Transdisciplinarity and the Future of Engineering
B.R. Moser et al. (Eds.)

doi:10.3233/ATDE220642

Usability Evaluation of Elder-Friendly Design: Application to Take Alipay App

Yan XIANG[a], Danni CHANG[a,1],Ying YAO[a], Leqi WANG[a], Anqi CHEN[a] and Jiajie LI[b]
[a] *School of Design, Shanghai Jiao Tong University, Shanghai, China*
[b] *MIT Media Lab, Massachusetts Institute of Technology, Cambridge, USA*

Abstract. In the context of the increasing aging population, design for the elderly has become more and more important. Many products have taken into consideration the elder-friendly design. However, the actual usability in real practice has not been sufficiently investigated. This study aims to tackle the problem and examine the practical usability of such a design through an experimental way. In particular, the Alipay app was selected as the target product to test. Its normal mode and elder-friendly mode were compared. For this purpose, we designed a set of experiments with 16 participants who are older than 45 years old and randomly divided them into two groups. One group used the normal mode, and the other used the elder-friendly mode. The eye tracker of Tobii Pro Glasses 2 was employed to collect participants' eye movement data on these interfaces. Moreover, user interviews, user behavior observation, and System Usability Scale were jointly adopted to collect related user behavior information and subjective experience evaluation. Based on the qualitative and quantitative data, the differences between the two modes in the dimensions of information architecture, interface design, and task flow were identified. The results show that the current elder-friendly design cannot effectively facilitate usability for elderly users. Furthermore, the corresponding transdisciplinary design strategies were proposed to help with the improvement in design for the elderly.

Keywords. Design for the elderly, usability evaluation, elder-friendly mode, design strategy, transdisciplinary

Introduction

With the permeation of mobile intelligent devices, various applications have become necessary for people to obtain essential services. It can be observed that there are still some problems for middle-aged and elderly people using these apps, especially in finding the interface portals and implementing the required operations. Therefore, it has become critical to conduct elder-friendly design. Although many apps claim they have considered the use of seniors and developed elder-friendly modes. Research has shown that there are still significant problems for the elderly in actually using these apps.

This research will explore the usability problem of the elder-friendly design in mobile apps. As the most popular App in China, Alipay's normal mode and elder-friendly mode will be compared, and their usability and learnability will be emphatically analyzed. Particularly, the usability evaluation and comparison will be performed from the aspects of information architecture, interface design, and task flow. Based on the evaluation

[1] Corresponding Author, Mail: dchang1@sjtu.edu.cn.

results, some transdisciplinary design suggestions further facilitate the design for the elderly will be generated.

Generally, the research is expected to examine the usability problems of existing elder-friendly design and provide designers with specific suggestions to make further design improvements that fully consider the characteristics of the elderly in vision, hearing, cognitive ability, and dexterity, to assist seniors in better enjoying the products and services in such an information age.

1. Literature review

Previous studies have shown that when designing for middle-aged and elderly people, their physical characteristics and cognitive abilities need to be especially considered [1]. At the same time, designers have also paid attention to design norms and standards that meet the physical and mental health of the elderly. Amongst them, minimizing the user's memory load and matching the system to the real world for better understanding are the two most frequently used methods [2].

In terms of usability design and testing methods, improving the usability of social networking mobile applications is critical, especially for older adults with age-related problems [3]. According to the ISO 9241-11 standard, usability consists of the following measurable elements: effectiveness, efficiency, and satisfaction [4]. Nielsen's usability definition has a great impact on usability thinking overall. He divides usability into five elements, which are learnability, efficiency, memorability, errors, and satisfaction [5]. And usability issues are often divided into four categories: (1) appearance, (2) language, (3) conversation, and (4) information [2].

New technologies change the life of the elderly. However, the reality is that many digital products fail to take into account the special needs of elderly users in front size, color contrast, icon metaphor, and so on. Any inappropriate design may lead to poor usability for the elderly, and there is still a great need for improvement in transdisciplinary design strategies for the elderly.

