A METRIC-BASED EVALUATION MODEL FOR APPLICATIONS ON MOBILE PHONES

Azham Hussain[1], Nor Laily Hashim[2], Nazib Nordin[3] and Hatim Mohamad Tahir[4]

[1-4]School of Computing, UUM College of Arts and Sciences, Universiti Utara Malaysia 06010 Sintok, Kedah, Malaysia

Corresponding author: azham.h@uum.edu.my[1]

ABSTRACT

A number of methods and measures have been used by researchers in software evaluation. Usability metrics is one of the approaches that has been used as a guideline to evaluate the quality of the system in many mobile applications. However, the metrics used for evaluation method keep changing due to new inventions of mobile phones. Thus, there is a need to create a dynamic model for evaluation that can grow together with new inventions and technology. In this paper, a dynamic usability metrics model for evaluation of mobile applications was designed. The model comprises usability goals, questions and metrics for evaluation of applications on mobile phones. To ensure that the model is reliable and effective, a usability study was conducted on two applications installed in different mobile phones. this model proves to be applicable for evaluation of mobile phone applications whereby its questions and metrics can be dynamically changed accordingly to comply with the requirements of the evaluators.

Keywords: Usability, goal question metric, mobile application, iPhone, O2 orbit.

INTRODUCTION

Usability evaluation has grown into a well-established research area. The first guidelines to evaluate the application system as well as today's mainstream is ISO 9241 – 11 standards (ISO, 1998). For several years, this research area which has focused on generic usability metrics has been countered by others who argue in favour of using specific usability metric. The discussion of this difference between generic and specific has mostly been a matter of opinions, and it has not been prominent in the literature on the comparison of usability metric, e.g. Bertoa, Troya and Vallecillo (2006) and Ahmed, Mohammad, Rex and Harkirat (2006).

New inventions and current features on mobile phones will reflect existing evaluation metrics and guidelines. An application that uses a built-in digital compass on mobile phones can find a location quickly and accurately by determining the direction as well as show the user the direction he or she is facing. on the other hand, the GPS receiver inside the mobile phones will benefit many users by using an application linked to the GPS receiver, for example, SatNav application and find-me application. All these applications require new metrics for evaluation method. Published work on the method(s) used to evaluate these applications with these new features on mobile phones are very limited and focus only on mobile devices and not the application itself.

A number of evaluation methods and tools are readily available to examine software usability. However usability measures that are specifically intended for mobile devices are very limited indeed. Limitations on current measures to evaluate mobile applications include: 1) their ability to generalize to other domains, 2) the focus on mobile devices instead of applications and, 3) they are not designed to measure applications that use new features on mobile devices. In addition to limited usability measures, evaluation has become more challenging due to the unique features of mobile devices such as limited bandwidth, unreliability of wireless networks, changing mobile context (e.g., location), small screen size, and limited memory.

There are a number of models for usability measurement; for instance, Quality in Use Integrated Measurement (QUIM) developed by Ahmed et al. (2006). QUIM is a consolidated model for usability measurement and metric; It is also appropriate for a user who has little or no knowledge of usability. The model consists of 10 factors, which are subdivided into 26 criteria. For the measurement of these criteria, the model provides 127 metrics. The model is used to measure the actual use of working software and identify the problems. However, the model is not optimal yet and needs to be validated. Many current models and methods which aim to evaluate usability still have some limitations, for instance, they are not intended for developers who are not familiar with the field of Human Computer Interaction (HCI) and thus, are difficult to apply.

According to a report from the International Telecommunications Union (ITU), mobile phone users worldwide will be over 50 billion by the end of 2011. The increasing number of mobile users means that many businesses have deployed mobile applications to gain an advantage. Applications in mobile devices such as Mobile Facebook, Skype, Amazon and SatNav systems have become popular and well-accepted. The fast growth and high demand for mobile applications have attracted researchers to extend studies on any potential area in mobile applications. One of the most important research issues in HCI is how to measure usability (Hornbaek & Law, 2007). Measuring usability is an essential task to ensure the accuracy of the application and safety of the user from strain injury (Ahmed et al., 2006). Focusing on usability and user experience are key elements in creating successful, high-quality applications. Unfortunately, there are few clear guidelines on how various definitions of usability factors, rules and criteria are related and how the usability of mobile applications are measured. Instead, many developers tend to employ usability methods that they are familiar with, and some of the methods may not be appropriate to all applications. The ISO 9241 – 11 Guidance on Usability (Effectiveness, Efficiency and Satisfaction) is the most referred to document to measure usability. However, it is general and very difficult to apply to specific domains (Bertoa et al., 2006); Ahmed et al., 2006).

