Empirical Validation of Evaluation Model for Deaf People on Mobile Applications

Shelena Soosay Nathan[1], Azham Hussain[1], Nor Laily Hashim[1], Mohd Adan Omar[1], Ashok Sivaji[2], Ngip Khean Chuan[2], Shi Tzuaan Soo[2]
[1]*Human-Centered Computing Research Lab, School of Computing, Universiti Utara Malaysia*
[2]*MIMOS Berhad, Technology Park Malaysia*
sn.shelena@gmail.com

***Abstract*—Usability evaluation is a process that accesses any system or application functions and user experience on it. Many usability evaluation models available to conduct the usability evaluation with the users. Deaf user is usually isolated when concerning on mobile application since they are disabled. However, many mobile applications are available to cater this community which is left unused due to bad user experience towards the applications. The usability evaluation model for the deaf people mobile application is developed to address the needs of deaf people requirement for better user experience in any application. This paper presents empirical findings of validating the model developed through usability evaluation conducted on a selected mobile application for the deaf. The result shows that the developed model is useful to evaluate and reliable for data collection.**

***Index Terms*—Deaf People; Mobile Application; Usability Evaluation Model; Validation.**

I. INTRODUCTION

Usability evaluation is a process to access the user experience and system performance over any system application that provides service to users. Through usability evaluation, it is able to identify the issue regarding the usability of any system or application [1]. Many usability evaluation models have been developed and used in many system and application evaluation for many years. Most commonly used usability evaluation model is already in use. Models such as Nielsen [2], ISO 9241-11 [3] and QUIM [4] are to name few usability models being referred for usability evaluation for many years. However, for mobile application, only a few models are being proposed. Hussain [5] have introduced mGQM which involve application evaluation that based on GQM model. Besides that, Harrison et al [6] have proposed another model for usability evaluation for mobile application by a combination of ISO 9241-11 [3] with cognitive load. However, these models are focused on mobile application in general and none are available to evaluate any mobile application that is developed for specific users with specific needs.

A mobile application developed for the deaf people are commonly left underutilized and this usually leads to the wastage of cost and energy being through for development of this application [13]. This is because many applications are unable to fulfil the deaf user requirement and low user experience are recorded for the mobile applications [7, 8]. General usability model is unable to identify specific requirements needed by the deaf people that left the mobile application are not able to identify the issues occurs in this mobile application when actual user experiencing the application.

A mobile application developed for special people normally designed to cater the need of the specific requirements of these people. However, some application tends to ignore these issues and this could not be identified when the application is evaluated with common models. This study purpose on evaluating a usability evaluation model that has been developed specifically to cater deaf people mobile application. The model consists of six (6) usability dimensions to be measured with the deaf user which relates to the requirements of the deaf people. This paper presents the empirical findings to validate the developed usability evaluation model for the deaf mobile application. Section two describes the developed usability evaluation model. Section three and four discuss the findings of the evaluation and analysis in validating the developed model while section five concludes the paper.

II. USABILITY EVALUATION MODEL FOR DEAF MOBILE APPLICATIONS

Development of the usability evaluation model for the deaf mobile application started with Systematic Literature Review (SLR) process where articles related to the domain were collected and analyzed intensively to identify measurements that are appropriate [9]. Through SLR, usability models available are explored to identify these measurements and found that none are available for specific disability application. Thus, requirements were gathered from some sample of deaf people to identify the needs of an application and measurement for the usability model for the deaf mobile application are being identified intensely.

Through this requirement gathering process, deaf people needs for mobile application are identified and captured the most useful and related measurements to be used in the model. The purpose of both SLR and requirement gatherings process id to list most appropriate usability measurement for the deaf mobile application usability evaluation. The outcome of these two processes results in the developed model. The model consists of six (6) dimensions, () criteria's and () metrics that can be used in a mobile application for deaf evaluation and usability issue identification. The developed model was then have been reviewed for the applicability of it in the real environment through expert review process [10].

III. MODEL VALIDATION

This section explains the results of the usability model validation conducted with the sample user who is deaf and an

expert mobile application user. Statistical analysis conducted for the validation of the developed model.

A. Application Selection

The actual evaluation of the deaf mobile application was conducted among selected participants at Malaysia Federation for Deaf (MFD), Selangor. For this evaluation, DeafWorld application was selected. The main objective of the application was to provide a social media platform for the deaf people community around the world. Through this application, deaf people around the world are connected and enable to share their daily video and comments with other deaf people. Participants were given a total of seven (7) tasks to fulfil while using this mobile application and data were collected. Two types of data were collected namely objective data to identify system performance and subjective data to measure user experience with the mobile application.

