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

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.
The inaugural mHealth Summit in 2009, a partnership between the National Institutes of Health, the Foundation for the National Institutes of Health, and the mHealth Alliance, involved 800 people. One year later, the number of people joining the conference was extended to 2,400 (Qiang, 2011). This indicates a significant increase in interest 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 (Shareef, 2006). This type of health education consists of several areas, such as learning how to receive appropriate treatment, how to manage life in an acceptable way and how to manage health requirements without barriers (Glanz et al., 2008). These requirements from the patients’ side parallel those of the health professionals who seek to update their medical knowledge and are looking for specific information. According to Istepanian and Lacal (2003), “M-health was defined as wireless telemedicine involving the use of mobile telecommunications and multimedia technologies and their integration with mobile healthcare delivery systems”.

As mobile phone devices are small, they have various restrictions, such as battery life and resource storage limitations, which bring a lack of quality to the service and the user experience (Silva et al., 2013). Mobile health apps can create risks of failure equivalent to those of supplementary medical devices, due to mechanical failure, faulty design, poor manufacturing quality, and user error, among other safety issues (NHS Innovations South East, 2014). One consequence of these weaknesses is that most software developers of mobile health education apps do not apply an appropriately specific model/framework in the evaluation of usability for these apps to ensure that they meet certain requirements, such as health education for patients and health professionals. This highlights the need for an evaluation framework to measure and evaluate these apps and make sure they meet the requirements of the stakeholders (Aljaber et al., 2015).

2. THE IMPORTANCE OF MOBILE HEALTH EDUCATION APPS

Recent research has demonstrated that one of the central development areas in computing is mobile phone devices (Ziefluence & Bay, 2008). According to Hernandez Munoz and Woolley (2009), mobile phones have become essential and common devices among the general population, which suggests that people depend on mobile phone devices in many aspects of their daily lives, for instance in commerce, transportation and health education. The overall trend is that people are keen to use mobile phones for purposes other than the conventional aspects of using them as flexible replacements for desktop computers. (Traxler, 2009). A wide range of types of information can now be accessed from anywhere at any time to anybody with a mobile phone using wireless technology (Ally, 2009).

In general, there are two types of mobile phone device. The classic type, which contains two parts: a display screen and a keyboard below; and the touch screen type, which is split into either a full touch screen controller or half touch screen controller and a half keyboard under the controller. The first touch screen mobile phone was the Simon Personal Communicator, developed by IBM in 1992, while the first smart mobile phone on the market was the Ericsson R380 released in 2000 (Lobo et al., 2011). Seven years later, the iPhone was released by Apple and was the first smartphone mainly to be controlled by its touch screen (Lobo et al., 2011). Only one year later, the Android operating system was on the market for touch screen mobile phones (Lobo et al., 2011). According to Leonardi et al. (2010), a touch screen set-up is more flexible and easy to use by novice users than the keyboard and mouse set-up, which supports the expectation of increasing reliability on mobile phones.

In addition, a recent estimate has shown that one person in every five in the world owns a smartphone (Heggestuen, 2013). Recent estimates indicate that about 497 million mobile devices were added to the circulation in 2014 and that, of these devices, smart devices represented 26% (Cisco Visual Networking Index [VNI], 2015). According to Hussain et al. (2015), the latest estimates point to the number of mobile devices being enlarged to 1.4 billion in 2017, as reported by the Cisco Global Mobile Data Traffic Forecast 2013 Update (Hussain et al., 2015). This reflects a vast growth in mobile devices, which increases the potential for having more applications installed and used in our mobile phone devices. This brings to the fore the lack of a method for ranking and distinguishing useful apps from the not so useful.