2. Methodology and experiment design

- **Experiment purpose:** The purpose of our experiment is to verify whether the elder-friendly mode of the Alipay app significantly improves the usability and learnability of middle-aged and elderly people compared with the normal mode. Based on the evaluation results, design improvement strategies suitable for senior users will be proposed.
- **Experiment tools:** A separate lab is prepared with task instruction cards, a video recording appliance, and a mobile phone with Alipay installed. The eye tracker of Tobii Pro Glasses 2 will be used to collect participants' eye movement data on the interfaces.
- **Experiment subjects:** According to the World Health Organization (WHO) criteria for age classification, 45-59 years old are considered middle age or pre-geriatric (early old age) [6]. And based on the literature research and our preliminary experiments, this age group was found to be more likely to use smartphones compared to people aged 60 and above who hardly use

smartphones, and will still be considered as our target subjects [7]. So we decided to select the 45-59 years old people to be our experiment subjects. To make sure that all participants are in good physical condition, without severe visual impairment, and meet the basic conditions to perform our tasks, a filtering process was performed. Finally, 16 participants between the ages of 45 and 59 were recruited for the experiment.

- **Evaluation metrics:** Two categories of measurement indicators were formulated: objective and subjective metrics. Objective measurement indicators are obtained from eye-tracking data and user behavior observation: completion time of certain tasks, completion rate, number of mistakes, times of requests for operation clue, and heat maps of the area of interest (AOI) by eye-tracking. Subjective measurement indicators are obtained by the modified System Usability Scale (SUS) and interviews. The experiment's quantitative data and qualitative data will be jointly analyzed to reach a comprehensive understanding of users' performance in the usability test of the Alipay app.
- **Experiment process:** Before the experiment, a brief description of the experiment goals and specified tasks were given to participants and they will be assisted with the wearing of the eye tracker. In particular, five tasks were designed and presented on five A4 sheets as the guide and prompt during the experiment, including: (1) Show the health code (which is a kind of QR code to identify the user's health condition), (2) Show payment code, (3) Top-up the cellphone card, (4) Buy train tickets, and (5) Take a taxi. Before the test, participants can observe the home page for 30 seconds, and then proceed to perform the five tasks. After the experiment, the SUS rating scale for the Alipay app needs to be completed, in which ten usability questions are set [8]. Apart from the questionnaires, interviews will also be conducted to collect more details of users' opinions on this App.

3. Results analysis

16 users aged 45-59 were randomly divided into 2 groups, of which group A (GA) used Alipay's normal mode to complete the tasks, and group B (GB) used Alipay's elder-friendly mode. Among them, the users for task 1 and task 2 are all experienced users, and for task 3, task 4, and task 5, they are all novice users as shown in table 1.

Table 1. Grouping strategy of the experiment subjects.

Tasks	Group A (n=8)	Group B (n=8)
Task 1, 2	Experienced users	Experienced users
Task 3, 4, 5	Novice users	Novice users

Based on the qualitative and quantitative data collected by the methods explained in Section 2, the differences between the two modes will be analyzed regarding (1) information architecture, (2) interface design, and (3) task flow design.

3.1. Analysis of overall information architecture

Descriptive statistics, completion time of tasks, and SUS rating scores are used to compare the performance of the two interface modes.

3.1.1. Descriptive statistics

Basic statistical analysis was performed on The number of tasks completed, The number of mistakes, The number of hesitation (number of pauses, delays, and hesitation), The number of assistance needed (number of requests for operation clue), The number of encouragements needed (number of attempts encouraged to continue), and The number of frustration points (number of being upset/frustrated/mad) as shown in table 2 [9].

Table 2. Descriptive statistics.

Usability element, N(%)	**Tasks**				
	1-Show health code	**2-Show payment code**	**3-Top-up cellphone card**	**4-Buy train tickets**	**5-Take a taxi**
The number of tasks completed					
GA-Finished tasks	8 (100)	8 (100)	7 (87.5)	6 (75.0)	3 (37.5)
GB-Finished tasks	8 (100)	8 (100)	8 (100)	3 (37.5)	6 (75.0)
The number of mistakes					
GA-Mistakes	0 (0)	1 (12.5)	3 (37.5)	5 (62.5)	5 (62.5)
GB-Mistakes	0 (0)	2 (25.0)	4 (50.0)	7 (87.5)	5 (62.5)
The number of hesitation					
GA-Pauses/delays/hesitation	2 (25.0)	0 (0)	2 (25.0)	4 (50.0)	3 (37.5)
GB-Pauses/delays/hesitation	3 (37.5)	3 (37.5)	3 (37.5)	6 (75.0)	5 (62.5)
The number of assistance needed					
GA-Needed more instructions	0 (0)	0 (0)	0 (0)	3 (37.5)	3 (37.5)
GB-Needed more instructions	0 (0)	1 (12.5)	1 (12.5)	5 (62.5)	2 (25.0)
The number of encouragements needed					
GA-Needed encouragement	0 (0)	0 (0)	2 (25.0)	4 (50.0)	2 (25.0)
GB-Needed encouragement	0 (0)	1 (12.5)	0 (0)	2 (25.0)	0 (0)
The number of frustration points					
GA-Upset/frustrated/mad	0 (0)	0 (0)	1 (12.5)	2 (25.0)	1 (12.5)
GB-Upset/frustrated/mad	0 (0)	0 (0)	0 (0)	1 (12.5)	2 (25.0)

