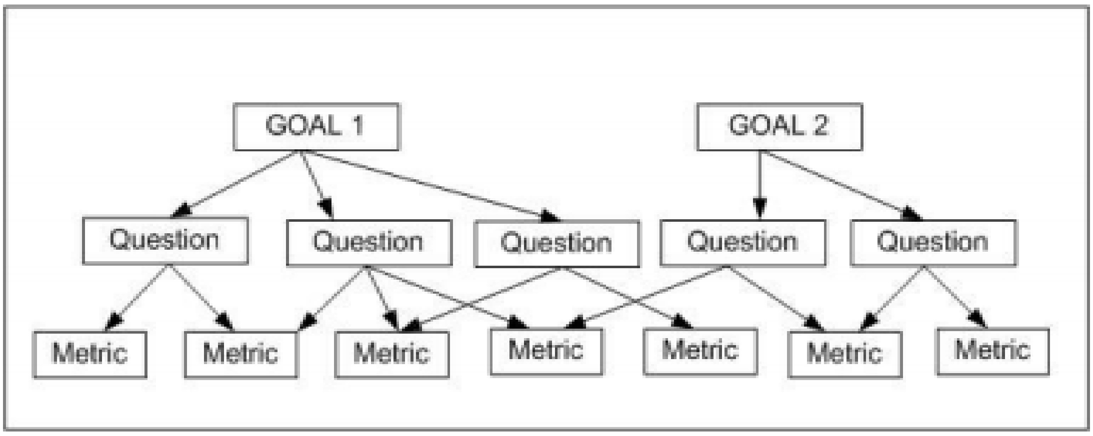


Figure 2.6: A GQM Model.
(Basili, *et al.*, 1994).

Hussain *et al*. (2012) developed a model mGQM from GQM. The goals of this model included measurement goals. Regarding usability of mobile phone apps and a group of question was posed to characterise to the likelihood of success. Finally, the metrics contained objective and subjective quantitative metrics, which were aggregated in order to answer the questions.

From these metrics, we can expand sub metrics such as mobile goal question metric (mGQM):

For example, taking attractive use as goal, we could formulate several questions, then take the answers to these question as metrics:

<u>Is the user happy with the interface?</u>

- Easy to find help.

- Satisfaction with signal indicator.
- Satisfaction with text.
- Satisfaction with system navigation.
- Satisfaction with touch screen.
- Satisfaction with interface.

Hussain *et al*. (2012) model has three limitations. First, they used only general mobile apps. Second, they did not identify the specific metrics to meet the requirements of different stakeholders. Third, they did not implement their model.

2.1.3.3 Quality in Use Integrated Measurement (QUIM)

Quality in Use Integrated Measurement (QUIM) is a consolidated model for usability measurement. (Seffah, et al., 2006). It is hierarchical in that it decays usability into factors, after that into criteria and in the end into specific metrics.

Seffah *et al*. (2006) suggested that the major application for Quality in Use Integrated Measurement (QUIM) is to supply a consistent framework and repository for usability factors, criteria and metrics for educational and research purposes.

According to Seffah *et al*. (2006), usability related to ease of learnability, without demands implying a high performance in task execution, e.g. regular or expert users might generally like to finish a large range and amount of tasks, with minimal obstruction, rather than fast completing a smaller range of tasks.

Seffah *et al*. (2006) based the QUIM consolidated model on ten usability factors. The inclusion of trustfulness as a usability factor in QUIM was based in part on the Cheskin Research and Studio Archetype/Sapient (1999). On user confidence in e-commerce sites. (Seffah, et al., 2006).

These factors are efficiency, effectiveness, productivity, satisfaction, learnability, safety, trustfulness, accessibility, universality and usefulness (Seffah, et al., 2006). These ten factors were not assumed to be independent; e.g. online clients could trust such a site only if they feel satisfied and safe when they use it.

These ten factors are divided into 26 sub-factors, called measurable criteria. Each criterion is measurable through at least one specific metric e.g. the factor called efficiency corresponds to the sum of criteria which are time behaviour, resource utilization, minimal action, minimal memory load, operability, feedback, navigability and loading time.

Seffah *et al*. (2006) used 127 specific usability metrics. By using some of their 26 criteria with some of their 10 factors, they define these metric in terms of formula and some simple countable data. The metrics that Seffah *et al*. (2006) use are shown in Table 2.1. These usability metrics have a number of limitations. For instance, the author's did not assess the metrics and some cannot be applied to other areas, like mobile settings.

	Factors									
Criteria	Efficiency	Effectiveness	Satisfaction	Productivity	Learnability	Safety	Trustfulness	Accessibility	Universality	Usefulness
Time behavior	+			+						
Resource utilization	+			+						+
Attractiveness			+						+	
Likeability			+							
Flexibility		+	+					+	+	+
Minimal action	+		+		+			+		
Minimal memory load	+		+		+			+	+	+
Operability	+		+				+	+		+
User guidance			+		+			+	+	
Consistency		+			+	+		+	+	
Self-descriptiveness					+		+	+	+	
Feedback	+	+							+	+
Accuracy		+				+				+
Completeness		+				+				
Fault-tolerance						+	+			+
Resource safety						+				
Readability								+	+	
Controllability							+	+	+	+
Navigability	+	+					+	+	+	
Simplicity					+			+	+	
Privacy							+		+	+
Security						+	+			+
Insurance						+	+			
Familiarity					+		+			
Loading time	+			+					+	+

Table 2.1: Relation between Factors and Criteria in QUIM
(Seffah, et al., 2006).

2.2 Introduction to Mobile Devices and Mobile Health Education Apps (MHEAs)

There are several types of technology that can be categorized as mobile. A mobile device is movable, portable and with a personal, as opposed to shared context of use. The expressions personal and mobile are frequently used interchangeably, but a device could be one without necessarily being the other (Rosas, et al., 2003).

Mobile devices, services and apps are of interested in people's day-to-day lives, on a private and professional level. Constructing a virtual environment of interaction is the main property of mobile devices (Kakihara & Sorensen, 2002; Pica, et al., 2004).

According to Fraunholz and Jung (2002) , mobile devices are characterized as follows:

A) Mobile devices are portable, which they are able to be carried comfortably.
B) Mobile devices allow some kind of communication. This communication may be either mobile communication, or offline data synchronization with a corporate database (McCormick & Schmidt, 2012).

In this research, we are going to focus on mobile apps and in particular mobile health education apps. There are several of Mobile Health Education Applications (MHEAs). Some of them are very useful for patients, health professionals and other stakeholders; and some of them are not. Appendix 10 explaining some of these applications.

This thesis is focuses on building an evaluation framework, to measure and evaluate MHEAs, depending on Usability Evaluation (UE) and Heuristic Evaluation (HE), discussed later in this chapter. The main aim from our project is to make sure that MHEAs are suitable, usable and satisfactory for patients, health professionals and other stakeholders.
The Global Observatory for e-Health (GOe) defined m-Health or mobile health as "*medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices"* (World Health Organization, 2011).

More recently, Istepanian and Woodward (2016) defined the mHealth as a combination of internet technologies and computing, with communication systems and information. Besides the sensors, it can form a wearable body area network with the patient's smartphone.

The study of the interaction between mobile users and mobile apps is called the mobile interaction. Mobile interaction study started in the last one and a half decades, beginning as a small study. For instance, Jorg (2002) examined the pattern of interaction for mobile devices, from popular software designs. However, he admitted the patterns not comprehensive, and required further investigation. Furthers study on mobile interaction become necessary due to the rise of modern mobile devices with several features, for instance built-in GPS, high-resolution camera, accelerometers and compass. Existing apps, for example apps in social networks in addition to enhanced network accessibility likewise prompted researchers to investigate mobile interaction (Jorg, 2002).

An example of a research to optimize the user interface for social networking to mobile users is the research done by Cui *et al.* (2010) , aiming to simplify social interaction on mobile devices. Cui *et al.* (2010) verified the interface with objective and subjective measures. However, they only used one measure, 'task completion time', for the objective measure.

Another similar research was that of Ruiz and Lank (2011) who examined motions and gestures by moving the device to raise commands. This research permitted mobile users to describe a set of motions and gestures for mobile interaction. They looked at whether the motions were reliable by requesting the users to offer feedback. However, this evaluation was built on qualitative feedback only (Ruiz & Lank, 2011).

According to Dix *et al.* (2004), evaluation assists the designers in addition to testing the systems to confirm that the systems behave as intended and meet the requirements of the user.

Moreover, the evaluation can differ significantly between the practitioner and the developer in interaction. Practitioners regularly evaluate interaction for usability bug. On the other hand, developers are expected to fix the bugs and the problems found. One aspect of the evaluation technique is the usability test, in which, the applications utilized by users are measured quantitatively to find if the applications fall within acceptable criteria (e.g. Time to complete a task, relative satisfaction, error rate). On the other hand, in mobile interaction the evaluation is quite infrequent and needs to be redesigned, as mentioned by Rukzio *et al*. (2007).

The number of health professional using tablet or smart phones in their professional capacity has increased during the last seven years. The GMC indicated that in 2011, 30% of health professionals used tablets or smart phone apps for reading in their professional capacity (Visser & Bouman, 2012). The number was expected to increase in 2012 and the expectation was that 83% of health professionals would use tablet or smart phone apps related to their professional capacity (Baumgart, 2011). More recently, a study presented that the total number of smartphone users worldwide would approach 2.87 billion (Statista, 2018).

Health professionals need to be able to evaluate MHEAs so they can make sure they are benefiting from using them to update their information and decide whether or not to recommend their use to their patients.

MHEAs increased greatly. This highlights the need to a type of methodology to evaluate this huge amount of applications. According to Molich and Nielsen (1990), any system designed for people to use should be easy to learn and remember, effective, and pleasant to use. (Molich & Nielsen, 1990).

2.3 Usability Metrics

Metric as a word utilized in several areas including mathematics, software engineering and HCI. Metrics in software engineering used to measure some property of a section of software intended at allocating quantitative keys of value to software (Goodman, 2004). According to Goodman (2004), software metrics is defined as:

"The continuous application of measurement-based techniques to the software development process and its products to supply meaningful and timely management information, together with the use of those techniques to improve that process and its products " (Goodman, 2004).

Metrics in software engineering are utilized to evaluate software and software reuse. Metrics for Software reuse are used to measure attributes using quantitative indicators (Frakes & Terry, 1996). Moreover, metrics are utilized to measure the performance of an organization's activities. An example is the work done by Gilligan and Golden (2009) which offers a conceptual framework for social impact and applying evaluative metrics to evaluate the framework. (Gilligan & Golden, 2009).

However, Pressman (2010) notes in the difference between measurement and metrics. Measurement comes about as the result of the collection of one or more data points, whereas software metrics try to relate the measures in some way, such as the average number of errors found per review (Pressman, 2010).

Metrics in HCI have been utilized to measure three things, which are:

1) User experience
2) Usability
3) Human interaction activities.

According to Holzinger*et al.* (2008), metrics are effective for measuring usability for specific groups of end-users, such as the elderly (Holzinger, *et al.*, 2008).

There are three established usability metrics dominant in usability evaluation: effectiveness, efficiency and satisfaction. These three attributes constitute the quality characteristics standard in ISO 9241-11 (Guidance on Usability) (ISO, 1998) .

Efficiency and effectiveness are objective characteristics but satisfaction is a subjective measure. Numerous studies, for instance Georgiev and Georgieva (2010) and Stickel *et al*. (2010) recommended metrics for mobile settings for instance, "time taken to read text". Nevertheless, the majority of these metrics faced limitations. (Georgiev & Georgieva, 2010; Stickel, et al., 2010)

From the above, we can tell that it is hard to find specific metrics in order to meet the requirements of measuring the usability of software. This suggests the need for using qualitative interviews, in order to determine the most suitable measuring metric to be used for measuring a specific software, with the purpose of matching the metrics with what we need to measure from the software. This will be explained in details in Chapter Four.

The study by Seffah *et al.* (2006), mentioned earlies in this chapter in section 2.1.3.3 followed IEEE_Std_1061 (1998) standard (Software Quality Metrics Methodology) for the purpose of deriving usability metrics by decomposing usability into factors, then into criteria, then finally into specific metrics. These usability metrics, however, have a number of limitations. For instance, he did not assess the metrics and some are unable to be applied to other areas, like mobile settings (Seffah, et al., 2006; IEEE_Std_1061, 1998).

Furthermore, research done by Nokia in particular Nokia lab (Nokia, 2008) produced usability guidelines based on a usability evaluation conducted on ten applications. The guidelines are:

1) Use the touch screen.
2) Remember mobility.
3) Use the key too.
4) Provide clear navigation.
5) Provide useful feedback.
6) Save automatically.
7) Use screen space.
8) Do not annoy the users.
9) Be consistent.
10) Provide help.

However, Nokia offered the most significant guidelines only for their own Smartphone. In addition, many guidelines could be considered aimed at usability metrics, particularly for mobile applications. Nevertheless, these usability guidelines need to be refined, and new guidelines must be added for the latest mobile phones.

Moreover, four items in a 'Likert scale' were used only for the subjective assessment developed by Finstad (2010) they are:

1) [This system's] capabilities meet my requirements.
2) Using [this system] is a frustrating experience.
3) [This system] is easy to use.
4) I have to spend too much time correcting things with [this system] (Finstad, 2010).

The metric was cut down from a ten-item scale System Usability Scale (SUS), to a four-item scale. The research did contain a definition according to usability by ISO 9241-11, however, the study supplied only subjective measures only and did not include usability issues for mobile settings (ISO, 1998).

Moreover, Holzinger *et al*. (2008) suggested usability metrics to assess apps aimed at the elderly. Nevertheless it would of more help if they could extend the metric to applications in mobile devices.

2.4 Mobile Human Computer Interaction (HCI)

The study and concept of the interaction between humans and computers called Human Computer Interaction (HCI), which is emphases how the interactions is affected by the design of technologies, and in what ways these designs might best be used. Human Computer Interaction is a wide research area in computer studies. Other terms referring to the same area are User-Computer Interaction (UCI), Computer-Human Interaction (CHI) and Man-Machine Interaction (MMI). HCI focuses on the interaction between humans and computers, for instance the interaction by user with the characters displayed on a computer screen. HCI is a multi-disciplinary subject, which includes not only ergonomics and software engineering, but also cognitive science, cognitive psychology, mathematics, social psychology, organizational psychology, computational linguistics, artificial intelligence and sociology.

Benbasat (2010) recommended study of HCI focusing on three characteristics: user base expansion, social networks and online services. The Apps on social networks are the highest in terms of user demands and provide rich potential, for HCI research. Several entities, for instance person to group, person to technology, person to person, involved in social networks, make for stimulating research in this area (Benbasat, 2010).

Several contexts for online services, for instance e- government, offer opportunities for researchers to discover. Moreover, these online services frequently include different types of users, such as beginners, professionals and the elderly. Benbasat(2010) also discussed expansion of HCI research focusing on specific users, to a broader perception of users. For instance designing interfaces for elderly users is demonstrated to be more challenging, especially in the areas of health care and e-government (Benbasat, 2010).

Nevertheless, Benbasat (2010) in his study did not include mobile HCI as a remarkable topic, although it can be realised that Mobile devices and Mobile apps are utilized by all of these groups in all of these areas (Benbasat, 2010). Moreover, other researchers for instance Lyytinen (2010) argued that there are three problems besides challenges in HCI research, which are:

A) Concern for environmental validity
B) Richer notions of cognition
C) Growth and access to new sets of data (Lyytinen, 2010).

Lyytinen (2010) states that these three issues are important to the information systems field. However, the interface on mobiles is also significant to information systems research, and should be included in his discussions. (Lyytinen, 2010).

Yeratziotis and Zaphiris (2018) said in their paper that the aspiration of Heuristic Evaluation (HE) is to help HCI experts and Web developers in designing the websites and evaluating them.

One of the significant issues in HCI is how to measure usability (Hornbask & Law, 2007) . Moreover, according to Seffah, *et al.* (2006) measuring usability is an important task to ensure the accuracy of the application and the safety of the user from strain injury. Meeting the usability and user experience is a key element in building successful and high-quality applications.

Caro-Alvaro *et al*. (2018) discussed in their study the growth of mobile devices, beside the enormous number of apps available in the stores. This calls attention to the lack of evaluating of the usability of such apps to provide a more satisfying user experience.

According to Booth (2014), is HCI defined as '*the study of the interaction between humans and computers*'. This is very similar to the definition of Kumar and Kumar (2005) , which is '*the study of interaction between people (users) and computers*'. Nevertheless, neither of these definitions signify the true complexity and multidisciplinary nature of the subject. HCI is a specialist field relating to computer science, cognitive science, human factors and information science and other fields (Cairns & Cox, 2008). The definition should follow the area and the interest of study, according to Helander (2014), they defined HCI to improve design and their definition is:

"In human-computer interaction knowledge of the capabilities and limitations of the human operator is used for the design of the systems, software, tasks, tools, environments and organizations. The purpose is generally to improve productivity while providing a safe, comfortable and satisfying experience for the operator" (Helander, 2014).

However, this research thesis focuses on applications inside mobile phone devices in fact specifically the Mobile Health Education Applications (MHEAs) so the definition can be precisely formulated as:

"The study of the interaction between humans and mobile health education applications (MHEAs)"

This is the definition which is intended whenever the term mobile HCI is used throughout the thesis.

Usability is one of the main issues in considerations in the evaluation of HCI. The term usability is commonly approved to be defined for HCI determinations by Shackel (1981), and referred to as:

"The capability in human function terms (of a product) to be used easily and effectively by the range of users, given specified training and user support, to fulfil the specified range of tasks, within the specified range of environmental scenarios" (Shackel, 1984) .

Moreover, the above definition few years later was developed into numerous international standards. ISO 9241 – 11 ISO (1998) 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" (ISO, 1998).

Usability includes a set of attributes which relate to the effort required for use, in addition to the singular assessment of such use, by an identified or implicit set of users and the ability of the software product to be learned, used, understood, and attractive to the user, when used under the specified conditions (Botella, *et al*., 2004) . ISO 9216 overcomes the apparent discrepancy between these definitions by introducing a further perception of Quality of Use as the capability of the software product to allow specified users to achieve specified goals with effectiveness, productivity, satisfaction and safety in specified contexts of use.

2.5 Usability Issues in Mobile Devices

Mobile devices are small sized computing devices that contains a display screen along with of touch input or a small keypad.

The substantial issue of m-health systems is usability, as it is the key factor assisting their success (Lacerda, *et al*., 2014)

According to research by Kumar & Mohite (2018), very few usability tests have been developed for mobile learning apps. However, the studies on mobile devices focusing on interaction have arisen over the last one and a half decades. The first phases of research into mobiles in around 2001 focused on minor areas, such as discussions on design issues on context aware systems and investigations into interaction patterns. However, recent mobile research now concentrates on several aspects, due to the expansion of mobile phone devices. Earlier research on mobiles is quit relevant but several areas have been discovered. There are some limitations on mobile devices, for instance the memory capacity and limited connectivity, which have been overcome now by the newest mobile devices, in addition to anytime everywhere network connections Wi-Fi, 3G or 4G. This situation has given rise to a new and remarkable research area into the little short battery life due to time-consuming network connection. Recently, study on mobiles devices has developed in several areas, for example social media and networks, mobile accessibility, activities in a mobile context, maps and navigation, multimodal interaction and mobile evaluation. The focus of the majority of the recent research is on how mobile devices can be utilized by the disabled community in mobile accessibility.

A study by Risald *et al.* (2018) concerned Mobile Application Design for deaf or hearing impaired patients. This research only focused on the deaf or hearing impaired patients and the design of the Mobile App For Emergency Calls. Moreover, they only focused on applying the User Centered Design (UCD) method, since they focused on designing mobile medical emergency call apps for deaf people. Another research done by Hussain *et al.* (2017) was about the UX of Amila pregnancy on mobile devices. This research focused only on pregnancy and pregnancy mobile apps and only applied Jakob Nielsen's five usability principles (effectiveness, efficiency, learnability, memorability and satisfaction) without considering specific metrics to meet the stakeholder's requirements. Another study by Zahra *et al.* (2018) focused on identifying the factors that affect chronic disease health apps. They focused only on usability factors affecting such apps. They did this by using interviews, without conducting any quantitative research. Another study by Spillers and Asimakopoulos (2014) was a diary study comparing mobile health applications. They considered only three different m-health running apps and they only performed a qualitative study. Another research by Liu *et al.* (2011) classified mobile-health applications for iOS devices into different groups. However, they only focus on classifying the apps according to their purposes, functions, and user satisfaction.

Moving into early research relative to the literature, one app was developed by Buttussi *et al.* (2010) to examine communication between deaf patients and emergency medical responders. In addition, they evaluated the system. However, they only used three evaluation characteristics: effectiveness, learnability and understandability without evaluation of mobile characteristics. Another similar situation arose in a research by Guerreiro *et al.* (2010) and Chittaro and Marassi (2010) with only one evaluation characteristic (error rate). Moreover; research in social media has increased significantly and repeatedly appears in the top conferences, for instance the Mobile-HCI 2010 and the CHI 2011 conference. This has occurred due to the growing number of users, on social media such as Twitter, Facebook and YouTube. According to Alison (2011), the networks and social media becomes more popular than entertainment in web destinations. Recent research on social media and networks is now developing on mobile settings, such as the study by Wagner et *al*. (2010) which mentioned uses of mobile devices to investigate location sharing on social networks. Wagner *et al*. (2010) collected mixed data, quantitative and qualitative, to stimulate user's location sharing practices in the context of online social networks.

A significance area of mobile devices, which has emerged is usability evaluation study in HCI (Schusteritsch, *et al*., 2007) . A number of researchers have engaged in mobile device evaluation, such as Hubert (2006). She did research on the usability and accessibility of mobile devices in the area of healthcare. Hubert's (2006) research outcome was usability guidelines for designing and testing mobile devices, utilized by older adults. Similarly, Gong and Tarasewich (2004) defined general usability guidelines for mobile devices. Moreover, Shneiderman interface (1998) created interface guidelines, which as a preliminary point for proposing a set of practical design guidelines for mobile device interfaces. However, it can be utilized only for designing interfaces for mobile devices not to evaluate the interface.

A comparison study by Duh *et al*. (2008) combined field tests and laboratory in order to evaluate the usability of mobile phone devices. They reported that there were several types and occurrences of usability problems discovered in the field compared to the laboratory test. A few of these problems were related to the devices actuality used in the field, while some others involved external factors linked with the environment of use, for instance the privacy issue when the device in a crowded place, noise, the mental and physical resources and the extent to which the users were affected by these factors. Another limitation to mobile devices is the difficulty of data entry in fieldwork (Duh, *et al*., 2008).

Moreover, the exceptionality of mobile phone apps, and the exclusive features of mobile devices have become key challenges in evaluation action. These unique features of mobile devices include unreliability of wireless networks, limited bandwidth, changing mobile context, limited memory and small screen size. Furthermore, contemporary technologies such as GPS receivers, high-resolution cameras and compasses are inserted into mobile phones, which effect of the usability of mobile apps. Numerous traditional usability metrics have been suggested. Nevertheless, these metrics might not be directly appropriate for mobile apps (Coursaris & Kim, 2007; Zhang & Adipat, 2005). Thus, this study explores the novelty of mobile apps, which influences the usability of health education applications in mobile phones.

A study done by Sauro and Kindlund (2005) suggested a method to make all the ordinary aspects of usability subject to a single standardized usability metric. This method supplies a continuous variable for usability evaluations, which could be in use for several things such as in testing, analysis and usability reporting. However, these metrics not take into account various aspects of measures, such as ease of use and learnability, which could be significant for some contexts (Sauro & Kindlund, 2005).

2.6 Evaluation of Mobile Apps

Nowadays evaluation has become a vast research area. One of the original rules used to evaluate applications was called the 'ISO 9241 – 11 standards'. In fact the 'ISO 9241 - 11 standards' has been used for a number of years. These guidelines concentration on general usability metrics, which are disputed by others, who argue in favour of using specific usability metrics. However; the difference between specific and generic has been a matter of opinion. In addition, it has not been noticeable in literature on the comparison of usability metrics, for instance (Seffah, et al., 2006; Bertoa, et al., 2006).

Moreover, there is an increasing number of mobile apps. Such as Mobile-Health (m-Health) (Adibi, 2012). The emergence of Mobile apps has been growing rapidly within the last six years, reaching more than 130000 apps by the end of 2012 (Strickland, 2012), and 259,000 apps in October 2016. (research2guidance, 2016). The number close to 325,000 apps in late 2017. (research2guidance, 2017). This great increase in mobile technology has led to an explosion in the number of mobile health (m-Health) apps and users.

The interest in these mobile health apps, from both health aspects and software development communities has arisen, as an outcome of the technology revolution. In 2009, the inaugural m-Health Summit, in cooperation among the National Institutes of Health, the Foundation for the National Institutes of Health, and the m-Health Alliance, involved 800 people. The following year this number of people who were attended the same conference reached 2,400 (Zhenwei Qiang, *et al*., 2011).

Moreover, user's perceptions are captured, by development of procedures for identifying the problems that users could have while trying the software system, or which evaluations is based. These results based on user perception are grounded on the real views of valid users on interface problems.

Nevertheless, user evaluation is time consuming and expensive. Expert based evaluation is involves design reviews of software projects and models. Based evaluation is based on expectations of performance from the model. Expert based the model based evaluations are less expensive. Therefore, several interface designs can be tested. Furthermore, whether evaluation is expert based, model based, or user based, so far there is no agreement among the relevant community nearby on which evaluation method is the most practical (Scholtz, 2004) .

Moreover, evaluation on mobiles apps is substantially different from desktops. This is because of the exclusive features of mobile phones, for instance small screen, limited bandwidth, changing mobile context (e.g., location) and limited memory (Harrison , *et al*., 2013).

On the other hand, numerous researchers have recommended methods and guidelines intended for evaluation of desktop applications, but these might not be suitable for mobile apps (Zhang & Adipat, 2005; Coursaris & Kim, 2007). Examples of researches on evaluation mobile include De Sá and Carriço (2008) and De Sá *et al*. (2008). They suggested a design of a framework as well as evaluations for mobile apps (De Sá & Carriço , 2008) , (De Sá, *et al.*, 2008) . In another research, Gong and Tarasewich (2004) produced a user interface guidelines for mobile devices. Research was also done by Gafni (2009) who focused on usability issues on mobile wireless systems. Nevertheless, none of the above studies provides usability metrics in the evaluation.

In another research done by Gafni (2008) her concern was with mobile quality metrics. However, her focus was on software issues. Also, she offered quality metrics, for instance 'Network throughput' and 'Display load', but the research did not deal with usability.

Usability metrics are commonly used by evaluators in order to determine what they are going to measure. One of the common usability measures is the 'ISO 9241 part 11'. The guidance on usability as presented in Table 2.2. This measure is widely utilized. The issue with this measure is the difficulty of applying it to specific domains. In addition, it is too general, which makes it not associate any quality characteristic to the measure (Bertoa, et al., 2006; Saffah, et al., 2006).

Effectiveness	Efficiency	Satisfaction
-Percentage of goals achieved	-Time to complete a task	-Rating scale for satisfaction
-Percentage of users successfully completing task	-Tasks completed per unit time	-Frequency of discretionary use
-Average accuracy of completed tasks	-Monetary costs of performing the task	-Frequency of complaints

Table 2.2: Usability Measures
(ISO, 1998)

Evaluating mobile apps can face a number of challenges. Nowadays multimodal mobile apps are developing increasingly. Multimodality combines touch and voice through a keypad or stylus as input with relevant spoken output. Also there are on screen visual displays for the purpose of improving the mobile user's experience. In addition, there is an increase in offering from the network operator service. A mix of multiple access channels offers new avenues of interaction to the users. Again, it pose however, they challenges to usability evaluation.

There are two approaches commonly employed by evaluators for the purpose of evaluating mobile apps. According to survey by Kjeldskov & Stage (2004), 71 percent of all mobile HCI evaluations were achieved in the laboratory and only limited predictable methods were customized in order to encounters the growing challenges of mobile apps evaluations (Kjeldskov & Stage, 2004). The data collection techniques could be the reason for this. For instance video recording, think aloud or observations are difficult in the field. However, evaluation in the lab has several advantages, for example, controllability and reproducibility. Nevertheless, the disadvantage of the lab experiments is their lack of realism.

However, the differences and similarities among field and lab based evaluations of mobile apps have begun to be discovered. Many comparisons have been made to observe the differences in the interaction behaviours in laboratory and field settings. For instance research by Baillie (2003) and Pirhonen et al. (2002). Led to the conclusion that it is useful to do an evaluation in the field, although it is problematic due to problems in capturing screen contents and the interaction between the user and the mobile device.

In the last decade, touch-based mobile devices have appeared, which has led to increased efforts to adapt professional prototyping tools to the mobile context. Nevertheless, one of the existing mobile prototyping solutions adapts sufficiently to meet the needs of different stakeholders (Hastreiter, *et al.*, 2014).

In research by Kumar and Mohite (2018), they note that research on Usability evaluation of mobile learning apps is an active area of research.

2.7 Mobile HCI: Current and Future Viewpoint

Mobile research takes place in several areas in HCI. One of these areas is mobile broadcasting, which allows the users to broadcast live video to websites on the Internet from mobile devices.

An investigation was done by Juhlin *et al*. (2010), regarding the potentials and barriers to the adoption of mobile broadcasting. In addition, they offer qualitative content analysis of a sample of four services. They note that how to evaluate mobile apps such as mobile TV is an area that could be expanded because of the unique characteristic of mobile broadcasting, for instance the battery life and the small screen.

Another novel research area in mobile apps, is navigation systems and mapping off mobile phones. Having a satellite navigation system in Mobile phones saves users from having to carry an extra device, i.e. a satellite navigation device.

Kratz *et al*. (2010) did a study on mapping and navigation systems. Their study illustrations that the only broad usability characteristics have been employed, for instance learnability, usefulness and satisfaction. Improving the evaluation to make a systematic evaluation could be a suggestion, by including a various measures such as safety, accuracy, or response time.

Studies on multimodal interactions have arisen quickly through many investigations into input techniques, for instance pen, touch, speech, gaze and head and body movements.

Exploration of new tag based interaction was done by Hardy *et al*. (2010) and Broll and Hausen (2010). This is in the area of Near Field Communication (NFC), which is like RFID (Radio Frequency Identification), which supports mobile phone interactions with tags simply by touching them (Hardy, et al., 2010; Broll & Hausen, 2010).

These new types of interaction will create additional space for researcher, to explore and evaluate them. NFC and additional specialised input techniques, for instance gaze and head and body movements are outside the scope of the present research. Nevertheless, as discussed later in the thesis, a key strength of an evaluation framework for MHEAs including a Medical Apps Selection Tool (MAST) is the ability to be adapted and so evaluation of these would be possible.

2.8 Conclusion

In this chapter, a description was provided of the research background on mobile HCI, in addition to an introduction to mobile devices and Mobile Health Education Apps (MHEAs) and some of the challenges involved in undertaking research on mobiles. Methods to evaluate mobile apps were clarified and we presented previous and recent studies that focused on mobile HCI and usability metrics. In addition, we provided an introduction to Heuristic Evaluation (HE) and Usability Evaluation (UE). Some issues in were identified related to mobile devices such as theoretical studies, those that focus in mainly on desktops and some that served for trade only. In addition, some previous studies of mobile phone evaluation and their limitations were reviewed.

Chapter three will discuss the designs stages for the framework. Moreover, it will clarify the reasons for moving from one design to another. Finally, it will show how the metrics were mapped onto the questionnaire questions.

Chapter 3: The Structure of the Framework

Chapter Two explained the core terms used in this thesis and reviewed previous research related to mobile devices in HCI as well as usability metrics. It provided an introduction to usability evaluation methods, Heuristic Evaluation and usability evaluation and HCI and the importance of MHEAs. Chapter Three will introduce information about structuring the evaluation framework for MHEAs. Which is allocating objective 1 in this thesis. Which is developing an evaluation framework for MHEAs by utilizing a mix of HE and UE metrics. Furthermore, it will explain the three main designs and the updating of the framework during the research stages. The framework is a project to reflect health education as MHEAs. It aims to satisfy the requirements of various stakeholders based on measuring the usability of existing MHEAs and offering the MAST Guidelines to help different types of stakeholders in selecting the most suitable MHEAs to meet their requirements.

3.1 The Research Methodology

This thesis uses a mix of quantitative and qualitative data collection and analysis to build an evaluation framework for MHEAs and MAST, in order to answer the research questions. This approach was evaluated using a case study of MHEAs. The framework has been implemented as a MAST Tool and this has been tested with some users.

The first step consisted of qualitative interviews about proposed selected hybrid metrics. Selected from Heuristic Evaluation and Usability Evaluation. These semi-structured interviews will be explored in depth in Chapter Four. The reason for these interviews was to identify appropriate selected hybrid metrics, which would meet most of the requirements of most stakeholders of MHEAs. These metrics were to be utilized based to measure the usability of MHEAs. Moreover, the selected hybrid metrics would be the basis of the questionnaire, which will be explained clearly in Chapter Five. Furthermore, they were to be deployed to measure usability of different MHEAs, with the aim of ranking these apps, as part of the evaluation framework.

The second step involved translating the outcomes from the first method, as is explained in Chapter Five, by analysing the feedback from different stakeholders who participated in the research. This lead to categorising different stakeholders with different MHEA needs and was do reflected it into the MAST app with the aim of matching different stakeholders with different MHEAs. This would save time and effort for different stakeholders in selecting the most suitable MHEAs to meet their requirements. The research process is presented in Figure 3.1.

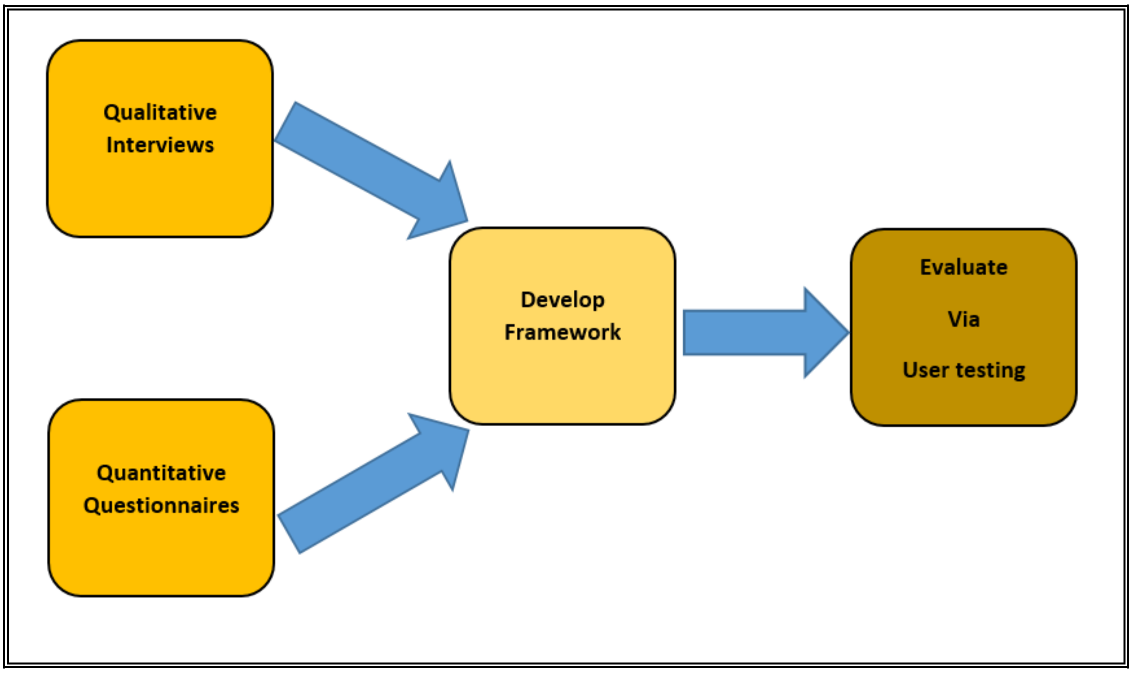


Figure 3.1: The Research Methodology

3.2 Designing the Framework

The evaluation framework went through three stages of designs as the research involved, the revealing need to include some elements or remove others. In order to improve the quality of the research, outcome and contribution.

3.2.1 First Design of the Framework

At the beginning of the research, the design of the evaluation framework was based on two sources, the questionnaire and the prototype. The questionnaire questions were developed depending on metrics that were obtained from the literature, from a selected hybrid metric, from Heuristic Evaluation (HE) and Usability Evaluation (UE) metrics. Appendix 2 represents the first design of the evaluation framework.

Moving forward in the research. It was discovered that this design was not a sufficient one to be considered for the research, since the research needed to use specific and appropriate selected hybrid metrics, in the questionnaire, in order to measure the usability of MHEAs. These MHEAs already existed and were utilized regularly by different stakeholders such as Health Professionals (HP), Patients (P) and Other Stakeholders (OS). Moreover, these different stakeholders would utilize these MHEAs for numerous aspects of their lives under the umbrella of health education. This revealed the lack of information about what these different stakeholders require to measure from these MHEAs. Therefore, in order to provide this information, it was necessary to include some qualitative semi-structured interviews, to obtain different stakeholder's requirements for measurement from these apps. This would enable the researcher design and finalize the selected hybrid metrics depending on the requirements, as explained clearly in Chapter Four. This led to the second design for the evaluation framework (Aljaber & Gordon, 2017a).

3.2.2 Second Design of the Framework

The second design of the evaluation framework was updated version of the first design, and reflected the two aspects of the first design; include the questionnaire and the prototype. The questionnaire questions were based on mapping between metrics obtained from the literature, from a selected hybrid metric from HE and UE metrics. However, after the researcher conducted the qualitative semi-structured interviews and identified the most appropriate selected hybrid metrics, it was possible to perform mapping between the selected hybrid metrics and the questionnaire questions, which helped to identify the specific usability issues in MHEAs. The questions were designed to be used for two purposes: firstly, ranking the existing MHEAs, which was one of the most important objectives of the research and secondly, for designing the prototype, which was part of the first and second framework designs. Appendix 3 presents the second design for the evaluation framework (Aljaber & Gordon, 2017a).

Nevertheless, this design has been modified to the third design, which had been considered in the research to swap the prototype into a MAST, which has replaced the third part of the framework, which been explained more in the next section.

3.2.3 Third Design of the Framework

This was the final design of the framework, which was the most critical design as it helped to answer the remainder of the research questions. This contained parts of the first design and the second design, but the last part was changed to a Medical Apps Selection Tool (MAST). In other words, this design reflected the first two investigations performed in the framework: firstly, the semi-structured interviews with different stakeholders (HP, P, and SD), in order to obtain the right selected hybrid metrics to meet their requirements; secondly, the questionnaire, which was designed depending on the outcomes from the semi-structured interviews. This questionnaire at the beginning of the framework was designed to support achievements of two aims, ranking MHEAs, and then, enhancing development of the prototype to reflect some of the features of MHEAs. However, for the third design of the evaluation framework, the aims of the questionnaire were slightly modified to include three aims, which were: firstly, ranking existing MHEAs; secondly, matching between different MHEA profiles and different stakeholder profiles; thirdly, providing guidance for developing a Medical Apps Selection Tool (MAST) to help different stakeholders to find different MHEAs, which meet their specific requirements. The MAST was designed depending on the outcomes from analysing the feedback of the questionnaire. Thus, the flow of the research entailed developing the MAST depending on the two investigations the semi-structured interviews and the questionnaire.

This third and final design was followed in the rest of the research, in order to achieve the rest of the contribution of the research, and, answer the research questions. The third design of the evaluation framework is presented in Figure 3.2, and the explanations for the filtering through the design are explored in appendix 12 .

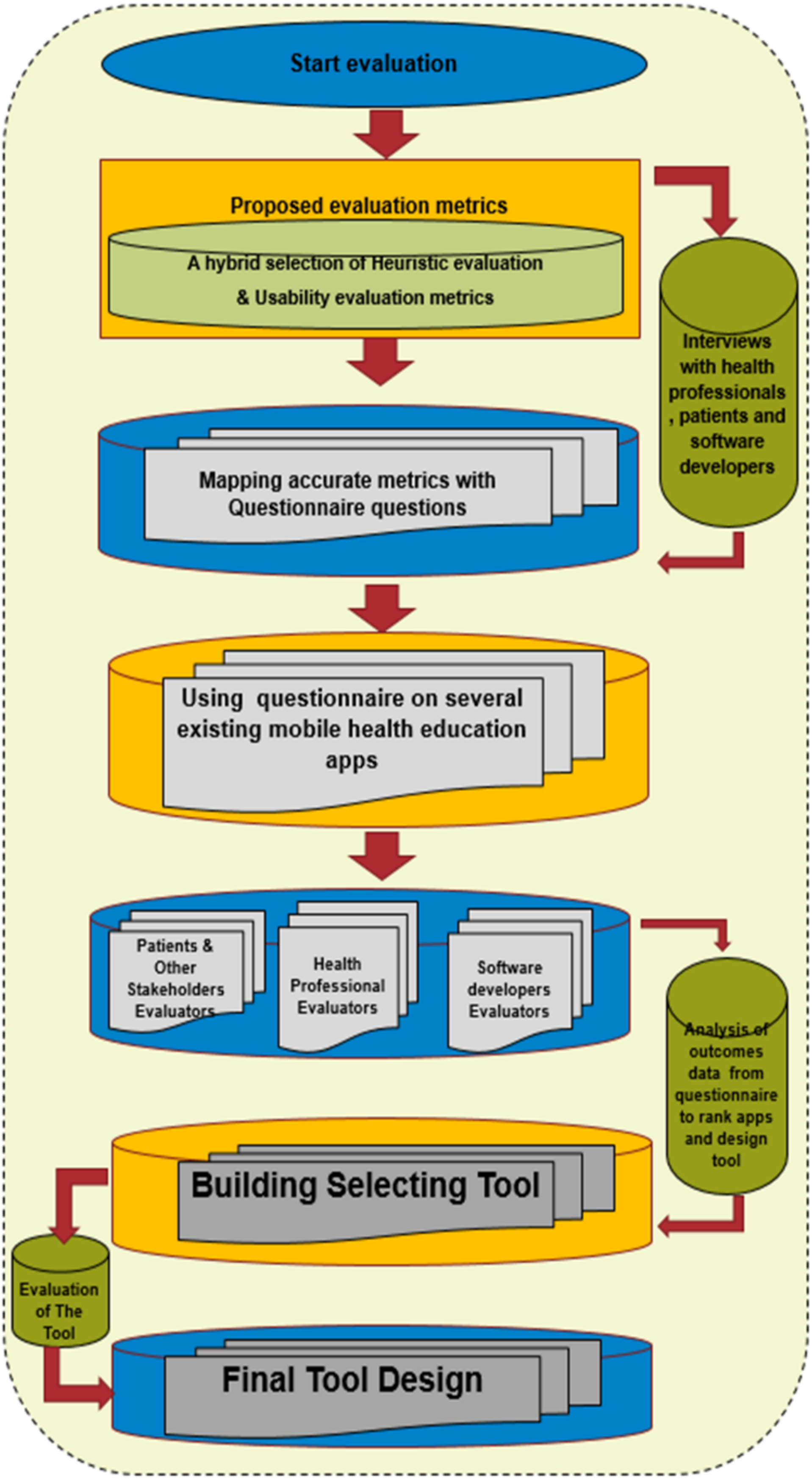


Figure 3.2: Third Design of the Framework

3.3 Medscape

We consider Medscape in our research, as Medscape from WebMD (medscape.com) is one of the most frequently downloaded medical apps by health professionals (Insurance, 2019; Healthcare, 2014). Medscape is popular with medical students, nurses and other healthcare professionals, as a source of clinical information. It is one of the most popular free apps for health professionals. Medscape has various features; it is easy to use for existing users, or new users Figure 3.3 illustrates the Medscape app user interface.

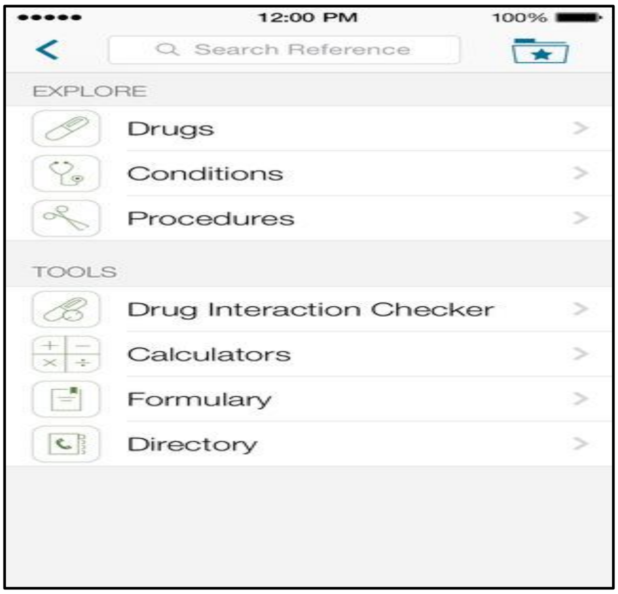


Figure 3.3: Medscape App
(WebMD, 2014)

3.4 Support Sources for Designing the Software and Tool.

Recent research showed that the most popular operating system for a smart mobile phone is Android, which in 2013- accounted for 78.4 % of the operating systems in the world (Gartner, 2014). In addition, according to Cass (2014), C# is one of the Top 10 programming languages.

Several meetings were held with the mobile platform team at Hull College in order to benefit from their experience, and they were asked for their advice on the most suitable programming language for developing a mobile application, as they develop mobile apps regularly, and have vast experience in this field. They suggested using C# as the programming language to develop the applications, and to use Android or windows 10 as the operating system to run the applications. The Impact from these meetings with the Hull college mobile platform team was to identify the programming language for developing the MAST tool in, namely C# for .NET. MAST was developed to run on desktop computers (Windows). The researcher decided to utilize C# to develop MAST, given the support for cross platform support and a range of platforms including ASP.NET. Whilst Android could have been used as the operating system to run the MAST, it was decided to select Windows 10 as the operating system. This was so that MAST could be run on a desktop computer. Future work would be to develop MAST to another operating system such as Android or iOS.