RELATED WORKS

HCI is a multi-disciplinary subject involving not only ergonomics and software engineering, but also cognitive psychology, cognitive science, social psychology, mathematics, organizational psychology, artificial intelligence, computational linguistics and sociology. In terms of current and potential research on HCI, Benbasat (2010) suggests that there is a potential for research to be conducted on HCI in three aspects, which include social networks, online services and 'user base' expansion. Applications on social networks have become the highest in terms of user demand and provide a rich possibility for HCI research. Various entities (such as person to person, person to technology, person to group) involved in social networks make interesting research in this area. The different contexts for online services, such as B2C, B2B and e-government, provide opportunities for researchers to explore. In addition, online services usually involve different types of users (professionals, the elderly and beginners). Benbasat (2010) also mentioned that to expand HCI research that focuses on certain users into a wider perspective of users, for example IN designing interfaces for elderly users, will prove to be more challenging, particularly in areas of e-government and health care. Technology should be treated as more than a static, objective, tool-like entity while evaluating interfaces. However, Benbasat (2010) did not include mobile HCI as an interesting topic to his study although it can be seen that mobile devices and applications are used by all these groups in all of the above domains. Furthermore, Lyytinen (2010) discusses three issues and challenges in HCI research: 1) concern for environmental validity, 2) richer notions of cognition, and 3) growth and access to new sets of data. Lyytinen claims that these three issues are essential to the information systems field but the interface on mobiles is also important to information systems research and should have been included in his discussions.

A mobile device is a pocket-sized computing device, usually having a display screen with touch input or a tiny keypad. This study used the term ‘mobile phone’ to represent all types of smart mobile devices similar to Ali Kattan, Rosni, Rosalina and Sureswaran (2009). Research on mobile phones that focus on interaction has emerged over the last decade. The early stages of research into mobiles (around 2001) only focused on a small area, for example on context aware systems, discussions on design issues and investigations into interaction patterns. Current research on mobiles focuses on numerous aspects due to the expansion in mobile phones technology. Previous research on phones is still relevant but various areas have been explored. A few limitations on mobile phones, for example the small memory capacity and limited connectivity, have now been overcome by the latest mobile phones and anytime everywhere network connections (Wi-Fi or 3G). However, this situation has given rise to a new and interesting area for research into the short battery life due to long network connection.

Current research on mobiles has emerged in many areas, such as in mobile accessibility, social media and networks, maps and navigation, activities in a mobile context, multimodal interaction and in mobile evaluation. In mobile accessibility, most of the current research focuses on how mobile devices can be used by the disabled community. Fabio, Luca, Elio and Marco (2010) developed an application to investigate communication between emergency medical responders and deaf patients. They also evaluated the system but used only three evaluation characteristics (learnability, effectiveness and understandability) without any evaluation on mobile characteristics. A similar condition occurred in a study by Tiago, Hugo, Joaquim and Daniel (2010) and Luca and Alessandro (2010) with only one evaluation characteristic (error rate).

Usability evaluation research in HCI has also emerged and gained importance in the area of mobile devices (Rudy, Carolyn & Mark, 2007). Several researchers engaged in mobile device evaluation include Rita (2006) who conducted studies on the accessibility and usability of mobile devices in the domain of healthcare. She outlined usability guidelines for designing and testing mobile devices used by older adults. In contrast, Jun and Tarasewich (2004) outlined general usability guidelines for mobile devices. Existing interface guidelines by Shneiderman (1998) have become a starting point to propose a set of practical design guidelines for mobile device interfaces. The guidelines can be used for designing interfaces for mobile devices but not to evaluate the interface. Henry Vivien and Chee (2008) made a comparison between laboratory and field tests to evaluate the usability of mobile devices. In the results section, they claimed that there were many more types and occurrences of usability problems found in the field compared to the laboratory test. Some of these problems were related to the devices being used in the field while others included external factors associated with the environment of use, such as noise, privacy of using the device in a crowded place, mental and physical resources, and the extent to which users were affected by these factors. Another limitation to mobile devices is the difficulty of data entry in field work. To overcome this limitation, Kondratova and Goldfarb (2006) explored the use of speech recognition technology on mobile devices in order to improve the usability of mobile applications. The combination of voice XML and XHTML will enable the human voice to be the input in mobile applications. The results show that the system can afford a speech recognition accuracy level of 97% for numeric data. The latest voice recognition features (Siri) introduced by Apple Inc. via iPhone 4s have the ability to use the voice to send messages, schedule meetings, make phone calls, and many more. This feature may affect the usability of mobile applications.