As is shown in the table above, the completion rates of the experienced users in tasks 1 and 2 are 100%, while the completion rate of the novice users in tasks 3, 4, and 5 is lower. Group B showed more errors, pauses, and need for help in Tasks 2, 3, and 4. Generally, the usability of Alipay's elder-friendly mode (group B) has not been improved compared with that of the normal mode (group A). The reasons for this will be explored and analyzed in detail as follow.

3.1.2. Completion time of the task

In this part, a T-test analysis was performed to make the statistical comparison between these two interface modes. The homogeneity of variances test for groups (task1, 3, 4, 5) was conducted, which indicates that the data meets a normal distribution, and non-parametric tests were performed for the group (task2) with the heterogeneity of variances. Then T-tests were executed, and the results are shown in Table 3.

Table 3. T-test analysis results of completion time of task 1, 3, 4, 5.

Time to finish tasks (seconds)	**Group (mean ± SD)**		**t**	**p**
	1.0 (n=8)	**2.0 (n=8)**		
Task 1's time (seconds)	14.13±26.25	17.13±27.53	-0.223	0.827
Task 3's time (seconds)	41.71±30.76	30.71±13.94	0.862	0.406
Task 4's time (seconds)	167.25±84.80	246.00±108.11	-1.146	0.295
Task 5's time (seconds)	100.67±44.74	79.83±42.79	0.680	0.519
*p<0.05**p<0.01				

From the above table, it can be seen that different groups (groups A and B) do not show a statistically significant difference in tasks 1, 3, 4, and 5 (all p values are larger than 0.05).

Then, a non-parametric test was performed on task 2. The analysis results of the non-parametric tests are shown in Table 4.

Table 4. Non-parametric test analysis results of completion time of task 2.

Task	Group median M (P25, P75)□		Mann-Whitney test statistic U value	Mann-Whitney test statistic z-value	p
	1.0(n=8)	2.0(n=8)			
Task 2's time (seconds)	0.500(0.0,1.0)	1.000(0.3,19.8)	22.000	-1.128	0.259
*p<0.05**p<0.01					

It can be seen that samples from different groups do not show a statistically significant difference on task 2 (all p values are larger than 0.05).

In conclusion, there is no statistically significant difference in users' performance on different interface modes regarding the completion time of the five tasks, indicating that Alipay's elder-friendly mode has not significantly improved the completion time.

3.1.3. SUS: System Usability Scale analysis

After the participants completed the five tasks, a questionnaire with SUS rating scales will be presented to them. They can use the five-point Likert scale ranging from "1" strongly disagree to "5" strongly agree to give their rating scores. With the rating scores, further statistical analysis was performed, as shown in Table 5.

Table 5. SUS questionnaire average score.

Descriptive statistics of SUS (n=8)						
Statistics	N	Total	Mean	Sd(S)	Min	Max
GA	8	572.5	71.5625	12.95166	60	100
GB	8	530	66.25	17.82855	45	97.5
Descriptive statistics of questions related to Usability (n=8)						
Statistics	N	Total	Mean	Sd(S)	Min	Max
GA	8	609.375	76.17188	13.03942	59.375	100
GB	8	587.5	73.4375	17.91286	50	96.875
Descriptive statistics of questions related to Learnability (n=8)						
Statistics	N	Total	Mean	Sd(S)	Min	Max
GA	8	425	53.125	27.34534	12.5	100
GB	8	300	37.5	26.72612	12.5	100

It can be seen that the SUS scores of the two modes are between OK and good, but in terms of learnability and usability in SUS, the score of the elder-friendly mode is lower than that of the normal mode. It indicates that the elder-friendly mode of Alipay has not significantly improved the usability and learnability for the elderly, and even causes unexpected difficulties for users.

To further explore the reasons for the poor user experience of Alipay, the interface design and task flow design were investigated.