3.5 The Strategies for Selecting the Stakeholders Type and Number in the Framework

According to Nielsen (1993), it is very important to identify the users who will test an app. since differences between users may affect the usability evaluation. Therefore, in this research, the plan was to send the questionnaire to three different categories of participants who were, as mentioned in the section in the framework architecture: Health Professionals (HP), Software Developers (SD) and Patients/General Public (P/GP).

The motivations for selecting Health Professionals (HP) was that they have a medical background, so their opinions and feedback would be very important in helping the research to cover as much as it possible the aspects of health education and Heuristic Evaluation.

Furthermore, as Software Developers (SD) come from a computer science background, their opinions and feedback would be very important in helping the research to cover as fully as possible the aspect of usability evaluation.

The aim of having a group of Patients / General Public (P/GP) to participate in the questionnaire was that they could reflect the opinions of ordinary people, such as patients, the students who deal with health education.

The numbers of participants, who took part in the interviews was **five** Health Professionals (HP), **five** Software Developers (SD) and **five** Patients/General Public (P/GP). According to Nielsen (2000) , *"five testing users are sufficient to find around 80 percent of usability problems. The best results come from testing no more than 5 users and running as many small tests as you can afford"*. Furthermore; Morse (2000), discusses sample size; and indicates the sample size for semi-structured interviews, which was close to the sample size in this research.

Moreover, as mentioned above, five testing users is sufficient to enable the research to discover **80%** of usability problems, which was the target of this research. In practical since the research interviewed same as in three different categories, with five users in each category this means the research had **15** testers in total. This enabled wider discovery of usability problems, reaching more than **90%,** substantially higher than the initial target of **80%**.

For the questionnaire, the research plan was to utilise the same categories of stakeholders as in the interviews, but with different numbers of stakeholders. Since this instrument was to collect quantitative data, the researcher selected 27 HP, 27 SD and 27 P/GP, because more data is needed for quantitative data to be analysed.

These stakeholders (27 HP, 27 SD and 27 P/GP for the questionnaire) and (5 HP, 5 SD and 5 P/GP for the interviews) were selected as a regular users for mobile health education applications for a minimum of two years. In addition, they depended on these applications to develop aspects of their health educations. The stakeholders were from different backgrounds. For instance, the health professionals for the interviews were working in hospitals in Jordan, and the health professionals for the questionnaires were working in different hospitals in the UK. The software developers for the interviews were working in software houses in Jordan, and the software developers for the questionnaire were a mix between working in software houses in the UK, and postgraduate students at University. The patients and general public for both the interviews and the questionnaire were from different working backgrounds.

3.6 Metrics Mapping with Questionnaire and Stakeholders

Three different categories of participants (HP, SD and P/GP) were involved in the investigation in order to map their opinions alongside the proposed efficient selected hybrid of usability and heuristic metrics. Three mapping techniques were applied alongside the metrics. These were mapping metrics with stakeholders, mapping metrics with the questionnaire and interviews and mapping metrics with the questionnaire questions. These are reviewed in the following sections.

3.6.1 Mapping Metrics with Stakeholders

Initially, the research used equally weighted mapping between all participants' categories and the proposed efficient selected hybrid of usability and heuristic metrics. Nevertheless, the researcher expected this weighting to be different, since the research considered different categories of participants, HP, SD and P/GP, who would have different opinions, as presented in Figure 3.4. For instance, HP have more experience to supply their opinions regarding criteria 1 and 2 as they are users of the apps, and experts in the Health education field.

Moreover, SD have different opinions. The researches expected them to have a wider range of opinions, as they are responsible for designing the software. Therefore, they might have opinions to cover criteria 1, 2 and 3.

However, the last category P/GP, might be unable to cover all the categories as they have unknown experience, so the researcher expected them to have the minimum opinions, which could be only about the criteria on 3.

From here, the researcher made a decision to give unequal weight to the mapping, between the participants and the efficient selected hybrid of usability and heuristic metrics, which could be 100/6, which is as follows:

1) HP = 33.3 %
2) SD = 50 %
3) P/GP = 16.7 %

Figure 3.4 shows the proposed unequal weighting besides the mapping between the different categories of participants and the researcher proposed selected hybrid of usability and heuristic metrics. A positive relationship is shown as 1 and a negative relationship as 0. However, this proposed unequal weighting and the mapping between the different categories of participants and the proposed selected hybrid of usability and heuristic metrics, was changed depending on the outcomes from analysing the data of the interviews, in Chapter Four as clarified in section 4.14 in Chapter Four.

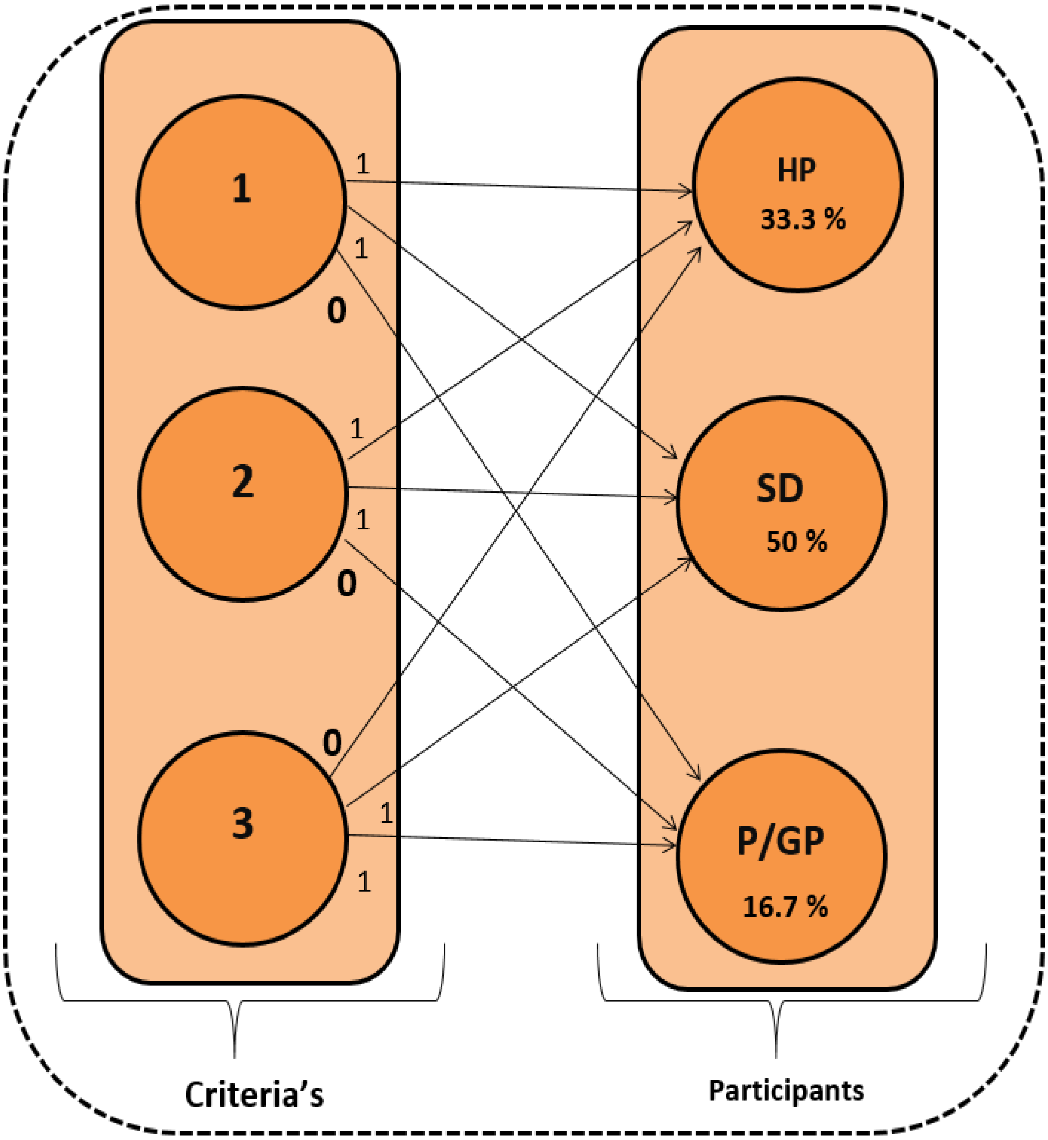


Figure 3.4: Factors and Criteria's _Mapping _ With Participants

Following the collection of the questionnaire feedback from all the participants (HP, SD and P/GP) the researcher determined the most important things for them, which is gave more confidence for the researcher to decide on giving unequal weight or not, in order to do the mapping between the participants and the proposed efficient selected hybrid of usability and heuristic metrics.

This weighting might amend the researcher's decision, to provide different and unequal weights to the mapping between the participants and our efficient selected hybrid of metrics. This is Clearfield in Chapter Four.

3.6.2 Mapping Metrics with Questionnaire and Interviews

This section is mapping metrics with questionnaires and interviews presents the filtering of the proposed hybrid select metrics to the final metrics, based on the outcomes from the semi-structured interviews, as well as the mapping between the final metrics and the questionnaire against different stakeholders groups consisting of (HP, SD, P/GP).

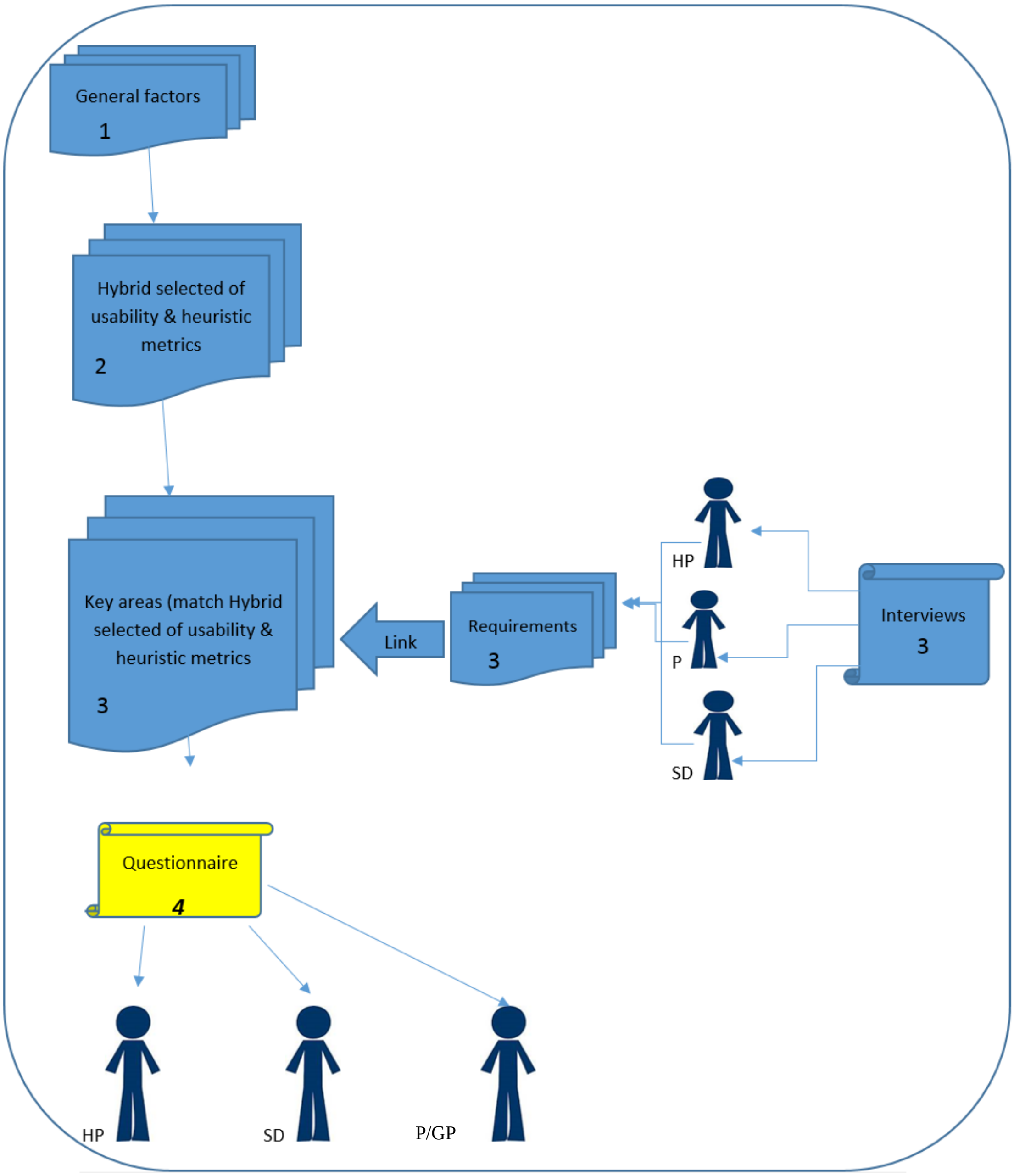


Figure 3.5: Metrics Mapping with Questionnaire and Interviews

3.6.3 Mapping Metrics with Questionnaire Questions

Following Figure 3.5, which shows the metrics mapping alongside the questionnaire and interviews, Table 3.1 explains the mapping between the proposed selected hybrid metrics and the questionnaire questions. The researcher collected in column 1 of Table 3.1 the general factors, some of which were shown to be useful to measure the usability of MHEAs. Column 2 demonstrates the selection of the hybrid proposed selection of heuristic metrics and usability metrics, in order to obtain the most relevant heuristic and usability metrics for MHEAs. Column 3 shows the key outcome of the interview pilot test; this test was carried out with three interviewees, one interviewee from each category of stakeholders. It indicated that there were some differences between the original proposed set of heuristic metrics and usability metrics, and the requirements for the different stakeholders from the interviews pilot test, which were matched into the questionnaire questions. Column 4 shows the mapping of the questionnaire question numbers, after matching them to the requirements for the different stakeholders from the interview pilot test.

This helped the researcher to identify most of the questionnaire questions. Nevertheless, to avoid missing some essential questions and to cover all the requirements of different stakeholders, in order to develop the most relevant questions for the questionnaire, finalise, the questionnaire questions and make it clear enough to use for measuring the usability in existing MHEAs and develop the MAST, the researcher conducted interviews with some HP, P, and SD. These had already been involved in using the MHEAs, for improving their health education or health condition. Another reason for doing this was to validate all the questionnaire questions, so they would most accurately reflect the requirements of the different stakeholders.

The interviews are clearly explained in Chapter Four. Analysis of the data from the interviews enabled the researcher to complete the final mapping of metrics to the questionnaire questions, as shown in Table 3.2. Table 3.2 shows the mapping between each metric from the selected hybrid metrics to the questionnaire questions. The researcher collected- in column 1 of Table 3.2 - the general factors from the HE and UE. Column 2 shows the final selection of heuristic metrics and usability metrics that make up the hybrid evaluation, based on matching each metric with the research questions. Column 3 shows the mapping for each metric with each questionnaire question.

1	2	3	4
General Factors Of A. Heuristic evaluation B. Usability evaluation metrics and methods	Hybrid Selected Of (Usability & Heuristic Metrics) from General Factors depending on previous researches (A) productivity (B) efficiency (C) effectiveness (D) attractiveness (E) responses (F) time and understanding	Key Areas (matching hybrid selected Usability & Heuristic metrics with Health Professionals requirements) 1) Organized information 2) Supporting photos 3) Simplicity design	Questionnaire Questions (reflecting the key Areas into questionnaire questions)
A. Heuristic evaluation: • Visibility of system status – do you know what state you are in? • Match between system and the real world – for example, using a metaphor like a desktop. • User control and freedom. • Consistency and standards. • Error prevention. • Recognition rather than recall. • Flexibility and efficiency of use. • Aesthetic and minimalist design. • Help users recognize, diagnose, and recover from errors. • Help and documentation. **B. Usability evaluation metrics and methods:** • Learnability • Efficiency • Memorability • Errors • Satisfaction 1) A USABILITY EVALUATION TOOL FOR ANDROID BASED MOBILE APPLICATION 2) The goal, question, metric (GQM) 3) The Quality in Use Integrated Measurement (QUIM).	C	3)	1)
	D	3)	2)
	B	3)	3)
	A	3)	4)
	C	2)	5)
	B	1)	6)
	F	1)	7)
	D	3)	8)
	C	2)	9)
	F	1	10)
	F	3	11)
	A	3	12)
	F	2	13)
	F	1	14)
	E	1	15)
	F	1	16)
	C	1	17)
	C	1	18)
	A	3	19)
	A	3	20)
	A + B + C + D + E + F	1 + 2 + 3	21)

Table 3.1: Initial Metrics mapping with Questionnaire

The questions designed for the interviews are shown in appendix 1. Also, the researcher applied for ethical approval by filling several forms through the School of Engineering and Computer Science, The University of Hull.

<table>
<tr><th>1</th><th>2</th><th>4</th></tr>
<tr><td>General Factors Of
A.Heuristic evaluation
B.Usability evaluation metrics and methods</td><td>Hybrid Selected Of
(Usability & Heuristic Metrics) filtered into five accurate metrics depending on hybrid interviews with Health professionals and patients
(A)Memorability
(B)Features
(C)Attractiveness
(D)Simplicity (containing learnability)
(E)Accuracy</td><td>Questionnaire Questions (reflecting the key Areas into questionnaire questions)</td></tr>
<tr><td>A. Heuristic evaluation:</td><td>C</td><td>1)</td></tr>
<tr><td rowspan="20">• Visibility of system status – do you know what state you are in?
• Match between system and the real world – for example, using a metaphor like a desktop.
• User control and freedom.
• Consistency and standards.
• Error prevention.
• Recognition rather than recall.
• Flexibility and efficiency of use.
• Aesthetic and minimalist design.
• Help users recognize, diagnose, and recover from errors.
• Help and documentation.
A. Usability evaluation metrics and methods:
• Learnability
• Efficiency
• Memorability
• Errors
• Satisfaction
1) A USABILITY EVALUATION TOOL FOR ANDROID BASED MOBILE APPLICATION
2) The goal, question, metric (GQM)
3) The Quality in Use Integrated Measurement (QUIM).</td><td>D</td><td>2)</td></tr>
<tr><td>B</td><td>3)</td></tr>
<tr><td>A</td><td>4)</td></tr>
<tr><td>C</td><td>5)</td></tr>
<tr><td>B</td><td>6)</td></tr>
<tr><td>E</td><td>7)</td></tr>
<tr><td>D</td><td>8)</td></tr>
<tr><td>C</td><td>9)</td></tr>
<tr><td>E</td><td>10)</td></tr>
<tr><td>D</td><td>11)</td></tr>
<tr><td>A</td><td>12)</td></tr>
<tr><td>B</td><td>13)</td></tr>
<tr><td>E</td><td>14)</td></tr>
<tr><td>E</td><td>15)</td></tr>
<tr><td>B</td><td>16)</td></tr>
<tr><td>C</td><td>17)</td></tr>
<tr><td>D</td><td>18)</td></tr>
<tr><td>A</td><td>19)</td></tr>
<tr><td>A</td><td>20)</td></tr>
<tr><td>A+B+C+D+E</td><td>21)</td></tr>
</table>

Table 3.2: Final Metrics Mapping with Questionnaire

3.7 Conclusion

Chapter Three, described the stages of the framework structure design during the research period, with on explanation for the changes of designs during the research. Moreover, it illuminated why Medscape was used in the interviews. It explains how the researcher selected the types and number of participants for the interviews, and how the questionnaire was developed with the mapping between the metrics and questionnaire questions.

Chapter Four reports on how the proposed selected hybrid metrics were mapped to the questionnaire, explain the priority of the metrics between different stakeholders, and calculates the weights for each of the metrics, for the different stakeholders.

Chapter 4: Metrics Selection Design and Qualitative Results

Chapter Three discussed the research methodology. It illuminated the strategies for designing the framework, selecting the stakeholders, and mapping between the metrics and the stakeholders. Chapter Four presents the fulfilment of objective 2, which was developing a ranking measurements system to assess and rank MHEAs. It explains the technique used in the research, to clarify the selection of hybrid metrics depending on qualitative interviews. In addition, explores the priorities among the selected hybrid metrics for several stakeholders (Health Professionals (HP), Patients (P) and Software Developers (SD)). The Chapter explores the similarities and differences between different stakeholders in their wears of the metrics. Finally, investigates the weighting in assigned to these metrics for the different stakeholders.

4.1 Introduction

At the beginning of the research, the researcher proposed some hybrid selection metrics from Heuristic Evaluation (HE) and Usability Evaluation (UE), depending on the literature, in order to determine if these selected hybrid metrics were the most suitable and optimal, to meet the needs of HP, P, and SD the major users of MHEAs.

Moreover, the researcher conducted semi-structured interviews which consisted of 11 questions, so as to gain a better understanding of the specific requirements of HP, P and SD users when using MHEAs. This was done to help construct the final hybrid metrics selected from H.E. and U.E. The interviews involved 15 participants: five from each category, namely 5 HP, 5 P, and 5 SD; in order to collect different opinions and feedback from the major users of the software and the developers of the software. The participants for the interviews ranged in age from 32 to 61 years old for HP, from 18 to 57 years old for P and from 25 to 35 years old for SD. Figure 4.1 presents the participant's categories.

These interviewees were in the Middle East, specifically in Jordan. The researcher conducted all the interviews by using the Viber app on a mobile phone. The researcher checked that the interviewees were setting in a quiet place, such as a closed room or office, to ensure they were free from distraction, which could affect their feedback.

The interviews were recorded; the purpose of recording the interviews was avoid losing any data during the interviews. In addition, the researcher checked that these 15 interviewees were regular users of the Medscape apps, and had been using them for a minimum of 2 years. Following the recording of the interviews, translations were entered into 15 Word documents. Creating these documents involved manual coding. Of the text in the 15 interviews and looking for the frequency of the metrics in the text. This made the data ready for analysis with the aid of Nvivo software, according to the coded transcripts; five nodes were created in Nvivo, representing the five metrics (A. Memorability, B. Features, C. Attractiveness, D. Simplicity and E. Accuracy), and each interview transcript's coding was mapped to these nodes. This simplified analysis of the data using Nvivo. The results are reported in the following sections in this Chapter.

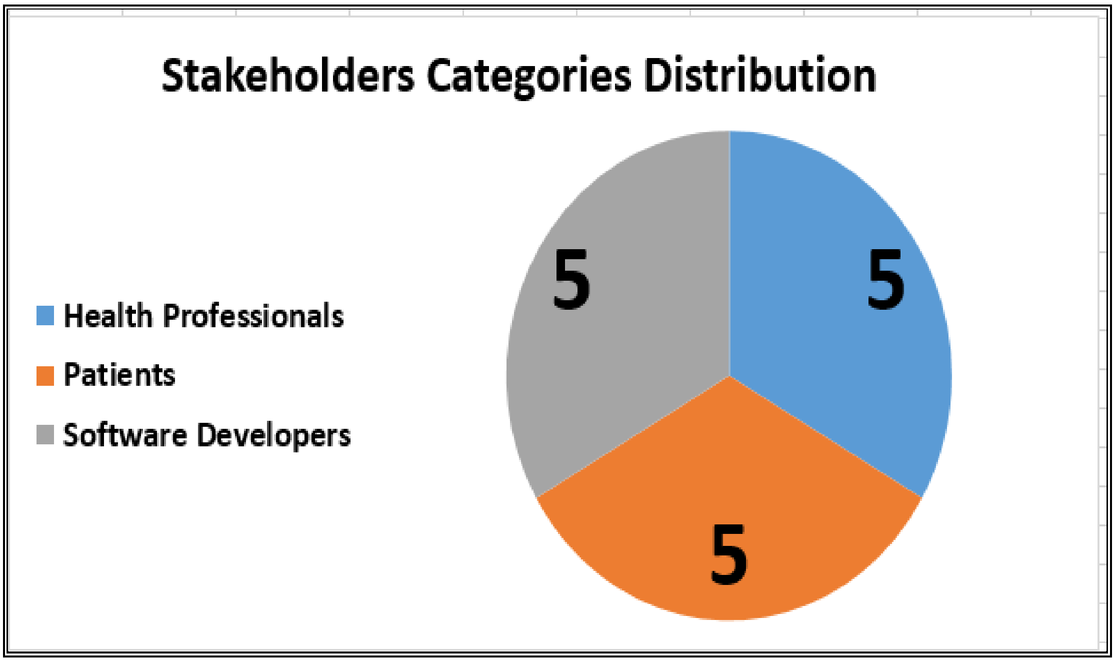


Figure 4.1: People who participated in the interviews separated into user categories

4.2 Proposed Hybrid Selection Metrics from HE and UE.

Initially, the proposed hybrid metrics selected from the HE and UE based on the literature were:

(A) Productivity: Is the application improving your productivity? Productivity is the ratio of output to inputs in production. Development Productivity = Development project size (in Functions) /Development effort (in hours) (Shivade & Sharma, 2014).

(B) Efficiency: How much effort do users require to do this? (Often measured in time) How fast are users in accomplishing different tasks once they have internalized the functionality of the interface design? (Nielsen, 1995; Shivade & Sharma, 2014).

(C) Effectiveness: Can users complete tasks, achieve goals with the product, i.e. do what they want to do? (Pedroli, et al., 2013).

(D) Attractiveness: Is the application interaction user friendly? (Shivade & Sharma, 2014).

(E) Responses: Is the response time of application satisfactory (Satisfaction – what do users think about the product's ease of use) (Shivade & Sharma, 2014; Pedroli, et al., 2013).

(F) Time and understanding: Can users understand how to select a software product that is suitable for their intended use and how it can be used for particular tasks? (Shivade & Sharma, 2014).

However, these selected hybrid metrics were potentially adjustable, as they were proposed metrics. The aim of proposing them was determine the most relevant metrics, to meet the requirements of different stakeholders. Fifteen semi-structured interviews were conducted with different stakeholders, five each of HPs, Ps and SDs.

4.3 Final Hybrid Selection Metrics Depending on HP, P and SD Requirements.

Following analysis of the semi-structured interviews data, the researcher determined that modification was needed for some of the metrics that had been selected as some of them did not meet all the requirements of HP, P, and SD. The semi-structured interviews reflected that some of the proposed hybrid metrics from the HE and UE were not necessary for HP, P and SD requirements, while other metrics needed to be added for the proposed metrics.

Therefore, the researcher finalised the hybrid selection of metrics from the HE and UE as:

1) Efficiency
    A) Memorability
2) Satisfaction
    B) Features
    C) Attractiveness
3) Effectiveness

D) Simplicity (containing learnability)

E) Accuracy

The definitions of these final selected hybrid metrics are as follows:

Efficiency: Is determined by the estimated costs (e.g., total time) of executing user Procedures (Seffah, et al., 2006). It measures the rate of task output, achieving maximum productivity with minimum wasted effort or expense (Press, Oxford University, 2016).

Memorability: When users have not worked with the interface for a long time, how much of its functionality do they remember and how fast can they regain their proficiency? (Shivade & Sharma, 2014; Md Samsur & SM Zabed , 2013).

Satisfaction: This refers to the subjective responses from users about their feelings when using the software (i.e., is the user satisfied or happy with the system?) (Seffah, et al., 2006). How satisfied are users after working with the interface concerning the time they had to invest? Was the interface pleasant to use? (Shivade & Sharma, 2014).

Features: Does the app provide extra like image, video, voice support/help -touch screen facilities, Voice guidance (Hussain, *et al*., 2012).

Attractiveness: Capability of the software product to be attractive to the user (e.g., through use of colour or graphic design (Seffah, et al., 2006; I.S.O, 2001).

Effectiveness: The capability of a software product to enable users to achieve specified tasks with accuracy and completeness in a specified context of use (Shivade & Sharma, 2014). Measures of the quantity and quality of task output and whether users succeed in achieving their goals when working with a system (Seffah, et al., 2006).

Simplicity: Ease to input the data, Ease to use output, Ease to install, Ease to learn (Hussain, *et al*., 2012).

Accuracy: Accurate (Hussain, *et al*., 2012). Capability to provide correct results or effects (Seffah, et al., 2006; I.S.O, 2001).

4.4 Analysing Health Professional's (HP) Interviews

Analysis was carried out of five interviews with health professionals (HP), to address their major requirements from using MHEAs. Figure 4.2, summarizes the relative importance of concepts, based on the interviews data in Table 4.1, by creating cross categories of usability for all five interviews. The figure was obtained from Nvivo. The colours inside each column correspond to individual interviews. For instance, the first column represents the first metric, which is A) Memorability, while the individual colours inside the first column represent the individual feedback (requirements) from each interviewee. The first column shows that four of the five interviewees considered A) Memorability as a part of their requirements. The second column shows all five interviewees considered B) Features. The third column shows that only one out of the five interviewees considered C) Attractiveness. The fourth column shows that four of the five interviewees considered D) Simplicity. Finally, the fifth column shows that all five interviewees considered E) Accuracy as a part of their requirements.

In addition, the height of the given columns is the total weighting given by all the 5 interviewees, to the categories of HP. For instance, the red colour representing interview number 10 with HP, out of the five interviews with HP. It represents the HP's requirements from the five metrics, which were B) Features and E) Accuracy, as represented in Figure 4.2. Furthermore, the numbers on the left-hand side of the figure represent the quantity of each metric; the colours on the right-hand side represent each interviewee. This reflects how many time they mentioned that metric. For example interview number 1 with HP reflects that the HP said A) Memorability in his interview 5 times, B) Features 5 times, C) Attractiveness 1 time, D) Simplicity 10 times and E) Accuracy 11 times. This type of figure is utilized in the rest of the Chapter, with a summary of the data in the figure, followed by an illustrated table.

In the first place, of the most important metric to be measured, to meet HP requirements is the D) Simplicity as shown by the highest column (36), in second place was B) Features, as shown by the second highest column (24). In third place was E) Accuracy as shown by the third highest column (20), in fourth place was A) Memorability as shown by the fourth highest column (12). In the last place was the C) Attractiveness, as shown by the lowest column (1). These results indicate that the most important metrics for HP were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

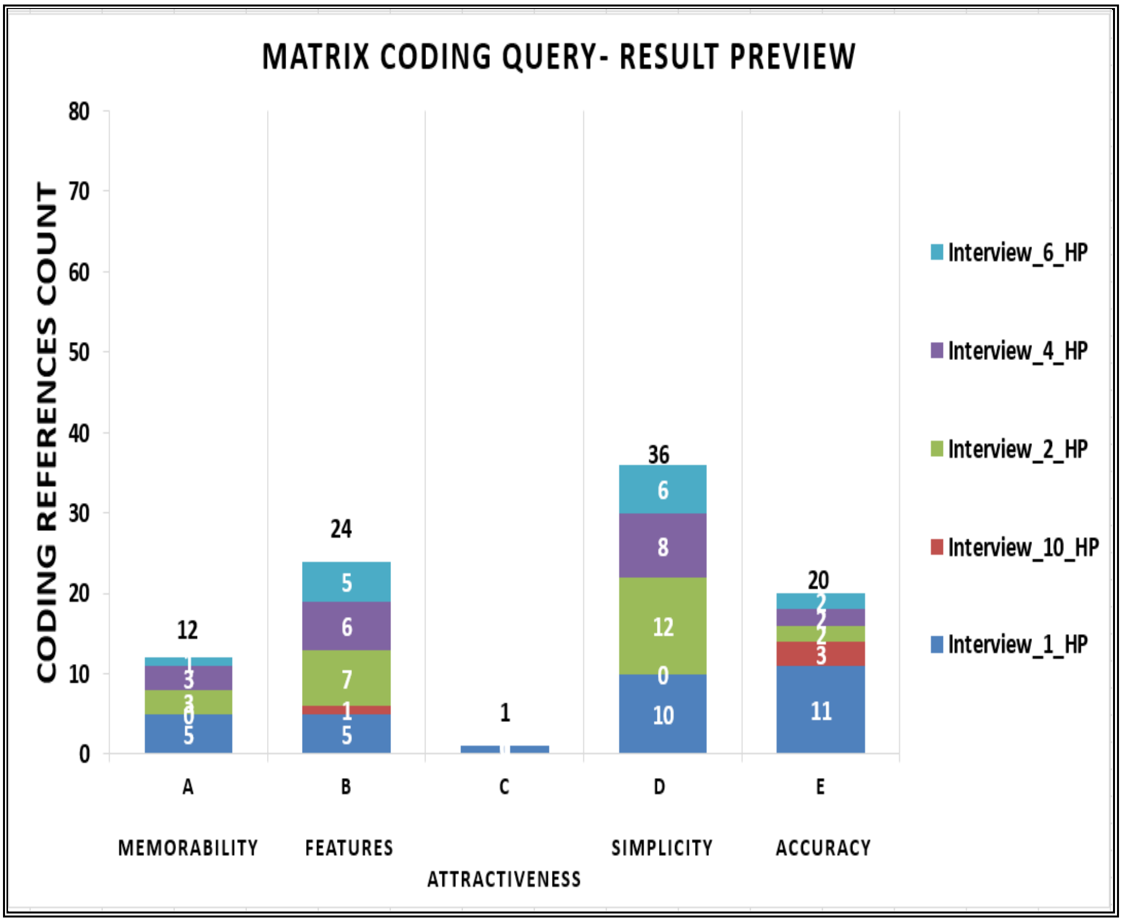


Figure 4.2: HP Metrics Requirements Results from HP Interviews

Metrics / Interviewees	A Memorability	B Features	C Attractiveness	D Simplicity (containing learnability)	E Accuracy
Interview_1_HP	5	5	1	10	11
Interview_10_HP	0	1	0	0	3
Interview_2_HP	3	7	0	12	2
Interview_4_HP	3	6	0	8	2
Interview_6_HP	1	5	0	6	2
Total	12	24	1	36	20
Average	2.4	4.8	0.2	7.2	4

Table 4.1: HP metrics requirements percentage result from interviews

4.5 Analysing Patients (P) Interviews

The second part of analysing the data collected was the analysis of five interviews with patients (P). The result is summarised in Figure 4.3,and explored in table 4.2 The numbers in the table cells reflect the number of times patients mentioned each metric. For example in interview number 5 with P, the interviewee said A) Memorability in his interview 3 times, B) Features 5 times, C) Attractiveness 0 times, D) Simplicity 9 times and E) Accuracy 3 times.

The most important metric to be measured to meet P requirements was D) Simplicity, as shown by the highest column (32). In the second place was B) Features as shown by the second highest column (21). In the third place was E) Accuracy, as shown by the third highest column (13). In the fourth place was A) Memorability as shown by the fourth highest column (12). In the last place was C) Attractiveness as shown by the lowest column (2). These results indicate that the most important metrics for P were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

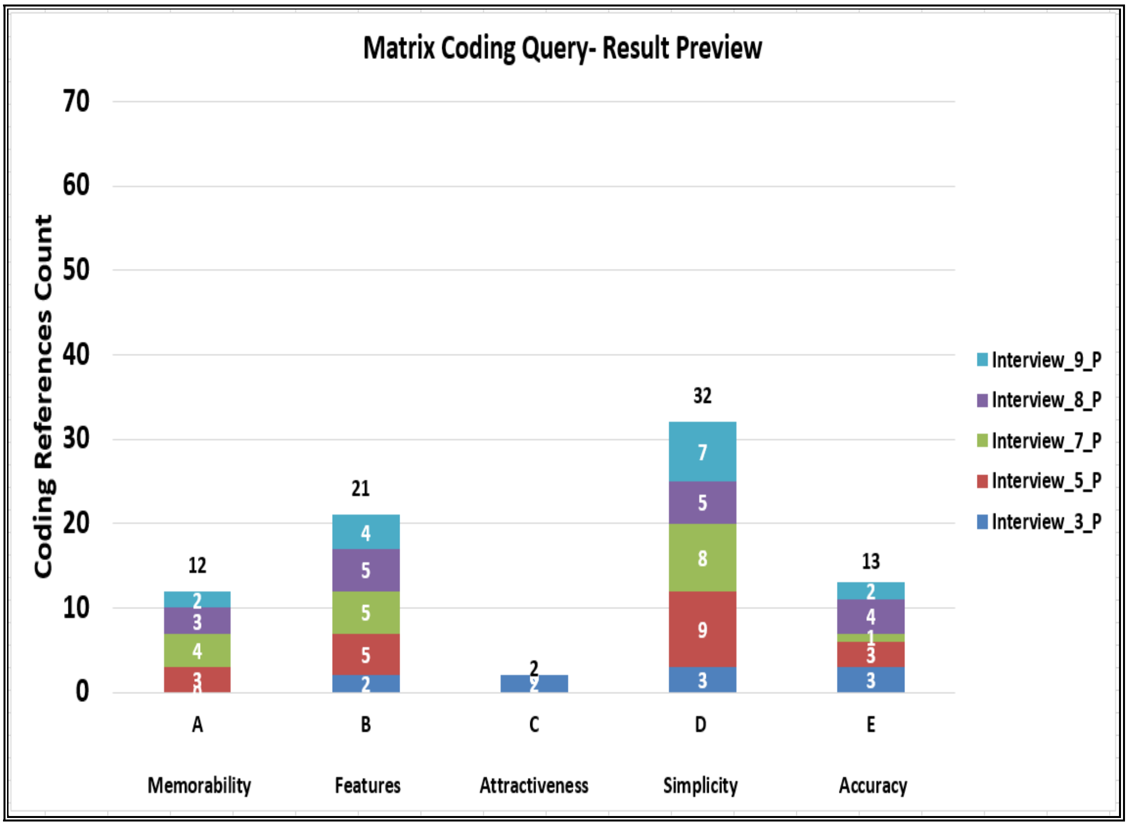


Figure 4.3: P Metrics Requirements Result from P Interviews

Interviewees \ Metrics	A Memorability	B Features	C Attractiveness	D Simplicity (containing learnability)	E Accuracy
Interview_3_P	0	2	2	3	3
Interview_5_P	3	5	0	9	3
Interview_7_P	4	5	0	8	1
Interview_8_P	3	5	0	5	4
Interview_9_P	2	4	0	7	2
Total	12	21	2	32	13
Average	2.4	4.2	0.4	6.4	2.6

Table 4.2: P metrics requirements percentage result from interviews

4.6 Analysing Software Developer (SD) Interviews

The third part of analysing the data was the analysis of five interviews with Software Developers (SD). The results obtained are presented in table 4.3 and in Figure 4.4.

The collected from the five interviews. With SD are summarised in Table 4.3. The numbers in the table cells reflect the number of times Software Developers mentioned each metric. For example in interview number 15, the SD said A) Memorability in his interview 1 time, B) Features 12 times, C) Attractiveness 0 times, D) Simplicity 5 times and E) Accuracy 2 times.

From these results, the researcher concluded that the most important metrics for SD were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

Interviewees \ Metrics	A Memorability	B Features	C Attractiveness	D Simplicity (containing learnability)	E Accuracy
Interview_11_SD	5	6	1	7	8
Interview_12_SD	0	5	2	4	6
Interview_13_SD	1	10	0	6	3
Interview_14_SD	4	6	0	6	3
Interview_15_SD	1	12	0	5	2
Total	11	39	3	28	22
Average	2.2	7.8	0.6	5.6	4.4

Table 4.3: SD metrics requirements percentage result from interviews

In the first place, of as the most important metric to be measured to meet SD requirements was B) Features as shown by the highest column (39) in the figure. In the second place was D) Simplicity as shown by the second highest column (28). In the third place was E) Accuracy as shown by the third highest column (22). In the fourth place was the A) Memorability as shown by the fourth highest column (11). In the last place was the C) Attractiveness, as shown by the lowest column (3). These results indicate that the most important metrics for SD were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

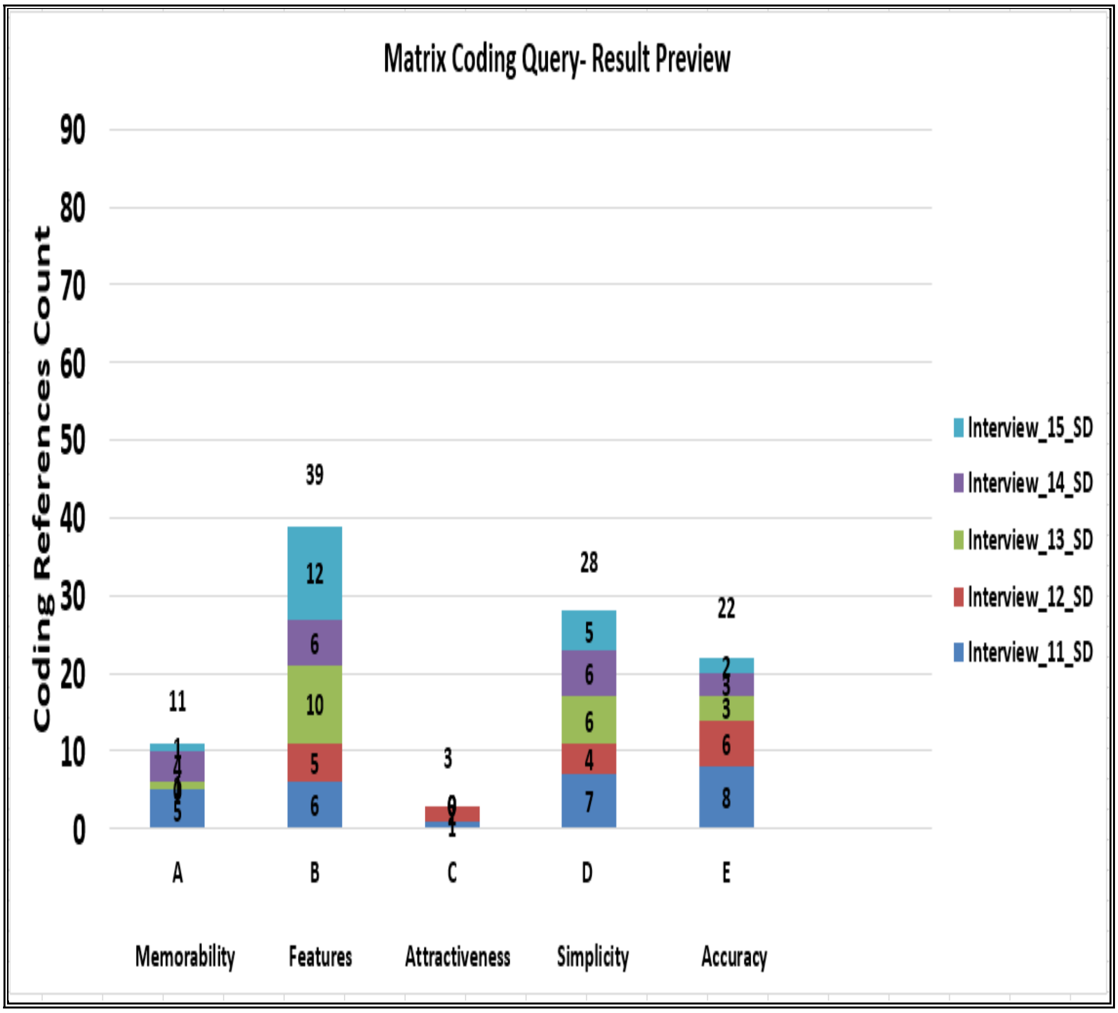


Figure 4.4: SD Metrics Requirements Result from SD Interviews

4.7 Comparision between HP and P Interviews

Following analysis of the data collected from the interviews with HP and P, a comparison between their requirements was applied, using the outcomes from Table 4.1 and Table

4.2, in order to determine the differences and similarities between them. The results were as follows:

Both HP and P focused on the metric D) Simplicity, as the most important metric to meet their requirements, as the HP score was 36 as the highest, P was 32, while the P score also the highest. Also, both HP and P related metric B) Features as the second most important metric with HP yielding a score of 24 and P, 21 both second highest in their groups. Moreover, both HP and P related metric E) Accuracy as the third most important metric, with on HP score of 20 as the third highest and P score of 13 also the third highest. Furthermore, both HP and P related metric A) Memorability as the fourth most important metric, with both groups producing the same score 12. Finally, both HP and P related the metric C) Attractiveness as the least important metric to meet of their requirements, as the HP score was 1 as and the score for P was 2, the lowest in each case.

This comparison between HP and P revealed that their requirements from using MHEAs are the same, although they use MHEAs for different reasons.

4.8 Comparison between HP and SD Interviews

Following analysis of the data collected from the interviews with HP and SD, a comparison between their requirements was made, based on the outcomes from Table 4.1 and Table 4.3, in order to determine the differences and similarities between them. The results were as follows:

In the first place, for HP as the most important metric to meet their requirements was D) Simplicity, with a score of 36. Whereas for SD the most important metric was B) Features with a score of 39.This shows some differences in priority in the requirements from using MHEAs between HP and SD. Also, the second most important metric for HP was B) Features with a score of 24, while for SD it was D) Simplicity, with a score of 28, again showing some a differences between HP and SD. However, both HP and SD related metric E) Accuracy as the third most important metric, as the HP score was 20 and that for SD was 22. Furthermore, both HP and SD related metric A) Memorability as the fourth most important metric, as the HP score was 12 and the SD score was 11. Finally, both HP and SD related metric C) Attractiveness as the least important metric as the HP score was 1, while the SD was 3, the lowest in each case.

This comparison between HP and SD revealed that HP and SD requirements from using MHEAs are mostly similar to each other, although they use MHEAs for different reasons.

4.9 Comparison between P and SD Interviews

Following analysis of the data collected from the interviews with P and SD, a comparison between their requirements was made. Based on the outcomes from Table 4.2 and Table 4.3, to identify the differences and similarities between them. The results were as follows:

For P the most important metric to meet their requirements was D) Simplicity with a score of 32, compared with SD, who related B) Features, with a score of 39, as the highest. This shows some differences in the first place requirements from using MHEAs between P and SD. Also the second most important metric for P was B) Features, with a score of 21, while for SD the second most important metric was D) Simplicity, with a score of 28, again showing some differences in the second place requirements from using MHEAs between P and SD. However, P and SD related the metric E) Accuracy as the third most important metric as P score was 13 while that for SD was 22. Furthermore, both P and SD related the metric A) Memorability as the fourth most important metric, as the P score was 12 and the SD score was 11. Finally, both P and SD related the metric C) Attractiveness as the least important metric, as the P score was 2 while that of SD was 3.

