Verification Process of Metric Based Usability Evaluation Model for Chronic Disease Management Mobile Applications

Fatima Zahra, Azham Hussain, Haslina bt Mohd

Abstract: ***This paper explains the verification process of a proposed model for evaluating the usability of chronic disease management (CDM) applications. This research aims to symbolize main practices in evaluating the usability of the CDM application. Twelve (12) usability experts from academia, medical and mobile application development from around the globe participated in examining the model components. The experts completed a verification form and questionnaire that measured the model in terms of consistency, understandable, ease of use, tailorable, verifiable and overall impression. Furthermore, the proposed model has been modified based on the comments and suggestions received from experts. Similarly, the experts' questionnaire result indicates that the proposed model is original, complete and acceptable. Therefore, this study will provide additional knowledge in both theory and practice towards model verification process, especially for usability evaluation of disease management applications.***

Index Terms: Chronic disease management apps, Dimensions, Mobile health, Usability.

I. INTRODUCTION

World-wide 4.68 billion people will use mobile phone by the end of 2019. Within this landscape, medical apps will see a 41% compounded annual growth rate between 2015 and 2020 [1], but unexpectedly, prevailing evidence indicates declining downloads of such apps and decreasing engagement of the intended end users. Reason for an increasing digit of medical apps is steadily increasing chronic diseases such as high blood pressure, diabetes, obesity, etc. As the apps for chronic disease management (CDM) proliferate, their effectiveness and little support for the user-friendly interface are becoming a question as the usability of these apps is less than the ideal. Thus, usability is found as a prerequisite for the success of mobile health. Current CDM applications are designed based on existing usability models who are too general and may not be as effective as those that involve end-users needs. CDM apps need to be produced with adequate consideration of the needs of their intended users so that they are easy to use and perceived as useful [2].

It has been shown that without considering usability, mobile apps are unable to retain users; tracking data has shown that users typically allocate less than thirty seconds to learn how to use the app before abandoning it for alternatives or give up using it.

Moreover, usability evaluation is paramount for the optimal design and development of mobile apps used in clinical cohorts. Several usability evaluation models and frameworks are available to measure the usability of any software, but limited work is found on mobile applications especially mHealth applications. Little research is found regarding usability testing of mHealth apps. Moreover, usability testing is performed by using basic usability model that are not able to explore the complex design of current mHealth apps as these models are not designed to evaluate the usability features specific to the mHealth applications. Therefore, a usability evaluation model for chronic disease management application is developed by using QUIM approach and establish usability dimensions, criteria, and metrics. Consequently, the developed model is based on the identified component of the usability related factors like technology, task, user, and environment [3, 4]. Moreover, these measurements of the usability are seen satisfactory enough to practice for estimating the adoption of CDM application and considered supporting towards Human-Computer Interaction principles [3, 5]. This paper aims to explain the model verification method. Little number of studies in literature does exist those can explain that how the model and its components have been confirmed by the professional experts [6]. The understanding regarding the modelling usability approach is the main contribution of this model. Thus, the anticipated model is reached the ultimate step of development and need to validate the model as it become essential.

II. PROPOSED USABILITY EVALUATION MODEL FOR CHRONIC DISEASE MANAGEMENT APPLICATIONS

The proposed model is constructed based on the requirement gathered from real users of chronic disease management and SLR that was conducted in an earlier phase. The literature reviewed provided a solid base for the proposed model.

Revised Manuscript Received on May 22, 2019.

Fatima Zahra, School of Computing, Universiti Utara Malaysia
Azham Hussain, School of Computing, Universiti Utara Malaysia
Haslina bt Mohd, School of Computing, Universiti Utara Malaysia