3.2. Interface design analysis

In this section, we conducted an in-depth analysis of the function portals of the Alipay homepage. The heat map generated from the 30s-observation will be analyzed, and the fixation rate on AOIs of Alipay's homepage will be emphatically examined.

3.2.1. Superimposed heat map

In the experiment, 16 users were required to observe Alipay's homepage for 30 seconds before starting the tasks. The results of the analysis are shown below.

a) home page b) area of interest(AOI) c) overall heat map

Figure 1. Comparison of normal mode(the left side of each) and elder-friendly mode(the right side of each) of Alipay's home page.

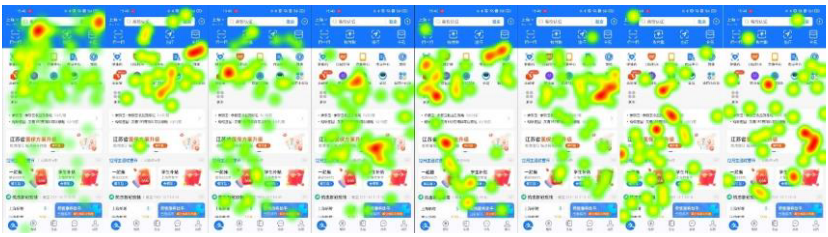

Figure 2. Heat map of fixations of 8 users using Alipay's normal mode.

Figure 3. Heat map of fixations of 8 users using Alipay's elder-friendly mode.

Comparing the heap maps, it can be seen that the fixations of users are mainly attracted by the function portals at the white central part of the page, due to the amplified icons on the elder-friendly interface, which bring about a problem that the attention to the top functional area has been reduced. However, the functions located at the top area are normally the most important and frequently used ones.

3.2.2. AOI fixation rate of Alipay's homepage

Firstly, the AOI areas of the homepage were defined according to different functions and services. Then the user's fixation rates on different AOIs of each task were calculated. The formula is the total number of fixations on each AOI divided by the total fixation duration.

The non-parametric test was used to compare the AOI fixation rates between the two interface modes in tasks 3, 4, and 5, which are all performed by novice users. The analysis results are shown in Table 6.

Table 6. Non-parametric test analysis results of fixation rate of AOI.

Task	Group median M (P25, P75)		Mann-Whitney test statistic U value	Mann-Whitney test statistic z-value	p
	1.0(n=6)	2.0(n=5)			
Task 3	4.609(0.0,13.3)	5.618 (1.2, 16.7)	14.000	-0.184	0.854
Task 4	5.005(1.6,18.8)	13.333 (6.7, 62.4)	8.000	-1.278	0.201
Task 5	0.435(0.0,4.4)	0.000(0.0,0.0)	7.500	-1.742	0.082

*p<0.05**p<0.01
Nemenyi's method for pairwise comparison of results

It can be seen that there was no significant difference regarding fixation rate (all p values are larger than 0.05).

We further analyzed why the fixation rates of the two interface modes were not significantly different, so we used the superimposed eye movement heat map, in which all the fixations of the three tasks are presented on one map.

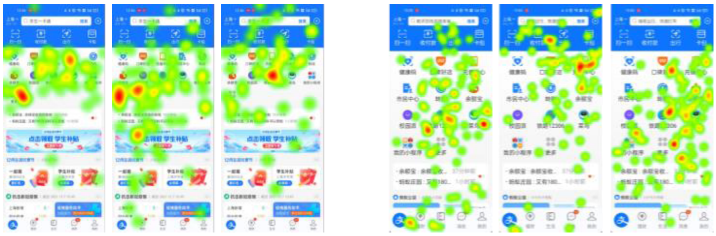

Figure 4. Superimposed eye-tracking data for task 3 4 5 in normal mode(left) and elder-friendly mode(right).

Through the comparison, it can be found that the amplification of icons and text size can effectively attract the attention of participants to important function information. Therefore, reducing the advertising panel and enlarging the function icons may help the elderly improve their information screening efficiency.

On the other hand, although larger icons are more likely to attract users' attention, if the total number of function portals remains the same, there will still be too many portals that distract the user and make it impossible to find what they want quickly. Moreover, there are some problems in the semantic understanding of some icons. For example, in task 3 and Task 4, although the average AOI fixation rate is slightly improved due to the larger icons in the normal mode, which is easier to attract users' interest. However, the overall fixation rate did not significantly increase because the interface entrance was not reduced so the attention was still distracted, indicating that simply expanding the icon area was not enough to improve usability. From the analysis of task 5, it can be seen that the "travel" icon is largely ignored by both modes. Conversely, the "buying train tickets" icon below has a higher fixation rate. It shows that "travel" has difficulty in users' semantic understanding, while "buying train tickets" is easier for the elderly to understand, which shows that no special consideration is given to semantic understanding problems in elder-friendly mode. All the problems analyzed above may lead to insignificant differences between these two modes.