This comparison between P and SD indicates that both P and SD requirements from using MHEAs are mostly similar to each other, despite their using MHEAs for different reasons.

4.10 Overall Analysing HP and P Interviews

Following analysing of HP and P feedback data, which was collected from the 10 interviews which included 5 interviews for HP and 5 interviews for P, an overall analysis of the total feedback from both of HP and P was applied, in order to address their main requirements from using MHEAs. The result obtained is presented in Figure 4.5.

The most important metric to be measured to meet HP and P requirements was D) Simplicity, as shown by the highest column. In the second place was B) Features, as shown by the second highest column. In the third place was E) Accuracy, as shown by the third highest column. In the fourth place was the A) Memorability as shown by the fourth highest column. In the last place was the C) Attractiveness as shown by the lowest column. From these results, the researcher concluded that the most important metrics for HP and P were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

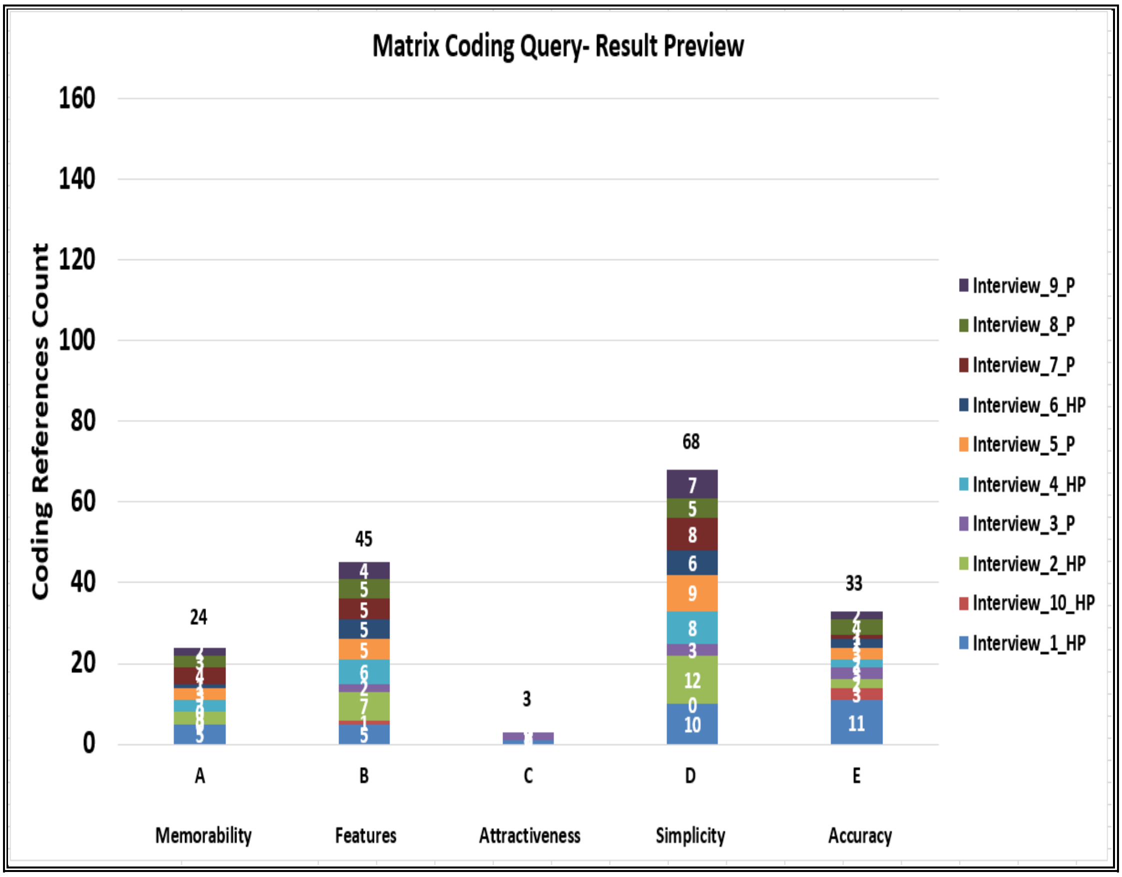


Figure 4.5: HP and P Metrics Requirements Result from HP and P Interviews

The data, collected from the 10 interviews is summerised in Table 4.4. The numbers in the table cells reflect the number of times HP and P mentioned each metric. For example interview number 1 with HP reflects that HP said A) Memorability in his interview 5 times, B) Features 5 times, C) Attractiveness 1 time, D) Simplicity 10 times and E) Accuracy 11 times. And from the overall results the most important metric to be measured to meet HP and P requirements was D) Simplicity, with the highest number of 68. In the second place was B) Features with the second highest number, 45. In third place was E) Accuracy, with the third highest number, 33. In fourth place was A) Memorability, with the fourth highest number, 24. In the last place was C) Attractiveness, with the lowest number, 3.

From these results, the researcher concluded that the most important metrics for HP and P were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

Metrics / Interviewees	A Memorability	B Features	C Attractiveness	D Simplicity (containing learnability)	E Accuracy
Interview_1_HP	5	5	1	10	11
Interview_10_HP	0	1	0	0	3
Interview_2_HP	3	7	0	12	2
Interview_3_P	0	2	2	3	3
Interview_4_HP	3	6	0	8	2
Interview_5_P	3	5	0	9	3
Interview_6_HP	1	5	0	6	2
Interview_7_P	4	5	0	8	1
Interview_8_P	3	5	0	5	4
Interview_9_P	2	4	0	7	2
Total	24	45	3	68	33
Averag	2.4	4.5	0.3	6.8	3.3

Table 4.4: HP and P metrics requirements percentage result from interviews

4.11 Overall Analysing HP and SD Interviews

The result of the overall analysis of the feedback from both HP and SD, in order to address their main requirements from MHEAs, is presented in Figure 4.6.

The most important metric to be measured, which would meet HP and SD requirements was D) Simplicity, as shown by the highest column. In second place was B) Features, as shown by the second highest column. In third place was E) Accuracy as shown by the third highest column. In fourth place was A) Memorability, as shown by the fourth highest column. In last place was C) Attractiveness, as shown by the lowest column. From these results, the researcher concluded that the most important metrics for HP and SD were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

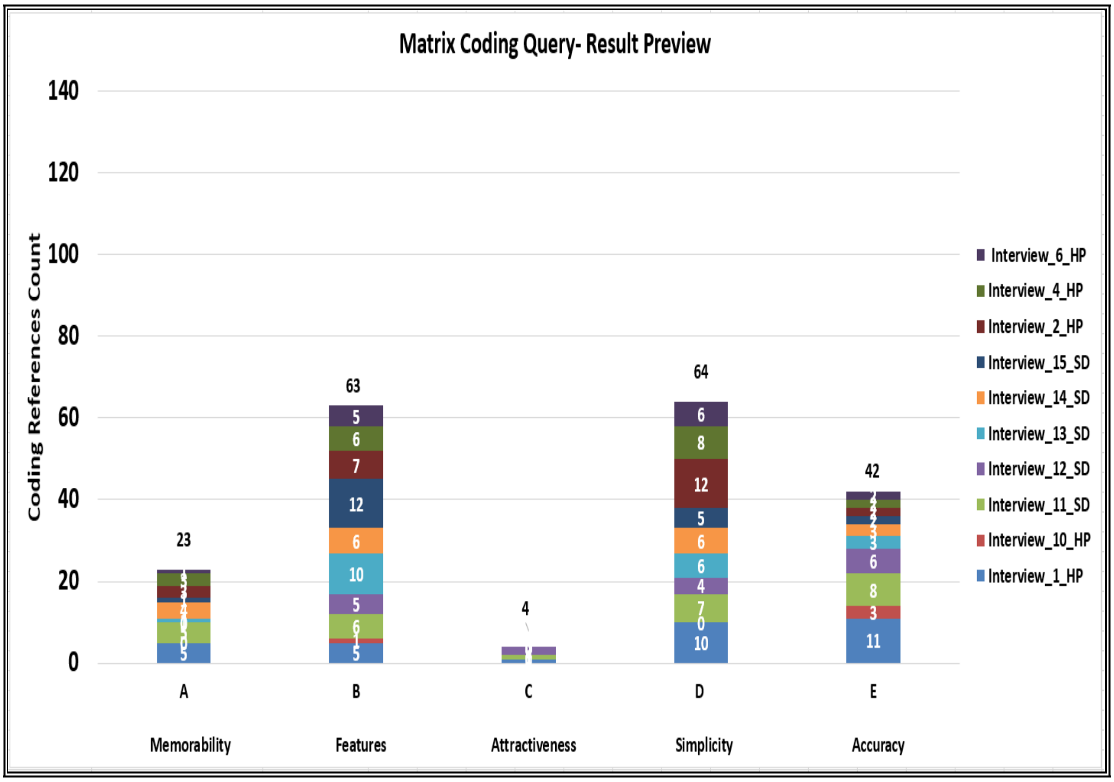


Figure 4.6: HP and SD Metrics Requirements Result from HP and SD Interviews

The data collected form the 10 interviews is summarised in Table 4.5. The numbers in the table cells reflect how many times HP and SD mentioned each metric. For example interview number 1 with HP reflects that HP said A) Memorability in his interview 5 times, B) Features 5 times, C) Attractiveness 1 time, D) Simplicity 10 times and E) Accuracy 11 times. The overall interviews result can be summarised as follows:

The most important metric to be measured, which would meet HP and SD requirements was D) Simplicity. As shown, the highest number is 64. In second place was B) Features, with the second highest number, 63. In third place was E) Accuracy, as shown by the third highest number, 42. In fourth place was A) Memorability, as shown by the fourth highest number, 23. In last place was C) Attractiveness, with the lowest number, 4.

From these results, the researcher concluded that the most important metrics for HP and SD are D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

Metrics / Interviewees	A Memorability	B Features	C Attractiveness	D Simplicity (containing learnability)	E Accuracy
Interview_1_HP	5	5	1	10	11
Interview_10_HP	0	1	0	0	3
Interview_11_SD	5	6	1	7	8
Interview_12_SD	0	5	2	4	6
Interview_13_SD	1	10	0	6	3
Interview_14_SD	4	6	0	6	3
Interview_15_SD	1	12	0	5	2
Interview_2_HP	3	7	0	12	2
Interview_4_HP	3	6	0	8	2
Interview_6_HP	1	5	0	6	2
Total	23	63	4	64	42
Average	2.3	6.3	0.4	6.4	4.2

Table 4.5: HP and SD metrics requirements percentage result from interviews

4.12 Overall analysing P and SD Interviews

An overall analysis was applied of the feedback from P and SD; in order to address their main requirements from using the MHEAs. The result is presented in Figure 4.7.

The most important metrics to be measured which would meet P and SD requirements were D) Simplicity and B) Features, as shown by the highest two columns. In second place was E) Accuracy as shown by the second highest column. In third place was A) Memorability, as shown by the third highest column. In the last place was C) Attractiveness, as shown by the lowest column. From these results, the researcher concluded that the most important metrics for P and SD were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

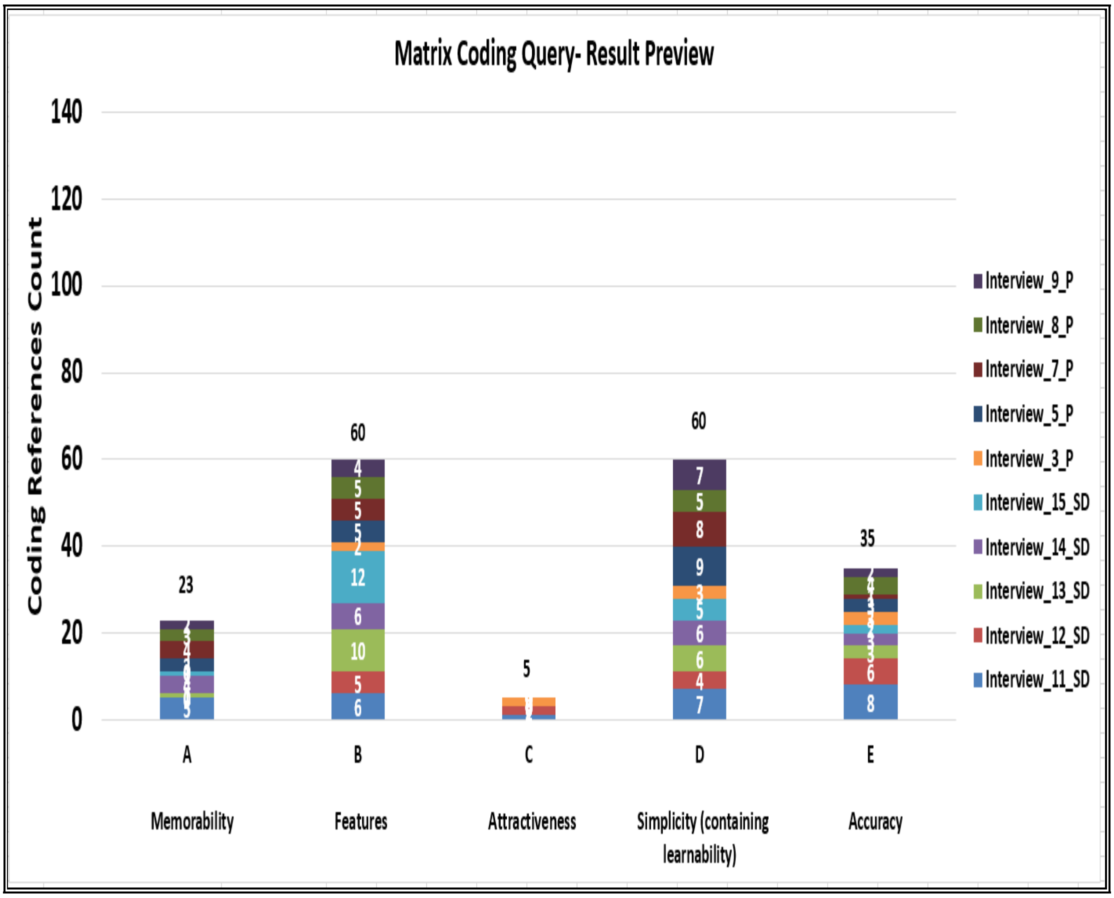


Figure 4.7 : P and SD Metrics Requirements Result from P and SD Interviews

The result of analising the data, been collected from the 10 interviews is summarised in Table 4.6. The numbers in the table cells reflect the number of times P and SD mentioned each metric. For example interview number 9 with P reflects that P said A) Memorability in his interview 2 times, B) Features 4 times, C) Attractiveness 0 time, D) Simplicity 7 times and E) Accuracy 2 times. The 10 interviews yielded a total result as follows:

The most important metrics to be measured, which would meet P and SD requirements were the D) Simplicity and B) Features, shown by that joint highest score of 60. In second place was E) Accuracy, as shown by the second highest number 35. In third place was A) Memorability, as shown by the third highest number, 23. In last place was C) Attractiveness, as shown by the lowest number 5.

From these results, the researcher concluded that the most important metrics for P and SD were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

Metrics / Interviewees	A Memorability	B Features	C Attractiveness	D Simplicity (containing learnability)	E Accuracy
Interview_11_SD	5	6	1	7	8
Interview_12_SD	0	5	2	4	6
Interview_13_SD	1	10	0	6	3
Interview_14_SD	4	6	0	6	3
Interview_15_SD	1	12	0	5	2
Interview_3_P	0	2	2	3	3
Interview_5_P	3	5	0	9	3
Interview_7_P	4	5	0	8	1
Interview_8_P	3	5	0	5	4
Interview_9_P	2	4	0	7	2
Total	**23**	**60**	**5**	**60**	**35**
Average	**2.3**	**6**	**0.5**	**6**	**3.5**

Table 4.6: P and SD metrics requirements percentage result from interviews

4.13 Overall Analysing HP, P and SD Interviews

Following analysis of all of the HP, P and SD feedback data, collected from the 15 interviews, (five with each group) an overall analysis was conducted of the total feedback from all of HP, P and SD, in order to address their main requirements from using MHEAs. The result is explained below and presented in Figure 4.8.

The most important metric to be measured which would meet HP, P and SD requirements was D) Simplicity, as shown by the highest column. In second place was the B) Features, as shown by the second highest column. In third place was E) Accuracy as shown by the third highest column. In fourth place was A) Memorability, as shown by the fourth highest column. In last place was C) Attractiveness, as shown by the lowest column. From these results, the researcher concluded that the most important metrics for HP, P and SD were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

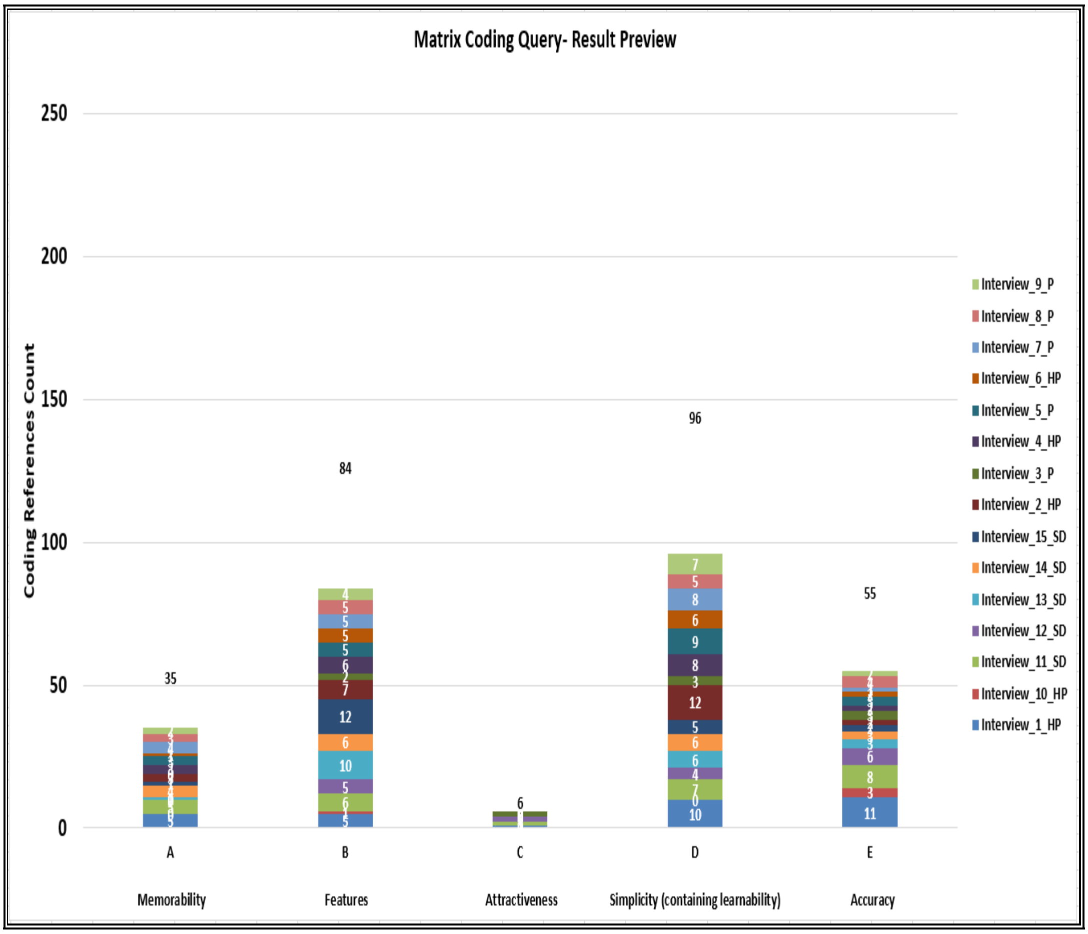


Figure 4.8: HP, P and SD Metrics Requirements Result from All Interviews

The analysis of the data collected from the 15 interviews is summarised in Table 4.7. The numbers in the table cells reflect the number of times the interviewee HP, P and SD, mentioned each metric. For example interview number 1 with HP reflects that HP said A) Memorability in his interview 5 times, B) Features 5 times, C) Attractiveness 1 time, D) Simplicity 10 times and E) Accuracy 11 times. Overall from the 15 interviews, the result can be summarised as follows:

The most important metric to be measured which would meet HP, P and SD requirements was the D) Simplicity, as shown by the highest number, 96. In second place was B) Features, as shown by the second highest number, 84. In third place was E) Accuracy, as shown by the third highest number, 55. In fourth place was A) Memorability, as shown by the fourth highest number, 35. In the last place was C) Attractiveness, as shown by the lowest number, 6.

From these results, the researcher stated that the most important metrics for HP, P and SD were D) Simplicity and B) Features, which would meet their main requirements from using MHEAs.

Metrics / Interviewees	A Memorability	B Features	C Attractiveness	D Simplicity (containing learnability)	E Accuracy
Interview_1_HP	5	5	1	10	11
Interview_10_HP	0	1	0	0	3
Interview_11_SD	5	6	1	7	8
Interview_12_SD	0	5	2	4	6
Interview_13_SD	1	10	0	6	3
Interview_14_SD	4	6	0	6	3
Interview_15_SD	1	12	0	5	2
Interview_2_HP	3	7	0	12	2
Interview_3_P	0	2	2	3	3
Interview_4_HP	3	6	0	8	2
Interview_5_P	3	5	0	9	3
Interview_6_HP	1	5	0	6	2
Interview_7_P	4	5	0	8	1
Interview_8_P	3	5	0	5	4
Interview_9_P	2	4	0	7	2
Total	35	84	6	96	55
Average	2.3	5.6	0.4	6.4	3.7

Table 4.7: Stakeholders metrics requirements percentage result from interviews

4.14 Mapping between Stakeholders and Metrics

The research applied two kinds of weighting. One of them was Individual Weighting between the stakeholders and the metrics and the second one was Overall Weighting between stakeholders and the overall metrics.

For Health Professionals (HP) Individual Weighting (HPIW) a calculation was applied, between HP and the metrics by dividing HP Total of Requirement for Specific Metric TRSM by the overall of HP Total for all of the Requirement for All the Metrics TRAM from Table 4.1:

HPIW = TRSM / TRAM

The results were:

A. Memorability = 12 / 93 = 12%

B. Features = 24 / 93 = 26%

C. Attractiveness = 1 / 93 = 1%

D. Simplicity = 36 / 93 = 39%

E. Accuracy = 20/ 93 = 22%

Figure 4.9 shows the mapping of the relationship between Health Professional (HP) and the Final Hybrid Selection Metrics.

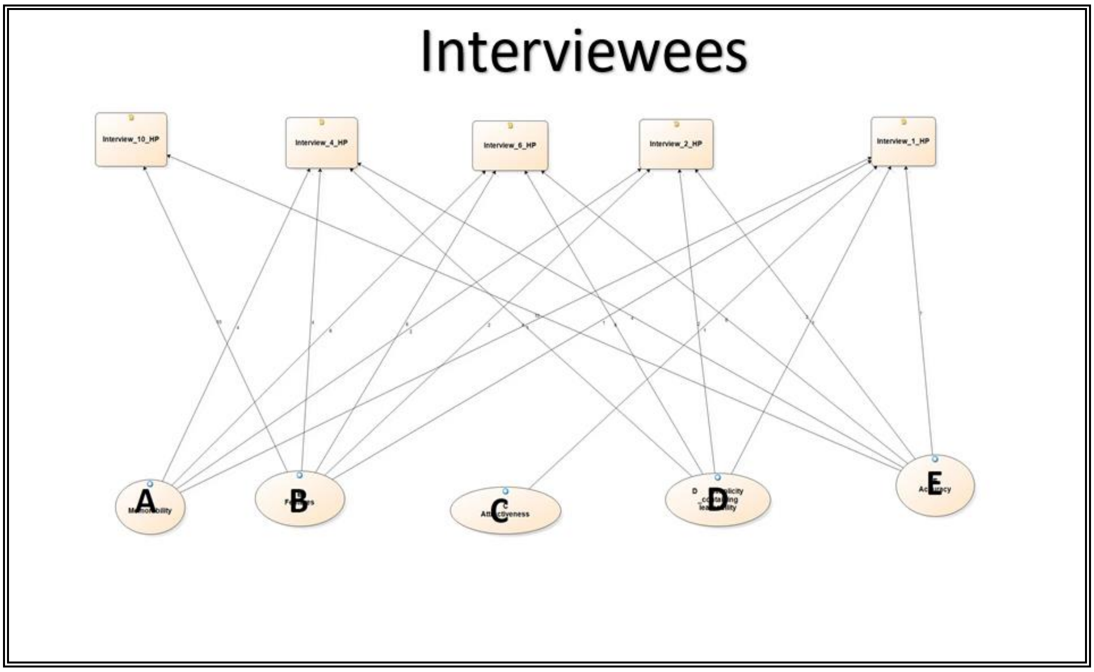


Figure 4.9: Mapping Between HP Requirements and Final Metrics

For Patients (P) Individual Weighting (PIW) a calculation has been applied, between P and the metrics by dividing P Total of Requirement for Specific Metric TRSM by the overall the P Total for all of the Requirement for All the Metrics TRAM from Table 4.2:

PIW = TRSM / TRAM

The results were:

A. Memorability = 12 / 80 = 15%

B. Features = 21 / 80 = 26%

C. Attractiveness = 2 / 80 = 3%

D. Simplicity = 32 / 80 = 40%

E. Accuracy = 13 / 80 = 16%

Figure 4.10 shows the mapping of the relationship between the Patients (P) and the Final Hybrid Selection Metrics.

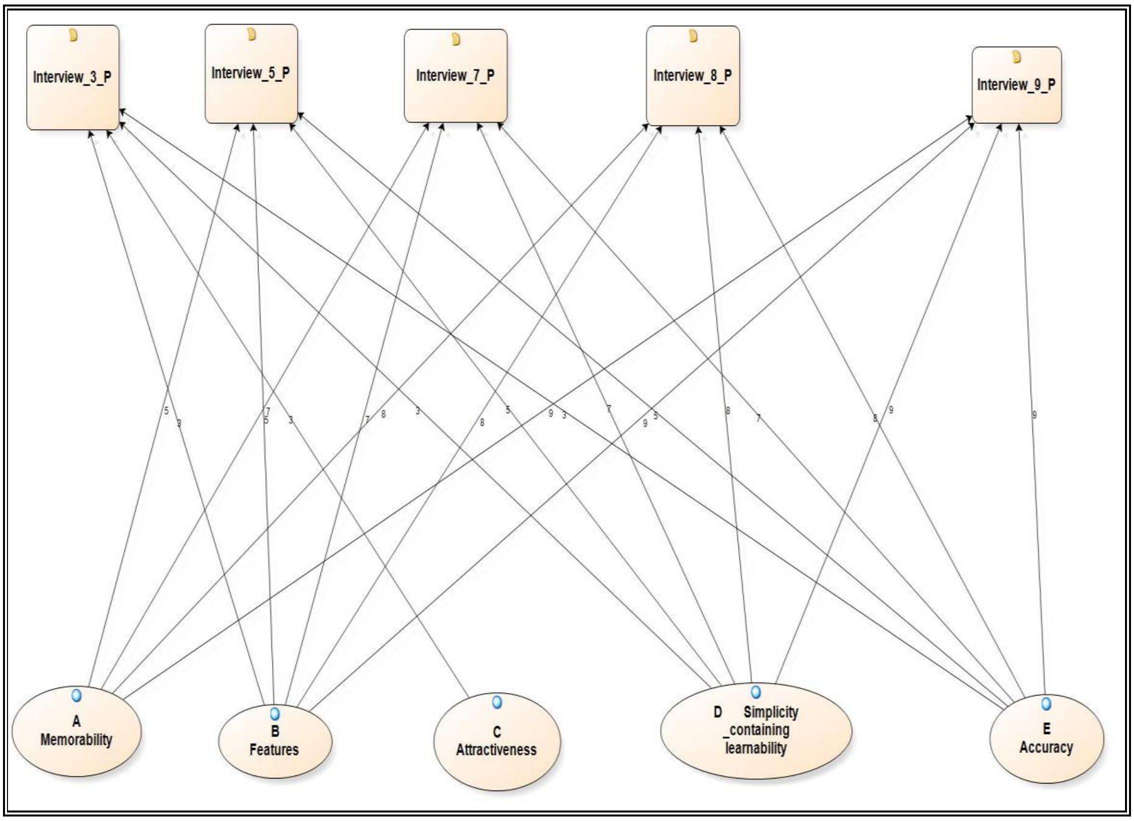


Figure 4.10: Mapping Between P Requirements and Final Metrics

Moreover, Figure 4.11 shows the mapping of the relationship between both the Patients (P) and Health Professional (HP) with the Final Hybrid Selection Metrics.

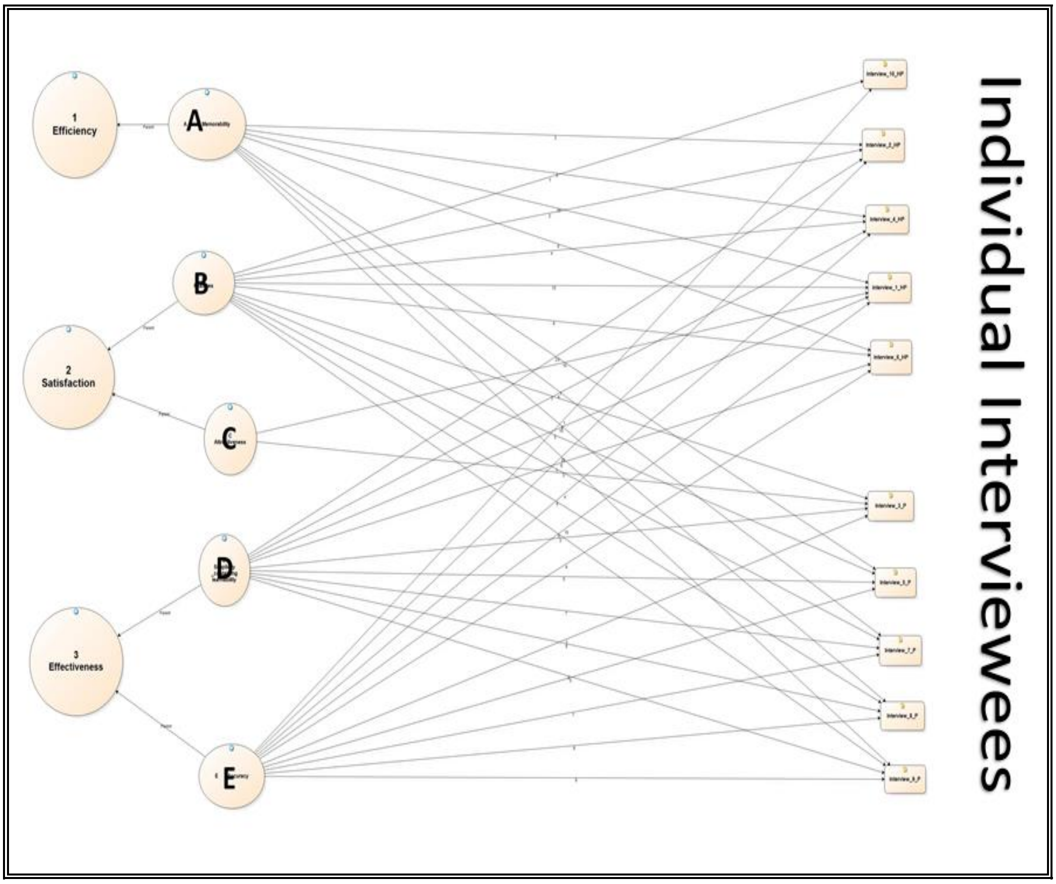


Figure 4.11: Mapping between HP and P Requirements And Final Metrics

For Software Developers (SD) Individual Weighting (SDIW) a calculation was between SD and the metrics by dividing SD Total of Requirement for Specific Metric TRSM by the overall SD Total for all of the Requirement for All the Metrics TRAM from Table 4.3:

SDIW = TRSM / TRAM

The results were

- **A.** Memorability = 11 / 103 = 11%
- **B.** Features = 39 / 103 = 38%
- **C.** Attractiveness = 3 / 103 = 3%
- **D.** Simplicity = 28 / 103 = 27%
- **E.** Accuracy = 22 / 103 = 21%

For Health Professionals (HP), Patients (P) and Software Developers (SD) Overall Weighting, following calculation of the Individual Weighting for each category of stakeholders' aggregation was applied with the results shown in Table 4.8 and Figure 4.12. This helped in obtaining the overall percentage and weighting for all stakeholders, for all the overall metrics requirements. The weightings and percentages were:

- **A.** Memorability = 38% with an overall weighting average of 13%
- **B.** Features = 90% with an overall weighting average of 30%
- **C.** Attractiveness = 7% with an overall weighting average of 2%
- **D.** Simplicity = 106% with an overall weighting average of 35%
- **E.** Accuracy = 59% with an overall weighting average of 20%

In first place for the overall weighting for all stakeholders was D) Simplicity, with a percentage of 106% as the highest percentage. In second place was B) Features with a percentage of 90% as the second highest percentage. In third place was the E) Accuracy, with a percentage of 59%, as the third highest percentage. In fourth place was A) Memorability, with a percentage of 38%, as the fourth highest percentage. In last place was C) Attractiveness with a percentage of 7% as the lowest percentage.

Metrics / interviewees	A Memorability	B Features	C Attractivness	D Simplicity (containing learnability)	E Accuracy	Total percentage
Health Professional (HP)	12%	26%	1%	39%	22%	100%
Patients (P)	15%	26%	3%	40%	16%	100%
Software Developers (SD)	11%	38%	3%	27%	21%	100%
Total	38%	90%	7%	106%	59%	300%
Average	13%	30%	2%	35%	20%	100%

Table 4.8: Stakeholders overall metrics requirements percentage result from interviews

Figure 4.12 reflects these overall weightings as a chart to explain and clarify the result more.

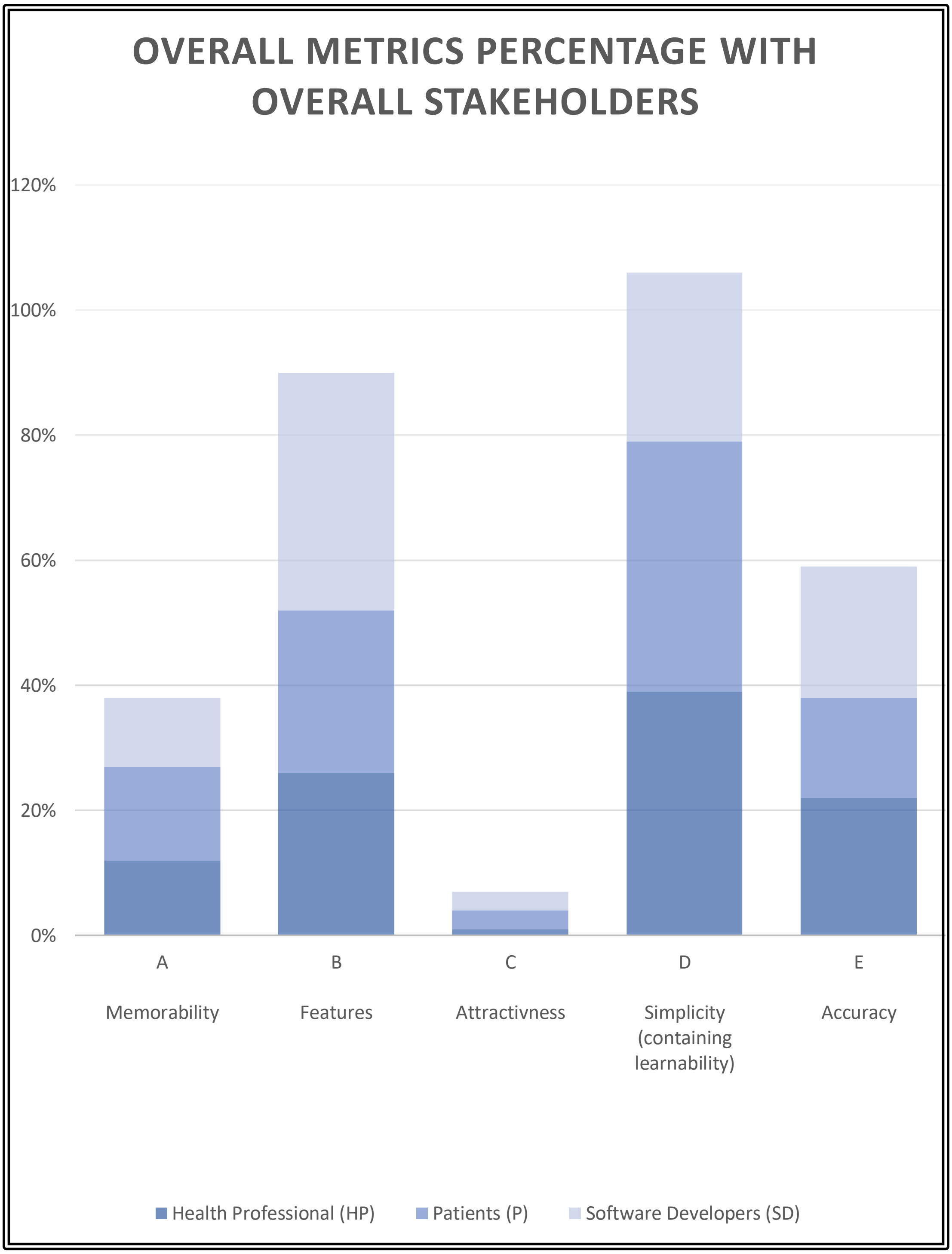


Figure 4.12: HP, P and SD Overall Metrics Requirements Result from all Interviews

4.15 Conclusion

In this Chapter, a description of the method utilized to filter the selection of hybrid metrics has been provided. Semi-structured interviews were applied with three categories of stakeholders, HP, SD and P. Analysis of the interviews yielded a selection of final metrics as: A. Memorability, B. Features, C. Attractiveness, D. Simplicity and E. Accuracy for MHEAs. These metrics would become a part of the framework; developed to rank and measure the usability of MHEAs. Another outcome from the interviews was to identify the major requirements for HP, P and SD for using MHEAs focusing on Simplicity (more simple design, simple to use, GUI, less complexity), and Features (more images to explain, more video/audio to help to understand). In addition, a mapping between the stakeholders and the metrics were described.

Chapter Five will apply the selected hybrid metrics by mapping them to the questionnaire questions, and it will be shown how the questionnaire was used to rank MHEAs.

Chapter 5: Designing the Questionnaire and Quantitative Results

In order to develop an effective and efficient way of choosing apps, the previous chapter discussed how the selected hybrid metrics were selected, depending on qualitative semi-structured interviews, in order to reduce the proposed metrics from 6 to 5 categories of metrics, to make it more efficient. The relative importance of categories of metrics was analysed in section 4.14. The research has a weighting for each category, based on the needs of each stakeholder group. In addition, a comparison was made between different stakeholders in terms of the metrics requirements besides weighting between metrics and stakeholders. Chapter Five presents the process of addressing objectives 2, 3 and 5; namely, developing a ranking measurement system to assess and rank Mobile Health Education Applications (MHEAs), identifying issues in MHEAs identified by the rankings to improve the design of MHEAs and characterizing the profiles of different stakeholders in order to determine their different requirements. In order to meet these objectives the selected hybrid metrics were converted into questionnaire questions. The chapter clarifies the stages of designing the questionnaire questions. Besides given the quantitative nature of the data, this chapter outlines the statistical approaches used to validate the data before analysing it, As well as the statistical approaches used to analyse the data and test the hypothesis. In addition, it explores the outcomes from analysing the questionnaire feedback from all stakeholders, and shows how the outcome from this analysis was used to measure and rank MHEAs on each individual metric rank. It also, finds each stakeholder group's individual metric rankings and matches MHEAs profiles with stakeholder profiles. Finally, the chapter describes how the outcome from analysing the questionnaire was used in designing the Medical Apps Selection Tool (MAST), which will be discussed in Chapter Six.

5.1. Generating the Number of Questionnaire Questions

This section describes the method utilized for deciding the number of questionnaire questions, in order to design the questionnaire. In the beginning, a proposed questionnaire design was developed in order to reflect all the selected hybrid metrics, with four questions per metric, giving 20 questions for the five metrics, and one overall question to give an overall view of all the metrics. In order to create a more effective approach, weightings were used to enable a focus on the key metrics to reflect the stakeholder requirements. Following the results reported in Chapter Four from the stakeholder interviews, it emerged that there are differences in priority among these metrics, and that the most important metrics to be measured for the stakeholders are B) Features and D) Simplicity (containing learnability).

Next in priority come A) Memorability and E) Accuracy. The least important metric to focus on, in terms of relevance to stakeholders is C) Attractiveness. Also, the weightings of these metrics differed from each other.

From these outcomes, the questionnaire was amended and the number of questions used to reflect each metric was changed. Instead of four questions per metric, the below formula shown in Table 5.1 was followed, based on the average of the overall weightings for the stakeholders:

METRIC	STANDARD QUESTIONS NUMBER	AVERAGE OF THE OVERALL WEIGHTING	TOTAL QUESTIONS NUMBER
A) Memorability	**20**	**13%**	**2.6**
B) Features	**20**	**30%**	**6**
C) Attractiveness	**20**	**2%**	**0.4**
D) Simplicity(containing learnability)	**20**	**35%**	**7**
E) Accuracy	**20**	**20%**	**4**

Table 5.1: Calculating the Relative Importance of Metrics

The original total number of questions in the initial design of the questionnaire, which was 20 was multiplied by the overall weighting for the stakeholders per metric, as shown above in Table 5.1, in order to obtain the total number of questions for each metric. Where, the results obtained did not provide a whole number, they were rounded up to the nearest ceiling number, to give the appropriate number of questions for each metric, as below in Table 5.2:

METRIC	STANDARD QUESTIONS NUMBER	AVERAGE OF THE OVERALL WEIGHTING	TOTAL QUESTIONS NUMBER	Approximate TOTAL QUESTIONS NUMBER
A) Memorability	20	13%	2.6	≈3
B) Features	20	30%	6	6
C) Attractiveness	20	2%	.04	≈1
D) Simplicity(containing learnability)	20	35%	7	7
E) Accuracy	20	20%	4	4

Table 5.2: Calculating the Questionnaire Accurate Questions Number

5.1.1 Questionnaire Architecture Review

This section is reviews the questionnaire architecture, which was started in Chapter Three and was modified as described in section 5.1. This showed how the number of questions was derived, by multiplying the standard number of questions by the average of the overall weighting per metric, in order to obtain an accurate number of questionnaire questions. This was illustrated in Chapter Three section 3.6.3 in Tables 3.1 and 3.2. The researcher demonstrated the mapping between the generally proposed metrics and the specific selected hybrid metrics, and these key areas were reflected in the questionnaire questions. The design of the questionnaire started from 21 questions to reflect the metrics A. Memorability, B. Features, C. Attractiveness, D. Simplicity (containing learnability), and E Accuracy.

A 5-point Likert scale response format was used for these questionnaire questions (Bertram, *et al*., 2007), to measure each metric, as well as one overall Likert scale question to reflect the overall metrics. Besides an open question was included to obtain responses to the questionnaire overall.

In order to check the suitability of the questionnaire questions, a pilot test of the questionnaire was conducted, with 9 participants, 3 from each category of stakeholders

(HP, SD, and P). This was done in order, to determine whether the questionnaire questions were clear, easy to understand and suitable for all categories of stakeholders, and whether the questionnaire helped to generate reliable data. The pilot test indicated that the questionnaire reliable.

5.1.2 Data Collection

In this section, the data collection procedure will be described, including the stakeholder categories and the numbers of participants. For any research, the number of participants involved typically depends on the effort needed to analyse data, plus money and time (Dumas & Redish, 1999). After the pilot test showed that, the questionnaire was valid, the actual data collection procedure started by having a mix of three categories of stakeholders as shown in Figure 5.1, which were:

1. Health Professionals (HP) involving 27 participants.
2. Software Developers (SD) involving 27 participants.
3. Patients/General public (P/GP) involving 27 participants.

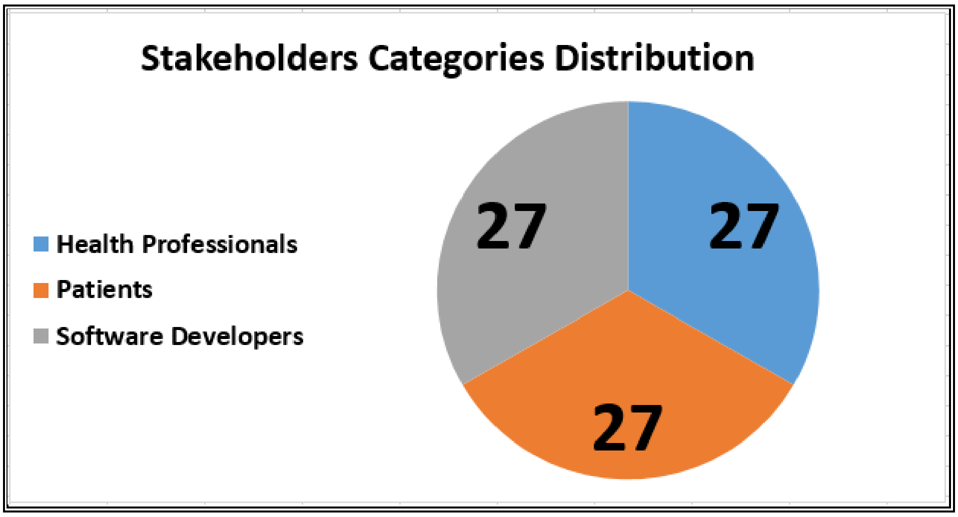


Figure 5.1: Participants in the Questionnaire Survey by User Categories

Thus, the total number of participants (stakeholders) was 81. They were divided among users of three categories of MHEAs: Medscape, Epocrates and Other Apps. All these stakeholders had been regular users of MHEAs for two years, so, they were familiar with MHEAs, and capable of responding to the questionnaire in a trustworthy manner. The reason for having 81 participants is that, according to the literature, a sample size of 25 participants is a good size, but for as this study involved three categories, the sample needed was 75 participants. Therefore, the research already used 81 participants, distributed in 27 per stakeholder category in the questionnaire (InterQ Research, 2017). The questionnaire was sent to the participants as a self-administrated online questionnaire, and was verified that they were free from any distractions.