Usability dimension, criteria, and metrics were reported in SLR, and identified dimensions and criteria were then scrutinized to avoid repetition or duplication of similar measurements [3]. Then usability metrics were generated from the literature reviewed such as from [7] and [8]. QUIM is chosen as guidance for proposing the usability evaluation model for this study because it unifies existing usability standards into a single consolidated, hierarchical model of usability measurement and declared its reliability and relationship. Studies also claimed that QUIM bring impact in usability measurement towards software testing by incorporating quality measurement in an arranged and proper way for the practitioner [9]. Metrics generated were then placed according to corresponding criteria and were explained further. According to [10], some criteria can be linked to many usability dimensions depending on the features and functions of the application which similar to [11] described how criteria are linked directly to usability dimensions. Besides that, the categorization of the dimension to the related criteria and corresponding metrics for the proposed model are as agreed and used in many usability evaluation studies [3, 7]. Total of eighteen (18) dimensions were collected from eighty-five (85) selected papers through SLR developed on the frequency being stated. In the literature these were used for usability evaluation in general and mobile platform for particularly health concerned or chronic disease. The [4] method of summarization was employed to group the dimensions that were similar in meaning to maintain consistency and avoid redundancy of dimension. Literature shows that efficiency, effectiveness, user satisfaction, learnability and ease of use are the most common usability dimensions in the recent usability studies such as [3, 4, 11]. These attributes are widely considered due to their direct relationship to the technical capabilities and ease of measurement. Therefore, the purposed model includes learnability, security, and accessibility in the context of chronic disease management apps and requirements of intended users. Selected usability dimensions are further broken down into measurable criteria (sub-factors). Moreover, it is precisely measurable through a specific metric that is related to a usability factor [12], [10], [13]. In other words, the measurement of usability dimension depends on the corresponding criteria, whereas, metrics are used to measure criteria that are assigned to a usability dimension. Metrics are defined in terms of formula or countable metrics which are extracted from raw data such as video observation or experiment depending on the application type [12], [14]. The usability metrics are classified into two main categories: testing and predictive metrics [10, 13], [7]. The proposed model is designed based on testing metrics that comprises of preference and performance metrics. See Appendix A for the first version of the proposed model of CDM application usability evaluation. Through systematic literature review, all sixteen criteria are developed. Each criterion is placed to its consistent usability dimension developed on the subsidiary literature. These groups of usability dimensions have been used in several usability evaluations studies in literature, particularly in software domain and mobile applications, such as [10],[8, 11]. Moreover, to produce significant usablity metrics to the generated criteria, the discribed metrics in QUIM [10], GQM model [15], and other usability researches such as mGQM [8], [16] were systematically analyzed and engaged to costume the need of CDM applications. Therefore, all sixty (60) metrics are derived into two categories of objective and subjective measurements. However, thirty-eight metrics measure objective data, and twenty-two metrics measure subjective measurements.

III. RESEARCH METHODOLOGY

In order to enhance the quality, originality and richness of the measurement, author used the expert review and verification method, and also take valuable inputs from the field of HCI medical application developers as well as from the academia of the industry [17]. The model verification is a method in which a model justified that its components are complete, sufficient and accurate according to the intention and approach of the specific model [6]. Similarly, verification process confirmed that all components relating to the model successfully met with the satisfactory range of consistency, completeness, and accuracy with the intended application. However, the inputs and comments from the domain and usability experts were positively contributed and enhanced the quality of model design [18].

Instrument development

For the expert reviewers' instruments were designed in the form of questionnaire and expert's verification form. This instrument is divided into three sections; section A is related to expert profile; section B is related to metrics with usability dimensions and corresponding criteria and section C is based on questionnaire. The questionnaire [6] is presented with little modification, and it comprises five measurable factors with two scale choices; "Agree," or "Disagree". The questionnaire is consisting of five measuring factors; 1) Ease to use, 2) Verifiable, 3) Understandable, 4) Tailorable and 5) Consistency [19]. However, the questionnaire also possessed the overall impression regarding the measure of experts' opinion about model. The experts are used five dimensions in order to judge the model for the expression of its acceptability and originality in practice for research purposes and usability practitioners [6]. The verification form comprises comprehensive details regarding the model. It includes the model's grammatical formation, the relationship of the model components that happens between each entity and also covers the five dimensions of the expert's questionnaire through which the model is judged.

Data collection

All the experts were communicated, through phone calls and emails. A breif discussion has been made with reference to the proposed model improvement. Forty experts were contacted; however only fifteen agree to participate in the research. Overall 12 experts were selected who comprise of four knowledge experts, four usability experts, and four domain experts. However, [20] stated that only three to five experts are adequate enough to verify and review a newly developed model. Among them (6) experts were contacted through emails due to a remote location, and other six agree on face to face meeting straight in their individual offices to get critical suggestions and comments.