3. USABILITY METRICS
Initially, we proposed a hybrid selection of metrics taken from heuristic evaluation (HE) and usability evaluation (UE) metrics to be filtered by conducting 10 hybrid selected interviews with health professionals and patients. The outcome from analysing the data from the 10 hybrid selected interviews demonstrated that we needed to modify some of the metrics we had selected, as some of the metrics we had proposed and selected did not meet the requirements of the interviewees. The 10 hybrid selected interviews with health professionals and patients reflected that some of the proposed hybrid metrics selected from HE and UE were not essential to their requirements, although some of the metrics needed to be supplementary to our proposed hybrid selection of metrics taken from HE and UE. Moreover, we systematized and categorized our hybrid selection metrics from HE and UE as: A) Memorability, B) Features, C) Attractiveness, D) Simplicity (containing learnability) and E) Accuracy. The definitions of these metrics are:

Memorability: When users have not worked with the interface for a long time, how much of its functionality do they remember and how quickly can they regain their proficiency (Shivade & Sharma, 2014)?

Features: Does the app provide extra features, such as images, video, voice support/help, touch screen facilities or voice guidance (Hussain et al., 2012)?

Attractiveness: The capability of the software product to be attractive to the user, e.g., through the use of colour or graphic design (International Organization for Standardization/International Electrotechnical Commission [ISO/IEC] 9126-1, 2001).

Simplicity: Ease of inputting the data, ease to use output, ease to install and ease to learn (Hussain et al., 2012).

Accuracy: The degree to which it is accurate (Hussain et al., 2012). The capability to provide correct results or effects (ISO/IEC 9126–1, 2001). Figure 1 shows these metrics.

![Figure 1. A hybrid of five selected metrics](image)

### 4. SOFTWARE EVALUATION METRICS AND METHODS

Previous research, such as Shivade and Sharma (2014), has considered that, as a rule, usability is essential 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 can be utilized to help because competition is increasing in mobile app marketing and effective and usable apps are needed. Mobile apps are frequently developed as reasonably small projects, which do not usually support usability testing. Moreover, developers are not always able to test mobile apps.

When we need to evaluate a user interface, we can adopt numerous methods, each method employing different assumptions and reflecting different a priori philosophical approaches and techniques. Broadly, we can decide among the following: first, formal methods, e.g., using a mathematical model; second, automatically, such as a black-box checking algorithm; third, empirically, which is by experiment, using separate cases to study statistical variance; fourth, heuristically, which is by informed opinion; and, finally, socially or ethnographically, by which a systemic approach is preferred and the whole is considered, in a social or cultural context, and empiricism avoided.

In the context of this paper, we need a practical evaluation methodology that we can apply to the present mHealth setting. Formal or automated approaches are not accessible with regard to the present work. Empirical work is prized when the occasion allows, although the quickly advancing and technologically challenging setting of the current work means that other additional pragmatic and heuristic tools are adopted here.

In our paper we conduct a hybrid approach selected from HE and UE.
First, HE is an informed method of usability analysis in which a series of domain experts are presented with an interface solution and mandated to use their specific expertise to comment upon it. The method uses a series of heuristics or rules of thumb about what generally constitutes good practice in interface design (Nielsen & Molich, 1990). There are several advantages to utilizing HE: first, HE is cheap; second, it is easy to motivate people to do it; third, it does not require advanced planning in the sense that an experimental evaluation would do; and, finally, it can be used at every stage of a project e.g., from initial design, to rolling prototype, to final project delivery (Brayshaw et al., 2014).

The most established approach is found in Nielsen (Nielsen & Molich, 1990; Nielsen, 1993), in which he proposes 10 rules concerning good design, which include consistency and standards, error prevention, flexibility and efficiency of use, and aesthetic and minimalist design. Implementing the 10 rules, however, comes at a price and we must be watchful when utilizing a method or technique such as heuristic evaluation. Science in its most empirical structure depends on solid replicable results and a heuristic methodology is not going to convey such exploratory exactitude. We have to be aware of two points: first, some of the time, indicating usability issues without supplying direct suggestions related to solving these issues; and, second, HE is hard for a single individual to do, since one person will never be able to discover all the usability problems in an interface.

Fortunately, experience gained from several different projects has shown that different people find different usability problems. Subsequently, it is possible to enhance the effectiveness of a method significantly by involving multiple evaluators (Nielsen & Molich, 1990). Given the rapidly advancing nature of the mobile application context, HE provides both adaptability and high quality evaluation to advise our research. For health education, it is likely that consistency and the avoidance of mistakes will be key.