As an example, the first participant was a female HP, in the age group 25-34 years old and she had been a regular user of Medscape for 2 years. Her overall Likert scale questionnaire question was:

Did you find the overall usability [Remembering, Features, Attractiveness, Simplicity and accuracy] of the health application useful?						
No at all useful	**1**	**2**	**3**	**4**	**5**	**Extremely useful**

Her answer was 5 (Extremely useful). Furthermore, to verify her response her answer was followed by the overall open question, to which she replied:

"*Overall good useful easy to pick up necessary information especially when I am in a place where medical books are not available. Giving accurate medical information for health professionals*"

However, before starting the procedure of analysing the data obtained from the 81 participants, the researcher verified that the data was valid to be considered in the framework, as explained in the next section 5.2.

5.2 Verifying Questionnaire Outcomes Data

The phase of verifying questionnaire outcome data is a critical phase, since if the data is not reliable; it cannot be trusted, to be used for further analysis. This would mean there is a need to collect the data again. In this phase, the Cronbach's Alpha (α) test was applied, using the SPSS program, before the start of analysis. The reasons for using Cronbach's Alpha, is that it is the most common measure of internal consistency ("reliability"). Moreover, it is the most commonly used for multiple Likert –scale format questions in a questionnaire (Divya , et al., 2016). Furthermore, Kumar and Mohite (2018) stated in there research that the reliability of the data was assessed using Cronbach's alpha.

Lee Cronbach established the Cronbach Alpha in 1951, as a measure of the internal consistency of a test or scale. Cronbach's Alpha is specified as a number between 0 and 1. The internal consistency describes the extent to which all the items in a scale measure, the same concept and therefore it is linked to the inter-relatedness of the items within the scale. Determining internal consistency should be done before a test can be employed for research purposes, to ensure validity. Furthermore, reliability estimates indicate the amount of measurement error in a test. Put simply, this analysis of reliability is the correlation of a measure with itself. Squaring this correlation and subtracting from 1.00 produces the key to measurement error (Tavakol & Dennick, 2011). For instance, if a scale has a reliability of 0.80, that means there is 0.36 error variance (random error) in the totals (0.80×0.80 = 0.64; 1.00 – 0.64 = 0.36). As the estimation of reliability increases, the proportion of a scale score that is attributable to error will drop. However, it should be noted that the reliability of a test reveals the effect of measurement error on the group of participants rather than on an individual. The standard error of measurement (SEM) must be calculated to determine the effect of measurement error on the observed score of an individual. The value of alpha is higher if the items in a scale are correlated to each other. However, a higher coefficient alpha does not always mean a high degree of internal consistency (Tavakol & Dennick, 2011). This is because alpha is affected by the scale length. The alpha value is reduced if the scale is too short. Hence, the more correlated items there are testing the same concept to a scale, the higher the alpha will be. Moreover, if the value of alpha is very high (> 0.90), this could indicate redundancies and suggest that the scale length should be shortened (Tavakol & Dennick, 2011).

Table 5.3 displays the value of Cronbach's alpha, for the data collected via the questionnaire, which is 0.774. This indicates a high level of internal consistency for our scale with this specific sample.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.774	.783	6

Table 5.3: The Reliability Statistics Table

Therefore, the researcher can be confident that the data collected is good and reliable(Manerikar & Manerikar, 2015), according to the citation in Table 5.4.

Cronbach's alpha	Internal consistency
α = 0.9	Excellent (High-Stakes testing)
0.7 = α < 0.9	Good (Low-Stakes testing)
0.6 = α < 0.7	Acceptable
0.5 = α < 0.6	Poor
α < 0.5	Unacceptable

Table 5.4: Acceptance Stages for Cronbach's alpha
(Lund Research, 2013)

The various measures of Cronbach's alpha for each construct in the questionnaire are presented in Table 5.5.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
A)MEAN	19.191	7.146	.398	.382	.782
B) MEAN	18.976	8.622	.249	.430	.798
C) MEAN	19.204	7.655	.525	.400	.740
D) MEAN	18.949	6.950	.648	.529	.707
E) MEAN	19.270	7.040	.541	.674	.734
Over-All	18.991	6.519	.849	.783	.659

Table 5.5: The Item Total Statistics Table

5.3 Ranking of MHEAs between All Stakeholders

The phase of ranking MHEAs entailed ranking three types of MHEAs, in order to justify the different types of MHEAs among all stakeholders. These three types of MHEAs were Epocrates, Medscape and Other Apps. These were ranked by three types of stakeholders, which were HP, P/GP and SD, who had already been using the MHEAs for two years. The following sections illustrate these ranking results, besides the outcomes from these rankings.

5.3.1 Ranking of EPOCRATES App among All Stakeholders:

To start the procedure of ranking Epocrates, a data sheet on SPSS software was created. Then the data collected from the 81 responders to the questionnaire were entered into this sheet in SPSS. Analysis revealed that users of the Epocrates app were 25 participants (6 HP, 8 SD and 11 P/GP). Figure 5.2 shows these stakeholder's rankings of the Epocrates app. Based on their experience of using Epocrates for the last two years. It reveals that different stakeholder groups gave different rankings to the same app. The research revealed that the highest ranking for Epocrates was by P/GP, with a score of 4.5. In second place was the ranking by HP, which was 3.3. The ranking by SD was the lowest, at 3.

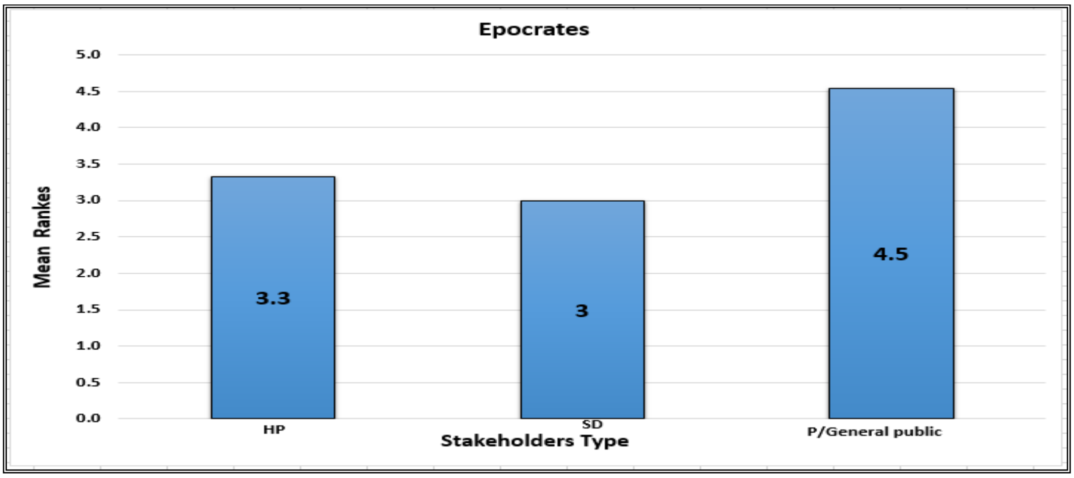


Figure 5.2: Epocrates Overall Ranking by HP, SD and P/GP

5.3.2 Ranking of Medscape App among All Stakeholders:

Afterward obtaining the ranking result for Epocrates from all stakeholders, the same procedures was applied to Medscape. The result revealed that Medscape was used by 30 participants (14 HP, 7 SD and 9 P/GP). Figure 5.3 illustrate the different rankings of the same app by different stakeholders. Moreover, the figure shows that the highest ranking for Medscape was by HP, with a score of 4.1, while the second highest ranking was by P/GP, of 4. In the last place were SD, who gave the app a ranking of 3.9.

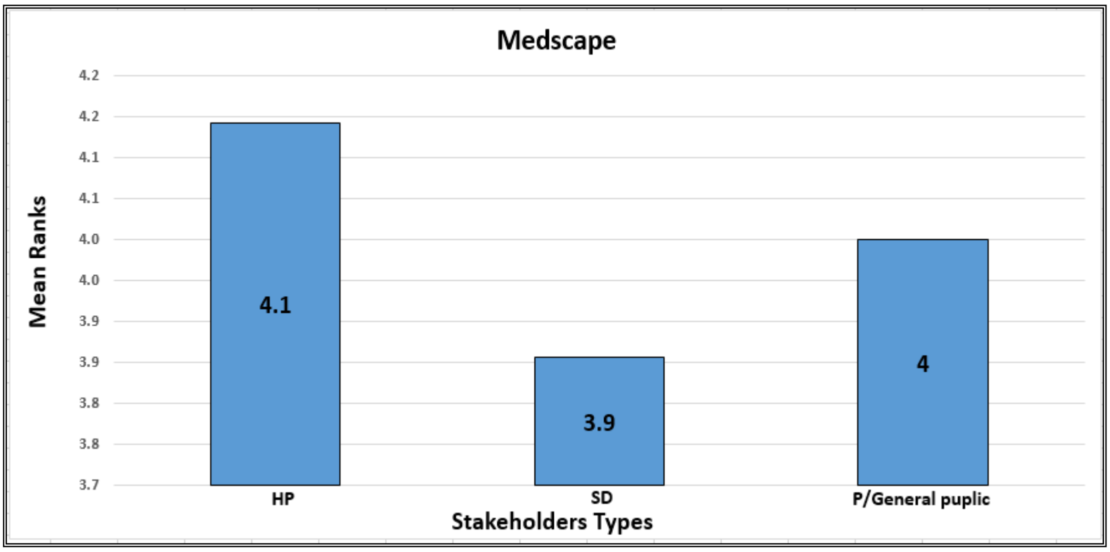


Figure 5.3: Medscape Overall Ranking by HP, SD and P/GP

5.3.3 Ranking of Other Apps among All Stakeholders:

The last MHEAs package to be ranked in the research was Other Apps, as described in this section. The analysis procedure applied was the same as for the other two MHEAs packages. The result revealed that Other Apps were used by 26 participants (7 HP, 12 SD and 7 P/GP). Figure 5.4 shows the different rankings for the same apps (Other Apps) by different stakeholders. It shows that the highest ranking was by SD, which was 4.5, in second place was the ranking by HP, which was 3.7, while the lowest ranking was by P/GP, which was 3.1.

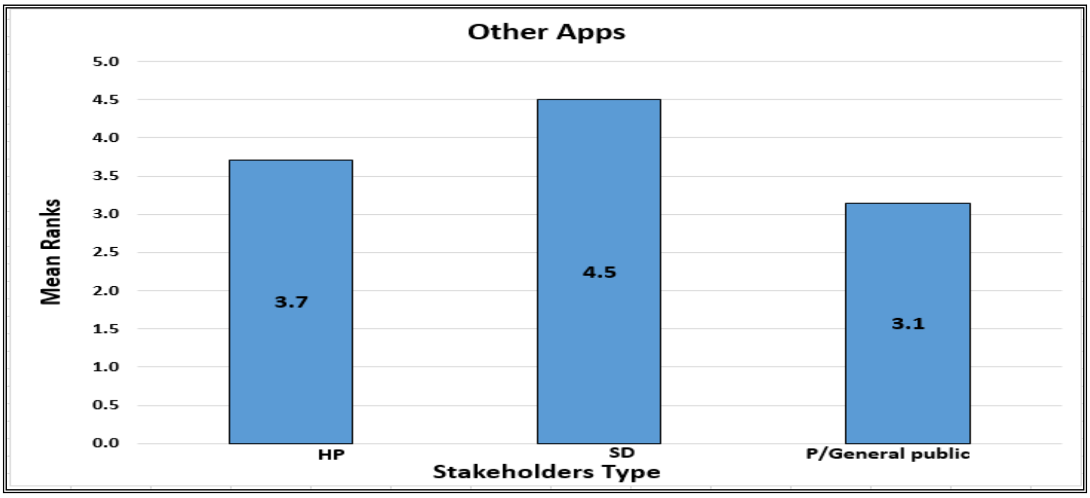


Figure 5.4: Other Apps Overall Ranking by HP, SD and P/GP

Overall, this phase revealed that different stakeholders ranked the MHEAs package differently and that each MHEAs package had a different ranking of stakeholders, which were need as a guide to developing the MAST in Chapter Six.

5.4 Ranking of MHEAs on All Metrics

The phase of ranking MHEAs on metrics involved looking at how the three types of MHEAs, (Epocrates, Medscape and Other Apps) were ranked on each of within the five selected hybrid metrics: Memorability, Features, Attractiveness, Simplicity and Accuracy. The purpose of ranking these metrics individually in each package of MHEAs was justify each metric profile, which would provide on the accurate profile for each MHEAs package. The following sections illustrate these ranking results and the outcomes from these rankings.

5.4.1 Ranking of Epocrates App on All Metrics:

This section describes how satisfied users were with the Epocrates app, based on the five-selected hybrid metrics included in our framework. Following the analysis of the questionnaire data obtained from 81 participants who had been using the MHEAs for 2 years, the result for the 25 users of Epocrates apps (6 HP, 8 SD and 11 P/GP) is shown in Figure 5.5. Figure 5.5 clarified that there are a differences in the ranking of the individual metrics in the same app. It can be seen that among the five selected hybrid metrics, the highest ranked metric was B) Features, with a score of (3.9), while, in second place was D) simplicity with (3.7). In third place was C) Attractiveness, with a score of (3.6), and joint last place are A) Memorability and E) Accuracy, which each scored (3.3).

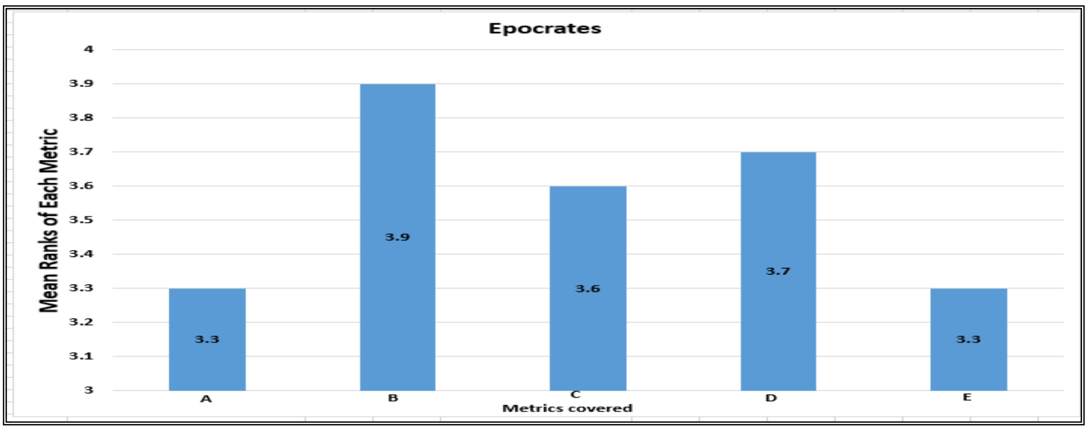


Figure 5.5: Epocrates App Ranking for All Metrics by HP, SD and P/GP

5.4.2 Ranking of Medscape App on All Metrics:

This section follows the same procedure as described in section 5.4.1, for Epocrates, to show the ranking of the five selected hybrid metrics for Medscape. The results for the 30 Medscape users (14 HP, 7 SD and 9 P/GP) are presented in Figure 5.6. The outcomes show how the stakeholders (HP, SD and P/GP) ranked the five selected hybrid metrics for the Medscape apps. Figure 5.6 illustrated that there are differences in the ranking of each individual metric in the same app. It shows that the highest ranked metric was for D) Simplicity (3.9). Followed, in the second place by B) Features that has approached (3.8), in joint third place A) Memorability and E) Accuracy, which each score (3.7), and, in the last place is C) Attractiveness, with a score of (3.6).

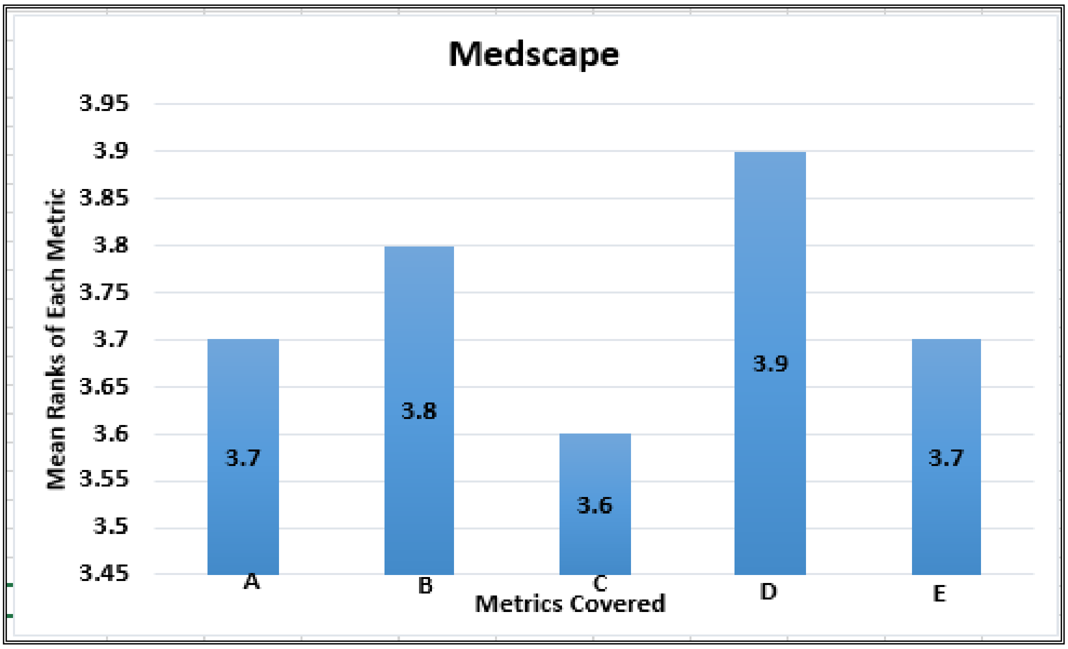


Figure 5.6: Medscape App Ranking for All Metrics by HP, SD and P/GP

5.4.3 Ranking of Other Apps on All Metrics:

This section focuses on analysing the questionnaire data obtained from 26 participants (7 HP, 12 SD and 7 P/GP), who were using Other MHEAs (e.g., WebMD, UpToDate). Figure 5.7 presents these stakeholder's ranking of the five selected hybrid metrics for the Other Apps. The outcomes show differences in the rankings of the individual metrics in the same app. The highest ranked metrics were A) Memorability and D) Simplicity, which scored (3.9), while, in second place was B) Features, which scored (3.8). In third, place was C) Attractiveness, which scored (3.7) and in last place was E) Accuracy, with a score of (3.6).

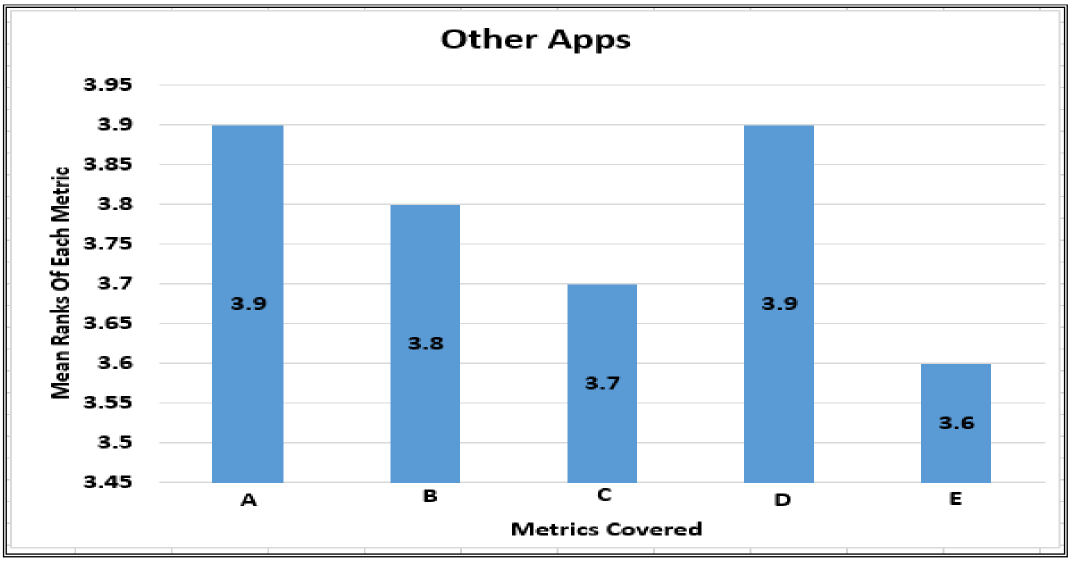


Figure 5.7: Other Apps Ranking for All Metrics by HP, SD and P/GP

5.5 Stakeholder's Priorities among Requirements of All Metrics

This section describes how the framework helped to prioritize the metrics requirements, among different stakeholders, this as a base for ranking three types of MHEAs, (Epocrates, Medscape and Other Apps), according to the five selected hybrid metrics (Memorability, Features, Attractiveness, Simplicity and Accuracy). The following sub-sections illustrate the priorities among the metrics requirements for each group of stakeholders in turn.

5.5.1 Health Professionals Priorities among Requirements of Metrics:

This section describes the requirements priorities among the metrics for health professional's based on the analysis of the questionnaire data, obtained from 27 HP participants, who had been using MHEAs for two years. The result is illuminated in Figure 5.8. Figure 5.8 illustrates that there are a differences in the requirements of HP for the individual metrics in MHEAs. The highest requirement expressed by HP was for D) Simplicity, with a score of (3.99). In second place was A) Memorability (3.89). In the third place was B) Features, with a score of (3.87). In fourth place was E) Accuracy, which scored (3.45) and in the last place was C) Attractiveness, which scored (3.37).

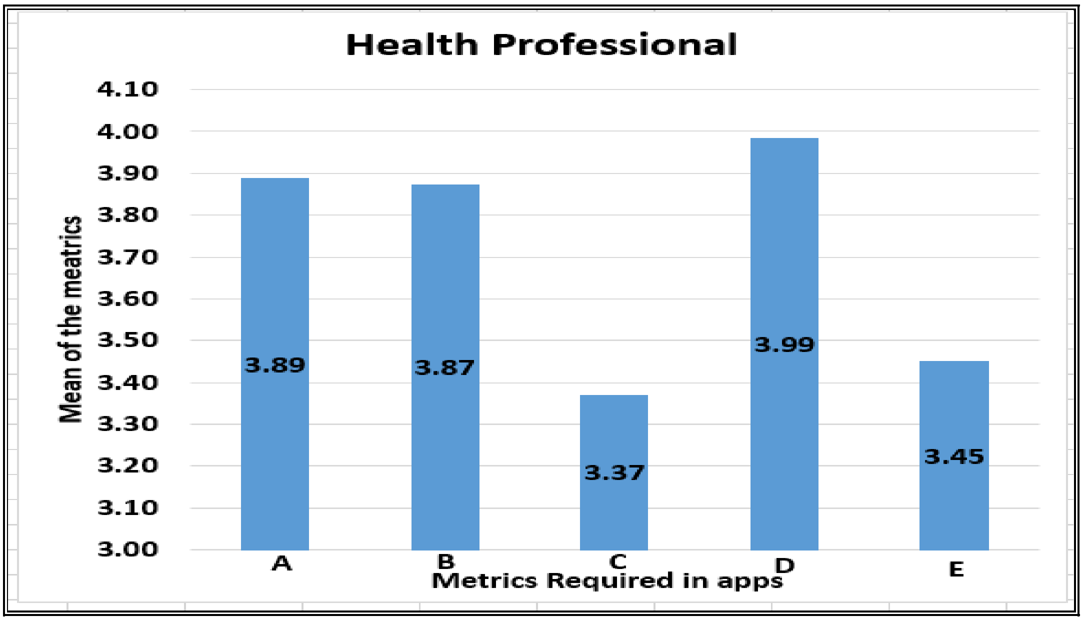


Figure 5.8: Metrics Required in Mobile Health Education Apps by HP

5.5.2 Software Developers Priorities among Requirements of Metrics:

This section adopts the same procedure as above, for the second stakeholder category, by describing the requirements priorities among metrics for software developers (SD). The analysis is based on the questionnaire data, from 27 SD participants, who had been using MHEAs for two years. The result is shown in Figure 5.9, which reveals that SD have different requirements for the individual metrics in MHEAs. It shows that the highest requirement was for D) Simplicity, with a score of (4.01). In second place was B) Features with a score of (3.93). In third place was C) Attractiveness, scoring (3.81). In fourth place was E) Accuracy with a score of (3.69) and in the last place was A) Memorability, which scored (3.39).

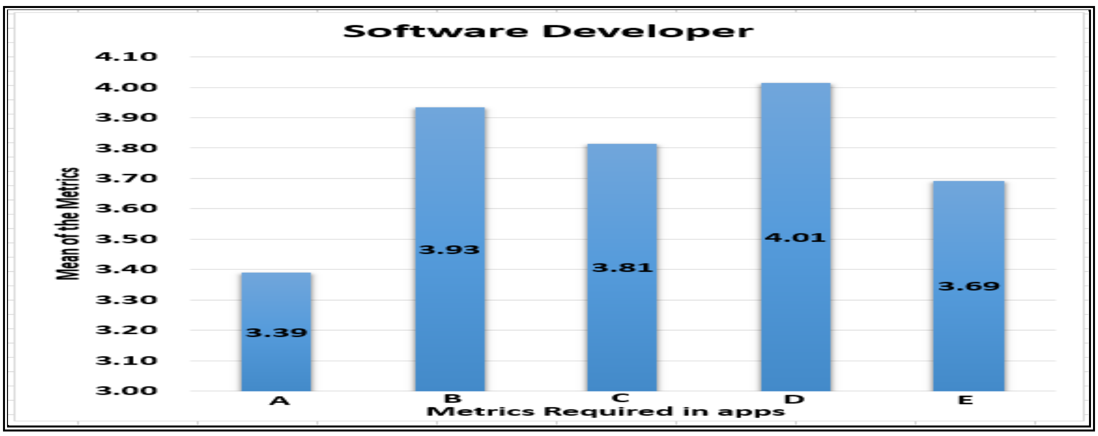


Figure 5.9: Metrics Required in Mobile Health Education Apps by SD

5.5.3 Patients and Other General Users Priorities among Requirements of Metrics:

This section is addresses the third stakeholder category, by describing the requirements priorities for patients (P) and Other General Users among all metrics. It is based in analysis of the questionnaire data obtained from 27 Patients and General Public (P/GP) participants, who had been using MHEAs for two years. The result is illustrated in Figure 5.10. The figure indicates that there are differences in the requirements of P/GP among the individual metrics in MHEAs. The highest requirement was for B) Features, with a score of (3.98). In second place was C) Attractiveness, which scored (3.93). In third place was D) Simplicity, which scored (3.90). In fourth place was A) Memorability, which scored (3.87), while in last place was E) Accuracy, with a score of (3.79).

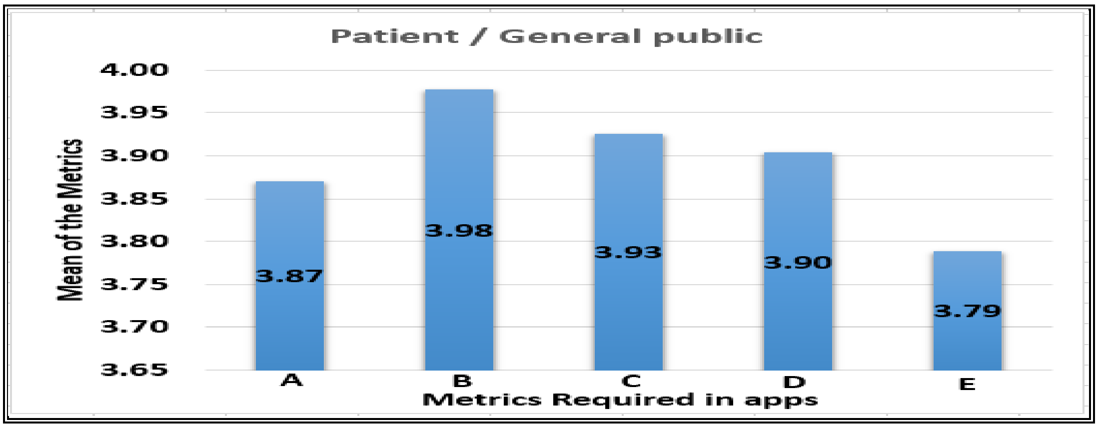


Figure 5.10: Patients and General Public Metrics Required in MHEAs

5.6 Comparisons among All Stakeholder Groups within MHEAs

This section describes how the framework helped, in comparing the metrics requirements, within the same MHEAs, among different stakeholders. This is done by ranking each type of MHEAs on all metrics, for the different users (stakeholders) at the same time. The three types of MHEAs (Epocrates, Medscape and Other Apps) are ranked on the five selected hybrid metrics (Memorability, Features, Attractiveness, Simplicity and Accuracy) for each group of users (stakeholders). The following sections illustrate the priorities among metrics requirements for each type of MHEAs in turn.

5.6.1 Comparisons among All Stakeholders for Epocrates App

This section is describes the metrics requirements priorities among different stakeholders, for the Epocrates App, based on the questionnaire data, obtained from 81 participants, who had been using MHEAs for two years. The result for the users of Epocrates apps is based on 25 participants (6 HP, 8 SD and 11 P/GP). Figure 5.11 is reflects these stakeholders rankings of the metrics for the Epocrates apps. The figure shows that there are differences in the rankings of the individual metrics among different stakeholders. It can be seen that the highest ranking for metric A) Memorability was by patients, with a score of 4.2. Health professionals came in second place, with a ranking of 3.3, while, in last place were software developers, with a ranking of 2.4. The highest ranking for metric B) Features was by patients, with a value of 4.3, in the second place came health professionals, with a value of 3.7. While the lowest ranking was for software developers, with a value of 3.6. The highest ranking for metric C) Attractiveness was for patients, with a value of 4.4, while health professionals and software developers came in joint second place with a value of 3.3. The highest ranking for metric D) Simplicity was for patients, with a value of 4.5; in second place was software developers, with a value of 3.4, while the lowest ranking (3.3) was for health professionals The highest ranking for metric E) Accuracy was for patients, with a value of (4.3). In second place was health professionals, whose ranking was (3), while, in last place was software developers with a value of (2.6). The highest Overall ranking of the metrics was for patients, at 4.5. In second place, was health professionals, with a value of (3.3), while the lowest ranking was for software developers, with a value of (3). From these data, the research indicates that wither on individual metrics or the overall metrics. Epocrates is most suitable for patients; this is the group that ranked it highest. However, considering the individual metrics, the Epocrates app is the most suitable for patients for all of the individual metrics. The Epocrates app is the second most suitable for health professionals considering individual metrics (except D) Simplicity), as it came as the third most suitable metric for health professionals. Finally, the Epocrates app is the third most suitable for software developers considering individual metrics except D) Simplicity and C) Attractiveness, as D) Simplicity came as the second most suitable metric for software developers and C) Attractiveness came as joint the second most suitable metric for software developers and health professionals.

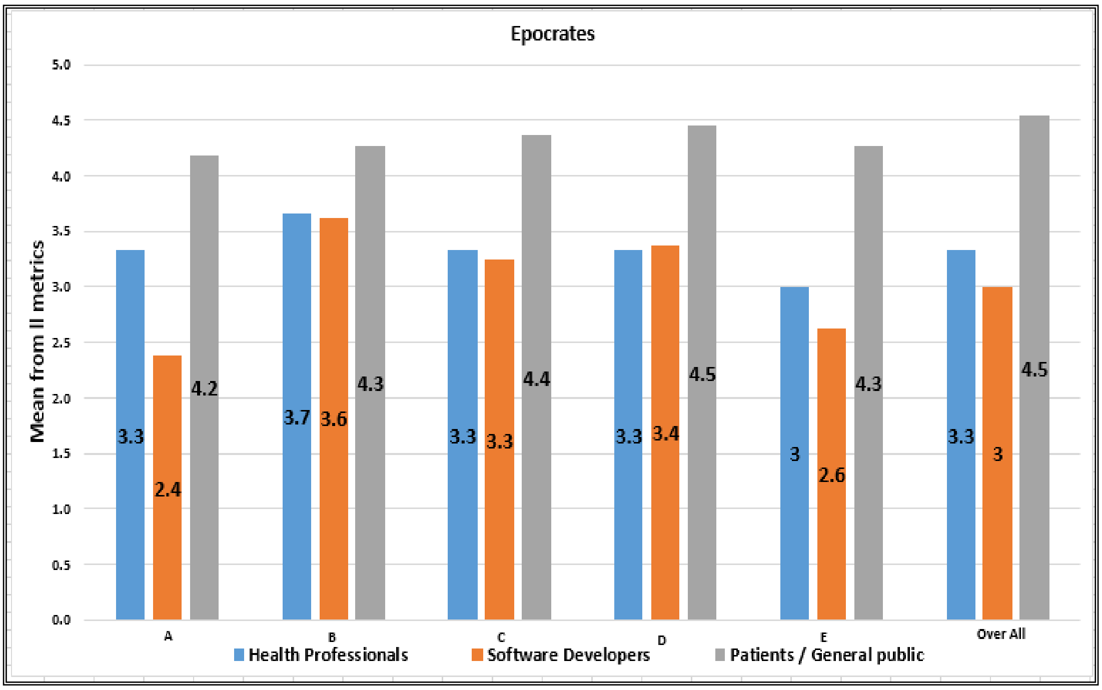


Figure 5.11: Epocrates App ranking comparison for all metrics by HP, SD and P/GP

5.6.2 Comparisons among All Stakeholders for Medscape App

This section follows the same procedure as in the previous section, for the second, MHEAs package, Medscape, showing the rankings of different stakeholder's categories for all metrics. The result for Medscape apps is based 30 participants (14 HP, 7 SD and 9 P/GP). Figure 5.12 shows these stakeholder's (HP, SD and P/GP) ranking of the metrics for the Medscape apps. It shows differences in among the groups in their rankings of the individual metrics. The highest ranking for metric A) Memorability was for HP, with a ranking of 4, while P/GP, came in second place, with a value of 3.7 and, in the last place were software developers, with a value of 3.6. The highest ranking for metric B) Features was for health professionals at 4.2, in second place, was that of patients, 4, while, in last place was the ranking of software developers, which was 3.3. The highest ranking for metric C) Attractiveness was for patients, at 3.7, in second place, came software developers, with a value of 3.6, while, in the last place came health professionals, with a value of 3.4. The highest ranking for metric D) Simplicity was for health professionals, at 4.4, followed in second place, by patients, with a value of 3.7, while the value was for software developers, at 3.6. The highest ranking for metric E) Accuracy was for software developers, with a value of 3.8. In joint second place, were health professionals and patients, who both yielded of 3.7. The highest rank for the Overall metrics was for health professionals, at 4.1. In second place was the patient's, with a value of 4. In the last place came software developers, with a value of 3.9. From these data, the research indicated that both in individual metrics and overall metrics Medscape is most suitable for HP, as received the highest ranking from HP. However, considering individual metrics, Medscape is the most suitable for health professionals for individual metrics except metrics C) Attractiveness and E) Accuracy. As C) Attractiveness is the third most suitable metrics for health professionals in addition E) Accuracy is joined as the second most suitable metrics for health professionals and. Patients. Nevertheless, Medscape apps is the second most suitable for Patients considering individual metrics except C) Attractiveness, as it came as the first most suitable metric for Patients. Finally, Medscape apps is the third most suitable for software developers considering individual metrics except C) Attractiveness and E) Accuracy, as C) Attractiveness came as the second most suitable metric for software developers and E) Accuracy came as first most suitable metric for software developers.

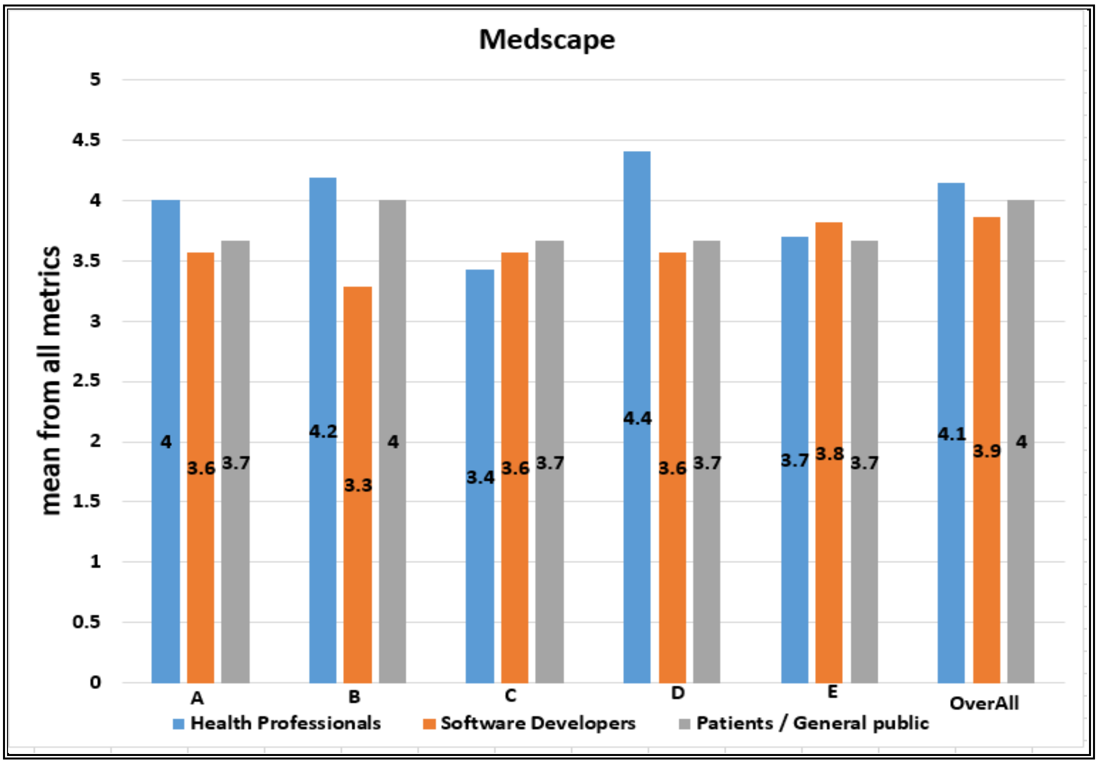


Figure 5.12: Medscape App ranking comparison for all metrics by HP, SD and P/GP

5.6.3 Comparisons among All Stakeholders for Other Apps

This section adopts the same approach as in the previous two sections, for the third MHEAs package Other Apps. The result for Other Apps is based on the data from 26 participants (7 HP, 12 SD and 7 P/GP). Figure 5.13 is reflects these stakeholder's rankings of the metrics for the Other Apps, revealing differences in the way they rankings each individual metric in the same apps. It shows that the highest ranking for metric A) Memorability was for health professionals, with a value of 4.1, while software developer's ranking came in second place, at 4. In last place was Patients, with a value of 3.6. The highest ranking for the metric B) Features was for software developers, at 4.5, in the second place, came patients, with a value of 3.5, while, in last place came health professionals, with a value of 3.4. The highest ranking for metric C) Attractiveness was for software developers, with a value of 4.3. In second place came patients, with a value of 3.6, while, in last place came health Professionals, yielded a value of 3.3. The highest ranking for metric D) Simplicity was for software developers, with a value of 4.7. In second place came health professionals, with a value of 3.7, while the lowest ranking was for patients, with a value of 3.3. The highest ranking for metric E) Accuracy was for software developers, at 4.3, followed by health professionals, whose value was (3.3), while, in the last place came patients, with a value of 3.2. The highest ranking for the Overall metrics was for software developers, at 4.5. In second place came health professionals, with a value of 3.7. In last place came patients, whose value was 3.1. From these data, the research indicates that both for in individual metrics and overall metrics Other Apps as the most suitable for software developers as this type gained the highest rankings from software developers. However, considering individual metrics only, the "Other Apps" was the most suitable for software developers for individual metrics, except metric A) Memorability, where it was the second most suitable metrics for software developers. Nevertheless, Other Apps is the second most suitable for health professionals considering individual metrics except A) Memorability, B) Features and C) Attractiveness, as A) Memorability came as the first most suitable metric for health professionals, B) Features and C) Attractiveness came as the third most suitable metric for health professionals. Finally, Other Apps is the third most suitable for patients considering individual metrics except B) Features and C) Attractiveness. As both B) Features and C) Attractiveness came as the second most suitable metric for patients.

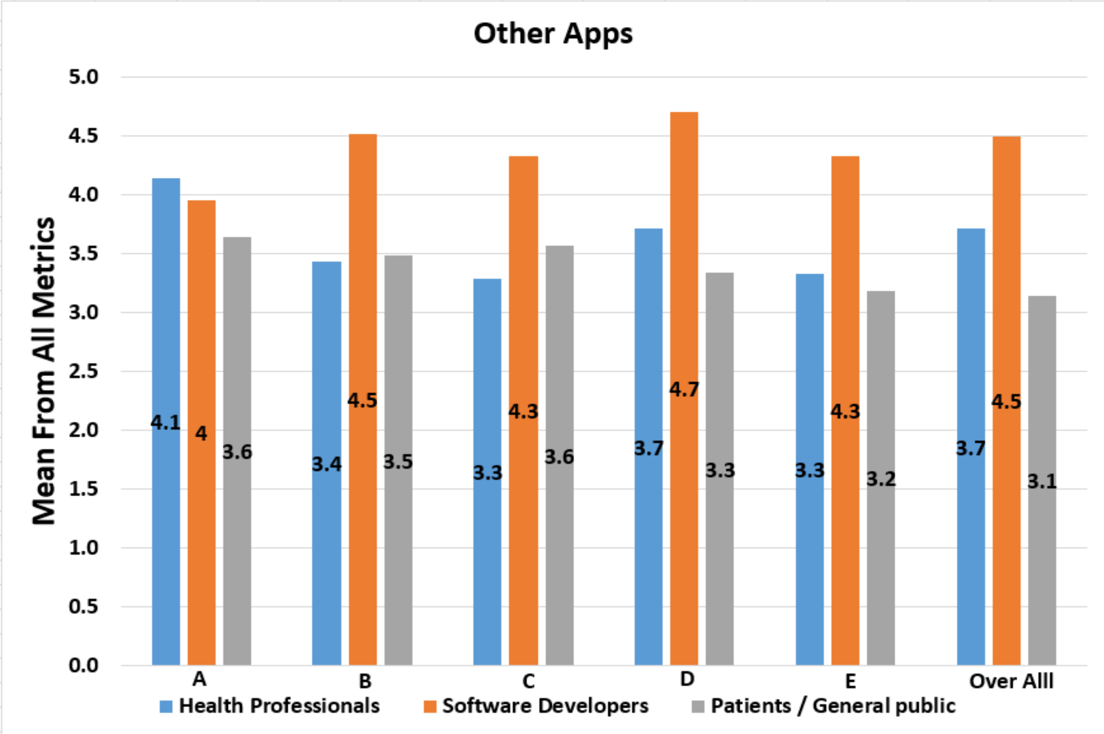


Figure 5.13: Other Apps Ranking Comparison for All Metrics by HP, SD and P/GP

5.7 Matching between Stakeholder Priorities and MHEA Metrics

From the results of analysing the questionnaire data, the researcher identified the match between different stakeholder's profiles and the profiles of different software packages.

The profile for the health app Medscape is the best match to the profile for the user group of health professionals. Figure.5.6 in section 5.4.2 illustrated that in the Medscape profile the highest-scoring metrics score were A=3.7, B=3.8, and D=3.9. These are the best match for the health professional's profile in Figure.5.8 in section 5.5.1, which showed that health professional's highest metrics scores were A=3.89, B=3.87, and D=3.99. The differences between these profiles are A=0.19, B=0.07 and D=0.09. In contrast the score for the Medscape health app for metric E was 3.7, which is low. However, this matches the health professional's profile for E=3.45, which is the lowest among the metrics with a differential for E=0.25. This as can be seen in Figure 5.14, the profile of Medscape is the best match for the profile of health professionals (Aljaber & Gordon, 2017b). The matching between health professionals' profile and the Medscape profile is useful in two ways. Firstly, in term of identifying the best match between a category of stakeholders profiles and the different mobile health education applications packages; this shows how certain packages suit certain groups and demonstrates that this can be identified through the profiling. Secondly, this matching is applied as a one of the matching options in the MAST Tool, which is explained in chapter 6.