Data analysis

For the analysis purpose all data collected from the experts stored and sorted in the statistical package of SPSS. In the expert's instrument, the SPSS was used to find out the MEANS of individual measurement. The result was analyzed and presented in the following section. For the improvement and increasing the reliability of the proposed model, all the suggestions and comments were acknowledged from the experts.

IV. DISCUSSION AND RESULTS

The verification process offers a healthy experience and has given substantial improvements to the developed model. The comments and suggestions given by the experts were addressed to the metrics of the model which some were repeated and not related to the application. To evade complication, repeated and unrelated metrics have been removed from the model. However, experts appreciated that the definition of usability factors, its criteria and most of the metrics are excellently produced and capable of collecting data. Additionally, all the experts agreed that proposed model will give satisfactory results and also agreed that it is clearly explicit to the chronic disease management application. Comments given by expert are presented below. However, the comments are rephrased due to limited space.

Table. 1 Experts' comments/suggestions for the proposed model

Experts	Comments /Suggestions
Expert A	Good initiative for medical development. However, include completeness under effectiveness which is necessary to measure the successful tasks and achieved goal include under self-tracking.
Expert B	Expired notification could be removed. Moreover, try to merge availability with accuracy as there are redundancy in metrics this may confuse user. Since it is measuring how efficient application work. You can also include number of undo action as a new metric in your model.
Expert C	Interesting research, overall impression is good. Concentrate more on security and accessibility as these are more focused towards medical. Discuss your model with usability experts as well.
Expert D	Helpful for patients, most of them are reluctant because most of them are naive and available apps do not focus on the usability. There should be an option to switch on and off the notification
Expert E	Overall model is good and helpful for medical patients. You can include how user feel after using the application if they feel happy or annoyed. This you can include under satisfaction
Expert F	Learnability dimension and Security is very important for patients. I would suggest for the app developer to add a glossary to explain medical jargon, self-explanatory as it will help patients to better understanding their disease and medication.
Expert G	Metrics could be modified such as time taken during task selection could be number of clicks for task selection, number of steps during task selection could be change to number of clicks to find call to action button. Some metrics need to be merge such as malfunction could be merge into satisfaction with the accuracy of the functions performed. Include navigation structure and help menu.
Expert H	There is redundancy and you need to rephrase such as time taken to response could be changed into time take to respond to command. You need to readjust metrics under most relevant criteria. Over all model is good it will be very helpful for the usability practitioners.
Expert I	This model is flexible, and changes could be made. Some metrics are untestable until usability testing is done in live lab. Moreover, I cannot see how 'numbers of clicks to sign in' relates to 'integrity'. Some metrics are not practical and irrelevant or need to replace such as metrics under non-repudiation. Non-repudiation it-self could be replaced with insurance. QUIM map used it under trustfulness that include security as well. It will enhance simplicity in your model. Metric stated as Ratio of all interaction entry in log and total number of interactions' is this question specific to the 'number' of interactions? if so, OK. But if the differences are on the requirements themselves, then it is too ambiguous & difficult.". Another metric "Number of validation check per total user input." it would be very challenging to figure out the difference between them and 'accurateness'. Or 'Validation check for data source '. again, quite subjective and not very practical as a topic-question-and interpretation of answers. how practical is such a question? I suggest that you try to answer this question in various contexts such as data source from the developer end, that would be difficult. Changes need to be made so that you can test this model in usability testing.

Expert J	Overall good model, However the model should be evaluated to examine its accuracy
Expert K	Evaluation metrics are fine however, If some metrics are redundant and cannot be evaluated correctly as it would make result inconsistent. More aspects need to consider on completeness under effectiveness such as completing a goal.
Expert L	This model is more towards usability evaluation of mobile health applications and all the dimensions and criteria are defined well. it could provide help both medical and research community. It is good that security and accessibility dimensions have been included to identify usability issues in clear and depth for medical applications.