EVALUATION OF MOBILE APPLICATIONS

Evaluation has grown into a well-established research area. The first rule to evaluate the applications system is ISO 9241 – 11 standards which, in fact, is today's mainstream. For several years, these guidelines focus on the generic usability metric that has been countered by others who argue in favour of using the specific usability metric. Evaluations based on user perception are accomplished by developing procedures for capturing the problem that users have while trying the software system. The results on user perception are based on the actual views of valid users on interface problems. However, user evaluation is expensive and time consuming. Expert-based evaluation is similar to design reviews of software projects while model-based evaluation is based on predictions of performance from the model. As expert-based and model-based evaluations are less expensive, many interface designs can be tested. Whether the evaluation is user-based, expert-based or model-based, there is no agreement yet in the community about which evaluation method is the most practical (Scholtz, 2004).

The evaluation on mobiles is slightly different from desktops due to the unique features of mobile phones such as limited bandwidth, small screen, changing mobile context (e.g. location) and limited memory. However, many studies which proposed methods and guidelines for evaluations on desktop applications may not be directly applicable to mobile applications (Constantinos & Dan, 2007; Zhang & Adipat, 2005). Examples of studies on mobile evaluation include those by Marco and Luís (2008) and Marco et al. (2008) who proposed design guidelines and evaluations for mobile applications, Jun and Tarasewich (2004) who outlined user interface guidelines for mobile devices, Jin and Ji (2010) who provided usability evaluations on physical user interfaces for mobile phones but not for applications, and Gafni (2009) who focused on usability issues on mobile wireless systems. The above studies did not include usability metrics in the evaluation part. However, the study by Gafni (2008) discusses mobile quality metrics and focuses on software issues while providing quality metrics, for example, “Display load” and “Network throughput”, but do not deal with usability.

Usability metrics are usually used by evaluators to identify what they are going to measure. Among the popular usability measures is ISO 9241 part 11, guidance on usability as shown in Table 1. This measure has been used widely but is too general, difficult to apply to specific domains and does not associate with any quality characteristic to the measure (Ahmed et al. 2006; Bertoa et al., 2006).

Table 1

Usability Measures

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

Source: Adopted from ISO (1998).

There are several challenges in evaluating mobile applications. Today, multimodal mobile applications are emerging. Multimodality combines voice and touch via a keypad or stylus as inputs with relevant spoken outputs (users are able to hear synthesized, pre-recorded streaming or live instructions, sounds and music on their mobile devices) and on-screen visual displays in order to enhance mobile users' experience and expand the network operator services offered. A combination of multiple access channels provides new avenues of interaction to users, but poses dramatic challenges to usability evaluation.

There are two methods usually employed by evaluators to evaluate mobile applications. A survey by Kjeldskov and Stage (2004) shows that 71% of all mobile HCI evaluations were performed in the laboratory while only a few conventional methods were customized to meet the rising challenges of mobile application evaluations. This may be due to difficulties faced in data collection techniques such as think aloud, video recording or observations. Evaluation in the lab has many advantages such as controllable conditions in situations and reproduction. However, the drawback of lab experiments is the lack of realism.

The similarities and differences between fields and lab-based evaluations of mobile applications have begun to be explored. Several comparisons have been made to observe the differences in interaction behaviors in the laboratory and field settings, for example, studies by Baillie (2003) and Pirhonen, Stephen & Chirstopher (2002). They conclude that it is worthwhile to carry out evaluations in the field, even though it may be problematic due to difficulties in capturing screen contents and the interaction between the users and mobile devices.