During the evaluation, user's activities were recorded and captured the mobile screens for further analysis for the validation. Time is recorded as participant starts the task and stopped when participant finished all the task given. Participants were then given set of mobile application feature satisfaction form to be evaluated which was developed with the subjective measurement from the developed model. Participants were allowed to comment throughout the evaluation process being conducted and their responses were observed and recorded.

B. Participants

Participants were gathered from MFD, Selangor through convenience sampling method. Convenience sampling means the participants are commonly involved or identified randomly [11]. All the participant is from MFD workers and students. Participants are actual deaf people who use the mobile application for more than three (3) years. Total of 20 participants was gathered at MFD, Selangor. Out of 20 participants, 45% were male and the rest are females. Among these 20 participants, a total of 52% is from Malay ethics, 40% Chinese while 8% are Indians. This detail is shown in the graphical form of Figure 1 and Figure 2.

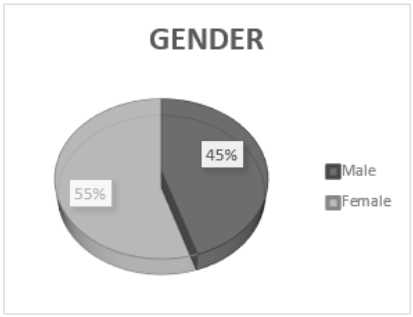

Figure 1: Percentage of participants by gender

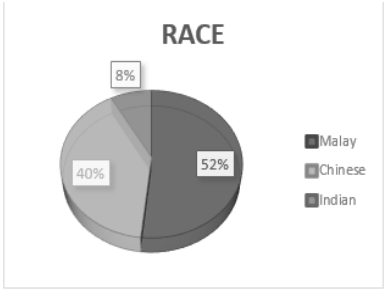

Figure 2: Percentage of participants by race

Participants were followed by two (2) translator in each session. The translators have been a great help in conducting this evaluation since the participants are deaf people and researchers are not well versed in the sign language. Thus, the researcher made sure all the information given to participants are translated slowly by the translator and the task can only be started once the participant agreed and ready.

Figure 3: Participants was briefed by translator

Total of seven tasks was given to the participants to conduct the evaluation. However, before the evaluation starts, researchers ensure that all the participant is agreed to be recorded on their activity and the task starts once receiving acknowledgement that was translated through the translator.

IV. Application Analysis

Analysis of the mobile application evaluation was conducted with the implementation of the tasks and data collected were analyzed. Statistical Package for the Social Sciences (SPSS) software tool used to analyse all the data collected. Collected data were then analyzed into three (3); task completion rate, total error rate and time to complete the task for all the objective data collected while finally satisfaction rating was also calculated in terms of identifying the level of participant satisfaction of using the mobile application.

A. Task Completion Rate

Task completion rate is one of the important measures in assessing the system or application performance [12]. The data for task completion rate were collected through objective data metrics that were pre-defined in the developed model. The data collected were either the task was successfully conducted by the participant or failed to complete. Each task completion rate was recorded by the researcher.

Table 1
Task Completion Rate

Task	Status	No of participant	Percentage (%)
Task 1	Success	10	50
	Failure	10	50
Task 2	Success	20	100
	Failure	0	0
Task 3	Success	19	95
	Failure	1	5
Task 4	Success	0	0
	Failure	20	100
Task 5	Success	20	100
	Failure	0	0
Task 6	Success	20	100
	Failure	0	0
Task 7	Success	5	25
	Failure	15	75

Table 1 shows the task completion rate by all the participants for all 7 tasks which are given during the evaluation. All the task was having completion rate except for Task 4 where none of the participants is able to find the menu for the named task. This ends with task failure for all the participants and was not included in the Table 1. Overall, all the task was able completed by the at least 50% of participant except Task 4 which is a total failure and Task 7 which only 25% were able to complete the task. This shows Task 4 and Task 7 having some functional difficulty that needs to be identified by the researcher. Since the rate of the failure is high for both the task, it might lead to dropped on the satisfaction level of participant and may cause discontinuity of using the application.