According to the comments and suggestions passed by the experts, all those metrics which are not correlated to the defined standards have been removed, while for the purpose of improvement and corresponding criteria of the model, some metrics have been added as shown in Table 4 and 5. Experts were requested to judge the proposed model by using five dimensions; understandability, verifiable and overall impression, consistency, easy to use and tailorable. The completeness, acceptability and completeness of the model is measured by these dimensions. In the questionnaire instrument two options were given "Agree" and "Disagree" [19]. The result of the questionnaire shows that all experts revitalized towards "Agree." Please, refer to Table 6 for mean scores of individual experts.

Table. 2 Expert's profile

Experts	**Profile**
Expert A	Qualification: PhD, Position: Associate Professor, Experience: 13-year, Country: Bahrain, Expertise: Internet of things and its Applications/ Evaluation of Authentication Applications for Smartphone
Expert B	Qualification: PhD, Position: Associate Professor, Experience: 20-year, Country: Bahrain, Expertise: Software testing and its verification
Expert C	Qualification: PhD, Position: Associate Professor, Experience: 20-year, Country: Malaysia, Expertise: Empirical Usability Evaluation/ User Experience
Expert D	Qualification: PhD, Position: Associate Professor, Experience: 12-year, Country: China, Expertise: Software Engineering/ Information User Behavior
Expert E	Position: Medical director, Experience: 30-year, Country: Malaysia, Expertise: Mobile health and Hematology
Expert F	Position: Medical doctor, Experience: 9- year, Country: Malaysia, Expertise: Patient safety and family physician and medicine.
Expert G	Position: Medical doctor, Experience: 9- year, Country: Malaysia, Expertise: health promotion/ Medical (general medicine)
Expert H	Position: Medical Director, Experience: 25+year, Country: Malaysia, Expertise Mobile health/ Primary care physician
Expert I	Qualification: PhD, Position: Software engineer, Experience: 40+ year, Country: Canada, Expertise: Software engineering/ Testing/ Software Measurement
Expert J	Qualification: PhD, Position: Software engineer and tester, Experience: 5-year, Country: Spain, Expertise: Usability and mobile applications
Expert K	Qualification: PhD, Position: Associate Professor, Experience: 10-year, Country: Australia, Expertise: Usability engineering
Expert L	Qualification: PhD, Position: Software engineer, Experience: 17- year, Country: Malaysia, Expertise: User Experience/ Usability Tester

Table. 3 List of Dropped metrics

Dropped metrics
Availability of self-tracking history
Satisfaction with goal achievement
Total number of sent notifications
Total number of dismissed and expired notification
Validation check of data source
Number of times application crash
Time taken during task selection
Number of steps taken during task selection
Satisfaction with the application immunity against malfunction
Ratio of all interaction entry in log and total numbers of interactions
Number of validation check per total user input

Table. 4 List of Added metrics

Added metrics
Time taken to respond to a command
Total number of interactions
Total time spend on executing the tasks
Time spend on errors by users
Number of times user can undo actions that may lead to errors
Satisfaction with the self-explanatory information provided to understand medical jargon
satisfaction with the video help provided
Time spend while key in data.
Total number of times help menu used
Time spends on successful login.
Number of clicks to find call to action button
Number of errors during navigation
Total number of failed commands

Table. 5 overall score for individual dimension for model verification

Dimension	Mean
Consistency	0.87
Understandable	0.75
Easy to use	0.83
Tailorable	0.87
Verifiable	0.83
Overall Impression	0.97

Understandable score 75% whereas ease of use and verifiable score 83% which could be due to the irrelevancy or inconsistency of some metrics identified by the experts as they suggest removing some metrics that are difficult to calculate with usability testing. However, those metrics were removed, and a few were added based on the comments and suggestions received from the experts. However, all the dimensions score above (75%) reveale that the developed model for evaluating Therefore, an amendment has been made concerning the metrics and their corresponding criteria from the first version of the model based on the expert's feedback. For instance, previously the model contains a total of Sixty (60) metrics in which twenty-two (22) are subjective data whereas thirty-eight (38) for objective data (refer Appendix A). Therefore, a total of fifty-three (53) metrics were deployed in the revised model, which consist of twenty-seven (27) objective metrics and remaining twenty-six (26) are subjective metrics (refer Appendix B).