METHOD

The model was developed by analyzing the journal articles related to HCI. A total of 409 journals articles were reviewed based on the keywords, "usability", "evaluation" and "metric". Only 26 out of 409 journals articles were selected for further review in obtaining the guidelines for mobile application development. Table 2 summarizes the journal articles that were reviewed.

The review is based on a conception of usability, similar to Hornbæk (2006). This conception merely discusses studies related to usability evaluation instead of the broad concept of usability. The quality characteristic of each measure was analysed to ensure there is no duplication. The measures were refined to simplify the guidelines and ensure that the model is not too complex. interestingly, the findings of this study depict that most of the studies employed effectiveness, efficiency and satisfaction as quality characteristics. Thus, this study decided to use these three characteristics as a basis for the guidelines while other characteristics were considered as sub-guidelines. From 17 popular guidelines, six usability characteristics were selected as the become a goals for the model shown in Table 3.

Table 2

Journal Papers/Articles Reviewed

Journal	Year range	No of articles reviewed	No of journal articles selected
TOCHI	2006-2011	54	8
HCI	2006-2011	36	2
IJHCI	2006-2011	97	5
IJHCS	2006-2011	222	11
Total		409	26

Table 3

Usability Characteristics

Quality characteristic	Goal	Guidelines
Effectiveness	Simplicity	-Ease to input the data -Ease to use output -Ease to install -Ease to learn
	Accuracy	-Accurate -Should be error-free -Successful
Efficiency	Time taken	-To respond -To complete a task
	Features	-Support/help -Touch screen facilities -Voice guidance -System resources info. -Automatic update
Satisfaction	Safety	-While using the application -While driving
	Attractiveness	-User interface

Goal Question Metric (GQM) approach by Basili, Caldeira and Rombach (1994) was employed to develop the model because the approach allows the model to be enhanced at any time. For the first step, the goals from Table 3 were entered into the model. Second, questions were designed to assess each goal described in the first step. The questions were designed carefully by refining each goal into several questions which are measurable. Finally, a set of metrics were developed to provide the information to answer these questions. In this case, all the questions were refined into metrics. The model below shows that it consists of 17 objective and 19 subjective measures.

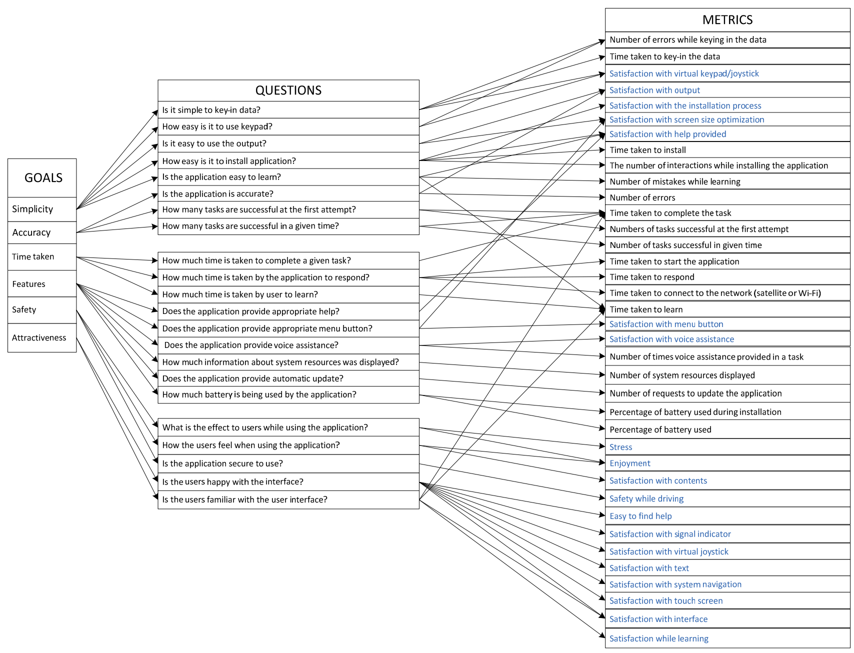
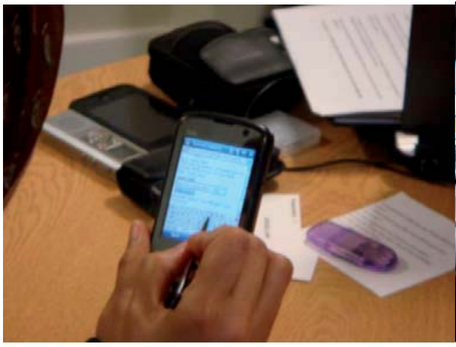


Figure 1. Evaluation model for mobile application.