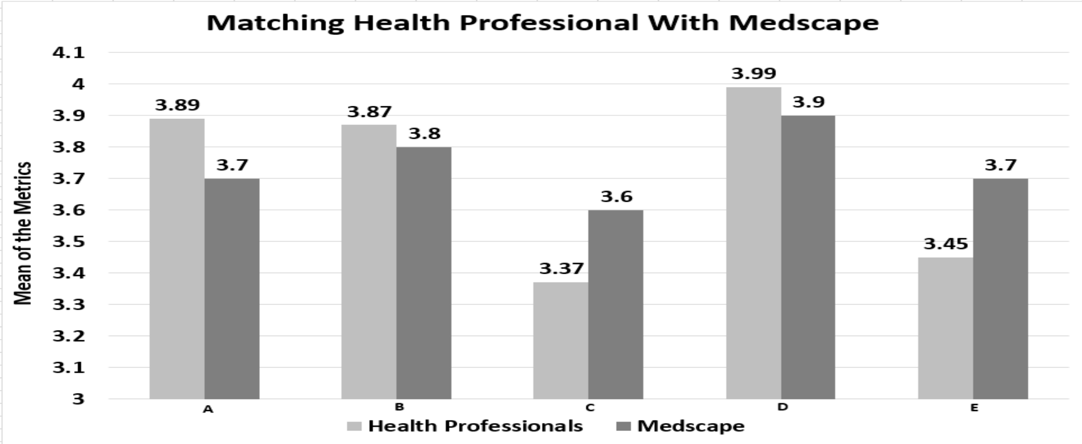


Figure 5.14: Matching Health Professionals profile with Medscape Profile

The profile for the health app Epocrates that the best match in the user group profile of patients. Figure.5.5 in section 5.4.1 illustrated that in the Epocrates profile, the highest metrics scores were B=3.9, C=3.6, and D=3.7. These are the best match for the patients profile presented in Figure.5.10 in section 5.5.3, which showed that the patients highest-scoring metrics were the same three, with B=3.98, C=3.93, and D=3.9. Moreover, the differentials between the patients and apps profiles for these metrics are B= 0.08, C= 0.33, and D= 0.2. Meanwhile the score for the health app Epocrates for metric A=3.3, which is low, matching the profile of patients for A=3.87, is the best match for the low score with a differential of A=0.57. By looking at Figure 5.15 it can be seen that the profile for Epocrates is a good match for the profile of patients (Aljaber & Gordon, 2017b).

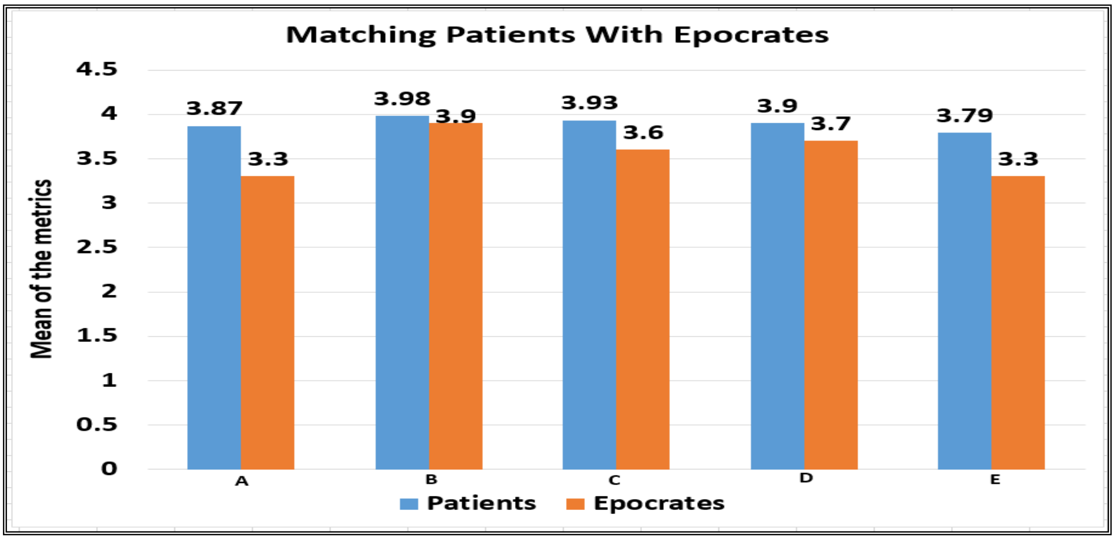


Figure 5.15: Matching Patients Profile with Epocrates Profile

The profile for the health app Other Apps is the best match for the profile user group, software developers. Figure.5.7 in section 5.4.3 illustrated that, Other App's highest metrics scores were for (B, C, D) and with B=3.8, C=3.7 and D=3.9. This is the best match for the software developer's profile in Figure.5.9 in section 5.5.2. The software developer's profile showed that the highest-scoring metrics were (B, C, D) with B=3.93, C=3.81 and D=4.01. Moreover, the differentials between the profiles in the scores for these metrics are B= 0.13, C= 0.11 and D= 0.11. In addition, the Other Apps profile showed that the highest metric score match is E=3.6, which is the best matches the profile of software developers, where is E=3.69 and the differential is 0.09, although this metric is not a priority software developers requirements. On the other hand, the profile for Other Apps for metric A=3.9, which is low. This which matches the profile of software developers where A=3.39, which is the best match for the low score and the differential is A=0.51. By looking at Figure 5.16, it can be seen that the profile of Other Apps is the best match for the profile of software developers (Aljaber & Gordon, 2017b).

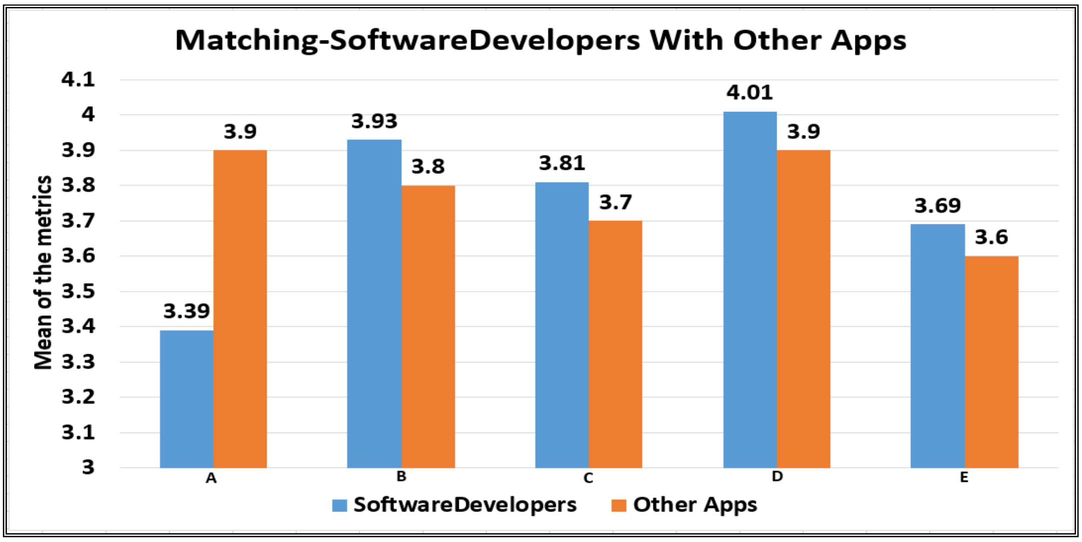


Figure 5.16: Matching Software Developers Profile with Other Apps Profile

It can be concluded that the selected hybrid metrics demonstrate how different MHEAs, would match the needs of different stakeholders. It showed that our measurements (A, B, C, D, E) produce some scores for metrics and that these suit these types of users, and match their preferences. The outcomes of this Chapter demonstrate that our selected hybrid metrics help in identifying that different users have different needs, and these metrics can help in designing tools, to match between what different stakeholders need and what is available. This constitutes our contribution, of “selecting a relevant collection of metrics developing the framework and applying them". Further information about MAST and designing it will be given in Chapter Six.

From the above analysis, we can refer to three categories of stakeholders (HP, P, SD) and match them with three categories of MHEAs (Medscape, Epocrates and Other Apps) as follows:

1. Medscape is the most suitable MHEAs for (HP) Health Professionals.
2. Epocrates is the most suitable MHEAs for (P/GP) Patient /General Public.
3. Other Apps are the most suitable MHEAs for (SD) Software Developers.

In addition, we could refer to another point, which has helped us to design and built the MAST for matching between different stakeholder's requirements and different mobile health education app's suitability, which can be summarised in these three points:

1. Medscape is the most suitable for stakeholders who are looking for (D) Simplicity.
2. Epocrates is the most suitable for stakeholders who are looking for (B) Features.
3. Other Apps are the most suitable for stakeholders who are looking for (A) Memorability and (D) Simplicity.

5.8 The Methods Used to Measure and Analyse the Questionnaire Data.

In this research, we converted the questionnaire feedback (data) from Google forms to a data table in SPSS, in order to analyse the feedback (data) from the questionnaire and apply several tests to obtain accurate results. After having the data ready in the SPSS table, the researcher used a Two-Way ANOVA test to analyse the respond to the questionnaire questions. This was appropriate because the data have two different factors: a group of three categories of mobile health education apps (Medscape, Epocrates and Other Apps), and a group of three categories of stakeholders (HP, SD and P/GP). Also as SPSS version 24 does not support the preparation of clear graphs. The results were converted for use in Microsoft Excel 2016. This enabled the same results obtained from SPSS to be reflected in clear charts. In addition, the mean for questions dealing with the same metric was applied, assuming the weight is equal for these questions.

5.9 Hypothesis testing.

The researcher applied the Kruskal-Wallis Test using SPSS, to test for a significant difference in needs between the three categories of the stakeholders (Health Professionals, Patients/General Public and Software Developers) regarding their needs from MHEAs, and views of their usefulness. The Kruskal-Wallis Test was appropriate since the data were not normally distributed.

The Kruskal-Wallis test is suitable for use for the following circumstances:

1. When there are three or more conditions to be compared.

2. When each condition is executed by a different group of participants; i.e., in an independent-measures design with three or more conditions.

3. If the data do not match the requirements for a parametric test. (i.e., it can be used if the data are not normally distributed; if the variances for the different conditions are significantly different; or if the data are measurements on an ordinal scale) (Graham Hole Research Skills, 2011).

The test is calculated according to the formula:

$$H = \left[\frac{12}{N(N+1)} * \sum \frac{Tc^2}{n_c} \right] - 3 * (N+1)$$

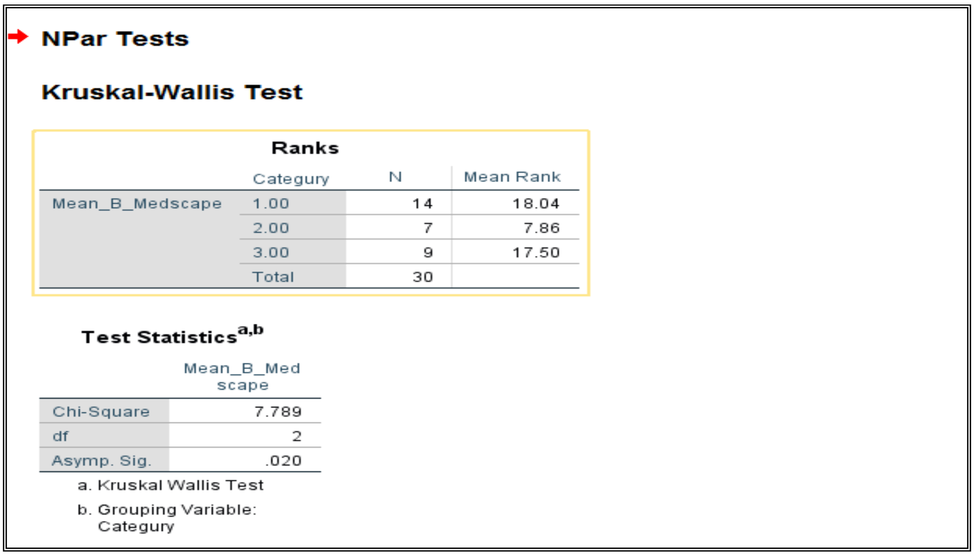
Where H = Kruskal-Wallis Test statistic

N = Total number of observations in all samples (We have 81 participants (3 groups of 27).

Tc = Sum of the ranks for each group.

nc = Total number of each group (we have 27) (Statistics Solutions, 2017).

The outcome revealed significant differences in the group's requirements regarding the selected hybrid selected metrics. Specifically three significant differences were found. One was for (B) Features, within the Medscape MHEAs for 30 participants including all the different types' of stakeholders. The outcomes showed that the p-value within all categories of stakeholders was 0.02, which is less than the criterion of 0.05, proving that there is a significant difference in the requirements between different stakeholders. Figure 5.17 reflects this result.



NPar Tests

Kruskal-Wallis Test

Ranks

	Categury	N	Mean Rank
Mean_B_Medscape	1.00	14	18.04
	2.00	7	7.86
	3.00	9	17.50
	Total	30	

Test Statistics[a,b]

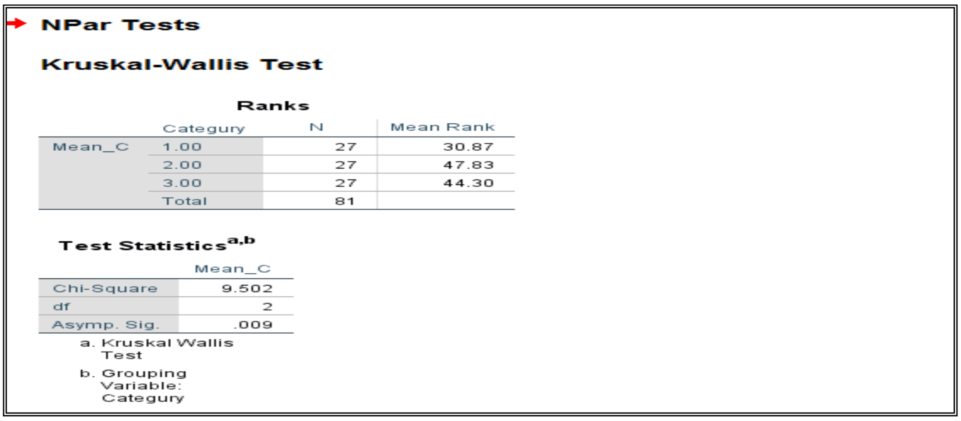
	Mean_B_Medscape
Chi-Square	7.789
df	2
Asymp. Sig.	.020

a. Kruskal Wallis Test

b. Grouping Variable: Categury

Figure 5.17: The P-Value for Medscape Apps for Metric B by HP, SD and P/GP

The second one is C) Attractiveness, within all of the MHEAs for all 81 participants. The outcomes showed that, the p-value obtained was 0.009, which is less than 0.05, proving that there is a significant difference in the requirements between different stakeholders. Figure 5.18 reflects this result.



NPar Tests

Kruskal-Wallis Test

Ranks

	Categury	N	Mean Rank
Mean_C	1.00	27	30.87
	2.00	27	47.83
	3.00	27	44.30
	Total	81	

Test Statistics[a,b]

	Mean_C
Chi-Square	9.502
df	2
Asymp. Sig.	.009

a. Kruskal Wallis Test

b. Grouping Variable: Categury

Figure 5.18: The P-Value for all Apps for C Metrics by HP, SD and P/GP

The third instance is for D) Simplicity, regarding the Medscape MHEAs, for 30 participants from all stakeholder groups. The outcomes showed that, the p-value obtained was 0.018, which is less than 0.05, proving that there is a significant difference in the requirements between different stakeholders. Figure 5.19 reflects this result.

NPar Tests

Kruskal-Wallis Test

Ranks

	Category	N	Mean Rank
Mean_D_Medscape	1.00	14	20.07
	2.00	7	10.86
	3.00	9	12.00
	Total	30	

Test Statistics[a,b]

	Mean_D_Medscape
Chi-Square	7.991
df	2
Asymp. Sig.	.018

a. Kruskal Wallis Test

b. Grouping Variable: Category

Figure 5.19: The P-Value for all Apps for D Metrics by HP, SD and P/GP

From all the above as the p-Value is less than 0.05, this lead to acceptance of our null hypothesis, that there is a significant difference between the requirements of different stakeholders for different MHEAs.

5.10 Conclusion

This Chapter described how the number of questionnaire questions was decided depending on the weighting obtained from the qualitative interviews explored in Chapter Four. In addition, it illuminated for different overall rankings of different MHEAs (Medscape, Epocrates and Other Apps) by different stakeholders (HP, P/GP, SD). Furthermore, the Chapter reported the use of the Cronbach's alpha test, to validate the data before further analysis. Data were valid to be utilized and analysis. Based on using Two-Way ANOVA, it reported the statistical analysis needed, in order to address objectives 2, 3 and 5. These were: developing a ranking measurement system to assess and rank MHEAs, identify issues in MHEAs identified by the rankings, to improve the design of MHEAs, and characterize the profile of different stakeholder group in order to identify their different requirements. Also the Kruskal-Wallis Test was used to test the hypothesis, which was supported since the p-value was 0.018, which is less than the 0.05, criterion of significance. It reflected the differences in ranking of the five metrics in the same and different MHEAs. As well, it reflected the differences among stakeholders in their views on the five metrics and their requirements for them. The Chapter showed the matching between the profiles of MHEAs (Medscape, Epocrates and Other Apps) with the profiles of different stakeholders (HP, P/GP, SD), and reported differential for the same metric requirements between different stakeholders for MHEAs. Finally, the Chapter explored the suitability of each MHEAs (Medscape, Epocrates and Other Apps) to different stakeholders, according to the metrics.

Chapter Six, will explore the MAST parts for MHEAs. It will describe the design of the MAST; explain how MAST works besides and investigate how the MAST helps different stakeholders to select their most suitable MHEAs.

Chapter 6: Developing Medical Apps Selection Tool (MAST)

The previous Chapter discussed how the selected hybrid metrics were converted into questionnaire questions. Additionally, it clarified how the questionnaire feedback from all types of stakeholders (HP, P/GP, SD) was analysed, and explored the outcomes from the analysis to measure and rank MHEAs and find the individual metrics ranking for MHEAs and each stakeholder group. Also, matched between each MHEAs profile (Medscape, Epocrates and Other apps) and each stakeholder group's profile (HP, P/GP, SD). This chapter explore the fulfilment of objective 4, which was achieved by developing a MAST to select suitable MHEAs for the different Stakeholder types. It discusses how the results from analysing the questionnaire were applied when designing the MAST. Moreover, the chapter clarifies the steps involved while designing the tool. It illustrates the details of this tool how it works, what it offers, the purposes of using it, and the contribution made by building it. User examples and some screenshots of this tool are presented by way of illustration.

6.1 Using the Questionnaire Outcomes in MAST

Chapter Five described the outcomes of the questionnaire. It explored how the data were analysed in SPSS software. Moreover, it discussed how to assist different stakeholders in matching their requirements to suitable MHEAs. This section will describe the method of utilizing the results from the questionnaire for designing the MAST.

Many algorithms and metrics exist which could be used for performing the matching procedure in MAST. However, in this study, the Euclidean Distance metric was selected for a number of reasons. The Euclidean Distance originates from the Pythagorean Theorem, which describes the physical world. In addition, it has been known since antiquity. Also, it reflects a natural notion of distance from a historical perspective. Moreover, Euclidean Distance is the only metric that is the same in all directions, that is, rotation invariant. This fits well with the general qualities of our universe, which is also rotation invariant.

The consideration of the Euclidean Distance metric for the tool was for five dimensions as the research has five metrics, which are A) Memorability, B) Features, C) Attractiveness, D) Simplicity, E) Accuracy. Therefore, the formula used for Euclidean Distance is the (n dimensions formula):

$$d = \sqrt{\sum_{i=1}^{p} (v_{1i} - v_{2i})^2}$$

(pbarrett.net, 2005)

Where p is the number of the dimensions (attributes) and V1i and V2i are respectively (i) the attributes (components) or data objects V1 and V2.

As an example of matching between the user profile and MHEAs profile. Table 6.1 presents a sample user profile. The user selects from MAST the his/her priorities for the user profile which are five selected metrics {X, Y, Z, V, W} as presented in Table 6.1. These are scored X=3.6, Y=3.7, Z=3.5, V=3.8, W=3.6. MAST will bring all the MHEAs profile data from the Array as shown in Table 6.2, and compare each metric from the user profile with the corresponding metric from each MHEAs profile, in order to find the best matching profile based on the metrics of the Euclidean Distance, which provides the shortest distance between each metric.

User Profile	**X ↓ A**	**Y ↓ B**	**Z ↓ C**	**V ↓ D**	**W ↓ E**
Patient	**3.6**	**3.7**	**3.5**	**3.8**	**3.6**

Table 6.1: Example for the User Profile Data Generated From Survey in MAST

Table 6.2 presents the means for each of the metrics for all MHEAs packages, obtained from the data collected from 81 participants, as discussed in detail in Chapter Five.

MHEAs profile /Metrics	A	B	C	D	E
Medscape	3.7	3.8	3.6	3.9	3.7
Epocrates	3.3	3.9	3.6	3.7	3.3
Others	3.9	3.8	3.7	3.9	3.6

Table 6.2: MHEAs Profile Generated from Questionnaire Outcomes Array

6.1.1 Verifying the Appropriateness of using Euclidean Distance in MAST

This section explains the method of verifying the suitability of the Euclidean Distance in the MAST, in order to utilize the outcomes from the questionnaire and carry out the mathematic calculations in MAST.

These mathematical calculations for the Euclidean Distance were utilised, based on finding the shortest distance between the user's profile and MHEAs profile, in order to identify the most suitable match between the user profile and the MHEAs profile. This was utilized in the coding part of the MAST in C#.

This calculation for the outcome from the questionnaire, by using the Euclidean Distance metric, was accomplished using SPSS software. Figure 6.1 shows the calculation for the Euclidean Distance between HP profile and Medscape, Epocrates and Other Apps profiles. It can be seen that the Euclidean Distance calculation yields three results, which are firstly 0.293 between the HP profile and Medscape profile; secondly, 0.675 between the HP profile and Epocrates profile; thirdly, 0.495 between the HP profile and Other apps profile. This shows that the profile and most closely matching the HP profile is the Medscape profile, since it shows the shortest Euclidean Distance result, which is 0.293. This illuminates that the profile matching between the HP profile and Medscape profile is the most suitable match.

Case Processing Summary

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
5	100.0%	0	0.0%	5	100.0%

Proximity Matrix

	Euclidean Distance			
	HP	Medscape	Epocrates	Others
HP	.000	.293	.675	.495
Medscape	.293	.000	.549	.320
Epocrates	.675	.549	.000	.713
Others	.495	.320	.713	.000

This is a dissimilarity matrix

Figure 6.1: The Euclidean Distance between HP profile and MHEAs profile using SPSS

6.2 Designing the Skeleton of MAST

This describes the method of designing MAST, with the purpose of matching between different stakeholder's profiles and different MHEAs profiles. Chapter Five explained how the data was obtained, reflecting different MHEAs profiles and different rankings for these apps. This data was utilised to build the array and the prototype in MAST, as presented in Figure 6.2. For this purpose, the questionnaire data was placed in the array of the MAST. The user of MAST would insert his\her data to reflect his\her profile when using the MAST. Then the MAST will make a decision depending on the user profile, in order to provide a suitable MHEAs matching profiles, besides suggestions for other software for the user as presented in Figure 6.2.

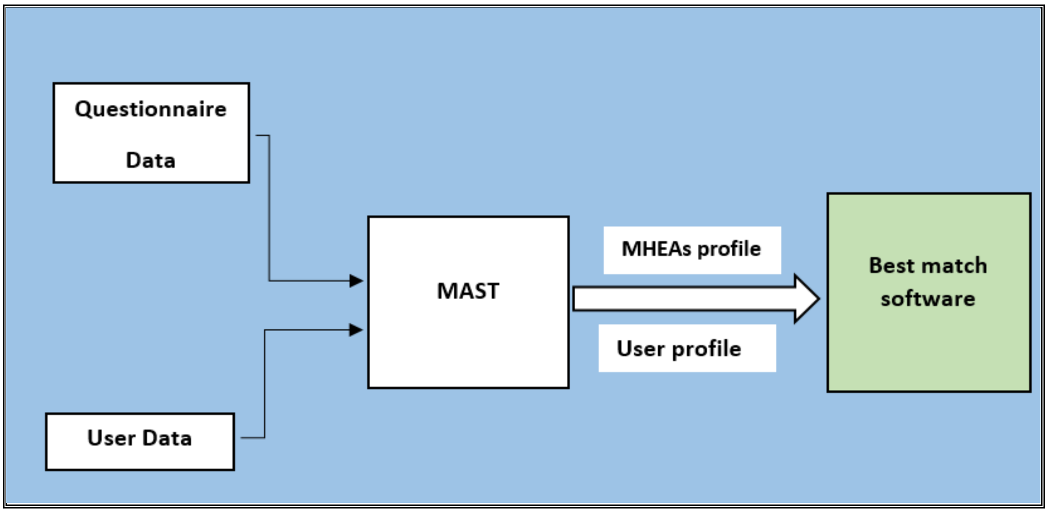


Figure 6.2: The Flowchart Design for MAST

6.3 The Work Flow for MAST

This section illuminates the flow of MAST, in order to explain the MAST steps, which are intended to meet the needs of the of MAST user. The flow of MAST is presented in Figure 6.3.

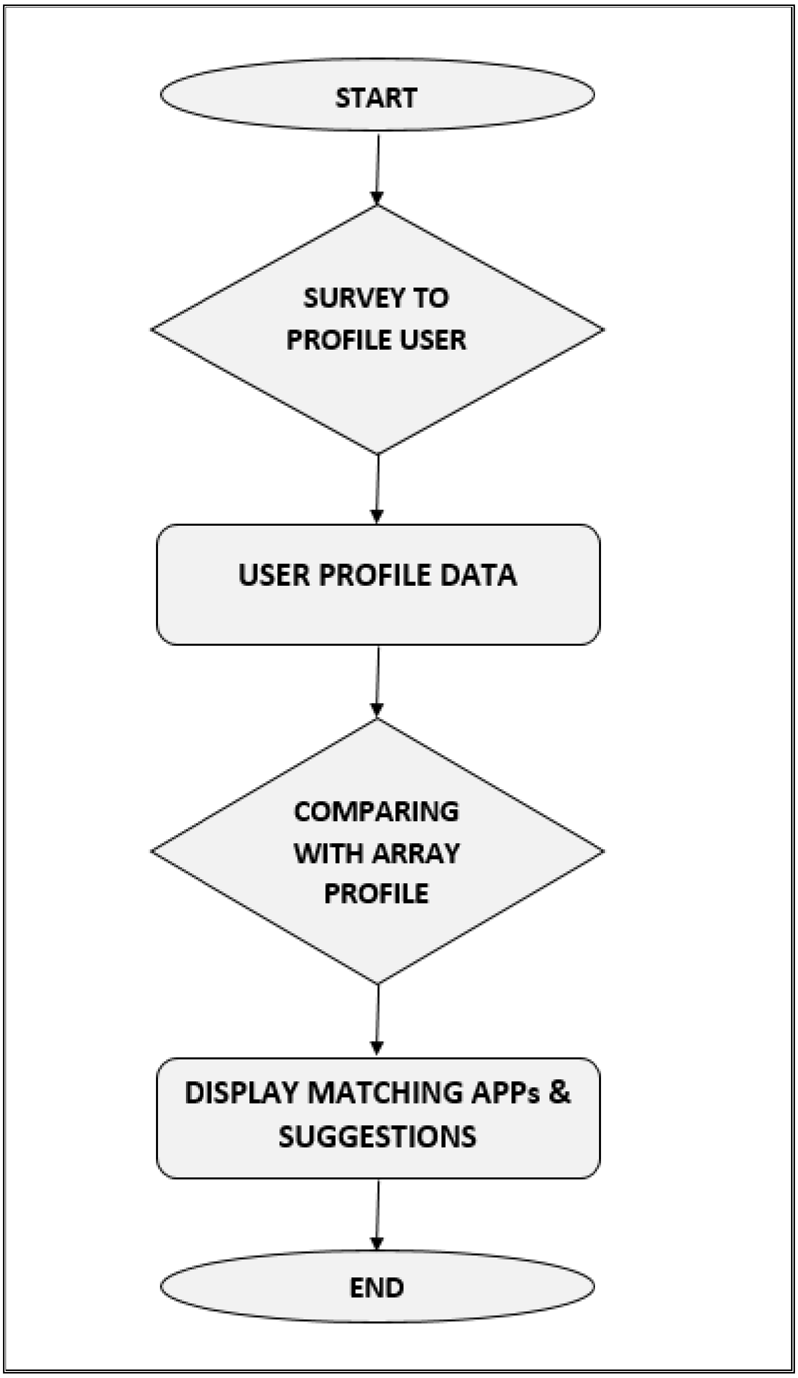


Figure 6.3: Flowchart of MAST Process

The user will be involved in answering the following survey, in order to select one of the three types of user profiles:

The questions for Survey user to profile users:

Are you a Software Developer?

Are you doing this as a part of your job?

Are you interested because you are patient?

This will identify what type of user is using the MAST. Then the user will be involved in providing his/her requirements from MHEAs, by answering the survey, to classify their profile metrics. This survey below consists of five questions, which reflect the five metrics (A. Memorability, B. Features, C. Attractiveness, D. Simplicity, and E Accuracy). The user will assign a score from 1.00 to 5.00 for each metric, in order to create their profile.

X→A: How interested you are in finding the Apps easy to remember how to use it?
Y→B: How interested you are in finding the Apps providing the features to use it?
Z→C: How interested you are in finding the Apps contain attractiveness to use it?
V→D: How interested you are in finding the Apps simplifying the way to use it?
W→E: How interested you are in finding the Apps responding accurately to use it?

Once the user profile is determined, MAST will connect to the Array, in order to call up the MHEAs profiles and execute the comparison procedure, between the user profile and MHEAs profiles, by using the Euclidean Distance metric. The final step is to provide the most suitable/ recommended MHEAs package, besides suggestions for the user for the other MHEAs packages options, depending on the Euclidean Distance between the user profile and MHEAs profile.

6.4 Addressing the Software Coding for MAST

This section describes the procedure of addressing the software coding for the MAST. The researcher decided to use C# programming language, as the language for developing the coding for MAST. The MAST could have been developed in any other programming language. However, the reason for utilising C# is that C# is one of the most familiar languages adopted for developing apps (Cass, 2014).

Coding in C# depended on using arrays, rather than using a database. The reason for using arrays is that the data outcomes from the quantitative results are not large. So there is no need to use a database at the current stage of the system. However, in the future, the system will require the use of a database to store the data. Moreover, the researcher defined two types of arrays: a one-dimension array to store the user profile data, which is the size (1 x 5), and a two-dimensional array to store the MHEAs profile data, which is the size (3 x 5). Additionally, the Euclidean Distance metric has been used, in order to compare between the user profile array and each row of the MHEAs array, in order to measure the Euclidean Distance between the user profile and each MHEAs.

6.5 Description of MAST

This section indicates the specification of the MAST and how the MAST works. It also provides some examples of the users for the MAST, including what kind of users they are, what their main requirements are in using MHEAs, what type of match they request, and what is the best MHEAs profile matching their requirements and their user profile, according to the MAST decision.

The MAST (prototype) was designed as an application running on a desktop computer, not on a mobile. Future work could include porting MAST to another operating system and platform, such as Android or iOS. Figure 6.4 presents the user interface for the MAST. It consists of four parts. The first part presents the user type. The second part is the user data (user requirements), which is constitute the user profile. The third part reflects the matching options, which as follows:

1) **Matching with the Group Profile**. This involve the MAST in using the Euclidean Distance metric, in order to measure, the Euclidean Distance, between the user profile and existing user profiles in the Array, according to the five metrics, described in Chapter Five.
2) **Matching with MHEAS Profile**. This involves the MAST in using the Euclidean Distance metric, in order to measure the Euclidean Distance, between the user profile and existing MHEAs profiles in the Array, for each of the five metrics described in Chapter Five.
3) **Matching with the Subset Profile**. This involves the MAST in using the Euclidean Distance metric, in order to measure, the Euclidean Distance, between the user profile and existing MHEAs profiles in the Array, for the three best matching metrics described in Chapter Five.

The fourth part presents the best matching software, besides some other recommended software, according to the MAST decision.

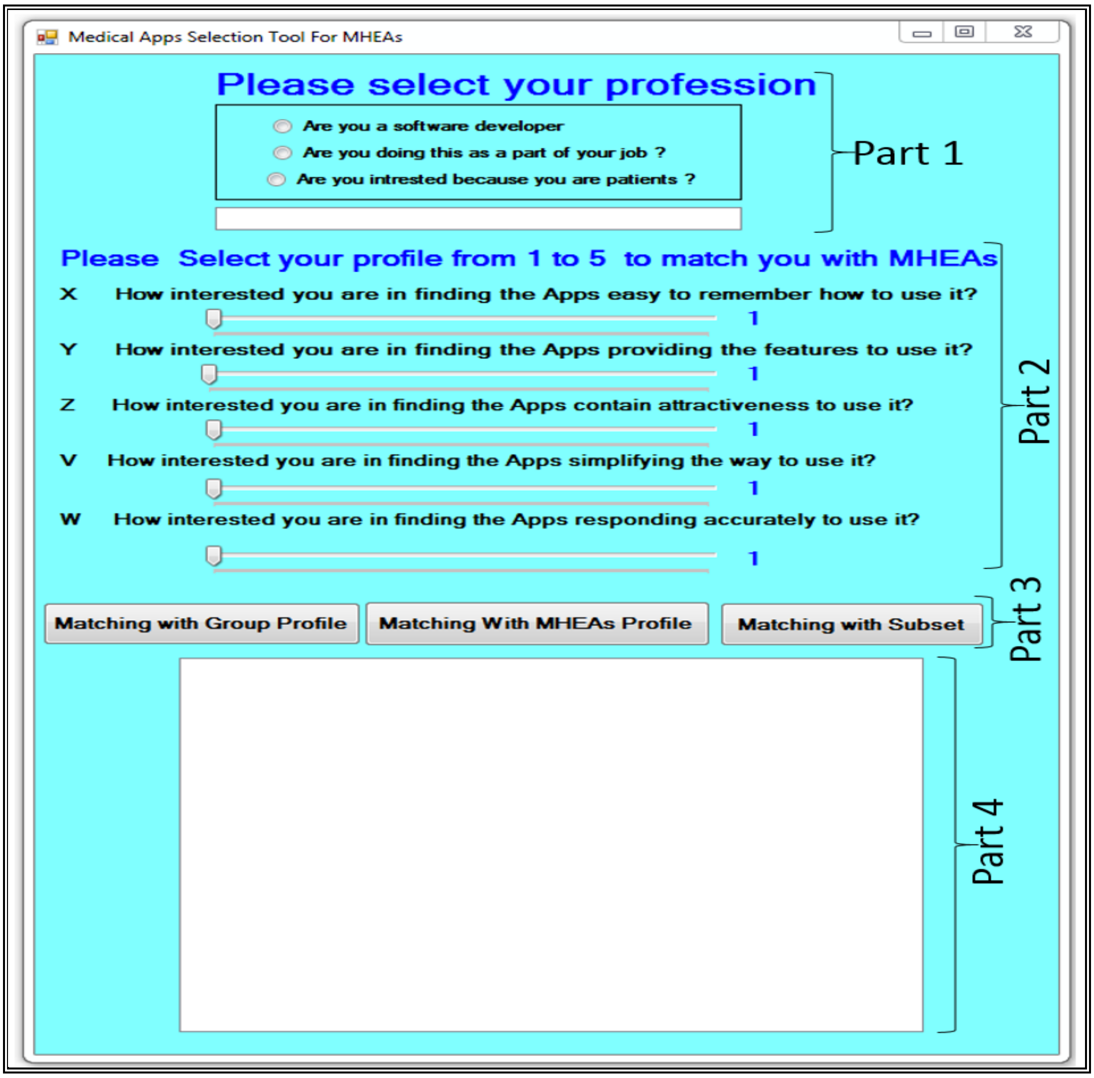


Figure 6.4: The Medical Apps Selection Tool (MAST) for MHEAs

The first user type example for MAST is HP, which is shown in Figure 6.5. The user selected the second option in the user type part, which is:

"Are you doing this as a part of your job?"

This respond identifies the user type of the MAST, presenting a Health Professional user type.

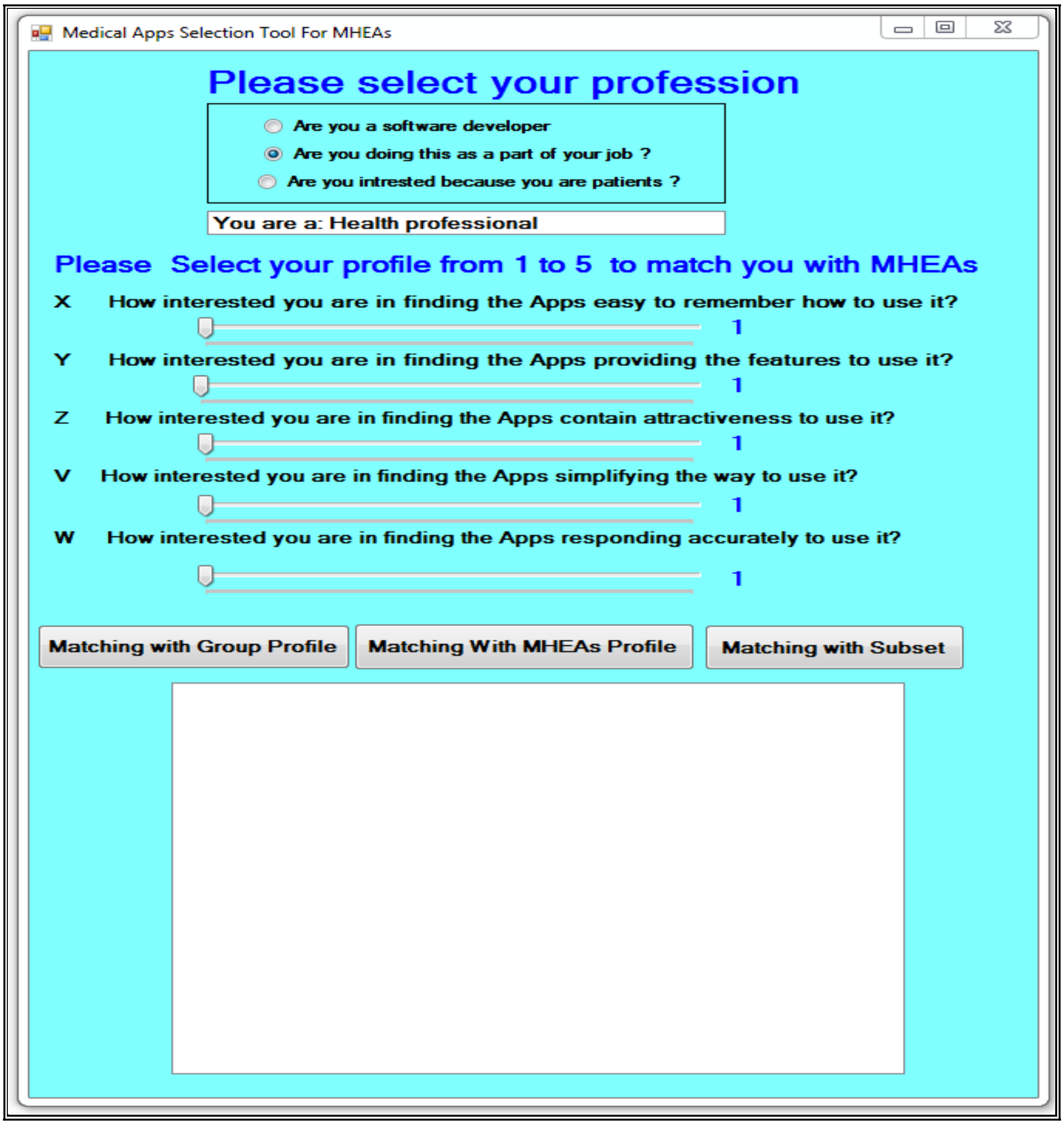


Figure 6.5: The First Type of Stakeholders Health Professional

Once the user type has been determined, the user will be require to move to the second part of the MAST, which is to create the user profile. The user responds by scrolling left to right from 1.00 to 5.00, in the second part of the MAST, in order to indicate his/her view of five selected hybrid metrics for the user profile. This will create the user profile, which presented in Figure 6.6.

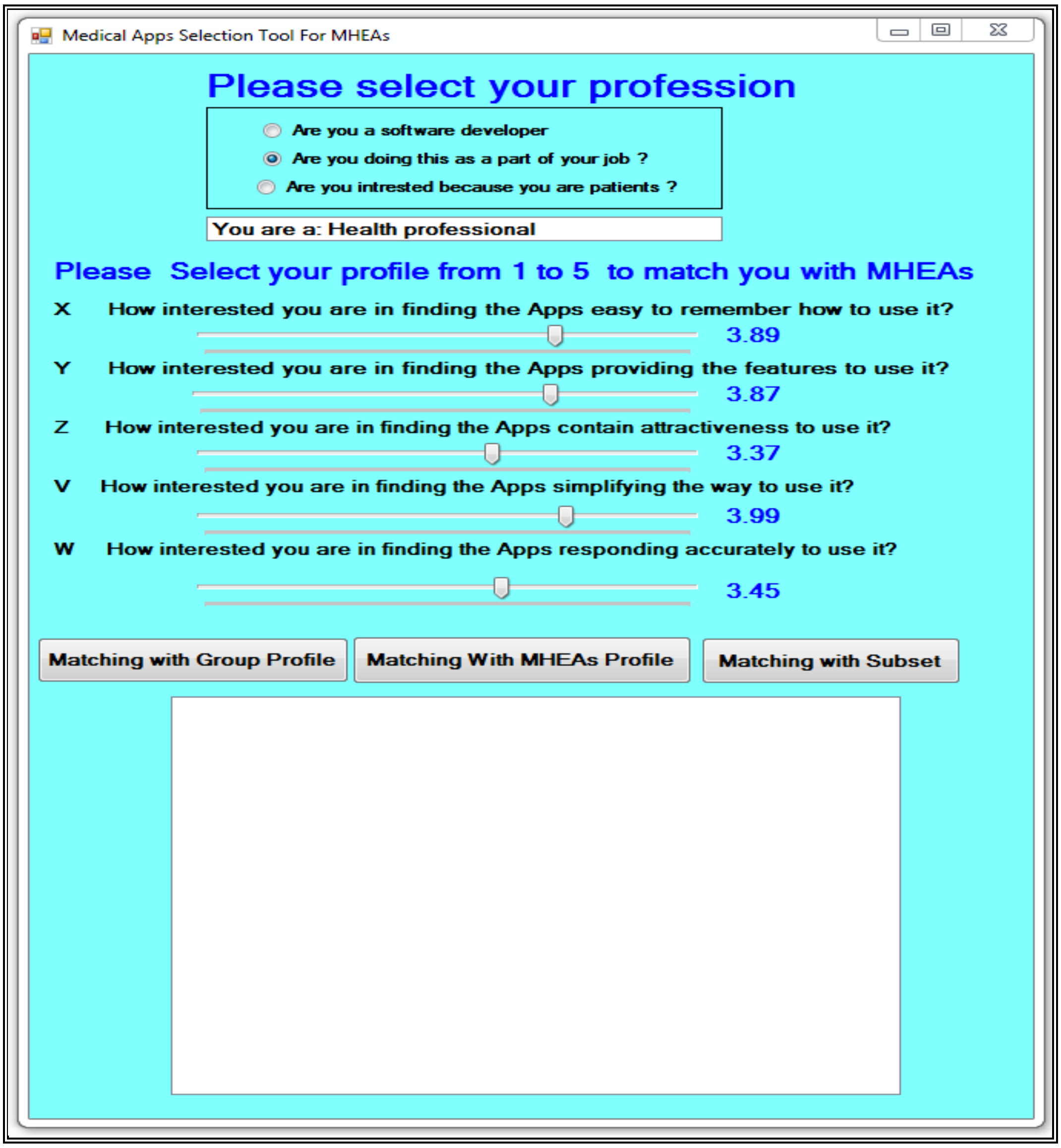


Figure 6.6: The Health Professional Profile

The final step for the user is to select one of the three matching options offered by MAST. In our example, the user selected the second option, which is matching with MHEAs profile. When the user clicks on the second matching button, The MAST will carry out the comparison procedure, in order to match between the selected user profiles and with the MHEAs profile stored in the Array, by using the Euclidean Distance metric. This comparison is will identify the best matching software, and make other recommendation software to the user, which are displayed in section four, of the tool parts. In the example gives, it shows that the best match for the user type profile of a health professional is Medscape. In addition, the tool presents some choices for the recommended software matching, namely, option 1 Medscape, option 2 Other Apps (WebMD, UpToDate) and option 3 Epocrates, depending on the closest result for the Euclidean Distance. The numbers displayed in MAST, next to the name of the mobile health apps in the Matching

Option section, display the full value of the Euclidean Distance. This is primarily for the software developers, whilst testing and evaluating the MAST Tool, and would be simplified (e.g. to a 0 to 10 scale) in a release version. The outcomes are presented in Figure 6.7.