Second, UE metrics and methods. Usability is a very broadly held metric for evaluating the learnability and attractiveness of software (Shivade & Sharma, 2014). UE is a critical approach to assessing the function of human-computer interaction (HCI). According to Nielsen (1993), usability is defined as follows. Learnability: the user should be able to find the system easy to learn. Efficiency: the user should be able to use the system efficiently and, once learned, the user must have the highest level of productivity potential. Memorability: the user should be able to remember how to use the system easily, so that if the user does not use it for any length of time, he or she should still be able to use the system without learning everything again. Errors: the user should make few errors during the use of the system, so it should have a low error rate so that the user can easily recover from them. Catastrophic errors must not occur. Satisfaction: the user should be satisfied with the system. Software UE is a methodology for evaluating the usability of each kind of software, including mobile applications (Shivade & Sharma, 2014). Usability metrics include response time, accuracy, operability, attractiveness, learnability, satisfaction, productivity, effectiveness and efficiency (Seffah et al., 2006).

5. RATIONALE FOR CONSTRUCTING A FRAMEWORK

There are at least four major reasons for constructing a framework for mobile health education apps. First, in recent decades, several researchers developed UE for systems in general, although the dominant part was for web applications by and large, not for mobile apps or mobile health education apps (Alva et al., 2003; Ivory & Hearst, 2001). Second, recently usability turn into a significant area for smartphones, as it is vital to avoid issues from practicing mobile applications (Baharuddin et al., 2013). Third, according to Constantinos and Kim (2011), the lack of empirical research into the impact of the environment on mobile usability and the significance of user characteristics is a central point. In addition, they specify in their paper that there has been no subjective study on the usability scopes considered in such mobile studies (Constantinos & Kim, 2011). This demonstrates the absence of the construction of a subjective study to enhance, rank and measure the usability of mobile apps in general, and mHealth education apps specifically. Fourth, and most importantly, according to Constantinos and Kim (2011), there is no usability evaluation framework that has yet been created in the context of a mobile computing environment. This provides support for the necessity of our research, which involves building an evaluation framework for mHE education apps. Furthermore, according to Smith (2013), smart mobile phone device ownership saw a progression of 10% between 2012 and 2013. The use of mHealth applications began to increase markedly from 2013: over 36% of all mHealth applications in 2014 had been released in the previous year (research2guidance, 2014). According to Dubey et al. (2012), usability
inquiries utilize different methods, such as surveys, logging, or interviews. Moreover, UE methods are one of the top approaches in terms of usability to evaluate different aspects, such as systems, applications and prototypes (Bernhaupt, 2009).

6. AN EVALUATION FRAMEWORK FOR MOBILE HEALTH EDUCATION APPS

Initially, we designed our proposed framework to evaluate and measure the usability in mHealth education applications. We had an early design which we then updated but, as time spent on the research passed, we discovered that our framework required a number of amendments to make it more efficient. The early framework design consist: The first step for in early design was based on building and mapping questionnaire questions, based on a hybrid selection from HE and UE. Second, the questionnaire questions had to be used on existing mHealth education apps, such as Medscape. Third, we developed a basic prototype for the purpose of reflecting on basic mHealth education apps. Fourth, we applied the questionnaire to a basic prototype. Fifth, we compared the result between the outcomes from the questionnaire applied to Medscape and the outcome from the questionnaire applied to the basic prototype. Six, we modified the basic prototype depending on the comparison outcome from step 5. Seven, we used the questionnaire in a new basic prototype and, finally, compared the result from this with the results from Medscape in step 2. These steps and the proposed framework are illustrated in Figure 1 (Aljaber et al., 2015).