USABILITY STUDY

To ensure that the model is reliable, effective and optimal experiments were carried out to test the usability of applications in two different mobile phones. The experiments were conducted to test whether the metrics can be used to collect the usability data. Besides, this study also analysed the data to compare the usability problem of applications being installed in current and traditional mobile phones. The experiments were divided into two parts: First, objective data was collected through observation, second, subjective data was collected via an interview to assess the perceptions of participants on mobile applications, as recommended by Nielsen (1994).



Picture 1: Participant using Mobile Facebook

Picture 2: Participant using CoPilot inside car

To test the model, this study used Mobile Facebook and CoPilot Live SatNav system installed in an iPhone and O2 Orbit mobile phone device. The experiment was conducted in a comfortable and quiet room for the Mobile Facebook while for the CoPilot system, the experiment was conducted inside a car in order to mirror the use of such applications in real practice. However, participants did not drive the car during the study for safety reasons. The participants comprise a mixed variety of professions, gender and age (E.G. novices, experts, men, women aged between 20 and 35), and all participants were requested to complete five tasks; They were given sufficient time to explore and learn the application before proceeding to complete all the tasks. Pictures 1 and 2 above are among some taken during the usability test. The test sessions were recorded using a video camera which captured the screen and keyboard of the mobile phone. Participants were given a brief introduction about the mobile phone, for example the basic uses of the mobile phone and how to operate the virtual keyboard. They were also encouraged to “think aloud” during the test. Tasks were given in sequence using card-based instructions.

All research which involves interaction with human subjects brings forth ethical issues concerning safety, acquisition of opinions, and the use of any knowledge gained. It is of the utmost importance that research involving human participants is carried out to the highest possible ethical standards. The key principles taken into consideration in all studies in this research were:

- not to harm,
- to ask for consent,
- to allow freedom of participants to withdraw at any time, and
- to ensure confidentiality.

All participants were over 18 years of age and had volunteered to take part in the study. Information about the mobile device and how it operates was given during the recruitment process. Briefing of what the participants should expect and a clear account of the purposes of the research was given before and upon their arrival to the lab. All necessary steps were taken to ensure that the participants felt comfortable and were aware that they were entitled to withdraw at any stage of the data collection session. The results of the study were accessible only to the researcher as stated on the consent form and are anonymised. Upon completion of the study, participants were asked to describe any negative experience they had encountered and provide feedback for further improvement of the data collection session.

RESULTS AND DISCUSSION

In this section, the objective and subjective results for both phones were compared. This study used SPSS software version 17 to run t-test for each metric to show the differences in the findings between the current and traditional mobile phone. The differences on subjective metrics were compared using NVIVO 8. Nineteen nodes were created similar to the number of subjective metrics in THE GQM model to check whether the data could be collected using GQM model.

Objective data

Most of the data can be collected. However, some of the data were not able to be collected due to the following reasons, for instance, the metric 'time taken to install', 'the number of interaction while installing the application' and 'percentage of battery used while installation'. facebook's application on 02 orbit is a wireless application and there was no installation process for facebook. moreover, this study is also unable to obtain data related to automatic update and influenced the metric - 'number of request to update the application'. sometimes an automatic update alert was received from the owner of the application. unfortunately it did not show while the experiment was being conducted. This researchers ran the t-test for all metrics and below is an example of the t-test for one of the metrics:

Metric 1: Time taken to input the data

H_0: $\mu_{iPhone} = \mu_{O2}$	H_1: $\mu_{iPhone} < \mu_{O2}$

Where
H_0 = the null hypothesis
μ_{iPhone} = time taken to input the data using iPhone
μ_{O2} = time taken to input the data using O_2
($t_8 = 0.018$, $p < .05$)

The above results shows that at 95% confidence level, time taken to input the data for an iPhone is shorter than O2.