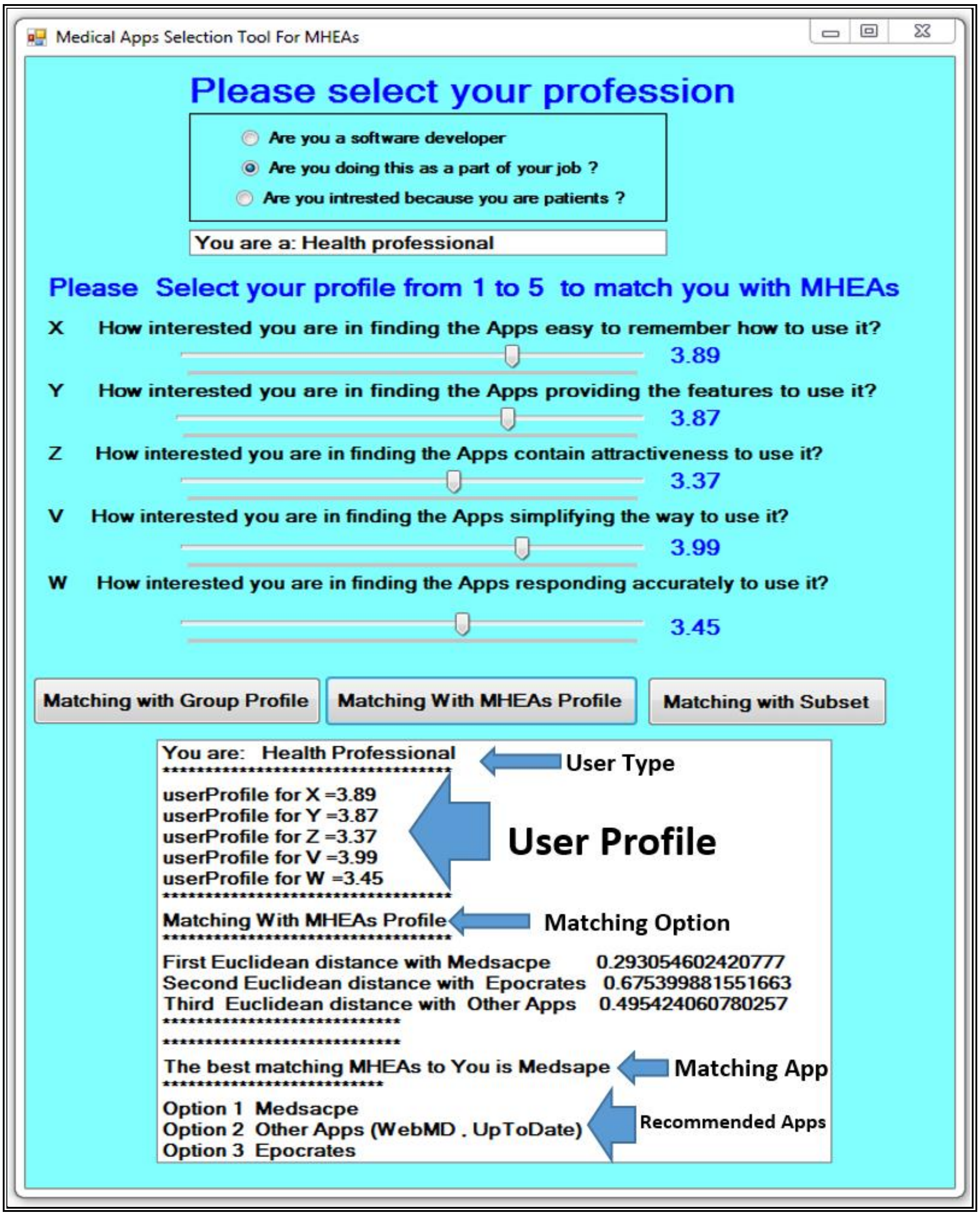


Figure 6.7: Matching for HP Profile with MHEAs Profile

The next example of using MAST is similar to the first example in that the user type is HP. However, the user choice was the first option button, which is matching with the group profile. This aims to match between the user profile and profile of other user of the same type. This reflects that the best match for the user type of HP is Medscape, and also presents some options for the recommended software matching. These are option 1 Medscape, option 2 Epocrates and option 3 Other Apps (WebMD, UpToDate). These outcomes are presented in Figure 6.8.

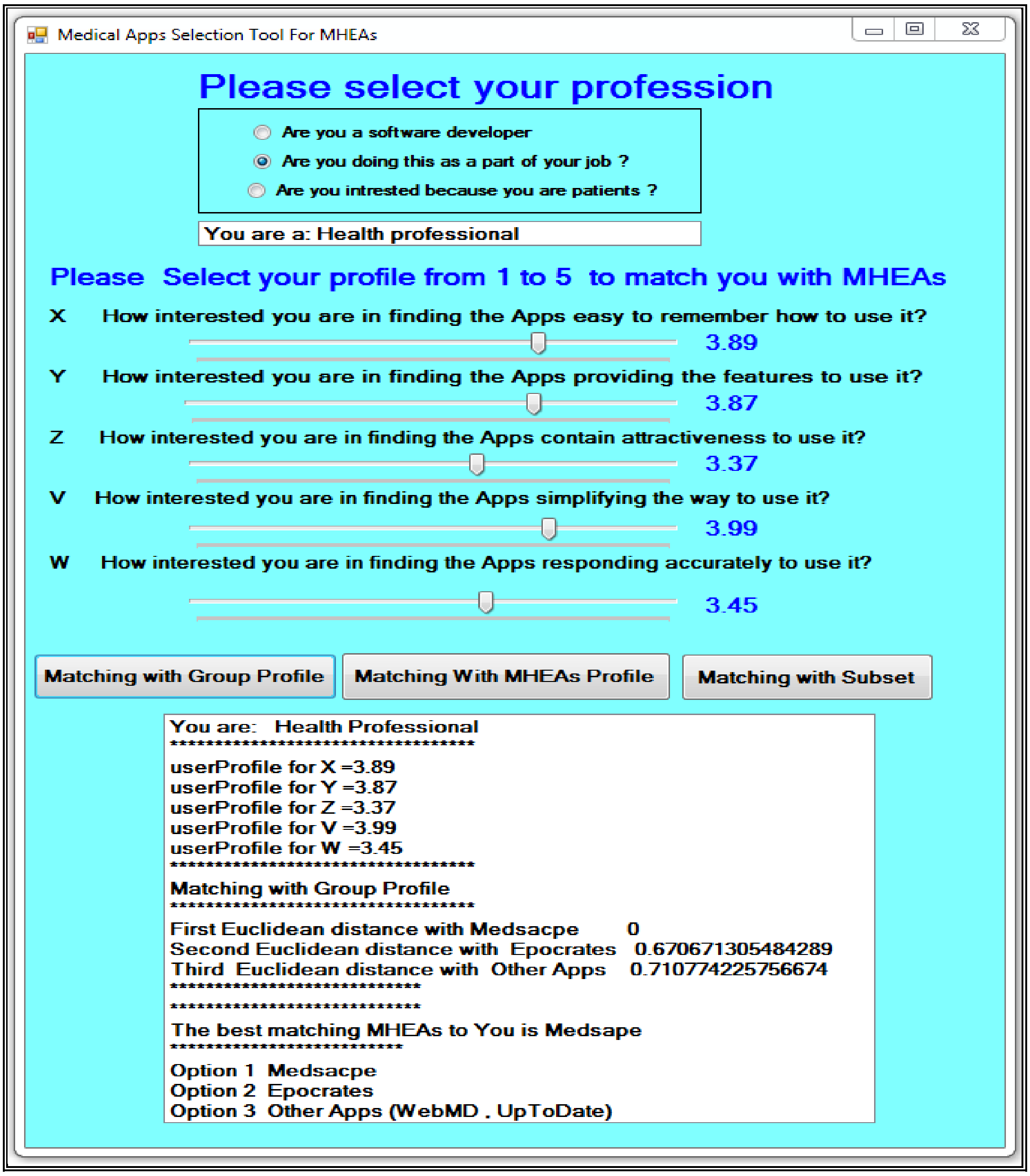


Figure 6.8: Matching for HP Profile with HP Array Profile

In the next example of using MAST, similar to the previous example, the user type is HP. However, the user chooses the third option button, which is matching with the Subset profile. This aims to match between the user profile and the closest three metrics from the MHEAs profile. This reflects that the best match for the user type profile of HP is Other Apps (WebMD, UpToDate). It also presents some options for the recommended software matching, which are option 1, Other Apps (WebMD, UpToDate), option 2, Medscape and option 3, Epocrates. These outcomes are presented in Figure 6.9.

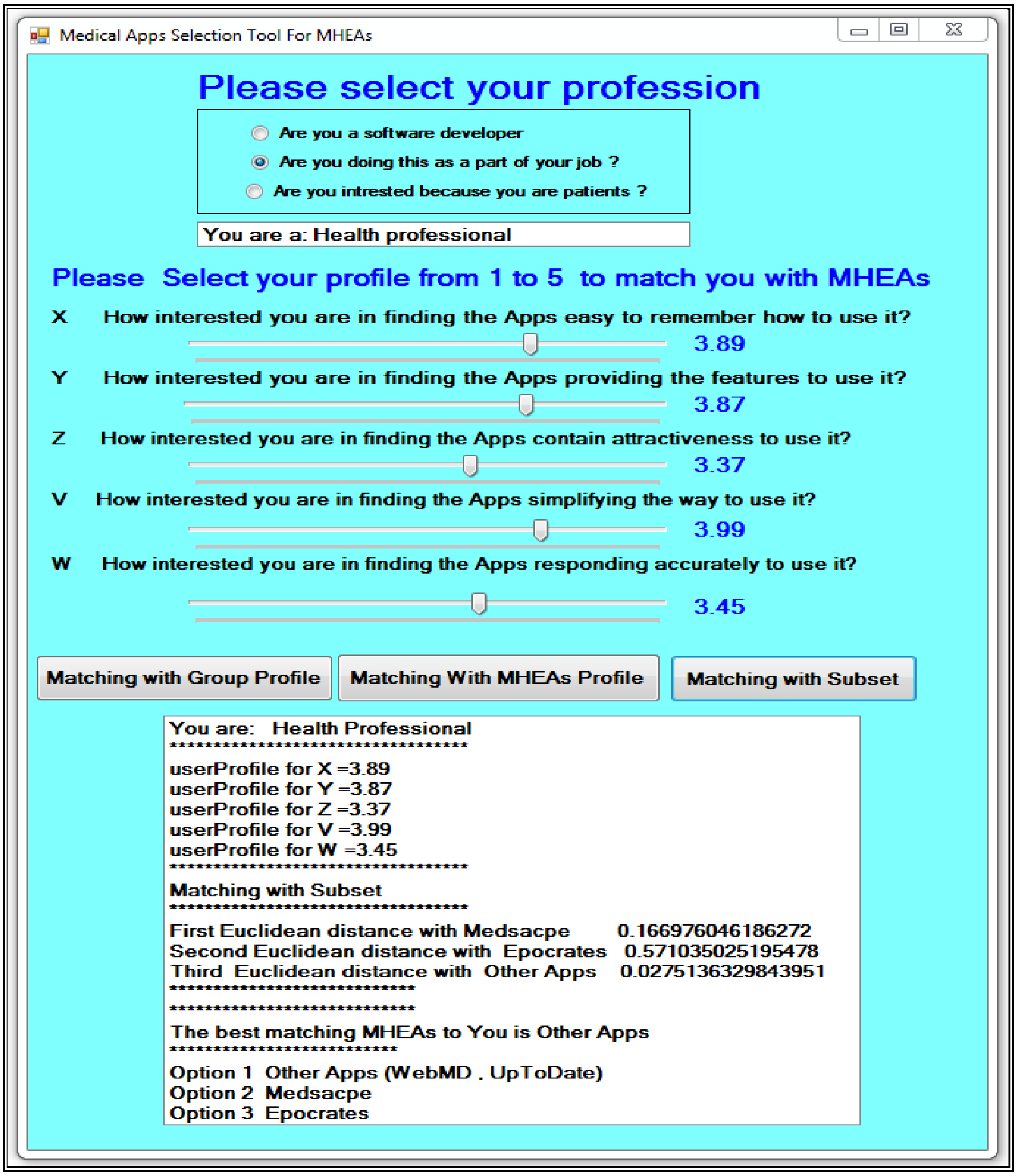


Figure 6.9: Matching for HP Profile with MHEAs Subset Array Profile

In the next example of using MAST, the user type is a patient. In this example, the user chooses the first option button, which is matching with the group profile. This aims to match between the user profile and the profile of the same type of user. This reflects that the best match for the user type of patient is Epocrates, besides presenting some options for the recommended software matching. These are option 1 Epocrates, option 2 Other

Apps (WebMD, UpToDate) and option 3 Medscape. These outcomes are presented in Figure 6.10.

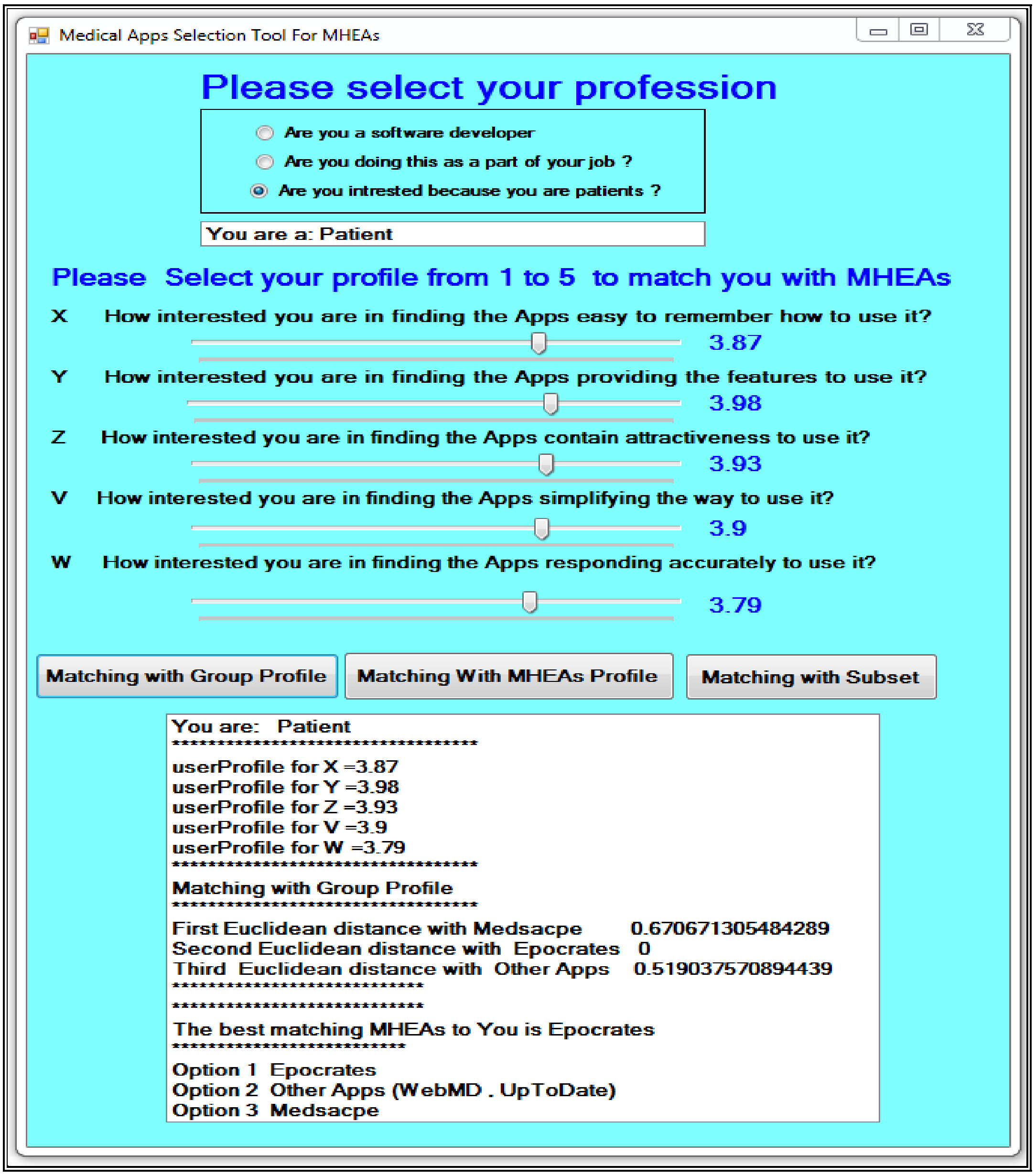


Figure 6.10: Matching for P Profile with P Array Profile

Another example of using MAST is for the user type of a software developer. In this example the user chooses, the first option button, which is matching with the group profile. This aims to match between the user profile and that of the same type of user. This reflects that the best match for the user type of SD is Other Apps (WebMD, UpToDate). In addition, some options for the recommended software matching are presented. These are option 1 Other Apps (WebMD, UpToDate), option 2 Epocrates and option 3 Medscape. These outcomes are presented in Figure 6.11.

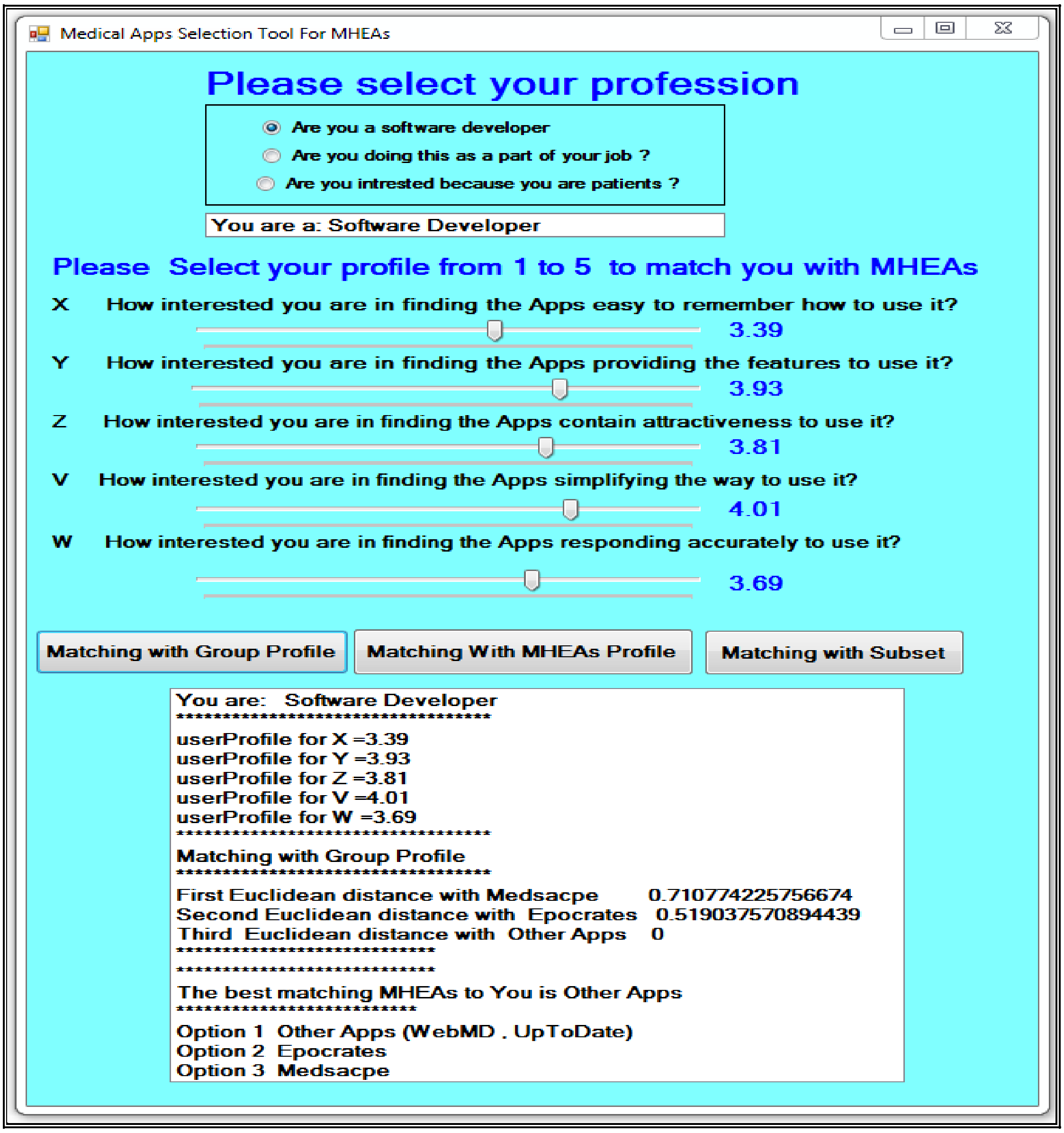


Figure 6.11: Matching for SD Profile and SD Array Profile

6.6 Testing the MAST

This section explains more about testing the MAST, including characteristics of the test, the tasks designed for the test and the rationale for selecting these tasks.

MAST is the third part of the evaluation framework. It was built to convert the outcomes from the first two parts of the evaluation framework, which were gathering the data and analysing it, into a software tool. MAST is a software tool (prototype) to match between different stakeholder profiles and different MHEAs profiles. The prototype provides a 'simple' interface with four parts. The first part is to identify the type of user of the prototype (software developers, health professionals, Patients). The second part is to identify the user profile (what the user requires from using MHEAs), by prioritizing their requirements for MHEAs, depending on the five selected hybrid metrics. The third part is offer for the user to request one of three types of matching option suggested by the prototype. The last part is the outcome of the matching, by the prototype with one of the MHEAs, and provision of some other recommended MHEAs. The MAST uses Euclidean Distance, to show the precise matching, and provide some other recommended MHEAs. The MAST was tested with nine participants, from three categories, three participants from each category (health professionals, software developers, patients).

Before starting the test, participants were introduced to the prototype (MAST). Particularly for novice participants, who were not very familiar with the tool applications, a short demonstration session was conducted, in order to provide an overall picture of the prototype and to ensure the participants had a basic familiarity with the prototype before continuing with the actual tasks.

For the actual experiment, the tasks chosen represent the whole prototype, and the main parts of the user interface (Nielsen, 1994). Five tasks were designed for participants to complete during testing for MAST. The tasks participants were asked to perform included, selecting their professions, changing the metrics unit of measurement, entering a matching option, rating how easy was the matching represented in part 4 and rating how easy it was to use the prototype overall. The tasks were designed to represent the overall features of the MAST. All nine participants were involved in doing this test.

On the aspect of determining how easy is it to use MAST, the outcomes of the test were positive, as all of the participants liked using the MAST. They said it is very simple to deal with, and very clear to use. In particular, the health professionals and patients expressed the new that it is very useful to have this type of tool, as it could save them a lot of time, besides helping them to find the most suitable MHEAs for their usage. (Nielsen, 1994)**.**

Moreover, further analysis for testing the MAST functionality was applied, as presented in Table 6.3; It can be seen that 9 out of 9 participants from the three categories (health professionals, software developers and patients) managed to perform Task_1 in MAST successfully form a functionality's aspect. Also, 9 out of 9 participants managed to perform Task_2 successfully. Furthermore, 8 out of 9 participants managed to perform Task_3 successfully, and only one participant was not successful in this task. Moreover, 9 out of 9 participants completed Task_4 successfully. Finally, 9 out of 9 participants completed Task_5 successfully.

Task	Task discription	Number-successful	Number-Unsuccessful
1	Selecting their professions	9	0
2	Changing the metrics unit of mesurment	9	0
3	Entering a matching option	8	1
4	Refereeing how easy the matching represented in part 4	9	0
5	Refereeing how easy is it to use the prototype over all	9	0

Table 6.3: Users Views from MAST Functionalities

Furthermore, the testing of the MAST functionality is presented in Figure 6.12, which summarises the result in Table 6.3. It shows that in 4 out of 5 columns in the chart, 9 out of 9 participants successfully performed Tasks 1, 2, 4, 5 within the MAST functionalities. Thus, only one column has the value 8 out of 9 participants who successfully performed the Task_3, within the MAST functionalities.

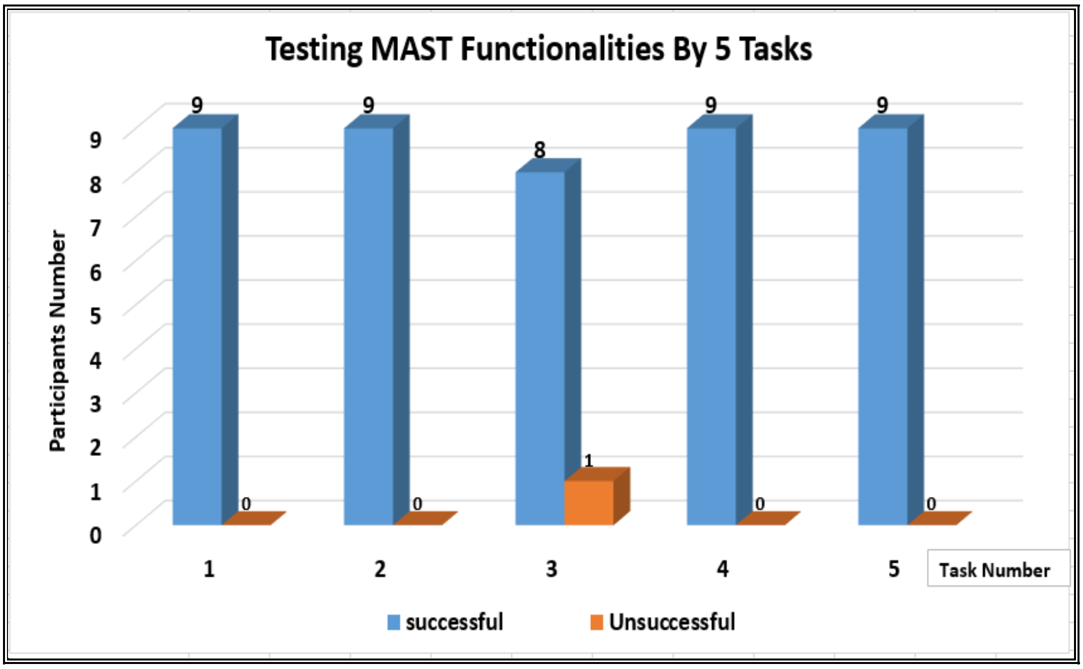


Figure 6.12: Users Views from MAST Functionalities

Further analysis for testing of the MAST output was applied and is presented in Table 6.4. This shows that 2 out of 3 professional participants determined that the output of the MAST fully matched their requirements, while one of the participants from this group decided that the output of the MAST partially matched his preferences. From the patient side, 3 out of 3 determined that the output of the MAST fully matched their requirements. Also, among the software developers, 3 out of 3 decided that the output of the MAST fully matched their needs. This in total, 8 out of 9 participants from the three categories were fully matched and only one partially matched. This means 89% of requirements were fully matched and only 11% partially matched.

Participant Type	Full Match	Partial Match	No Match
Professionals	2	1	0
Patients	3	0	0
Developer	3	0	0
Total	8	1	0
Percentage	89%	11%	0

Table 6.4: User's Views of the MAST Output with their Requirements

A further illustration of the testing of the MAST output is presented in Figure 6.13, which summarises the result in Table 6.4. It shows that the matching between the outputs from the MAST with the requirements from the three categories of participants was 2 professionals, 3 patients and 3 developers having their needs fully matched, and only one professional who determined that MAST partially matched his requirements. Finally, none of the participants from three categories of participants reported that the MAST did not match their needs.

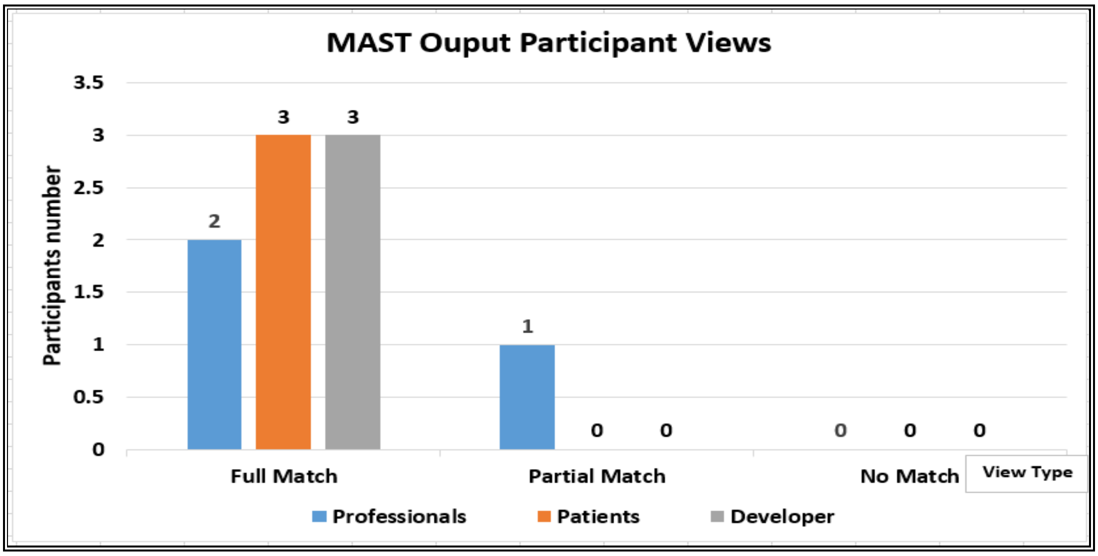


Figure 6.13: Users Views from MAST Output with their Requirements

For the outcome of the testing is further presented in Figure 6.14, which summarises the result in Table 6.4. As shown by the total and percentage, the matching between the outputs from the MAST with the requirements from three categories of participants was that 8 out of 9 participants had their needs fully matched with a percentage of 89%, while only 1 out of 9 had his needs partially matched, with a percentage of 11% and none out of 9 had the no match.

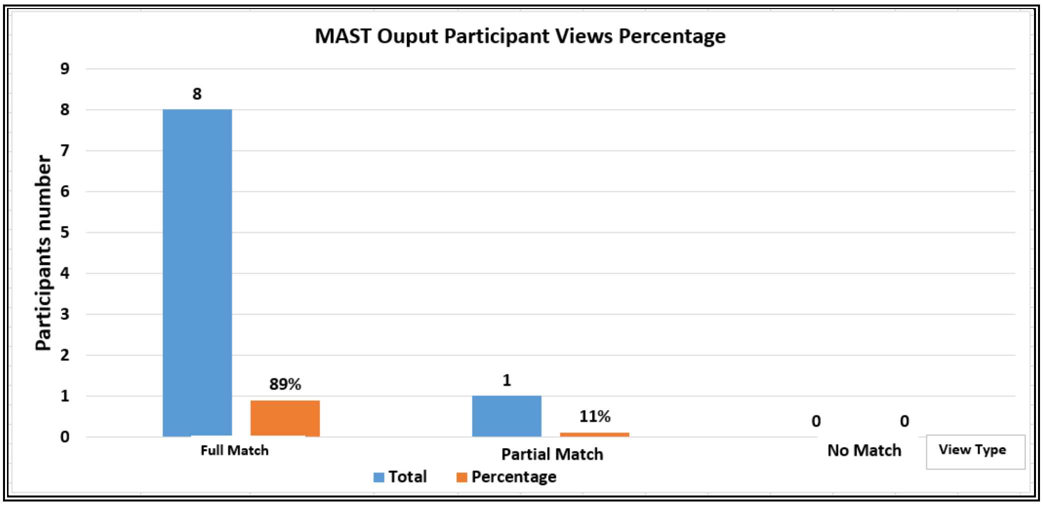


Figure 6.14: Users Views Percentage from MAST Output with their Requirements

Additionally, further analysis of the testing of MAST, in terms of MAST's output recommendation with all categories' matching preferences is presented in Table 6.5. This shows that for first professional, the matching of recommended apps, which came as in the first place Medscape, in second place Other Apps and third place Epocrates, matched the professional's preference. For the second professional the matching of recommended apps came as in first place Other Apps, in second place Medscape and in third place Epocrates, which partially was matched to the professional's preference. Since the professional expected to have in first place Other Apps, second Epocrates and third Medscape. The reason that this match came as a semi-match appears to be that the professional did not carefully select his/her profile from the 5 metrics. Finally, for the rest of the participants, the output recommendation was matched with their matching preference. These outcomes are presented in Table 6.5, which reeds from left to right with the first match, second match and third matching apps.

Users Type	Recommended Apps	Matching Preference	Reason For the Matching
Professionals	Medscape / Other apps / Epocrates	Match	Carefully select user profile from 5 metrics
Professionals	Other apps / Medscape / Epocrates	Semi Match	Not carefully select user profile from 5 metrics
Professionals	Medscape / Epocrates / Other apps	Match	Carefully select user profile from 5 metrics
Patients	Epocrates / Other apps / Medscape	Match	Carefully select user profile from 5 metrics
Patients	Epocrates / Other apps / Medscape	Match	Carefully select user profile from 5 metrics
Patients	Epocrates / Other apps / Medscape	Match	Carefully select user profile from 5 metrics
Developers	Other apps /Epocrates / Medscape	Match	Carefully select user profile from 5 metrics
Developers	Other apps /Epocrates / Medscape	Match	Carefully select user profile from 5 metrics
Developers	Other apps /Epocrates / Medscape	Match	Carefully select user profile from 5 metrics

Table 6.5: User's Views of MAST Output Recommendation with their Matching

6.7 Guideline for Use of the Framework and Guidance Tool

The evaluation framework and guidance tool is a dynamic framework where evaluators can adapt the framework in order to utilize it to assess MHEAs. Users of the evaluation framework and guidance tool are also able to change the selection of metrics, for usability testing if they want to focus on measuring usability for a different area of evaluation.

For instance, if they want to evaluate mobile marketing applications, they can select the participants for the interviews from a marketing background, which will generate a different set of selected hybrid metrics for mobile marketing apps. Then these new selected hybrid metrics can be used to generate questionnaire questions. Respondents to the questionnaire would also be selected from a marketing background, in order to measure usability for mobile marketing applications.

Moreover, the evaluation framework and guidance tool is a motivated framework, which covers all aspects of MHEAs usability. The metrics can be used to evaluate any MHEAs, on a range of mobile devices. The framework and guidance tool have been tested on three different types of mobile apps, and also verified by experienced evaluators. A few changes to the evaluation framework and guidance tool were suggested. It is clear that the framework can be adapted to a specific application. The researcher suggests adding more MHEAs to the framework, if the evaluators intend to assess more MHEAs, and expanding the data collection categories for more MHEAs.

6.8 Conclusion

This Chapter described the method utilized to exploit the outcomes from the questionnaire obtained in Chapter Five, in order to allocate the outcomes in MAST depending on using the Euclidean Distance metric. In addition, it provided a justification for using Euclidean Distance metric in MAST, by demonstrating that it is an appropriate and effective method. In addition, the chapter explained the design of skeleton of the MAST, beside the workflow for MAST. Moreover, it illustrated the working of the MAST in with the aid of some examples of users of MAST, showing the different matching results for different user profiles (HP, P/GP, SD), with different MHEAs profiles. Furthermore, the chapter explored the suitability of the MAST by reporting testing in the MAST. It indicated the contributions from the building MAST and using it, in order to achieve objective 4, which was to develop a MAST to select suitable MHEAs for different stakeholder types. Finally, it provided a guideline for using the framework and MAST tool.

Chapter Seven, will explore the methods considered and used to indicate and describe investigates of the evaluation of the framework.

Chapter 7: Discussion

Chapter Six described the MAST tool prototype, which was developed and used to match different stakeholders with different MHEAs packages. This fulfils object 4, which was to develop MAST to select suitable MHEAs for different stakeholders. The chapter presented some examples of different user's for the tool. This chapter explains the impact of the results on the framework. The discussions begins by providing an overview of the evaluation framework and guidance tool, used to gather and analyse data to rank MHEAs and build the MAST. After that, contributions against objectives are provided by evaluation of the research questions and objectives, and mapping them together, showing how the 5 research questions were addressed through 5 specific objectives. Afterwards an evaluation of the techniques of the framework is provided by describing the technique used for the interviews, questionnaire and MAST, tool. Subsequently the limitation of the framework are acknowledged and suggestions offered on how to improve the framework by increasing the MHEAs packages. Besides the usability study methods, an explanation of how to enhance the framework and MAST tool is provided, by adding more MHEAs packages and more data, which can be explored in future work. The chapter closes with a summary of the discussion.

7.1 Overview of the Evaluation Framework and Guidance Tool

The evaluation framework is a framework based on a usability metric and used for evaluation of apps on mobile devices, specifically MHEAs. Two methods informed the development of the evaluation framework, which are gathering the data from interviews and a questionnaire and analysing this data to rank the MHEAs and build the MAST.

The evaluation framework was used to measure and rank the usability of mobile apps, which influence the usability of MHEAs. The evaluation framework was also used to derive selected hybrid metrics, while the selected hybrid metrics were used to develop the MAST tool. The outcome of the combination of the two methods above was the evaluation framework and MAST tool, which consists of usability metrics to measure the usability of MHEAs.

The MAST consists of different matching options, to match between different stakeholders and different MHEAs. In addition, it provides recommended MHEAs packages for different stakeholders. The MAST was tested by conducting a usability study with different user groups. The results of the study were reported as well as feedback and suggestions obtained from the participants regarding the MAST, during the tutorial session. The next section will discuss the reliability of the evaluation framework, along with the effectiveness of the evaluation framework in the evaluation.

7.2 Contributions against Objectives

This section will explore the contributions of the research against the five objectives by presenting the five research questions, and showing how each of them has been addressed in different locations and sections through the thesis. These research questions and objectives were set out in Chapter One, sections 1.3 and 1.4. The mapping between these research questions and objectives is presented below in Figure 7.1. However, the tool used to create figure 7.1 was manually by determining each research question location through the thesis and mapping it to where has it been addressed through the thesis with the objectives. Besides determining, each subsection inside the thesis and page numbers through the thesis in term of consideration the accuracy for the mapping between the research questions and the objectives.

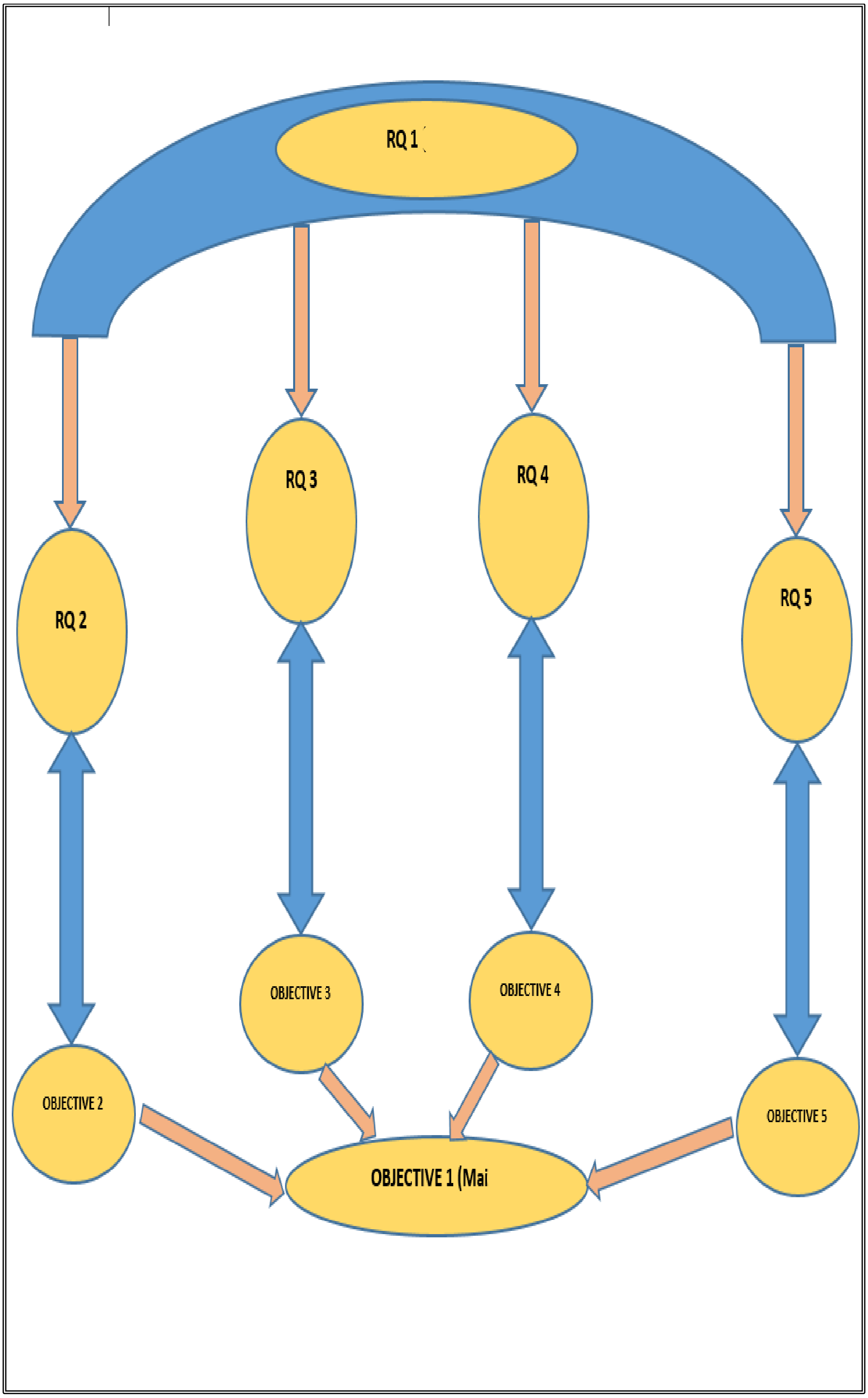


Figure 7.1: Mapping Research Questions with Objectives

Figure 7.1 highlights that the thesis proposed one main research question and four subsidiary research questions, to give a total of five research questions, which have been answered by one main objective and four subsidiary objectives.

The first and main Research Question of this thesis was:

How can we construct an effective and efficient hybrid evaluation framework for medical health apps utilizing an amalgam of Heuristic Evaluation (HE) and more general Usability Evaluation (UE) metrics?

The following objective was derived in order to address the first main research question.

Objective 1 Main: Developing an effective and efficient hybrid evaluation framework for Mobile Health Education Applications by utilizing a mix of Heuristic Evaluation (HE) and Usability Evaluation (UE) metrics.

The intention behind this objective was to gather the components of an effective and efficient hybrid evaluation framework, in order to develop an effective and efficient hybrid evaluation framework and MAST tool. This objective was mainly addressed in Sections 3.2, 3.3, 3.4 and 3.8, and partly addressed in Sections 4.14, 4.2, 4.3, 5.1, 5.1.1, 5.9 and 6.2.

As described in Chapter Three, the framework was developed in three stages. The three design components are similar to the first design, which was the first proposed design for an effective and efficient hybrid evaluation framework. For instance, in terms of efficiency, the framework reduced the number of factors from a large general factor set down to five selected hybrid metrics. Also in terms of effectiveness, the framework has been shown to have positive outcomes, which demonstrated that the framework worked well. The second stage was an updated version of the first stage of design to determine the selected hybrid metrics for MHEAs. The third stage of the design was then updated from the second stage, in order to develop the MAST tool.

To determine the minimum selected hybrid metrics, the framework proposed a larger set of potential hybrid metrics and utilised this larger set in order to structure the interviews, which were the first component of the framework. These interviews were analysed carefully, in order to decide on the final set of selected hybrid metrics.

Once the final set of selected hybrid metrics had been decided, it was possible to map them onto the questionnaire questions, in order to construct the evaluation questionnaire. The number of questions number in the questionnaire was calculated depending on the priority of each final selected hybrid metric, which was the second part of the evaluation framework. Furthermore, the outcomes from the questionnaire helped to explain the different stakeholder's requirements from different MHEAs. This information was utilized to develop the MAST tool, which was the third part of the framework.

Furthermore, the evaluation of this objective was achieved by involving the participants from three different categories, to participate in a pilot test, to determine whether the questionnaire was reliable to be used, as a part of the framework. The outcomes included confidence; these were explored in Chapter Five, section 5.1.1.

Additionally, as further evaluation of this objective, hypothesis testing was employed to validate the framework. For example, the thesis proposed the hypothesis that "there is a significant difference in the requirements for MHEAs between different stakeholders", by using the Kruskal-Wallis Test, using SPSS, the researcher found that the p-value for metric B, Features, was 0.02, which is less than the 0.05 threshold. This indicated a significant difference in the requirements for MHEAs between different stakeholders, which was explored in Chapter Five section 5.9.

To summarise, the evaluation of this objective was conducted from the perspective of efficiency, since the framework helped to minimize the number of questionnaire questions, which were explored in Chapter Five section 5.1. From the aspect of effectiveness, the framework resulted, after the procedure of analysing the questionnaire, in verification of the hypothesis, as explored in Chapter Five section 5.9.

The second Research Question of this thesis was:

How effectively can such as evaluation framework provide a suitable measurement in order to assess and rank Mobile Health Education Apps?

The following objective was derived in order to address the second research question.

Objective 2: Utilising the framework to develop a ranking measurements system so we can assess and rank Mobile Health Education Applications.

The intention behind this objective was to utilize the components of the effective and efficient hybrid evaluation framework in order to develop a ranking measurements system, to assess and rank MHEAs. This objective was mainly addressed in sections 4.13, 5.1.1, 5.2, 5.3, 5.7 and 5.9.

Initially, the interviews detected the stakeholders groups' priorities for the final selected hybrid metrics. This was one of the major outcomes of the interviews, which was followed by reflecting and mapping the information obtained in the questionnaire questions. In other words, the analysis of the interviews, which have been illustrated in Chapter Four section 4.13 provided suitable metrics to be mapped and used in the evaluation questionnaire questions, in order to rank and assess MHEAs.

The evaluations questionnaire is the second part of the evaluation framework. This questionnaire was distributed among three types of stakeholders, who had already been using several types of MHEAs for the purpose of developing their health educations. The outcomes from the questionnaire identified different rankings for different MHEAs. Moreover, the outcomes from the questionnaire identified that several MHEAs could be matched with the needs of these different types of stakeholders, which is would assist different types of stakeholders in selecting different MHEAs.

Finally, the evaluation for this objective applied by involving the participants from three different categories, to participate in a pilot test, to determine if the questionnaire was reliable enough to be utilized, as a part of the framework. The outcomes were positive; these were explored in Chapter Five, section 5.1.1. Furthermore, the evaluation of this objective was pursued by applying the Cronbach's alpha test, to validate the questionnaire items, before analysing the questionnaire outcomes. To provide measurements by which to asses and rank MHEAs; as explored in Chapter Five, section 5.2. Also, the objective was further evaluated by hypothesis testing, as explored in Chapter Five, section 5.9.

The third Research Question of this thesis was:

How can the proposed evaluation framework be used to identify issues and to aid in improving the design of Mobile Health Education Apps by supporting software developers?

The following objective was derived in order to address the third research question.

Objective 3: Using the framework to identify issues in Mobile Health Education Applications identified by the rankings, to aid in improving the design of Mobile Health Education Applications.

The intention behind this objective was to utilize the components of the effective and efficient hybrid evaluation framework, to identify issues in MHEAs detected by the rankings, so as to provide develops with information which could in improving the design of MHEAs. This objective was mainly addressed in section 5.4 and partly addressed in Sections 5.2, 5.6 and 5.9.