The updated framework design: The updated framework design depended on the same principle as the early design for the framework, but we modified step 1 of the early design (the building and mapping of questionnaire questions based on a hybrid selection from HE and UE) and mapped it instead with accurate requirements of
patients and health professionals, as they are the most frequent users of these apps. In addition to meeting these specific requirements, a hybrid of five selected metrics from a larger set of heuristic and usability evaluation was filtered from the output of 10 interviews with patients and health professionals. According to Nielson (2000), dealing with five test users is sufficient to obtain around 80% of the results. In our research, we dealt with 10 users, which should bring our results to more than 80% of the accurate requirements for health professionals and patients using mHealth education apps (Nielsen, 2000). The five metrics selected 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. (The updated design for the framework is shown in Figure 2.)

7. INITIAL RESULTS FOR THE FRAMEWORK

After analysing the feedback data collected from the 10 interviews (five interviews with health professionals and five with patients), we analysed the feedback from both sets of interviewees to address their most important requirements when using mHealth education applications. Our results are shown in Figure 3 and are detailed below.

The most important metric to be measured in meeting patients’ and health professionals’ requirements is D) Simplicity, shown as having the highest score with an average of 6.9. In second place is B) Features, with an average of 4.6. Third is E) Accuracy, with an average of 3.2. In fourth place is A) Memorability, with an average of 2.4. In last place is C) Attractiveness, which has the lowest score with an average of 0.3. From these results, we can state that the most important metrics for health professionals and patients when using mHealth education applications are D) Simplicity and B) Features.

![Average requirements for the five metrics](image)

Figure 3. Results from the interviews measuring patients’ and health professionals’ requirements.

To further explain the results we gained from analysing the data collected from the 10 interviews, we have summarized our findings in Table 1. The numbers in the table cells reflect the number of times an interviewee (a health professional or a patient) mentioned a particular metric. For example, interview 1 with a health professional reflects that the interviewee mentioned A) Memorability in his interview five times, B) Features five times, C) Attractiveness once, D) Simplicity 11 times and E) Accuracy 10 times. An overall summary of the results, by taking the aggregated score of each of the five metrics across the 10 interviews, is given below:

The most important metric to be measured in terms of meeting the health professionals’ and patients’ requirements is D) Simplicity, showing the highest score (69). In second place is B) Features (46). Third is E) Accuracy (32), in fourth place is A) Memorability (24) and the lowest number was given to C) Attractiveness (3).

From these results, we can state that the most important metrics for health professionals and patients are D) Simplicity and B) Features in meeting their main requirements when using mHealth education applications.
8. CONCLUSION

Health professionals and patients utilize mHealth education applications for the critical reasons of saving time and effort and to simplify their understanding of matters connected with health education. If such a mobile health education application does not offer clients these components, the application will not be utilized effectively as part of the wider goal of satisfying the prerequisites of health professionals, patients, and different partners. What is more, this will not encourage an application to be set up in the first place. An effective way to avoid these issues is by employing HE and UE metrics. This will help to improve the memorability, features, attractiveness, simplicity (including learnability) and accuracy of mHealth education applications, as well as reducing the potential for the failure of these kinds of apps. Moreover, a mHealth education application can be enhanced by evaluating it in the early stages of software design. In this case, our hybrid of five selected usability metrics - Memorability, Features, Attractiveness, Simplicity (containing learnability) and Accuracy - can be improved to increase the chance of these apps being successful and useful. This type of framework could be utilized as an evaluation tool for various sorts of mobile application. Moreover, mHealth education applications are critically important for health professionals and patients, as utilizing low-quality applications could, in extreme cases, lead to the death of patients.

This paper concentrated on highlighting the building of an evaluation framework for mHealth education applications that depends on a hybrid selection of metrics (heuristic evaluation and usability evaluation) and incorporated a hybrid of five selected metrics from a larger set of heuristic and usability evaluations, filtered based on interviews with patients and health professionals. These five metrics correspond to particular aspects of usability identified through a requirements analysis of typical users of mHealth apps. These metrics were decomposed into 21 specific questionnaire questions. The aims of this project are to build a set of tools for the evaluation of mHealth apps to enable them to be ranked, and to assist several of the stakeholders (health professionals, patients and others) in selecting appropriate apps. Our Evaluation Framework for Mobile Health Education Apps is intended to offer guidance on system specifications and different user requirements for software developers of mHealth apps.

REFERENCES