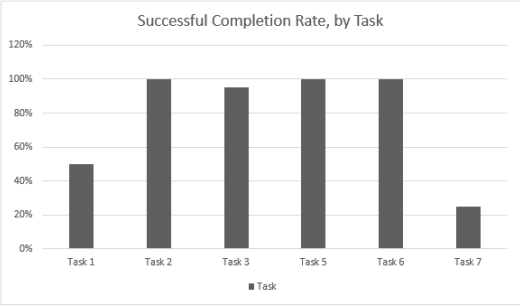

Figure 4: Successful completion rate percentage by task

B. Total Error Rate

The total error rate was another measure that needs to be counted in understanding how easy an application is to be used. This measure in the developed model calculated on a number of mistakes made by the deaf during each task and recorded. Through this, task difficulty is also able to be identified and issues related to usability of the application can be easily identified by the practitioners.

Table 2
Total Error Rate

Task	Total No of Error	Mean
Task 1	101	5.05
Task 2	56	2.80
Task 3	56	2.80
Task 4	55	2.75
Task 5	86	4.30
Task 6	14	0.70
Task 7	64	3.20

Table 2 shows the average of each task error rate by the participant. Task 1 recorded to have the highest number of error made by the participants with an average of 5.05 each participant. This shows the navigational structure of the task is rather confusing the participants and difficult to be used. While Task 6 recorded to have a lesser error made by the participants. This correlates with the task completion rate whereas all the participant has successfully completed the task. This shows the task is easy and straightforward for the deaf user to understand and completes.

C. Time to Complete Task

Time to complete the task is another important measure of the system or the application efficiency in the developed model whereas the total time to complete the task were calculated. These measures counted from the time a task started by the participants and ends with the time they completed each task. This will measure how much time is spent on each task by the participant to identify how easy an application to be used by the deaf participants.

Table 3
Time to Complete Task

Task	Mean Time	Std Dev.	Min Time	Max Time
Task 1	1.61	1.75	0.00	6.81
Task 2	0.76	0.53	0.12	2.15
Task 3	0.72	0.46	0.04	1.87
Task 4	0.29	0.25	0.10	0.96
Task 5	0.65	0.42	0.11	0.16
Task 6	0.11	0.18	0.01	0.70
Task 7	0.47	0.45	0.05	1.49

Table 3 shows the total time spent on each task by the participants. Task 1 shows the highest time recorded to complete with 6.81 minutes maximum time has been taken by participants. This shows the level of difficulty in the task. Besides that, the average time between task is 0.11 seconds until 1.61 minutes taken by each participant shows difficulty in some task which leads to an incomplete rate of the task. This is considered important usability issues. Users commonly looking for an easy and straightforward application to be used. Thus, lesser the time taken to complete task shows easier the application to be used.

However, as for this application, it is recorded some tasks taken more than a minute to complete which is considered long. Though so, it is also depending on the task difficulty level and other external factors that might lead to this time such as mobile phone specification that is old which is not compatible with the application or the connection of Internet if the task needs to complete it. Thus, these factors also need to be considered for each task time.

Many participants struggle in completing Task 1 and after considering the external factors, it is known that this is caused by the application itself. The application is having issues in handling big video to be uploaded and restricted to low-quality video. Thus, when participant trying to upload high-quality video time taken to complete the task become longer than usual as reported. This issue should be resolved so that participants are not distracted by these errors during usage of the application since current mobile phones is preset to high-quality videos nowadays. Mobile application should be reliable to accept any quality of video into their application to ensure satisfaction of user is considered.

D. Satisfaction Score

The satisfaction score for the application was evaluated with the measurement of subjective metrics defined in the developed model. Figure 5 shows the satisfaction level achieved in the application.

The overall level of satisfaction was obtained from a questionnaire provided to participants after the task was completed. It was divided by three to categorize the level of satisfaction of the user towards usage of the application into low satisfaction, moderate satisfaction and high satisfaction.

Figure 5 shows the number of level of satisfaction achieved by the participant in using the mobile application during evaluation. It is clear that 70% of participants are having low satisfaction towards the application and felt that the application failed to deliver many requirements deaf people are needing. Besides that, application evaluated are also rated very difficult to be used by deaf people and many usability issues in terms of navigation and interface needed to be resolved.