Initially, the outcomes from the questionnaire identified the positive and negative aspects of the MHEAs, which could help to improve the design and software engineering. Utilizing the evaluation questionnaire as a part of the framework identified a variety of issues in MHEAs. For instance, Chapter Five, section 5.4.1 presented that the Epocrates app has issues with memorability and accuracy, which need to be improved in ordered to satisfy different users of this app.

Moreover, the outcomes from the questionnaire described how the evaluation framework identified different stakeholders' opinions on different MHEAs, which would aid in improving the design of MHEAs.

Finally, the evaluation for this objective was applied by utilizing the Cronbach's alpha test, to validate the questionnaire items used to identify issues in MHEAs, as explored in Chapter Five, section 5.2. Also, the evaluation for this objective involved hypothesis testing to validate the framework used effectively, as explored in Chapter Five, section 5.9.

The fourth Research Question of this thesis was:

How far and in what ways can this evaluation framework be used to enable the specification and development of an appropriate tool to select suitable apps for different user groups? For this research, the different groups are Health Professionals, Patients and Other Stakeholders, including software developers/engineering?

The following objective was derived in order to address the fourth research question.

Objective 4: Using the framework to develop a Medical Apps Selection Tool to select suitable Mobile Health Education Applications for the different stakeholder types.

The intention behind this objective was to utilize the components of the effective and efficient hybrid evaluation framework, in order to develop a MAST, to select suitable MHEAs for the different stakeholder types. This objective was mainly addressed in sections 6.1, 6.2, 6.3, 6.4, 6.5 and 6.6.

Initially, the data from the evaluation questionnaire was utilized to develop the MAST tool. For instance, the matching section in Chapter Five section 5.7 was used to construct the profile data for MAST, in C# programming language and to develop the other parts of the MAST tool prototype. This would then help to match between the needs of different users of the MAST tool and different suitable MHEAs packages. A description was provided of the MAST tool, the coding, the flow process and a guideline for using it, with illustrative example.

Finally, the evaluation of this objective was performed by involving participants from the three different categories in using the MAST Tool. They were request to provide their opinions on it, which were positive, as reported in Chapter Six, section 6.6.

The fifth Research Question of this thesis was:

How effectively can the evaluation framework be used to characterize the different stakeholders and their different requirements for MHEAs?

The following objective was derived in order to address the fifth research question.

Objective 5: Utilising the framework to characterise the profile of different stakeholders in order to identify their different requirements.

The intention behind this objective was to utilize the components of the effective and efficient hybrid evaluation framework, to characterize the profile of different stakeholders, in order to identify their different requirements from a MHEAs. This objective was mainly addressed in section 5.5 and partly addressed in sections 5.6, 5.7 and 5.9.

Initially, the outcomes from the evaluation questionnaire described how the evaluation framework identified the characteristics of different stakeholders' requirements. For example in Chapter Five section 5.5.1, the framework identified the characteristics of HP priorities among requirements and it identified that the main requirement for HP in terms of using MHEAs is Simplicity. This indicates that HP would prefer MHEAs with a simpler user interface. Moreover, the outcomes from the evaluation questionnaire enabled a comparison between different stakeholder's requirements within the same MHEAs, to address their characteristics. Furthermore, the outcomes from the evaluation questionnaire enabled the matching between different stakeholder's requirements and different MHEAs.

Finally, the evaluation of this objective was performed by applying hypothesis testing, to confirm that there is a difference in the requirements of different stakeholders' from different MHEAs. This was explored in Chapter Five, section 5.9. In addition, the evaluation of this objective was performed in Chapter Five, section 5.7, in which the researcher specified the nearest characteristics for different stakeholders and their different requirements from MHEAs.

7.3 Evaluation of Techniques

In this section, firstly, an evaluation of the interviews applied in Chapter Four is made. Secondly, the questionnaire applied in Chapter Five is evaluated. Thirdly, the MAST proposed in Chapter Six is evaluated.

7.3.1 The Interview Approach

Interviews were applied in Chapter Four for filtering the proposed selected hybrid metrics in order to generate the final set of selected hybrid metrics. It was explained in Chapters Two and Four how the metrics were selected. Chapter Two reviewed in the literature and previous research in order to decide which were the most relevant metrics for MHEAs, to generate a proposed set of selected hybrid metrics. Then in Chapter Four, the proposed metrics were filtered to find the most relevant metrics for MHEAs and their stakeholders.

Fifteen interviews were conducted, for the reasons of identify the practicality issues. The interviews were effective, because they enabled the gathering of the data from three categories of stakeholders as the main users of MHEAs. Moreover, they were employed with participant's who had been regular user of MHEAs for at least 2 years, in order to gather data from users with a good background regarding MHEAs.

It was stated in Chapter Two that, according to Nielson (2000), dealing with five testing users is sufficient to find around 80% of the issues. Since the research involved fifteen users, divided among five HP, five P and five SD, this would be sufficient to identify more than 80% of the specific requirements for P, SD and HP from using MHEAs. Also, the research used various ways to gather the data from the interviews, such as voice recording and writing notes then converting the voice recording to word documents as raw data, followed by transcript coding of the raw data to NVIVO software. This supported the researcher to manage to gather and analyse the data in an effective way, besides making it meaningful.

There were same issues with these interviews, due to the need to include the three types of stakeholders (HP, P, and SD), and to ensure that all of users had been using MHEAs for at least 2 years. Therefore, finding these 15 interviewees was difficult and time-consuming task. Also even when it was possible to find prospective interviewees, conducting the interviews posed several issues. For instance, one of the interviewees was a male HP and the researcher managed to make an appointment with him. Nevertheless, as he was a very busy doctor, the researcher had to reschedule the reappoint with him four times. This delayed the research for two weeks, and at the end of the interview, it was found that his feedback was not reliable, so his interview had to excluded. This a major limitation of these interviews is the number of stakeholders, as it was not easy to find HP, P and SD, who had used the MHEAs for 2 years or more for health education, and moreover, finding stakeholders who could provide reliable feedback in their interviews. This was a very time- consuming.

On the other hand, these interviews were effective in the sense that they helped to specify the important requirements of different stakeholders, besides identifying the selected hybrid metrics, which were addressed depending on the ground truth of regular users of MHEAs.

7.3.2 Evaluation of the Questionnaire Technique

The questionnaire applied in Chapter Five for ranking MHEAs, identified issues in MHEAs. Besides, it characterized the different stakeholders and their different requirements for MHEAs. The questionnaire had a good response rate, it was sent to 128 participants, distributed among 60 x SD, 40 x HP and 28 x P/GP, and the researcher received 81 responses to the questionnaire, distributed among 27 x SD, 27 x HP and 27 x P/GP. This yields the following individual response rates:

SD = 27 ÷ 60 = 45%

HP = 27 ÷ 40 = 68%

P/GP = 27 ÷ 28 = 96%

These response rates can be considered as good percentages in this research for a small sample size (Sax, *et al.*, 2003). Moreover, for the overall response rate the research had the following response rate:

The Overall Response Rate = 81 ÷ 128 = 63%

Thus, although the overall sample size for the questionnaire was not huge, it was acceptable, as the overall response rate was 63%.

Designing the questionnaire was an important procedure. The research utilized two methods to design the questionnaire questions. The first method was illustrated in Chapter Five; it was reflecting and calculating the questionnaire questions depending on the metrics, as well as the priorities for each metric on the part of stakeholders. The second method was to avoid vague questions and ensure the questions clear for participants of all kind of background. Moreover, a pilot test was applied to determine whether the questionnaire questions were reliable.

Applying the questionnaires raised some issues, as the questionnaire was utilized with three types of stakeholders (HP, P/GP, and SD), and all participants had to have used the MHEAs for two years. Consequently finding these 81 participants was hard work, and time-consuming. Moreover, even after finding these participants, administration of the questionnaire has raised several problems. For example, one of the participants was a female HP. The researcher managed to send her the link to the questionnaire. However, as she was a very busy doctor, she took a week to respond to the questionnaire. Such experiences delayed the research.

The research also had some issues with the analysis of the data from the questionnaire. Determining whether the data was reliable for use before analysing it was important. Moreover, many types of software to analyse data are available, such as Excel, SPSS, MATLAB and R programming language. The researcher chose to analyse the data by using SPSS, as it is easy and fast to use. However, it was necessary to convert the analysis results from SPSS to Excel, as it provides more figure shape options (chart, pie) and colours than SPSS.

Despite these difficulties, the questionnaire produced useful feedback. It helped in ranking MHEAs and identifying issues in MHEAs, besides, characterizing the different stakeholders and their different requirements for MHEAs. All this information was valuable in helping to construct the MAST tool and its data.

7.3.3 Evaluation of the MAST Development

The MAST was applied, as reported in Chapter Six, for matching different MHEAs with different stakeholders and their different requirements for MHEAs. For the MAST tool, the researcher considered different ways to address the matching procedure between different stakeholders and different MHEAs. Eventually, the researcher decided to use the Euclidean Distance for several reasons, which were explained in Chapter Six.

Developing the MAST tool entailed some difficulties. For instance, as the tool was constructed based on data from the questionnaire, the sample size was small. With such a small sample, constructing a database for the MAST tool was not required at this stage. Thus, the database was replaced by an array and the task of developing the database has been left for future work. Moreover, applying the matching procedure in MAST was not an easy decision, there are many possible algorithms and metrics could be used for calculating the matching procedure. The researcher selected Euclidean Distance as it works well for variables with equal weighting, like usability. Thus, it was more suitable for this research. However, there are other algorithms, such as Cosine similarity, Jaccard similarity, CLUSTERING ALGORITHM and Pearson Correlation Coefficient, which could be used to implement the MAST and so another limitation of this research is the need to look at other measures other than Euclidean Distance, in order to find out which one is the best (Huang, 2008). In our opinion, Euclidean Distance is simpler to use. In addition, it might be the best and most relevant one for this research, in the sense that it is quick and intuitive, which suggests that Euclidean Distance is more suitable than other algorithms (D’Agostino & Dardanoni, 2009).

Moreover, comparing to other algorithms, the research picked the Euclidean Distance because of its benefits, because the researcher conducted some checking of its use and there was evidence that people thought it is the best match, so the research validated the Euclidean Distance in this way. This was clarified in Chapter Six section 6.1. The application of the Euclidean Distance metric for the MAST tool was for five dimensions, since the research has five metrics, which are A. Memorability, B. Features, C. Attractiveness, D. Simplicity, and E. Accuracy.

In order to verify that the Euclidean Distance is a suitable metric to consider for the MAST tool, Euclidean Distance was applied using SPSS software. The procedure was to find the shortest distance between the user profile and MHEAs profile, in order to arrive at the most suitable matching between the user profile and the MHEAs profile. Since the outcome was positive. This confirms that the Euclidean Distance metric is an appropriate metric to use. This was illuminated in Chapter Six section 6.1.1.

Furthermore, the MAST tool provided three types of matching for MHEAs, which are: Matching with group profile. Matching with MHEAs profile and Matching with the subset. These three types of matching help the user of the MAST tool, to select different types of matching according to their requirements. Nevertheless, the MAST tool has the limitation of having only three matching types, which could be increased in the future to more than three kinds of matching, as explained in Chapter Six, section 6.5.

The MAST tool was utilized with three types of stakeholders HP, P, and SD. Chapter Six section 6.5 showed different user cases for the MAST tool. These cases demonstrated that the MAST is a tool capable of addressing the matching of different MHEAs with different stakeholders and their different requirements for MHEAs.

7.4 Limitations and Future Work

As shown above, some of the contributions of this work – achieved through constructing an effective and efficient hybrid evaluation framework for MHEAs - are to enable the evaluation and ranking of MHEAs. This will make MHEAs easier to select and utilize by several types of stakeholders. Nevertheless, there remain a number of areas for further research.

Below are some examples of limitations related to the nature of the interviews, and the extent of their usefulness for gathering data, the nature of the questionnaire, and how far we can use it to gather data, and the different users of the tool. Moreover, it is discussed how the research opens the new paths for further research.

7.4.1 The Nature of the Interviews

One of the limitations is the nature of the interviews. Since the interviews were to be distributed among three categories of stakeholders, this required the researcher to develop a suitable design for the interview questions, in a way that would be suitable for all these categories, namely HP, SD and P. These interview questions were semi-structured, and the interviews consisted of 11 questions. To identify the suitability of the interview questions for all stakeholder categories, a pilot test was applied with three interviewees, one interviewee from each category of stakeholders. Following the pilot test and having confirmed the reliability of the interview questions, 15 interviews were carried out amongst the three categories of stakeholders.

Ideally, the research would have conducted more than 15 interviews in order to gather more data for the research and produce a wider range of results. However, due to lack of time and resource, it was not possible to conduct and analyse a larger number of interviews. However, whilst the number of interviews was only 15 - which seems a small sample size - nonetheless, the research managed to obtain trustworthy results, since the researcher selected the interviewees carefully and based on other research that shows that a sample of 15 can be valid. Since the participants had utilized MHEAs for at least two years, and depended on these apps for aspects of developing their health educations, this would provide the research with convincing results, and they provide useful feedback.

This has provided the research with trustworthy results, from which to identify an appropriate set of selected hybrid metrics with an efficient and reliable result, in addition capable of being used for the next step of the research (the evaluation questionnaire).

7.4.2 The Nature of the Questionnaires

Another limitation is related to the nature of the questionnaire The researcher needed to produce a questionnaire that would not take too long, whilst reflecting the selected hybrid metrics in the questionnaire questions. Moreover, as the questionnaire was to be distributed among three categories of stakeholders, this forced the researcher to adopt a suitable design for the questionnaire questions that would be suitable for all types of categories: HP, SD and P/GP. In order to identify the suitability for the questionnaire questions for all stakeholder categories, the research applied a pilot test with nine participants, three from each category of stakeholder (3 x HP, 3 x SD, 3 x P/GP). After obtaining the outcomes of the pilot test and confirming the reliability of the questionnaire questions, 81 respondents completed in the questionnaire, from the three categories of stakeholders.

Moreover, the appear to have provided good feedback, since the researcher verified the questionnaire outcomes freedom from errors by using Cronbach's Alpha (α) in SPSS. This yielded a value for Cronbach's alpha, of 0.774, which indicates a high level of internal consistency for our scale with this specific sample. Subsequently the researcher analysed these data by using SPSS software. The researcher used Two-Way ANOVA to analyse the questionnaire questions, as explained in Chapter 5. In addition, to test the hypotheses, the researcher used the Kruskal-Wallis Test using SPSS, to verify that there is a significant difference between the three categories of stakeholders. The outcome was significant since the p-value was significant at 0.02, which is less than the 0.05 criterion, thereby suggested acceptance of the proposition that there is a significant difference in the requirements between different stakeholders. This provided the research with confidence that the results have more than 70% accuracy, which indicated that the result was efficient and reliable. In addition, it was capable of being used for the next step of the research (designing the MAST Tool).

7.4.3 Different Users for the Tool

The MAST Tool has a limitation with regard to the sample of different users of the tool. The research included three types of users of the MAST tool, which were HP, SD and P. Having only three categories can be considered a limitation. In addition, the researcher could have considered other groups, such as students or teachers, which could have enabled the research to generate more options in wider contexts for the MAST Tool. However, the researcher selected these three categories because they are the main users of MHEAs, with relevant background in their uses.

Another limitation of the MAST Tool concerns the MHEAs packages available for the users of the tool, since the research had only three type MHEAs packages, namely Epocrates, Medscape, and Other Apps. The researcher selected these three MHEAs packages because they are the main utilized MHEAs packages. Besides, the researcher used them in the research background for the evaluation framework, as the researcher utilized these MHEAs packages as the focus of the questionnaire.

7.4.4 Future Work for Further Research

Even though a lot has been achieved during the Ph.D. research, nonetheless there are still some other valuable research areas that could be completed in the future. This work opens up a new paths for further research. These further works are described as follows:

- As this research focus on the MHEAs area, the researcher designed and tested the evaluation framework and MAST Tool only on MHEAs. In addition, we focused only on ranking and measuring MHEAs. However, a focus for the future research could be to consider mobile apps for other field, such as mobile marketing apps or games mobile apps. This could help to expand our framework and MAST Tool features to make it useful more widely, to rank and measure other mobile apps, as well as health apps.

- The research focused on three types of stakeholders, which are HP, SD and P/GP, so we designed and tested the evaluation framework and MAST Tool only on these groups. In addition, we focused on ranking and measuring apps just for HP, SD and P/GP. However, the future research could consider other stakeholders, such as teachers and students which could help to expand our framework features and MAST Tool, to make it useful to rank and measure apps for other groups.

- The research involved 81 participants who responded to the questionnaire, alongside 15 participants who participated in the interviews, distributed among three types of stakeholders, HP, SD and P/GP. This sample size seems small; since it was constrained by the time available within the Ph.D. as a three to four-year project. Thus, future research could increase the sample size. Instead of just, 81 participants for the questionnaire, recruiting a wider range of participants could generate a larger sample size. Correspondingly, we could increase the interviews sample size, instead of having only 15 participants. Applying a wider range of participants, and a larger sample size, could help to expand the features of our framework and the MAST Tool to make it more useful to rank and measure stakeholder requirements and app features.

- Because of the small sample size, in this research, it was not necessary to develop a database, to store the questionnaire data outcomes. Instead, we used stored the data within the program. However, for future work, having a larger sample size would produce a broader range of data content.This would benefit from the development and use of a database, in order, to store the data set for a larger data sample, and to enable users to access it online.

- The research focus on developing the evaluation framework and MAST Tool for the MHEAs area. Consequently, we designed and tested the evaluation framework and MAST Tool only to rank and measure these types of apps. Thus, a future area of focus could be to consider other areas, to expand our framework capability features, and to make it beneficial for other computer apps or areas such as smart cities, robots or spaceflight.

- Lastly, the research focused on developing the evaluation framework and MAST Tool in three parts, some of which have remain as future work, such as a fully controlled automatic system to rank and measure health applications. Hence, a following area of focus could be to consider, full implementation of our framework to produce a fully controlled automatic system to rank and measure health applications. This could help to expand our framework and MAST Tool characteristics to make it more useful to rank and measure apps.

7.5 Conclusion

This Chapter has provided an evaluative discussion of the research from different aspects. The chapter began by reviewing the achievement of the research against five objectives, mapped to the five research questions it was demonstrated in detail how our contributions, which are a collection of metrics, the framework and the MAST tool applying them, met each of these objectives, thereby addressing each research question. Cross-reference was made to the relevant thesis sections in these evaluations.

Moreover, the chapter provided an evaluation of the techniques used: which were Interviews, questionnaire and the MAST tool by illustrating some of the issues when utilizing these techniques. Furthermore, the chapter identified some of the limitations of the research and overviews for future work. These include issues related to the nature and samples for the interviews and the questionnaire, and the uses of the MAST tool. It was shown how these open up avenues for the further research.

Chapter Eight will encapsulate the whole thesis, providing an overall conclusion.

Chapter 8: Conclusion

Chapter Seven described the evaluations of the framework and MAST tool prototype. The chapter presented a discussion, and identified some limitations and suggested future work. This chapter provides a summary and overview of the evaluation framework and the MAST tool as a conclusion to the thesis and highlights the significance of the research in terms of its approaches and contributions.

8.1 Summary of the Thesis

This thesis has presented the rationale for conducting this research, which was presented in eight chapters. Chapter One explained the research problem, showing the need to explore further the usability evaluation area with specific reference to Mobile Health Education Apps (MHEAs). It also introduced the hypothesis of this thesis, that there are significant differences in the needs of distinct stakeholders for MHEAs and hence, the effective usefulness of such apps.

This was followed by Chapter Two, which explored the previous research's on usability evaluation for MHEAs. It described the background research on mobile HCI, and provided an introduction to mobile devices and MHEAs, highlighting some of the challenges involved in undertaking research's on mobile technology. It discussed methods of evaluating mobile apps, and presented previous and recent studies' focus on mobile HCI and usability metrics. In addition, it provided an introduction to Heuristic Evaluation (HE) and Usability Evaluation (UE). Chapter Two also raised some issues related to mobile devices and reviewed some previous studies for mobile phone evaluation, identifying their limitations.

Chapter Three explained the proposed evaluation framework for MHEAs. It described the stages of designing the framework, which went through three designs (first, second and third design). Moreover, Chapter Three clarified the reasons for moving from one design to another, with the reasons for modifying these designs. In addition, it investigated the proposed selected hybrid metrics of HE and UE and how they were mapped to the questionnaire questions, besides state the priority of the metrics for measurement among different stakeholder groups.

Chapter Four addressed the final selection of five hybrid metrics from a large set of HE and UE metrics, based on interviews with HP, P and SD. These led to the selection of critical hybrid HE and UE metrics: A. Memorability, B. Features, C. Attractiveness, D. Simplicity and E. Accuracy. These critical selected hybrid metrics were useful to form a part of the framework developed to rank and measure the usability of MHEAs. This constituted on effective framework for evaluation of MHEAs that could be used, in order to improve the design of MHEAs and identify current issues in MHEAs.

Chapter Five described the method used to develop the questionnaire questions, depending on the weighting obtained from the qualitative interviews. In addition, it illuminated the difference in the overall rankings of different MHEAs packages, by different stakeholders. Also, Chapter Five explored the methods used to analyse the questionnaire using SPSS software. As well, it reported on the differential in the same MHEAs within the five metrics and the rankings of these metrics in the same and different MHEAs. In addition, the chapter explored the differences among Stakeholders in their ranking of the five metrics and the metrics requirements of the same and different stakeholders. It presented the matching between the profiles of MHEAs packages and the profiles of different stakeholders. Furthermore, Chapter Five tested the hypothesis for the thesis. Finally, it explored the suitability of each MHEAs package according to the metrics, to suit different stakeholders.

Chapter Six described the method utilized to apply the outcomes from the questionnaire, reported in Chapter Five, in order to develop the MAST tool, depending on using Euclidean Distance, and provided a justification for using it. Moreover, Chapter Six reported on designing the skeleton of the MAST tool, besides the workflow of MAST tool. The Chapter also explored the working functionality of the MAST tool, supported by some examples of users, and presented various matching results for different user profiles and different MHEAs profiles. Finally, Chapter Six explored the suitability of the MAST tool by reporting on its testing and indicating the contributions from building the MAST tool and using it.

Chapter Seven contained a discussion of the thesis from different aspects. The chapter began with an overview of the evaluation framework and the MAST tool. In addition, it illuminated the research contributions against five objectives, highlighting them and the five research questions and mapping between them. It highlighted the research contribution, which is a collection of metrics and the framework and MAST tool applying them. Chapter Seven provided evaluations for all these objectives, by addressing each research question through the relevant objective, utilized to answer it. Moreover, Chapter Seven contained an evaluation of the techniques used Interviews, the questionnaire and the MAST tool. Furthermore, it identified some of the limitations of the research and related to different users of the tool and the nature of the interviews and questionnaire, and showed how these issues open a new path for further research.

Finally, Chapter Eight summarised the whole thesis, with a summary of the key findings regarding the evaluation framework and guidance tool. It concludes by highlighting the significance and contributions of research.

8.2 Summary of Key Findings for Evaluation Framework and Guidance Tool

This research has developed a suitable way to evaluate mobile apps, specifically MHEAs. Contemporary evaluation methods are too general. In addition, appropriate metrics to assist evaluation are limited. Moreover, current metrics were mainly established for desktop systems there has been and limited focus on mobile applications. The previous researchers on mobile settings who proposed metrics focused more on assessing mobile devices rather than evaluating mobile applications. In fact, there are millions of different kinds of mobile apps, but there are not millions of different types of mobile devices. (Statista, 2019).

The research outcomes of this study are on evaluation framework for MHEAs and the MAST tool. To the best of our knowledge, there is no published work using this research approach to develop usability metrics for mobile apps specifically MHEAs. This research provides usability metrics that can be used to evaluate mobile apps, in specific, MHEAs. It can help different stakeholders to choose among different mobile apps to meet their requirements by providing some recommended mobile apps. Three different types of stakeholders participated in using the framework and the MAST tool: Health Professionals, Software Developers and Patients / General Public. In addition, three different types of MHEAs were tested using the evaluation framework and the MAST tool: Medscape, Epocrates and Other Apps (WebMD, UpToDate).

The evaluation framework and the MAST tool were able to evaluate all the apps mentioned above, and to identify the usability problems of each app. One reason that makes the evaluation framework and the MAST tool suitable and effective is their dynamic capability. The potential to add or drop types of stakeholders or types of mobile apps is one of the advantages of the evaluation framework and the MAST tool. MAST tool users also have the opportunity to assign priority to the requirements for mobile apps. In addition, they have the option to select the type of matching they prefer, which can even adjust the requirement's goals. For instance, if the users want to find a suitable app focusing on ease of using the app, the level of the metric 'Simplicity' can be increased, to match with mobile apps matching these metrics, such as Medscape.

Moreover, the evaluation framework and the MAST tool can increase the matching of users and mobile apps by adding more stakeholders and more mobile apps packages. The evaluation framework and the MAST tool have shown an ability to collect subjective (interviews) and objective (questionnaire) data. Validation of the evaluation framework and the MAST tool was applied, by validating the data for the evaluation framework and testing the MAST tool, across three stakeholders categories with one participant per category (HP, SD, P). The reason for this was to verify whether the evaluation framework provides reliable data on different mobile apps and different stakeholders, and to check whether the MAST tool is suitable for different user different mobile apps.

Comparative evaluation of the MHEAs indicates that the characteristics of MHEAs influenced the usability for stakeholders. e.g. through 'how easy is it to use the mobile health apps'. The results from the usability study indicate that the evaluation framework and the MAST tool can be used to evaluate MHEAs. The evaluation framework enabled the collection and analysis of objective and subjective data.

Moreover, the MAST tool enabled the characterizing of different users and their matching with different MHEAs. This illustrated usability problems and permitted the suggestion of improvements to the MHEAs. However, the evaluation framework and the MAST tool are just a model with usability metrics. The model (the evaluation framework and the MAST tool) has some limitations, which were explained in Chapter Seven section 7.4. We are aware of these limitations and intend to address them in future study of the evaluation framework and the MAST tool, which was proposed in the section on future work in Chapter Seven, section 7.4.

8.3 Significance and Contributions of Research

The research presented in this thesis proposed five-research questions, which in summary are: How can we develop an evaluation framework for MHEAs? Can the framework provide a suitable measurement to rank MHEA? How can the framework be used to identify issues in MHEAs? How can the framework help in developing an appropriate tool to select suitable apps for different user groups? How can the framework be used to characterize the stakeholders and their requirements for MHEAs? These five research questions were answered through the following five objectives: Developing a framework for MHEAs by utilizing a mix of Heuristic Evaluation (HE) and Usability Evaluation (UE) metrics, Utilising the framework to rank MHEAs, Using the framework to identify issues in MHEAs, Using the framework to develop a Medical Apps Selection Tool, and Utilising the framework to characterise the profile of different stakeholders in order to identify their different requirements. In addition, the research assumed the hypothesis that there are significant differences in the needs of distinct stakeholders for - and the effective usefulness of – MHEAs. This was tested and explored in Chapter Five, by analysing the data from the questionnaire using Two-Way ANOVA in SPSS, and using the Kruskal-Wallis test in SPSS to check the p-value which was less than 0.05.

Moreover, the research presented in this thesis focused on the usability metrics, to evaluate MHEAs. Evaluation of MHEAs is different from that of other apps. The unique characteristics of MHEAs are the main challenge to evaluation and finding suitable metrics. We derived the selected hybrid metrics for mobile applications by applying semi-structured interviews with the main users of MHEAs. Our initial outcomes are the evaluation framework and MAST tool, which was validated in a usability study; besides, some selected hybrid metrics were validated. We have addressed some of the issues in a few MHEAs packages, and presented a refined version of the evaluation framework and MAST Tool. The evaluation framework is capable of assisting the design of MHEAs, by providing a ranking for these apps. In addition, the MAST tool is capable of matching between different stakeholders and with different MHEAs packages, besides providing the users of with a recommendation for MHEAs, which will meet their requirements. These contributions will save time and effort for different types of stakeholders, in finding suitable MHEAs, which will help them to improve their health education. A guideline for using the framework and MAST Tool is provided. In addition, the MAST tool has been tested, and methods employed to get feedback and have been evaluated and suggestion provided to improve the framework and MAST Tool. The evaluation framework and MAST tool presented in this thesis contribute an option to the field of evaluation related to mobile apps.

References

Adibi, S. (2012) Link technologies and BlackBerry mobile health. *IEEE Transactions on Information Technology in Biomedicine*, July, 16 (4), 586-597.

Ali, A. A. (2013) A *framework for measuring the usability issues and criteria of mobile learning applications.*

Alison. (2011) *Social Media overtakes entertainment as most popular web.*

Aljaber, T. & Gordon, N. (2016) *Evaluation of mobile health education applications for health professionals and patients*. Funchal, Madeira, Portugal, Proceedings of the 8th International conference on e-Health (EH 2016), 107-114.

Aljaber, T. & Gordon, N. (2017a) *A guidance and evaluation approach for mHealth education applications.* Vancouver,Canada, In International Conference on Learning and Collaboration Technologies. Springer, Cham, 330-340.

Aljaber, T. & Gordon, N. (2017b) *A Hybrid Evaluation Approach and Guidance for mHealth Education Applications.* Los Angeles,California, USA, In International Conference on Applied Human Factors and Ergonomics. Springer, Cham, 282-290.

Aljaber, T., Gordon, N., Kambhampati, C. & Brayshaw, M. (2015) *An Evaluation Framework for Mobile Health Education Software.* London, UK, Proceedings of Science and Information Conference (SAI), 786-790.

Alva, M. E. , Ch, T.H.S. & López, B. (2003) Comparison of Methods and Existing Tools for The Measurement of Usability in the Web. In International Conference on Web Engineering. Springer, Berlin, Heidelberg, 386-389.

Allam, A. H., Hussin, A. R. C. & Dahlan, H. M. (2013) User Experience: Challenges and Opportunities. *Journal of Information Systems Research and Innovation,* 3 (1), 28-36.

Ankeny, J. (2010) *Forecast: 50 billion mobile app downloads in 2012.* [Online] Available at: http://www.fiercemobilecontent.com/storv/forecast-50-billionmobile- [Accessed 15/02/2014].

Archetype/Sapient. (1999) *e-Commerce trust study.* [Online] Available at: http://experiencecurve.com/articles/ecommerceTrustCheskin.pdf [Accessed 10/10/2014].

Baharuddin, R., Singh, D. & Razali, R. (2013) Usability dimensions for mobile applications-a review. *Research Journal of Applied Sciences, Engineering and Technology,* 21 February, 5(6), 2225- 2231.

Baillie, L. (2003). *Future Telecommunication: Exploring actual use,* Paper presented at the International Conference on Human-Computer Interaction (INTERACT03), 697-700.

Basili, V., Caldeira, G. & Rombach, H. D. (1994) *The goal question metric approach.* Encyclopedia of Software Engineering, 528-532.

Benbasat, I. (2010) HCI research: Future challenges and directions. *AIS Transactions on Human-Computer Interaction,* 2(2), 16-21.

Bernhaupt, R. (2009) Usability and user experience evaluation methods. In: C. Mourlas (ed) *Mass customization for personalized communication environments: Integrating Human Factors.* New York: Information science reference, 232-233.

Bertoa, M., Troya, J. & Vallecillo, A. (2006) Measuring the usability of software components. *Journal of Systems and Software,* 79(3), 427-439.

Bertram, . D., Scales, . L. & Scale, . L. (2007) *Likert Scales… are the meaning of life,* Calagary: University of Calagary, Department of Computer Science.

Bevan, N., Carter, J., Earthy, J. & Geis, T. (2016) *New ISO standards for usability, usability reports and usability measures.* In International conference on human-computer interaction. Springer, Cham, 268-278.

Bobian, M., Kandinov, A., El-Kashlan, N., Svider, P. F., Folbe, A. J., Mayerhoff, R. & Raza, S. N. (2016) Mobile applications and patient education: Are currently available GERD mobile apps sufficient?. *The Laryngoscope,* 127(8), 1775–1779.

Booth, P. A. (2014) *An Introduction to Human-Computer Interaction.* Psychology Press . Hove: Lawrence Erlbaum Associates, Ltd.

Botella, P., Burgués, X., Carvallo, J. P., Franch, X., Grau, G., Marco, J., & Quer, C. (2004) *ISO/IEC 9126 in practice: what do we need to know.* Milan, In Software Measurement European Forum.

Boudreaux, E. D., Waring, M. E., Hayes, R. B., Sadasivam, R. S., Mullen, S., & Pagoto, S. (2014) Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. *Translational Behavioral Medicine,* 4(4), 363–371.

Brayshaw, M., Gordon, N., Nganji, J. & Wen, L. (2014) *Investigating heuristic evaluation as a methodology for evaluating pedagogical software: an analysis employing three case studies.* Heraklion,Greece, In International Conference on Learning and Collaboration Technologies. Springer, Cham, 25-35.

Brodkin, J. (2008) *Mobile phones to be primary Internet device by 2020.* [Online] Available at: http://www.networkworld.com/news/2008/121508-pew-report.html [Accessed 11/02/2014].

Broll, G. & Hausen, D. (2010) *Mobile and physical user interfaces for NFC-based mobile interaction with multiple tags,* In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services.ACM, 133-142.

Brown, B., Balatsoukas, P., Williams, R., Sperrin, M., & Buchan, I. (2016) Interface design recommendations for computerised clinical audit and feedback: hybrid usability evidence from a research-led system. *International Journal Of Medical Informatics,* 94, 191-206.

Silva, B. M., Rodrigues, J. J., Canelo, F., Lopes, I. C., & Zhou, L. (2013) *A novel cooperation strategy for mobile health applications.* [Online]
Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6585879
[Accessed 12/06/2014].

Buttussi, F., Chittaro, L., Carchietti, E. & Coppo, M. (2010) *Using mobile devices to support communication between emergency medical responders and deaf people*, In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, 7-16.

Cairns, P. & Cox, A. L. (2008) *Research methods for human-computer interaction.* Cambridge: Cambridge University Press.

Caro-Alvaro, S., Garcia-Lopez, E., Garcia-Cabot, A., de-Marcos, L., & Martinez-Herraiz, J. J. (2018) Identifying usability issues in instant messaging apps on iOS and Android platforms. *Mobile Information Systems.*

Cass, S. (2014) *Top 10 programming languages.* [Online]
Available at: http://spectrum.ieee.org/computing/software/top-10-programming-languages
[Accessed 22/ 05/2015].

Cisco. (2017) *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2016–2021 White Paper,* s.l.: Cisco and/or its affiliates.

Coursaris, . C. & Kim, D. (2007) *A research agenda for mobile usability,* In CHI'07 extended abstracts on human factors in computing systems, 2345-2350.

Coursaris, C. K. & Kim, J. D. (2011) A meta-analytical review of empirical mobile usability studies. *Journal of Usability Studies,* 6(3), 117-171.

Cui, Y., Honkala, M., Pihkala, K., Kinnunen, K., & Grassel, G. (2010) *Linked internet UI: a mobile user interface optimized for social networking*, In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, 45-54.

Chaudhuri, A. (2012) ICT for development: solutions seeking problems?. *Journal of Information Technology*, 27(4), 326-338.

Chen, C. (2018) Introduce to" Usability-quality Product Development Technology (UPDT)": Taking the Design of Desk and Chair as an Example. *International Journal of Application or Innovation in Engineering & Management,* 7(12), 10-21.

Chittaro, L. & Marassi, A. (2010) *Supporting blind users in selecting from very long lists of items on mobile phones*, In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, 27-30.

D'Agostino, M. & Dardanoni, V. (2009) What's so special about Euclidean distance? A characterization with applications to mobility and spatial voting. *Social Choice and Welfare,* 33(2), 211–233.

Baumgart, D. C. (2011) Smartphones in clinical practice, medical education, and research. *Archives of internal medicine,* 171(14), 1294-1296.

De Sá, M. & Carriço , L. (2008) Defining scenarios for mobile design and evaluation. *In CHI'08 extended abstracts on Human factors in computing systems,* 2847-2852.

De Sá, M., Carriço, L., Duarte, L. & Reis, T. (2008) A framework for mobile evaluation. *In CHI'08 extended abstracts on Human factors in computing systems,* 2673-2678.

Dix, A., Finlay, J., Abowd, G. D. & Beale, R. (2004) *Human-computer interaction.* 3rd edition. London: Prentice-Hall.

Dubey, S. K. & Mridu, A. R. (2012) Analytical Comparison of usability measurement methods. *International Journal of Computer Applications,* 39(15), 11-18.

Duh, H., Chen, V. & Tan, C. (2008) *Playing different games on different phones: an empirical study on mobile gaming*, In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, 391-394.

Dumas, . J. S. & Redish, . J. C. (1999) *A Practical Guide to Usability Testing.* Exeter,UK: Intellect Books.

East, N. I. S. (2014) *App Development An NHS Guide for Developing Mobile Healthcare Applications,* s.l.: NHS.

Finstad, K. (2010) The usability metric for user experience. *Interacting with Computers,* 22(5), 323-327.

Frakes , W. & Terry, C. (1996) Software reuse: metrics and models. *ACM Computing Surveys,* 28(2), 415-435.

Fraunholz, B. & Jung, J. (2002) *Evaluation of Mobile Application -Software-Technical and Human Espect.* Paris, Paper presented at the European Conference on Information Technology Evaluation (ECITE 2002), Universite Paris-Dauphine, Paris, France.

Gafni, R. (2008) Framework for quality metrics in mobile-wireless information systems. *Interdisciplinary Journal of Information, Knowledge & Management,* 3, 23-38.

Gafni, R. (2009) Usability issues in mobile-wireless information systems. *Issues in Informing Science & Information Technology,* 6(2009), 755-769.

Gartner. (2014) *Gartner Says Annual Smartphone Sales Surpassed Sales of Feature Phones for the First Time in 2013.* [Online]
Available at: http://www.gartner.com/newsroom/id/2665715
[Accessed 12/02/2015].

Georgiev, T. & Georgieva, E. (2010) *Module for investigation of the user's text reading speed on mobile devices*, In Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies, 347-352.

Ghasemifard, N., Shamsi, M., Kenari, A. & Ahmadi, V. (2015) A new view at usability test methods of interfaces for human computer interaction. *Global Journal of Computer Science and Technology*.

Gilligan, C. & Golden, L. (2009) Re-branding social good: Social profit as a new conceptual framework. *Academy of Marketing Studies Journal*, 13(2), 97-117.

Glanz, K., Rimer, B. K. & Viswanath, K. (2008) *Health behavior and health education: Theory, Research, and Practice.* 4th edition. San Francisco: Jossey-Bass Wiley.

Gong, J. & Tarasewich, P. (2004) *Guidelines for handheld mobile device interface design.* Boston, In Proceedings of DSI 2004 Annual Meeting, 3751-3756.

Goodman, P. (2004) *Software metrics: best practices for successful IT management.* s.l.:Rothstein Associates Inc.

Gordon , N., Brayshaw , M. & Aljaber , T. (2016) *Heuristic evaluation for serious immersive games and M-instruction*, In International Conference on Learning and Collaboration Technologies . Springer, Cham, 310-319.

Graham Hole Research Skills. (2011) *The Kruskal-Wallis test.* [Online]
Available at: http://users.sussex.ac.uk/~grahamh/RM1web/Kruskal-Wallis%20Handoout2011.pdf
[Accessed 05/01/2017].

Guerreiro, T., Nicolau, H., Jorge, J. & Gonçalves, D. (2010) *Assessing mobile touch interfaces for tetraplegics*, In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, 31-34.

Hardy, R., Rukzio, E., Holleis, P. & Wagner, M. (2010) *Mobile interaction with static and dynamic NFC-based displays*, In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, 123-132.

Harrison , R., Flood, D. & Duce, D. (2013) Usability of mobile applications: Literature review and rationale for a new usability model. *Journal of Interaction Science,* 1(1), 1-16.

Hastreiter, I., Krause, S., Schneidermeier, T. & Wolff, C. (2014) *Developing UX for Collaborative Mobile Prototyping*, In International Conference of Design, User Experience, and Usability. Springer, Cham, 104-114.

Internet Medicine Digital Healthcare. (2014) *TOP FREE MEDICAL APPLICATIONS.* [Online]
Available at: http://internetmedicine.com/best-free-medical-apps/
[Accessed 29/04/2019].

Heggestuen, J. (2013) *One In Every 5 People In The World Own A Smartphone, One In Every 17 Own A Tablet [CHART].* [Online]
Available at: http://www.businessinsider.com/smartphone-and-tablet-penetration-2013-10?IR=T
[Accessed 2/08/2015].

Helander, M. (2014) *Handbook of human-computer interaction.* Amsterdam: Elsevier Science B.V.

Hernadez, M. L. & Woolley, S. I. (2009) *A user-centered mobile health device to manage life-threatening anaphylactic allergies and provide support in allergic reactions.* In Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine, IEEE, 1-4.

Holzinger, A., Searle, G., Kleinberger, T., Seffah, A. & Javahery, H. (2008) *Investigating usability metrics for the design and development of applications for the elderly.* In International Conference on Computers for Handicapped Persons. Berlin: Springer, 98-105.

Hornbask, . K. & Law, E. L.C. (2007) *Meta-analysis of correlations among usability measures*, In Proceedings of the SIGCHI conference on Human factors in computing systems.ACM, 617-626.

Huang, A. (2008) *Similarity Measures for Text Document Clustering*, In Proceedings of the sixth new zealand computer science research student conference. Christchurch, New Zealand. Scimago Journal & Country Rank, 4, 9-56.

Hubert, R. (2006) *Accessibility and usability guidelines for mobile devices in home health monitoring*, ACM Sigaccess Accessibility and Computing, 84, 26-29.

Hussain, A., Kutar, M., Abdul Mutalib, A. & Md Kamal, F. (2012) Modeling subjective metrics for mobile evaluation. *Journal of Information System Research and Innovation,* 1, 11-20.

Hussain, A., Mkpojiogu, E., Fadzil, N. & Hassan, N. (2017) *The UX of amila pregnancy on mobile device*, In Proceedings of the AIP Conference, 1891(1), 020061(1-7).

Hussain, A., Raed Mohammed Hussein, Jasim Mohammed Dahr & Mohammed Abbas Neamah. (2015) EXISTING trends in usability evaluation method (UEM) for M-learning apps: A systimatic review. *Journal of Engineering and Applied Sciences,* 10(15), 6649-6653.

I.S.O. (2001) *ISO/Iec 9126-1: Software engineering-product quality-part 1: Quality model,* Geneva, Switzerland, International Organization for Standardization, 21.

IEEE. (1998) *IEEE Std 1061–1998 IEEE standard for a software quality metrics methodology*.: IEEE Standards Association.

Husain, Iltifat., Einerson, Brett. (2015). *Top 15 Free Android Medical apps for Healthcare professionals.* [Online]
Available at:
http://iims.uthscsa.edu/sites/iims/files/Top%2015%20Free%20Android%20Medical%20apps%20for%20Healthcare%20professionals.pdf
[Accessed 15/09/2015].

InterQ Research. (2017) *DETERMINING SAMPLE SIZE FOR QUALITATIVE RESEARCH: WHAT IS THE MAGICAL NUMBER?* [Online]
Available at: https://interq-research.com/determining-sample-size-for-qualitative-research-what-is-the-magical-number/
[Accessed 11/05/2017].

ISO 9241-11. (2018) *Ergonomics of human-system interaction —part 11: usability: definitions and concepts.* [Online]
Available at: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en.
[Accessed 05/01/2019].

ISO DIS, 9.-2. (2008) *Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems.* [Online]
Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075
[Accessed 13/09/2014].

ISO. (1998) *International Standard: ISO 9241-11 (Guidance on usability),* Geneva.

Istepanian, . R. S. & Lacal, J. (2003) *Emerging Mobile Communication Technologies for Health: Some Imperative notes on m-health,* In Proceedings of the 25th IEEE Annual International Conference: Engineering in Medicine and Biology Society, 2, 1414-1416.

Istepanian, . R. S. & Woodward, B. (2016) *M-Health: Fundamentals and Applications*, John Wiley & Sons.

Ivory, M. Y. & Hearst, M. A. (2001) The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys,* 33(4), 470-516.

Jokela, T., Koivumaa, J., Pirkola, J., Salminen, P. and Kantola, N. (2006) Methods for quantitative usability requirements: a case study on the development of the user interface of a mobile phone. *Personal and Ubiquitous Computing,* 10(6), 345-355.

Jorg, R. (2002) Patterns of mobile interaction. *Personal and Ubiquitous Computing,* 6(4), 282-289.

Juhlin, O., Engström, A. & Reponen, E. (2010) *Mobile broadcasting: the whats and hows of live video as a social medium,* In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services. ACM, 35-44.

Kakihara, M. & Sorensen, C. (2002) *Mobility: An extended perspective.* Hawaii, In Proceedings of the 35th annual hawaii international conference on system sciences, IEEE , 1756-1766.

Kjeldskov, J. & Stage, J. (2004) New techniques for usability evaluation of mobile systems. *International journal of human-computer studies,* 60(5-6), 599-620.

Kratz, S., Brodien, I. & Rohs, M. (2010) *Semi-automatic zooming for mobile map navigation*, In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services.ACM, 63-72.

Kronbauer, A., Santos, C. & Vieira, V. (2012) *Smartphone applications usability evaluation: a hybrid model and its implementation.* Berlin, Heidelberg, In International Conference on Human-Centred Software Engineering. Berlin, Springer, 146-163.

Kumar & Kumar, R. (2005) *Human Computer Interaction.* New Delhi: Firewall Media.

Kumar, B. & Mohite, P. (2016) Usability guideline for mobile learning apps: An empirical study. *International Journal of Mobile Learning and Organisation ,* 10(4), 223–237.

Kumar, B. & Mohite, P. (2018) Usability of mobile learning applications: a systematic literature review. *Journal of Computers in Education,* 5(1), 1-17.

Lacerda, T. C., Wangenheim, C. G. v., Wangenheima, A. v. & Giuliano, I. (2014) Does the use of structured reporting improve usability? A comparative evaluation of the usability of two approaches for findings reporting in a large-scale telecardiology context. *Journal of biomedical informatics,* 52(2014), 222-230.