V. CONCLUSION

Current model is developed to guide the CDM app developers and practitioners and to narrate procedures in evaluating application usability. The research further describes how the CDM apps usability model was evaluated and verified by the experts. Few metrics were added, while some were removed based on experts' feedback. The dimension of the model was evaluated in terms of understandable, consistency, tailorable, easy to use, verifiable and overall impression. The feedback from the experts for all the dimensions was satisfactory representing that the proposed usability evaluation original, acceptable and complete. However, in model development process, the verification by experts of a proposed model is merely the last stage of the first cycle. The future research aims the usability testing of the developed model on different CDM applications. Henceforth, this will observe subjective and objective metrices capability for collecting data.

REFERENCES

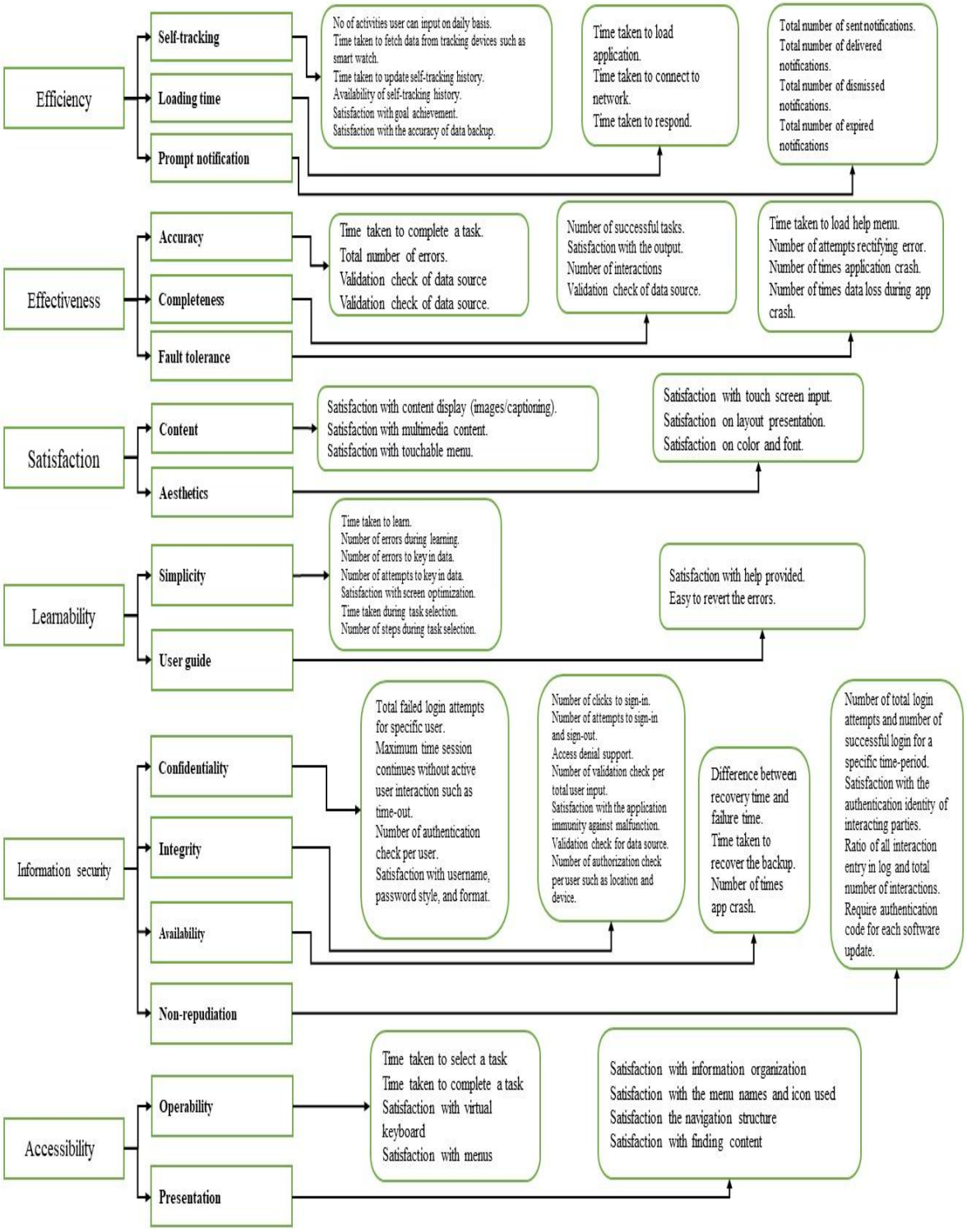
1. M.S. Liew, J. Zhang, J. See, and Y. L. Ong, "Usability Challenges for Health and Wellness Mobile Apps: Mixed-Methods Study Among mHealth Experts and Consumers," JMIR mHealth and uHealth, vol. 7, p. e12160, 2019.
2. R.Schnall, M. Rojas, S. Bakken, W. Brown III, A. Carballo-Dieguez, M. Carry, et al., "A user-centered model for designing consumer mobile health (mHealth) applications (apps)," Journal of biomedical informatics, vol. 60, pp. 243-251, 2016.
3. R.Baharuddin, D. Singh, and R. Razali, "Usability dimensions for mobile applications—A review," Res. J. Appl. Sci. Eng. Technol, vol. 5, pp. 2225-2231, 2013.
4. C. K.Coursaris and D. J. Kim, "A meta-analytical review of empirical mobile usability studies," Journal of usability studies, vol. 6, pp. 117-171, 2011.
5. A.Madan and S. K. Dubey, "Usability evaluation methods: a literature review," International Journal of Engineering Science and Technology, vol. 1, pp. 590-599, 2012.
6. S.Beecham, T. Hall, C. Britton, M. Cottee, and A. Rainer, "Using an expert panel to validate a requirements process improvement model," Journal of Systems and Software, vol. 76, pp. 251-275, 2005.
7. S.K.Dubey, A. Gulati, and A. Rana, "Integrated model for software usability," International Journal on Computer Science and Engineering, vol. 4, p. 429, 2012.
8. A Hussain, "Metric based evaluation of mobile devices: mobile goal question metric (mGQM)," University of Salford, 2012.
9. A.Fernandez, E. Insfran, and S. Abrahão, "Integrating a usability model into model-driven web development processes," in International Conference on Web Information Systems Engineering, 2009, pp. 497-510.