This study summarized all t-test results and the findings show that 13 metrics out of 17 were tested for the CoPilot application while 11 metrics were tested for Facebook. The t-test could not be run for the some of the metrics due to fact that the standard deviation for both groups were 0. Similarly, t-test for some metrics also could not be run for some metrics (time taken to install) on Facebook due to unavailability of data to test. For the CoPilot application, a significant difference between iPhone and O2 orbit was found for 7 usability metrics while the remaining 6 were found to have no differences. For facebook application, a significant difference was found between iPhone and O2 orbit for 6 usability metrics while for the remaining 5, no differences were found. The overall results show that the application installed on an iPhone is better than in O2 Orbit.

Subjective data

A semi-structured instrument for the interview session was designed to test whether the GQM model is able to collect subjective data. The questions were designed to minimize the use of technical terms, and the session was conducted in an informal manner. the overall aim was to obtain participants' opinions while using the mobile application. Examples of questions include feelings experienced after completing the task, comments on the menu arrangement, voice assistance, interface, screen, and satisfaction on system speed and safety. This study also requested participants to comment on the devices for both the iPhone and O2 Orbit in terms of screen size, speed and text size.

Participants were interviewed after they had used each of the applications on different mobile phones. Only one a metric was unavailable, that is, 'Virtual joystick'. This study did not request the participants whether they felt satisfied with the 'Virtual joystick' because they did not use the joystick for both applications. In the results section, the number of positive and negative comments for the applications inside the iPhone and O2 Orbit were compared. The interview transcripts were analyzed and the comments categorized based on the nodes created in NVIVO. This study checked the comments and identified whether they were it positive or negative comments. Table 4 shows the overall results for both applications in two different mobile phones. iPhone obtained 27 positive feedbacks and 13 negative feedbacks, whereas o2 orbit obtained only 12 positive feedbacks and 39 negative feedbacks.

From the interview transcript, most participants were very happy to use the CoPilot inside the iPhone except for one participant who expressed dissatisfaction with the virtual keypad. All participants were unhappy with the CoPilot on O2 Orbit and cited reasons cited including screen size, touch screen, tiny virtual keypad. Most participants stated that overall they did not enjoy using the CoPilot on the Orbit. For the Facebook on an iPhone, interestingly this study found a more equal balance of positive and negative feedback. Participants were unhappy using the virtual keypad on the iPhone, and they noted that the keypad was too sensitive. Most participants gave positive feedbacks about the contents. For Facebook on O2 Orbit, all participants mentioned that the virtual keypad was too small and they did not like to use the stylus. Some participants still made mistakes while using the stylus and suggested that a physical keypad for data entry was preferred. Participants were also unhappy with the overall navigation and interface design, and they suggested having one main menu for all sub menus on Facebook.

Table 4

Result for Subjective Measure

Application / Device	Positive feedback	Negative feedback
CoPilot / iPhone	16	8
Facebook / iPhone	11	5
Total for iPhone	27	13
CoPilot / O2 Orbit	6	15
Facebook / O2 Orbit	6	24
Total for O2Orbit	12	39

The results for objective and subject measures show that the application on an iPhone is better than O2 Orbit in terms of interaction. However, this comparison is not the main objectives of this study apart from validation purposes only. This study also found that the model could generated too many metrics and become too complex. Thus, having an optimal number of metrics are recommended by reducing or combining the metrics, for example, the metric 'virtual keyboard' and 'virtual joystick' can be combined into 'satisfaction with touch screen'.

CONCLUSION

The GQM model was developed to serve as measures to evaluate the usability of mobile applications and the experiments were conducted to prove that the model can be used to evaluate the application on mobile phones. The model can be edited to add or drop the goals, questions and metrics. This capability allows a new measure to be inserted into the model by developing a new goal or new questions. The model will benefit usability evaluators as well as a mobile application developers by serving as a guidelines when designing mobile applications. However, the model is only a list of usability metrics; an evaluator still needs to set up and plan the experiment method. Moreover, this model focuses merely on interactions between human-computer and could be used to enhance other areas, for instance, in terms of how the device handle memory load and load the content onto the screen. For future studies, the development of automated tool is recommended to evaluate mobile applications using the GQM model and the tool will have features to add or drop the metrics. For further tests on the model, the usability test on field is suggested to ensure that the model can be used in any condition.