Leonardi, C., Adriano, A., Fabio P & Massimo Z. (2010) *An Exploratory Study of a Touch-based Gestural Interface for Elderly.* [Online]
Available at: http://delivery.acm.org/10.1145/1870000/1869045/p845-lenoradi.pdf?ip=150.237.96.64&id=1869045&acc=ACTIVE%20SERVICE&key=BF07A2EE685417C5%2EB052FA0818D6DC5F%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=714536761&CFTOKEN=90292930&__acm__=1442652491_465ae974be4
[Accessed 20/01/2015].

Liu, C., Zhu, Q., Holroyd, K. & Seng, E. (2011) Status and trends of mobile-health applications for iOS devices: A developer's perspective. *Journal of Systems and Software,* 84(11), 2022-2033.

Lobo, D., Kaskaloglu, K., Kim, C. & Herbert, S. (2011) Web Usability Guidelines For Smartphones:A Synergic Approach. *International Journal of Information and Electronics Engineering,* 1(1), 33-37.

Ziefle, M. & Bay, S. (2008) Transgenerational Designs in Mobile Technology. In *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*. New York: InformatIon scIence reference, 122-141.

Lund Research. (2013) *Cronbach's Alpha (α) using SPSS Statistics.* [Online] Available at: https://statistics.laerd.com/spss-tutorials/cronbachs-alpha-using-spss-statistics.php
[Accessed 23/04/2017].

Lyytinen, K. (2010) HCI research: Future directions that matter. *AIS Transactions on Human-Computer Interaction,* 2(2), 22-25.

Manerikar, V. & Manerikar, S. (2015) Cronbach's alpha. *Aweshkar Research Journal,* 19(1), 117-119.

Fox, N., Hunn, A. & Mathers, N. (2009) Sampling and sample size calculation. *East Midlands/Yorkshire: the National Institutes for Health Research. Research Design Service for the East Midlands/Yorkshire & the Humber,* 19(1), 117-119.

McCormick, Z. & Schmidt, D. C. (2012) *Data synchronization patterns in mobile application design*. Tucson, Arizona, the Pattern Languages of Programs (PLoP) 2012 conference, 1-14.

Molich, R. & Nielsen, . J. (1990) *Improving a human - computer dialogue.* [Online] Available at: http://delivery.acm.org/10.1145/80000/77486/p338-molich.pdf?ip=150.237.42.81&id=77486&acc=ACTIVE%20SERVICE&key=BF07A2EE685417C5%2EB052FA0818D6DC5F%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=426030571&CFTOKEN=59780192&__acm__=1410729809_58d2191880b3bc5cf49
[Accessed 11/09/2014].

Morse, J. (2000) Determining sample size,*Sage,* 10(1), 3-5.

Nielsen, J. & Mack, R. (1994) *Usability inspection methods.* NY: John Wiley & Sons.

Nielsen, J. & Molich, R. (1990) Heuristic evaluation of user interfaces. *In Proceedings of the SIGCHI conference on Human factors in computing systems*, 1-5 April, 249-256.

Nielsen, J. (1993) *Usability engineering.* 2nd edition. San Francisco,: Morgan Kaufmann Publishers.

Nielsen, J. (1993) *Usability engineering.* California: Academic Press.

Nielsen, J. (1994) *Usability engineering.* San Francisco, CA, USA: Morgan Kaufmann.

Nielsen, J. (1994) *Usability Inspection Methods.* [Online]
Available at: http://www.nngroup.com/books/usability-inspection-methods/
[Accessed 01/09/2014].

Nielsen, J. (1995) *How to Conduct a Heuristic Evaluation.* [Online]
Available at: http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/
[Accessed 11/09/2014].

Nielson, J. (2000) *Why You Only Need to Test with 5 Users.* [Online]
Available at: http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/
[Accessed 22/04/2015].

Nielson, J. (1989) *Designing and Using Human-Computer Interfoxes and Knowledge Bused Systems.*, Amesterdam: Elsevevier.

Nokia. (2008) *Getting started with mobile design*.: Nokia Corporation.

Norman, F. & Shari, L. (1998) *Software metrics: A Rigorous and Practical Approach.* 2nd edition. London: International Thomson Computer Press.

Pagliari, C. (2007) Design and evaluation in ehealth: Challenges and implications for an interdisciplinary field. *Journal of medical Internet research*, 9(2), 1-17.

pbarrett.net. (2005). *Euclidean Distance raw, normalized, and double-scaled coefficients.* [Online]
Available at: http://www.pbarrett.net/techpapers/euclid.pdf
[Accessed 29/04/2019].

Pedroli, E., Cipresso, P., Serino, S., Riva, G. & Albani, G. (2013) *A virtual reality test for the assessment of cognitive deficits: usability and perspectives*, In Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare (pp. 453-458). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Pica, D., Sorensen, C. & Allen, D. (2004) *On mobility and context of work: exploring mobile police work.* Hawaii , In 37th Annual Hawaii International Conference on System Sciences,IEEE.

Pirhonen, A., Brewster, S. & Holguin, C. (2002) April. *Gestural and audio metaphors as a means of control for mobile devices*. In Proceedings of the SIGCHI conference on Human factors in computing systems, 291-298.

Predictive Analytics Today. (2015) *Top 12 Qualitative Data Analysis Software.* [Online]
Available at: http://www.predictiveanalyticstoday.com/top-qualitative-data-analysis-software/
[Accessed 02/10/2015].

Press, Oxford University. (2016) *POWERSAVE.* [Online]
Available at:
http://www.oxforddictionaries.com/us/definition/american_english/efficient#efficient__4
[Accessed 29/04/2019].

Pressman, R. S. (2010) *Software engineering: a practitioner's approach.*:McGraw Hill Higher Education, 68.

research2guidance. (2014) *mHealth App Developer Economics 2014.* [Online] Available at: http://research2guidance.com/r2g/research2guidance-mHealth-App-Developer-Economics-2014.pdf
[Accessed 20/08/2015].

research2guidance. (2016) *mHealth App Developer Economics 2016.* [Online] Available at: http://research2guidance.com/r2g/research2guidance-mHealth-App-Developer-Economics-2016.pdf
[Accessed 24 10 2016].

research2guidance. (2017) *mHealth App Developer Economics 2017.* [Online] Available at: http://research2guidance.com/r2g/research2guidance-mHealth-App-Developer-Economics-2017.pdf
[Accessed 01/08/2017].

Risald, R., Suyoto, S. & Santoso, A. (2018) Mobile Application Design Emergency Medical Call for the Deaf using UCD Method. *International Journal of Interactive Mobile Technologies*, 12(3), 168-177.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., Grau, V., Lagos, F., López, X., López, V. and Rodriguez, P. (2003) Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education*, 40(1), 71-94.

Ruiz, J. & Lank, E. (2011) *User-defined motion gestures for mobile interaction*, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 197-206.

Rukzio, E., Broll, G., Leichtenstern, K. & Schmidt, A. (2007) *Mobile interaction with the Real World:An evaluation and comparison of physical mobile interaction techniques.* In Proceedings of the European Conference on Ambient Intelligence. Berlin: Springer.

Sauro, J. & Kindlund, E. (2005) *A method to standardize usability metrics into a single score*, In Proceedings of the SIGCHI conference on Human factors in computing systems.ACM, 401-409.

Sax, L. J., Gilmartin, S. K. & Bryant, A. N. (2003) ASSESSING RESPONSE RATES AND NONRESPONSE BIAS IN WEB IAS IN WEB. *Research in Higher Education*, 44,(4), 409-432.

Scholtz, J. (2004) Usability evaluation. *National Institute of Standards and Technology*, 1.

Schusteritsch, R., Wei, C. & LaRosa, M. (2007) *Towards the perfect infrastructure for usability testing on mobile devices,* In CHI'07 Extended Abstracts on Human Factors in Computing Systems. ACM, 1839-1844.

Seffah, A., Donyaee, M., Kline, R. B. & Padda, H. K. (2006) Usability measurement and metrics: A consolidated model. *Software Quality Journal,* 14(2), 159–178.

Shackel, B. (1981) *The concept of usability.* NY, Paper presented at the Proceedings of IBM software and information usability symposium, Poughkeepsie.

Shackel, B. (1984) *The Concept of Usability, Visual display terminals: usability issues and health concerns.* New Jersey: Prentice Hall, 45-87.

Shareef, A. F. (2006) *Special issue on "Innovation in Distance Learning.* [Online] Available at: http://lttf.ieee.org/issues/july2006/lt_july2006.pdf [Accessed 15/04/2014].

Shivade, B. & Sharma, M. (2014) USABILITY analyzer tool: ausability evaluation tool forandroid based mobile application. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS),* 3(3), 247-252

Shneiderman, B. (1998) *Designing the User Interface Strategies for Effective.* 3rd edition. Michigan: Addison-Wesley Longman.

Sivaji, A., Abdullah, M., Downe, A. & Ahmad, W. (2013) *Hybrid usability methodology: Integrating heuristic evaluation with laboratory testing across the software development lifecycle.* Las Vegas, NV, USA, In 2013 Tenth International Conference on Information, 375-383.

Smith, A. (2013) *Smartphone Ownership – 2013 Update.* [Online] Available at: http://boletines.prisadigital.com/PIP_Smartphone_adoption_2013.pdf [Accessed 30/04/2015].

Smith, S. & Mosier, J. (1986) *Guidelines for Designing User Interface Software,* Bedford: Mitre Corporation.

van Solingen, D.R. & Berghout, E.W. (1999) *The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development,* Maidenhead, England. McGraw-Hill Publishing Company.

Solutions, H. E.(2016) *Online Medical Education and Certification.* [Online] Available at: http://www.healthedsolutions.com/ [Accessed 28/04/2016].

South, South 2016 NHS Innovations. (2016) *App Development: An NHS Guide for Developing Mobile Healthcare Applications,* NHS.

Spillers, F. & Asimakopoulos, S. (2014) *Does Social User Experience Improve Motivation for Runners?.* In International Conference of Design, User Experience, and Usability Springer, Cham, 358-369.

Statista. (2018) *Number of mobile app downloads worldwide in 2017, 2018 and 2022 (in billions).* [Online]
Available at: https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/
[Accessed 20/11/2018].

Statista. (2018) *Number of smartphone users worldwide from 2014 to 2020 (in billions).* [Online]
Available at: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/
[Accessed 03/12/2018].

Statista. (2019) *Mobile App Usage - Statistics & Facts.* [Online]
Available at: https://www.statista.com/topics/1002/mobile-app-usage/
[Accessed 05/01/2019].

Statistics Solutions. (2017) *Kruskal-Wallis Test.* [Online]
Available at: https://www.statisticssolutions.com/wp-content/uploads/wp-post-to-pdf-enhanced-cache/1/kruskal-wallis-test.pdf
[Accessed 06/01/2017].

Stickel, C., Ebner, M. & Holzinger, A. (2010) *The XAOS metric–understanding visual complexity as measure of usability.* Berlin, In Symposium of the Austrian HCI and Usability Engineering Group, Springer, 278-290.

Strickland, E. (2012) *The FDA Takes On Mobile Health Apps.* [Online]
Available at: http://spectrum.ieee.org/biomedical/devices/the-fda-takes-on-mobile-health-apps
[Accessed 29/04/2019].

Tavakol, M. & Dennick, R. (2011) Making sense of Cronbach's alpha. *International Journal of Medical Education,* 53-55.

Taylor-Powell, E., Steele, S. & Douglah, M. (1996) P*lanning a program evaluation.* [Online]
Available at: http://learningstore.uwex.edu/assets/pdfs/g3658-1.pdf
[Accessed 17/09/2014].

Theofanos, M. & Quesenbery, W. (2005) Towards the design of effective formative test. *Journal of Usability Studies,* 1(1), 27-45.

Todhunter, L. D., Leach, R. K., Lawes, S. D. A. & Blateyron, F. (2017) An analysis of type F2 software measurement standards for profile surface texture parameters. *Measurement Science and Technology,* 28(6), 1-21.

Traxler, J. (2009) Current State of Mobile Learning1. In: M. Ally, (ed). *Mobile learning transforming the delivery of education and training.* Edmonton: AU Press, Athabasca University, 12-13.

Tselios, N., Avouris, N. & Komis, V. (2008) The effective combination of hybrid usability methods in evaluating educational applications of ICT: Issues and challenges. *Education and Information Technologies,* 13(1), 55-76.

Tworek, J.K., Jamniczky, H.A., Jacob, C., Hallgrímsson, B. & Wright, B. (2013) *The LINDSAY Virtual Human Project: An immersive approach to anatomy and physiology.* [Online]
Available at:
http://apps.webofknowledge.com/full_record.do?product=WOS&search_mode=Refine&qid=2&SID=W2GSUYy7skzWP1mmDtV&page=1&doc=4
[Accessed 07/05/2014].

Ventola, C. (2014) Mobile devices and apps for health care professionals:. uses and benefits. *Pharmacy and Therapeutics,* 39(5), 356-364.

Visser, B. & Bouman, J. (2012) There's a medical app for that. *BMJ Careers.*

Wagner, D., Lopez, M., Doria, A., Pavlyshak, I., Kostakos, V., Oakley, I. and Spiliotopoulos, T. (2010) *Hide and seek: location sharing practices with social media,* In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services. ACM. 55-58.

WebMD. (2014) *Medscape.* [Online]
Available at: https://itunes.apple.com/us/app/medscape/id321367289?mt=8
[Accessed 29/04/2019].

World Health Organization. (2011) *mHealth,*: Organization, World Health Organization.

Xie, D. X., Wang, R. Y. & Chinnadurai, S. (2018) Readability of online patient education materials for velopharyngeal insufficiency. *International Journal of Pediatric* , 104, 113-119.

Yeratziotis, A. & Zaphiris, P. (2018) A Heuristic Evaluation for Deaf Web User Experience (HE4DWUX). *International Journal of Human–Computer Interaction,* 34(3), 195-217.

Zahra, F., Hussain, A. & Mohd, H. (2018) Factor Affecting Mobile Health Application for Chronic Diseases. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC),* 10(1-11), 77-81.

Zhang, D. & Adipat, B. (2005) Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction,* 18(3), 293-308.

Zhenwei Qiang, C., Yamamichi, M., Hausman, V. & Altman, D. (2011) *Mobile Applications for the Health Sector.* [Online]
Available at:
http://siteresources.worldbank.org/INFORMATIONANDCOMMUNICATIONANDTE

CHNOLOGIES/Resources/mHealth_report.pdf
[Accessed 12/05/2014].

Zuse, H. (1997) *A Framework of software measurement.* Berlin: Walter de Gruyter.

Appendix 1: Ethical Approval and Interviews Questions

Medscape (Heuristic & Usability) Evaluation Interview Questions

Dear Evaluator/Participant

I ’am doing research project about improving mobile health applications, to help to do that we need to ask you some questions about the usability of Medscape applications:

We would like to thank you for your time helping us to evaluate Medscape (one of the most popular mobile health application) which will help us to structure our framework which is aiming to evaluate mobile health education applications.

The interview questions is made out of 11 questions, which is focusing on the evaluating of the Medscape among selected criteria’s from both (heuristic evaluation & usability evaluation). So we would like to ask you please to answer all of them according to your best knowledge as the data will be used for serious manner .By having this interview questions completed by you (evaluator/participant) the outcomes (data) from this interview questions will be used to help us completing structuring our framework which is the major part of our research for our PhD degree; which is aiming to evaluate mobile health education applications. By having this research successful this research will open a wide range of approaches for (mobile health education, health professionals, software developers, software houses).

We would like to ask you please to familiarise yourself with Medscape, by using it twice at least before starting answering the interview questions. And we would like to remind you that the information you will be supplying will be under full confidentiality term. Also we would like to thank you again for your kindness.

With many thanks

Tareq Aljaber

Interviews questions to specify health professionals & patients requirements by using mobile health applications (MHA), the participant will be mixed of Health professionals, patients and software developers.

On a scale of 1 to 5 where 5 is strong agree and 1 is strongly disagree:

Close

1) Do you used MHA? If the answer is 1 then have you ever used MHA?
2) Do you use MHA regularly, (To make sure they use it)?
3) Is mobile health applications ever likely to become health educations essential, (to discover how much MHA will expand in the future)?
4) Have you ever used poorly designed MHA? (To see if they met useless MHA)?

Open

5) What do you need from using MHA?

6) What do you expect from using MHA to improve your Health Education.?

7) Why do you use MHA?

8) What make MHA easy to use?

9) What effect how you choice or select a particular MHA?

10) What make MHA poor?

11) What is your experience of using MHA in your everyday life, for the purpose of health education, (open question) (A, B, C, D, E, F)?

(A) Productivity

(B) Efficiency

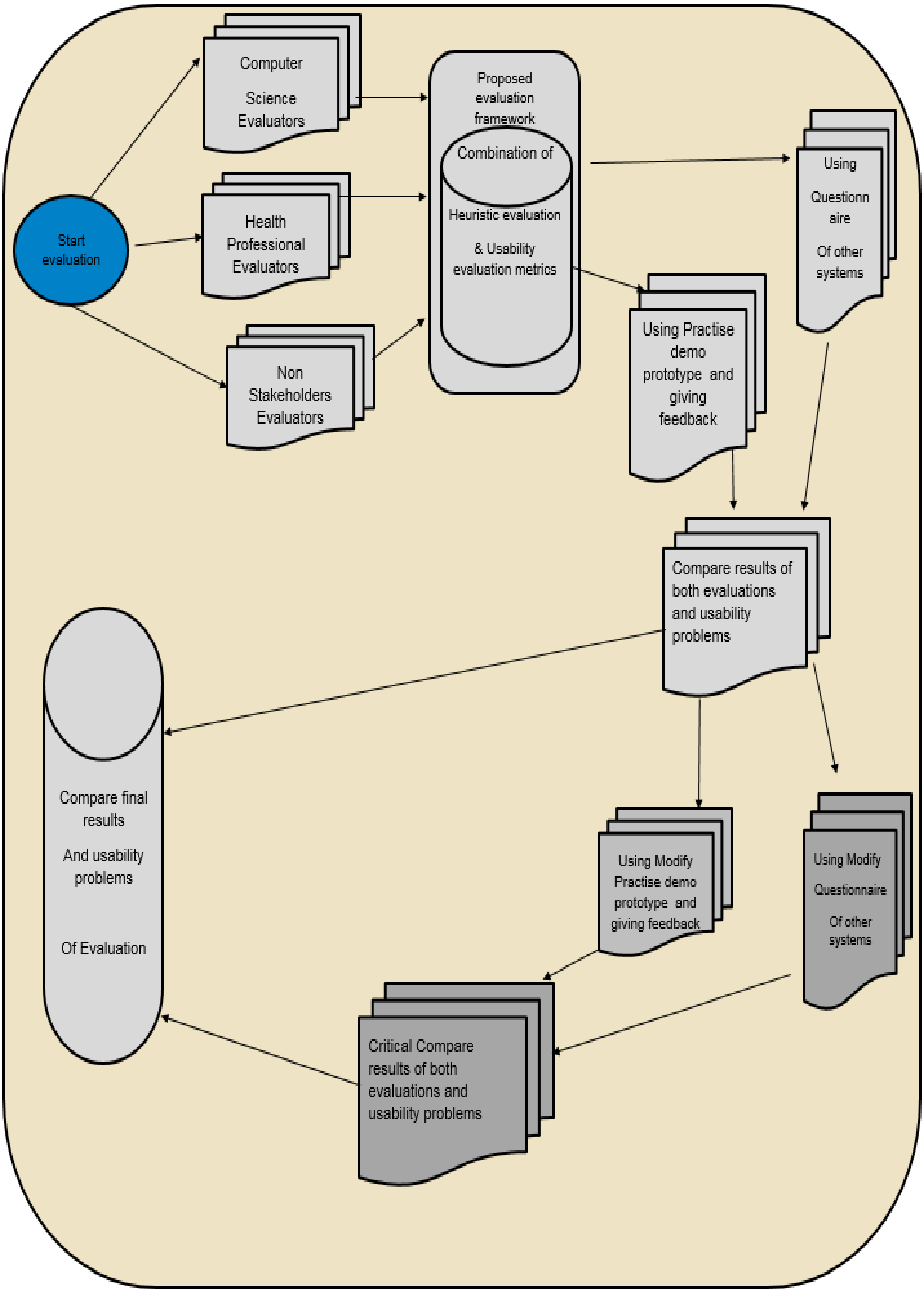
(C) Effectiveness

(D) Attractiveness

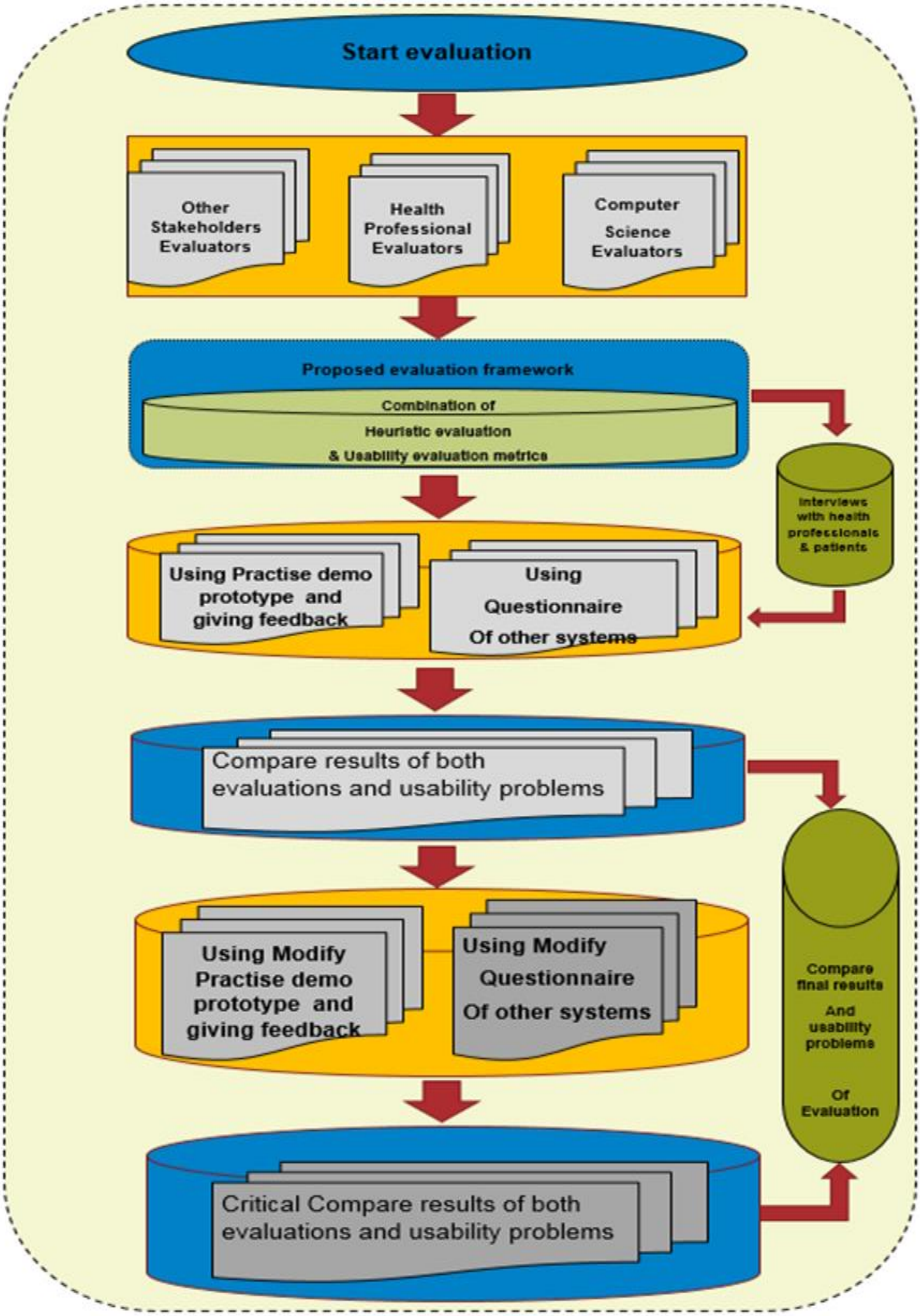
(E) Responses

(F) Time and understanding

Appendix 2: First Design for Evaluation Framework



Appendix 3: Second Design for Evaluation Framework



Appendix 4: Published Paper (for SAI Conference 2015)

Science and Information Conference 2015
July 28-30, 2015 | London, UK

An Evaluation Framework for Mobile Health Education Software

Tareq Aljaber, Neil Gordon, Chandrasekhar Kambhampati, Mike Brayshaw
Department of Computer Science
University of Hull
Hull, United Kingdom
T.ALJABER@2013.hull.ac.uk, n.a.gordon@hull.ac.uk, C.Kambhampati@hull.ac.uk, M.Brayshaw@hull.ac.uk

Abstract: **Mobile applications in general, and mobile applications for health education in particular, are commonly used to support patients, health professionals and other stakeholders. A critical evaluation framework is needed to ensure the usability and reliability of mobile applications for health education in order to save time and effort for the various stakeholders. This paper proposes a framework for evaluating mobile applications for health education. The intended outcome is to meet the needs and demands of different stakeholders and provide improvement for software engineering by creating new and more effective ways to evaluate such software. We conclude with some specific evaluation metrics that we applied in our evaluation framework: a hybrid utilizing heuristic evaluation (HE) and usability evaluation (UE).**

Keywords: ***Evaluation framework; mobile applications for health education; heuristic evaluation; usability evaluation metrics***

I. Introduction

Mobile phones now play a major role in many parts of human life. Improvements in mobile technology have allowed a wide range of applications to be developed, so that these can be used by people in various aspects of their life [1]. Everyday life in general is also becoming increasingly complex and in this context there is a focus on minimizing time wasted in nonessential activities. One of these complex elements is the large number of mobile applications that deal with healthcare or health education, which is steadily increasing.

The increase in mobile technology has led to explosive growth in the numbers of mobile health (mhealth) applications and users. Interest from both the health and software development communities has occurred as a result of the technology revolution. In 2009, the inaugural mHealth Summit, a partnership between the National Institutes of Health, the Foundation for the National Institutes of Health, and the mHealth Alliance, involved 800 people. A year later, the number of people attending the same conference reached 2,400 [2]. This demonstrates a huge increase in demand over a short period. Learning and training on health matters are increasingly important issues, particularly as more people are living longer and because of the scale and complexity of support for health and well-being [3].

This type of health education includes several areas, such as learning how to receive appropriate treatment, how to manage life in an acceptable way and how to employ and manage health demands without barriers [4]. These demands from the patient side parallel those from health professionals who seek to update their medical knowledge and are looking for specific information, as well as other health professionals' requirements.

Different healthcare scenarios/contexts demand a different approach to managing risk in both the development and use of medical applications. Mobile medical apps can pose the same risks of failure as other medical devices, due to mechanical failure, faulty design, poor manufacturing quality, and user error, among other safety issues [5].

The above-mentioned demands for the use of mobile phone apps for various requirements, such as health education, raise the absence of an evaluation framework to measure and evaluate these applications and make sure they meet the requirements, such as that of usability, of patients, health professionals and other stakeholders from several sides.

II. Software Evaluation Metrics and Methods

Previous research [6] has considered that, in general, usability is a very important condition for the utility and longevity of an application. If an app is not convenient to operate, people will not be interested in it and they will not use it, which will lead to the failure of the application. Usability metrics (usability testing) can be used to assist here, because competition is increasing in mobile app marketing so effective and usable apps are needed. Mobile apps are often developed as comparatively small projects, which do not usually support usability testing, and developers are not always able to test for - nor discover - the reasons for usability defects.

When we evaluate a user interface, we can adopt various methods, each embracing different techniques and reflecting different a priori about philosophical enquiry and technique. Broadly, we can distinguish between the following:

- Formal methods where we have some notion of correctness and can relate the design to this e.g., by a mathematical model.
- Automatically by a computerized procedure e.g., a back-box checking algorithm.
- Empirically by experiment, using differential variables and separate cases to study statistical variance.
- Heuristically by informed opinion.
- Socially or ethnographically, where a systemic approach is preferred: the whole is considered, in a social or cultural context, and empiricism avoided.

Appendix 5: Published Paper (for E-Health Conference 2016)

EVALUATION OF MOBILE HEALTH EDUCATION APPLICATIONS FOR HEALTH PROFESSIONALS AND PATIENTS

Tareq Aljaber
Department of Computer Science, University of Hull
Hull, HU6 7RX, United Kingdom
T.ALJABER@2013.hull.ac.uk

Neil Gordon
Department of Computer Science, University of Hull
Hull, HU6 7RX, United Kingdom
n.a.gordon@hull.ac.uk

ABSTRACT

Mobile applications for health education are commonly utilized to support patients and health professionals. A critical evaluation framework is required to ensure the usability and reliability of mobile health education applications in order to facilitate the saving of time and effort for the various user groups; thus, the aim of this paper is to describe a framework for evaluating mobile applications for health education. The intended outcome of this framework is to meet the needs and requirements of the different user categories and to improve the development of mobile health education applications with software engineering approaches, by creating new and more effective techniques to evaluate such software. This paper first highlights the importance of mobile health education apps, then explains the need to establish an evaluation framework for these apps. The paper provides a description of the evaluation framework, along with some specific evaluation metrics: an efficient hybrid of selected heuristic evaluation (HE) and usability evaluation (UE) factors to enable the determination of the usefulness and usability of health education mobile apps. Finally, an explanation of the initial results for the framework was obtained using a Medscape mobile app. The proposed framework - An Evaluation Framework for Mobile Health Education Apps – is a hybrid of five metrics selected from a larger set in heuristic and usability evaluation, filtered based on interviews from patients and health professionals. These five metrics correspond to specific facets of usability identified through a requirements analysis of typical users of mobile health apps. These metrics were decomposed into 21 specific questionnaire questions, which are available on request from first author.

KEYWORDS

Usability Evaluation, Heuristic Evaluation, Metrics, Health Professionals, Evaluation Framework.

1. INTRODUCTION

Mobile phones now play a major role in many parts of human life. Enhancements in mobile technology have allowed a wide range of applications to be developed, which can be utilized in various aspects of people's lives (Harrison et al., 2013). One example are the mobile health (mHealth) education applications which have been utilized by several stakeholders, such as patients and health professionals, to improve their knowledge of their own or others' health in certain aspects of their life (Aljaber et al., 2015). Everyday life in general is becoming increasingly complex and there is a need to assist patients in improving their health education to ensure corresponding improvement in their health condition by supporting their well-being via the use of mobile health education applications. Consequently, the accumulative effect of mobile technology has led to significant growth in the number of mobile health mHealth applications and users. Interest from both the health field and software development communities has been shown as a result of this technology revolution.

Appendix 6: Published Paper (for HCI International Conference 2016)

Heuristic Evaluation as applied to Serious Immersive Games and M-Learning

Mike Brayshaw, Neil Gordon and Tareq Aljaber

Department of Computer Science, University of Hull,
Hull, HU6 7RX, United Kingdom
m.brayshaw@hull.ac.uk;
n.a.gordon@hull.ac.uk;
t.aljaber@2013.hull.ac.uk

Abstract. Two fast growing areas for technology-enhanced learning are serious games and mobile instruction (M-instruction or M-Learning). Serious games are ones that are meant to be more than just entertainment. They have a serious use to educate or promote other types of activity. Immersive Games frequently involve many players interacting in a shared rich and complex – perhaps web-based - mixed reality world, where their circumstances will be multi and varied. Their reality may be augmented and often self-composed, as in a user-defined avatar in a virtual world. M-instruction and M-Learning is learning on the move; much of modern computer use is via smart devices, pads, and laptops. People use these devices all over the place and thus it is a natural extension to want to use these devices where they are to learn. This presents a problem if we wish to evaluate the effectiveness of the pedagogic media they are using. We have no way of knowing their situation, circumstance, education background and motivation, or potentially of the customisation of the final software they are using. Getting to the end user itself may also be problematic; these are learning environments that people will dip into at opportune moments. If access to the end user is hard because of location and user self-personalisation, then one solution is to look at the software before it goes out. Heuristic Evaluation allows us to get User Interface (UI) and User Experience (UX) experts to reflect on the software before it is deployed. The effective use of heuristic evaluation with pedagogical software [1] is extended here, with existing Heuristics Evaluation Methods that make the technique applicable to Serious Immersive Games and mobile instruction (M-Learning). We also consider how existing Heuristic Methods may be adopted. The result represents a new way of making this methodology applicable to this new developing area of learning technology.

Keywords: Heuristic Evaluation; Serious Games; M-Learning

1 Introduction

Throughout education, there is a growing focus ways to improve student engagement [1], which may be through utilising different pedagogic approaches [3] or tech-

adfa, p. 1, 2011.

Appendix 7: Published Paper (for HCI International Conference 2017)

A Guidance and Evaluation Approach for mHealth Education Applications

Tareq Aljaber and Neil Gordon

School of Engineering and Computer Science, University of Hull, Hull, HU6 7RX, United Kingdom
T.ALJABER@2013.hull.ac.uk
School of Engineering and Computer Science, University of Hull, Hull, HU6 7RX, United Kingdom
n.a.gordon@hull.ac.uk

Abstract. A growing number of mobile applications for health education are being utilized to support different stakeholders, from health professionals to software developers to patients and more general users. There is a lack of a critical evaluation framework to ensure the usability and reliability of these mobile health education applications (MHEAs). Such a framework would facilitate the saving of time and effort for the different user groups. This paper describes a framework for evaluating mobile applications for health education, including a guidance tool to help different stakeholders select the one most suitable for them. The framework is intended to meet the needs and requirements of the different user categories, as well as improving the development of MHEAs through software engineering approaches. A description of the evaluation framework is provided, with its efficient hybrid of selected heuristic evaluation (HE) and usability evaluation (UE) factors. Lastly, an account of the quantitative and qualitative results for the framework applied to the Medscape and other mobile apps is given. This proposed framework – an Evaluation Framework for Mobile Health Education Apps – consists of a hybrid of five metrics selected from a larger set during heuristic and usability evaluation, the choice being based on interviews with patients, software developers and health professionals.

Keywords: Heuristic Evaluation, Usability Evaluation, Evaluation Framework, Stakeholders, Metrics.

1 Introduction

Development of a wide range of applications has been enabled as a consequence of enhancements in mobile technology which can be utilized in numerous aspects of people's lives [1]. One example of these applications is mobile health education applications (MHEAs). These education applications have been utilized to improve the knowledge of different stakeholders, such as patients and health professionals, in addition to improving health [2, 3]. Patients need to develop their own health education to ensure corresponding improvement in their health, with support for their well-being

Appendix 8: Published Paper (for AHFE Conference 2017)

A Hybrid Evaluation Approach and Guidance for mHealth Education Applications

Tareq Aljaber and Neil Gordon

School of Engineering and Computer Science, University of Hull, Hull, HU6 7RX, UK
{T.ALJABER@2013., n.a.gordon@}hull.ac.uk

Abstract. Mobile health education applications (MHEAs) are used to support different users. However, although these applications are increasing in number, there is no effective evaluation framework to measure their usability and thus save effort and time for their many user groups. This paper outlines a useful framework for evaluating MHEAs, together with particular evaluation metrics: an efficient hybrid of selected heuristic evaluation (HE) and usability evaluation (UE) factors to enable the determination of the usefulness and usability of MHEAs. We also propose a guidance tool to help stakeholders choose the most suitable MHEA. The outcome of this framework is envisioned as meeting the requirements of different users, in addition to enhancing the development of MHEAs using software engineering approaches by creating new and more effective evaluation techniques. Finally, we present qualitative and quantitative results for the framework when used with MHEAs.

Keywords: Evaluation Framework · Usability Evaluation · Heuristic Evaluation · Metrics · Stakeholders.

1 Introduction

Enhancements in mobile technology have enabled the development of a wide range of applications that can be used in many aspects of people's lives [1]. An example of these mobile applications is mobile health education applications (MHEAs). MHEAs are widely utilized by a range of different stakeholders. Moreover, these different stakeholders vary in the type of knowledge and background they possess: some are specialists and some are not (mixed users). These different stakeholders need to be able to use the software easily, which is referred to as considering the usability and having an effective framework for the applications. MHEAs are utilized to develop the knowledge of health professionals and patients in order to improve health [2], [3]. Patients need to expand their own health education to ensure consistent development of their health, with support for their well-being coming with MHEAs. Learning and training around health difficulties are progressively substantial issues, mainly as more people are living longer; long life enlarges the scale and complexity of the care of health and well-being [5]. Several areas are covered by health education, such as learning how to manage life in an acceptable way, how to oversee health requirements without difficulty and how to obtain appropriate treatment [6]. In addition to requirements from

Appendix 9: Sampling Data Table Summarising the People Involved in Interviews

The table below summarises the demographic profile of the interview participants. Starting with the interviews durations: the average duration for the interviews was 14 minutes, with a minimum of 11 minutes and a maximum of 19 minutes. The total time utilized to apply these interviews were 217 minutes, and all these interviews were carried out in Jordan. The total number of participants in these interviews was 15 interviewees. Distributed as: 5 health professionals, 5 software developers and 5 patients. Regarding gender wise, they were 8 male and 7 female. All of these participants are utilizing mobile health applications for a minimum of two years for different aspects. For instance, health professionals are utilizing these applications to educate and update their health knowledge. Software developers are experienced at developing health ones. They utilize these applications to determining and developing the interface for these applications besides developing their own health education. The patients are utilizing these applications to develop their health education regarding their illness. These 15 participants sample represents a random selection of people who are utilizing the mobile health applications for a minimum of two years. The average age for these participants was 37, with a minimum age of 18 and a maximum age of 61. Furthermore, these participants were from different working background, for instance the health professionals were working in different hospitals, and the software developers were working in different software houses and the patients were from different working background. Finally, all these participants participated voluntarily to this study.

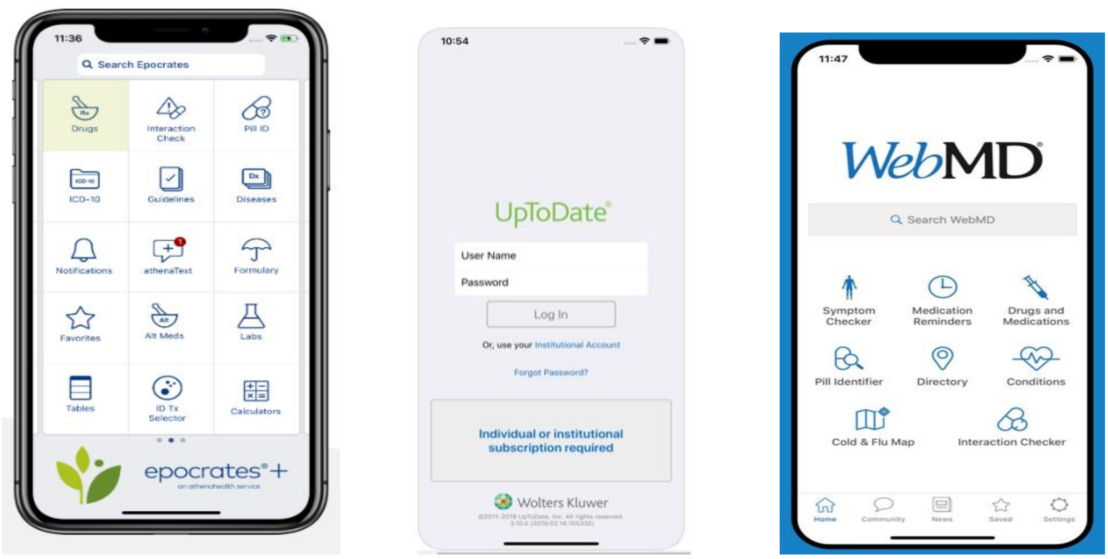
	Duration of Interview in Minuts:	Gender	Age
Average	14	7 Female/ 8 Male	37
Total	217	15	
Max	19		61
Min	11		18

Appendix 10: Example of MHEAs Considered in the Thesis

This thesis considered the use of several mobile health applications. Medscape was the main one – and described in section 3.3. Below is a summary of the other health applications considered.

Epocrates: Is one of the most popular mobile health education applications used by health professionals on mobile devices (athenahealth, 2019). It contains several features, which help health professionals in their health education and work, such as Drug Information, Interaction Check, Pill ID, Clinical Practice Guidelines, Disease Information, Formulary and other features.

Other Apps: This research considered WebMD and UpToDate and grouped them as Other apps. WebMD (Apple, 2019) is a healthcare app for Androids and iPhones devices. It provides several features used by health professionals and patients, such as support to learn about conditions and drugs, check symptoms, find doctors and specialists in your area, research treatments and diagnoses. The UpToDate healthcare app (Capterra, 2019) is also available for Androids and iPhones devices. It is a clinical decision support with medical resources and information for health professionals. It contains several features such as Patient Billing, Patient Portal, E-Prescribing and Inventory Management.



Screenshots from the 3 health apps

References for Appendix 10.

Apple, 2019. WebMD: Symptoms, Doctors, & Rx. [Online]

Available at: https://apps.apple.com/us/app/webmd-symptoms-doctors-rx/id295076329#?platform=iphone

[Accessed 10 08 2019].

athenahealth, 2019. Explore the #1 medical reference app. [Online]

Available at: https://www.epocrates.com/products/features?cid=headerExplore

[Accessed 11 08 2019].

Capterra, 2019. UpToDate. [Online]

Available at: https://www.capterra.com/p/176042/UpToDate/

[Accessed 09 08 2019].

Appendix 11: Updates since 1995 of Nielson's Heuristics

Nielsen is considered a key expert and leader in the field of Heuristic Evaluation. His heuristic approach is probably the most utilized usability heuristic approach for user interface design. His approach consists of 10 principles, which are: Visibility of system status, Match between system and the real world, User control and freedom, Consistency and standards, Error prevention, Recognition rather than recall, Flexibility and efficiency of use, Aesthetic and minimalist design, Help users recognize, diagnose, and recover from errors and Help and documentation. Wharton, et al. (1994) discuss another Usability Inspection Method, known as the Cognitive Walkthrough Method. It is a method which focusses on evaluating the design for ease of learning through exploration. One of the limitations for this approach is that, to perform a task, it evaluates each step necessary, trying to reveal any design errors that would affect learning by exploration. However, the Walkthrough analysis consist of two phases (preparatory and analysis). For the first phase - the preparatory phase - the analysts decide on the input conditions for the Cognitive Walkthrough, namely the user population, the task, action sequences for each task, and the interface that will be subjected to analysis. In the second phase, the main analytical work takes a place, throughout which the analysts work through each action of each task being analyzed. The walkthrough is a usability inspection method. It is used to evaluate a design for ease of learning via exploration. The walkthrough can be achieved after the specification of a relatively detailed design of the user interface, which arises after requisites analysis and definition of functionality of an application. It can also be executed as a paper simulation of the interface, or on a minima prototype created with a tool such as HyperCard, Visual Basic, or other similar tool. It could be performed on a fully functioning prototype of the design. In term of defining the input to the walkthrough method, four areas are essential to agree upon before the walkthrough analysis begins: Who will be the users of the system? What task (or tasks) will be analyzed? What is the correct action sequence for each task and how is it described? How is the interface defined?

Another set of cognitive engineering principles for enhancing human-computer performance was developed by Gerhardt-Powals (1996). Her principles are similar to Nielsen's heuristics but take a more holistic approach to evaluation. Her principles are: Automate unwanted workload, Reduce uncertainty, Fuse data, Present new information with meaningful aids to interpretation, This approach uses names that are conceptually related to the function, that group data in consistently meaningful ways, limit data-driven tasks, include in the displays only that information needed by the user at a given time and provide multiple coding of data when appropriate and Practice judicious redundancy). Her approach involved obtaining a set of significant cognitive design principles from the literature, utilizing them to develop an interface, then evaluating its performance. Her principles theorised that the interface cognitively engineered in this way is superior in satisfaction, performance and workload compared to interfaces that are not cognitively engineered. Weinschenk and Barker (2000) developed their classification for heuristic by creating a categorization of heuristics and guidelines through several major providers into twenty types which are: (User Control, Human Limitations, Modal Integrity, Accommodation, Linguistic Clarity, Aesthetic Integrity, Simplicity, Predictability, Interpretation, Accuracy, Technical Clarity, Flexibility, Fulfilment, Cultural Propriety, Suitable Tempo, Consistency, User Support, Precision, Forgiveness and Responsiveness). However, Nielsen heuristic still considered the most utilized Heuristic Evaluation. As well as Nielsen established his own company in 1998 called Nielsen Norman Group.

References for Appendix 11.

Wharton, C., Rieman, J., Lewis, C. and Polson, P., 1994. Usability inspection methods. *J. Nielsen & R. Mack (Eds.),* pp.105-140.

Gerhardt-Powals, J., 1996. Cognitive engineering principles for enhancing human-computer performance. International Journal of Human–Computer Interaction, 8(2), pp. 189-211.

Weinschenk, S. & Barker, D., 2000. Designing Effective Speech Interfaces. s.l.:John Wily.

Appendix 12: Explanations for the Filtering Through the Third Design

The third design of the framework is starting the evaluation by the first step, which was the proposed selected hybrid metrics from HE and UE. The second step was converting the proposed metrics to interviews, followed by applying them to three categories of stakeholders which were:(health professionals, patients and software developers). The third step was mapping the outcomes from the interviews (the accurate metrics) with the questionnaire questions, in term of generating the questionnaire questions. The fourth step was utilising the questionnaire on existing mobile health applications, followed by applying the questionnaire to three categories of stakeholders which were:(health professionals, patients and software developers), which was the fifth step in the framework. The sixth step was analysing the questionnaire to rank mobile health application and to design the MAST Tool in term of matching between different mobile health education applications and different stakeholders. Step seven was evaluating the MAST Tool to determine the impact behind the MAST Tool and the framework itself, to justify to step eight, which is the final design for the MAST Tool.