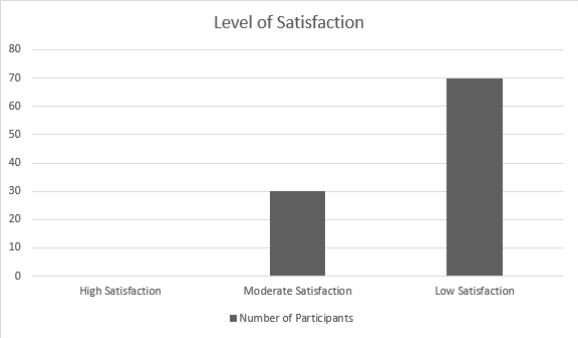

Figure 5: Satisfaction level

E. Overall Usability Score

The mean score for the overall system usability with major usability dimension from the developed model is given in Figure 6.

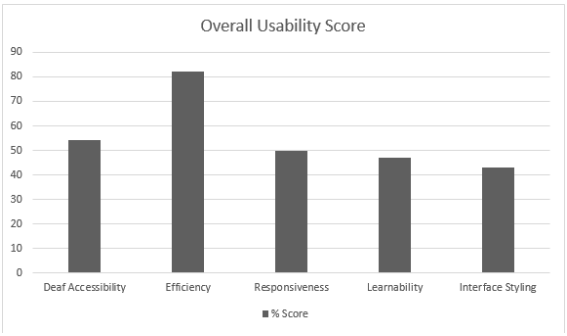

Figure 6: Overall usability level

Figure 6 shows overall usability score of the DeafWorld mobile application with 82% rare efficiency. However, it is observed that other dimension scores 54% and below. The lowest usability is recorded in Interface Styling where the layout of the application which seems dull and unattractive for the deaf users. While 47% scores for learnability shows the application is difficult for the deaf to use and it is the reason on why the total time taken is higher for some tasks. This shows that the overall usability of the Deaf World is low in satisfaction and there is room for improvement to improvise better and serve the deaf community.

V. CONCLUSION

Usability evaluation for the deaf are very crucial since their requirements for mobile application are different than for non-disabled people. Identifying their requirements are the first and foremost important thing to be considered by the developer in ensuring usefulness and continuity of the application developed among the community are higher. This also shows the lack in the general usability evaluation model which unable to identify such requirements related issue for this specific people. This paper presents the usability evaluation model validation result that was conducted with actual deaf participants. The results show that the model are able to be used in evaluation of usability of the deaf people mobile application. Future studies can incorporate the detail usability report by dimensions to further analyze the issue related to the deaf.

ACKNOWLEDGMENT

The author would like to thank to Universiti Utara Malaysia for supporting this research under Geran Penjanaan Universiti (SO CODE: 13620).

REFERENCES

[1] A. Dix, J. Finlay, G. Abowd and R. Beale, Human Computer Interaction, 3rd ed., *Prentice Hall*, 2004.
[2] J. Nielsen, "Heuristic Evaluation", *Usability Engineering (Vol. 44)*, doi:10.1145/1508044.1508050, 2014.
[3] ISO, "International Standard: ISO 9241-11, *Guidance on Usability*, Geneva, 1998.
[4] A. Seffah, M. Donyaee, R. B. Kline and H. K. Padda, "Usability measurement and metrics: A consolidated model", *Software Quality Journal*, 14(2), 159-178, 2006.
[5] A. Hussain, "Metric based evaluation of mobile devices: Mobile goal question metric (mGQM)", (Doctoral dissertation, University of Salford, 2012).
[6] R. Harrison, D. Flood and D. Duce, "Usability of mobile applications: literature review and rationale for a new usability model", *Journal of Interaction Science*, 1(1), 1-16, 2013.
[7] R. S. Hassan, "Mobile communication for people with disabilities: A case study on iPhone technology usage for deaf and mute Qatari adults*", International Congress on Communication*, 5, 587-596, 2011.
[8] D. Zhang and B. Adipat, "Challenges, methodologies, and issues in the usability testing of mobile applications", *International Journal of Human-Computer Interaction*, 18(3), 293-308, 2005.
[9] B Kitchenham,. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, *33*, 2004.
[10] M. Tory & T. Moller, Evaluating visualizations: do expert reviews work? *Computer Graphics and Applications, IEEE*, *25*(5), 8-11, 2005.
[11] J. W. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 3rd ed., *Sage Publications*, 2008.
[12] W. Albert, & T. Tullis, Measuring the user experience: collecting, analyzing, and presenting usability metrics. *Newnes*, 2013.
[13] E. G. Nilsson, Design patterns for user interface for mobile applications. *Advances in engineering software*, *40*(12), 1318-1328., 2009.