ACKNOWLEDGEMENT

The authors would like to thank the Universiti Utara Malaysia for the funding under LEADS grant.

REFERENCES

Ahmed, S., Mohammad, D., Rex, B., & Harkirat, K. (2006). Usability measurement and metrics: A consolidated model. *Software Quality Control*, *14*(2), 159-178.

Ali Kattan, Rosni Abdullah, Rosalina Abdul Salam, & Sureswaran Ramedas. (2009). Building distributed heterogeneous smart phone Java applications an evaluation from a development perspective. *Journal of Information and Communication Technology*, *8*, 67 - 83.

Baillie, L. (2003). *Future telecommunication: Exploring actual use.* Paper presented at the International Conference on Human-Computer Interaction, (INTERACT ´03).

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

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

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

Constantinos, K. C., & Dan, K. (2007). *A research agenda for mobile usability*. Paper presented at the CHI '07 Extended Abstracts on Human Factors in Computing Systems.

Fabio, B., Luca, C., Elio, C., & Marco, C. (2010). *Using mobile devices to support communication between emergency medical responders and deaf people*. Paper presented at the 12th International Conference on Human Computer Interaction with Mobile Devices and Services.

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

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

Henry, D., Vivian, H., & Chee, B. T. (2008). *Playing different games on different phones: An empirical study on mobile gaming*. Paper presented at the 10th International Conference on Human Computer Interaction with Mobile Devices and Services.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64*(2), 79-102.

Hornbæk, K., & Law, E. L.C. (2007). *Meta-analysis of correlations among usability measures*. Paper the SIGCHI Conference on Human Factors in Computing Systems.

ISO. (1998). *INTERNATIONAL STANDARD: ISO 9241-11(Guidance on usability)*. Geneva.

Jin, B. S., & Ji, Y. G. (2010). Usability risk level evaluation for physical user interface of mobile phone. *Computers in Industry, 61 (May), 350-363* .

Jun, G., & Tarasewich, P. (2004). *Guidelines for handheld mobile device interface design.* Paper presented at DSI 2004 Annual Meeting, Boston.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies, 60*(5 - 6), 599 - 620.

Kondratova, I., & Goldfarb, I. (2006). *M-learning: Overcoming the usability challenges of mobile devices.* Paper presented at the Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies, 2006. ICN/ICONS/MCL.

Luca, C., & Alessandro, M. (2010). *Supporting blind users in selecting from very long lists of items on mobile phones*. Paper presented at the 12th International Conference on Human Computer Interaction with Mobile Devices and Services.

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

Marco, S., & Luis, D. (2008). *Defining scenarios for mobile design and evaluation*. Paper presented at the CHI ‘08 Extended Abstracts on Human Factors in Computing Systems.

Marco, S., Luis, C., & Tiago, R. (2008). *A framework for mobile evaluation*. Paper presented at the CHI ‘08 extended abstracts on Human factors in computing systems.

Nielsen, J. (1994). *Usability engineering*. London: Morgan Kaufmann.

Pirhonen, A., Stephen, B., & Christopher, H. (2002). *Gestural and audio metaphors as a means of control for mobile devices*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves.

Rita, H. (2006). Accessibility and usability guidelines for mobile devices in home health monitoring. *ACM SIGACCESS Accessibility and Computing*, (84), 26-29.

Rudy, S., Carolyn, W., & Mark, L. (2007). *Towards the perfect infrastructure for usability testing on mobile devices*. Paper presented at the CHI '07 Extended Abstracts on Human factors in Computing Systems.

Scholtz, J. (2004). *Usability evaluation*. Retrieved from http://www.itl.nist.gov/iad/IADpapers/2004/Usability%20Evaluation_rev1.pdf

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd ed.). Michigan: Addison-Wesley Longman Publishing.

Tiago, J., Hugo, G., Joaquim, J., & Daniel, G. (2010). *Assessing mobile touch interfaces for tetraplegics*. Paper presented at the 12th International Conference on Human Computer Interaction with Mobile Devices and Services.

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.