10. A.Seffah, M. Donyaee, R. B. Kline, and H. K. Padda, "Usability measurement and metrics: A consolidated model," Software Quality Journal, vol. 14, pp. 159-178, 2006.
11. J. Tan, K. Ronkko, and C. Gencel, "A framework for software usability and user experience measurement in mobile industry," in Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2013 Joint Conference of the 23rd International Workshop on, 2013, pp. 156-164.
12. E.Frøkjær, M. Hertzum, and K. Hornbæk, "Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?," in Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2000, pp. 345-352.
13. R.Harrison, D. Flood, and D. Duce, "Usability of mobile applications: literature review and rationale for a new usability model," Journal of Interaction Science, vol. 1, pp. 1-16, 2013.
14. A.Hussain, M. Kutar, A. A. Mutalib, and F. M. Kamal, "Modeling subjective metrics for mobile evaluation," Journal of Research and Innovation in Information System (JISRI), vol. 1, pp. 11-20, 2012.
15. V. R. Basili, "Software modeling and measurement: the Goal/Question/Metric paradigm," 1992.

A. Hussain and M. Kutar, "Usability metric framework for mobile phone application," PGNet, ISBN, vol. 2099, pp. 978-1, 2009.

16. K. E. Wiegers, "Seven truths about peer reviews," Cutter IT Journal, vol. 15, pp. 31-37, 2002.
17. J. Nielsen, "Usability inspection methods," in Conference companion on Human factors in computing systems, 1994, pp. 413-414.
18. H.I. Abubakar, N. Hashim, and A. Hussain, "Verification Process of Usability Evaluation Model for M-banking Application," in Recent Advances in Computer Science Proceedings of the 14th International Conference on Applied Computer and Applied Computational Science (ACACOS'15), 2015, pp. 23-25.
19. J.Nielsen, Usability engineering: Elsevier, 1994.

Appendix A: Proposed model for usability evaluation of chronic disease management



Appendix B: Revised model for usability evaluation of chronic disease management