3.3. Task flow design analysis

In the task flow analysis, the points and descriptions of users' errors, hesitations, delays, and all other problems that users have experienced in the experiment will be analyzed. Firstly, similar errors will be clustered. Secondly, these errors will be prioritized according to the severity and occurrence frequency. In addition, the eye-tracking data will help to verify the reasons for these problems. Based on the analysis results, the improvement strategies for the elderly will be proposed.

3.3.1. Problem clustering

After qualitative analysis, five usability problems which are: Information perception problem, Operation problem, Graphic perception problem, Semantic perception problem, and Procedural perception problem were summarized and clustered into the following three types:

1. **Information perception problem:** For middle-aged and elderly people, information perception disorder often happens mainly caused by the decline of visual and auditory abilities [10]. Therefore, information density, graphic size, and color contrast, all affect the user's perception of the interface elements. If the ratio is improper, it will lead to the obstruction of information screening and understanding.
2. **Operation problem:** Since it may be difficult for elder people to always perform precise operations, accidental touch often happens, which may lead to errors and reduce their operation efficiency [11].
3. **Understanding problem:** Including graphic perception problem, semantic perception problem, and procedural perception problem. Due to the decline of cognitive ability, it happens that elder users cannot understand the meaning of the graphic icons and the description terms during the experiment. Moreover, they have no clear idea about what’s going on in the current interface and imagine what to do next, resulting in turning the interfaces back and forth.

3.3.2. Weight evaluation method

Quantifying the usability problems encountered by users is a key metric to measure the impact of usability activities [12]. The weight evaluation method used in this study refers to the method proposed by Lewis [13]. The weights of different problems are calculated by multiplying the severity of the problem and the frequency of the problem, the range of severity depends on the values assigned to each impact level. As 4 is assigned to the most serious impact level, so the formula is divided by 4 to create a scale ranging from 0 to 100%, as shown in formula (1):

$$Weight = \frac{severity \times frequency}{4} \times 100\% \tag{1}$$

In particular, severity is assessed using a rating scale of 1-4, 1=minor, 2=moderate impact on task completion, 3= significant delay or frustration, and 4=prevention of task completion.

Table 7. Weight evaluation results of the five problems.

Impairment category	**Information perception problem**	**Operation perception problem**	**Graphic perception problem**	**Semantic perception problem**	**Procedural perception problem**
Task 3					
Frequency	0.5	0	0.25	0.25	0.3125
Severity	3	1	2	2	3
Weight (%)	15	0	5	5	9.375

Impairment category	Information perception problem	Operation perception problem	Graphic perception problem	Semantic perception problem	Procedural perception problem
Task 4					
Frequency	0.75	0.125	0.4375	0.25	0.1875
Severity	3	1	1	1	3
Weight (%)	22.5	1.25	4.375	2.5	5.625
Task 5					
Frequency	0.25	0	0.1875	0.1875	0
Severity	1	1	1	3	2
Weight (%)	2.5	0	1.875	5.625	0

From Table 7, it can be seen that the information perception problem is the most severe problem for task 3 and task 4, followed by the procedural perception problem. Through analysis combined with heat map, we found that this was due to a disproportionate distribution of design elements and unclear task flow design, respectively. Participants' attention areas are scattered in these parts, which implies that they are confused about the sequence of operations and disoriented about the purpose of the operation. For task 5, the semantic perception problem is the most serious problem and the usage descriptions for some icons are vague and easily misunderstood. These are problems that need to be specially considered in the design for the elderly and need to be improved in the future.

4. Discussion and limitations

Through a series of analyses of Alipay's elder-friendly mode, some important problems can be identified. First, simply zoom-in on the icons on the page without any other elder-friendly improvements cannot effectively improve elder users' efficiency. Second, with many function portals located, the selection of the portal desired is still complicated for the elderly. Third, the homepage of the elder-friendly mode lacks necessary guidelines, and the graphic meaning of the icons is too concise, which easily leads to obstacles for users to understand and may become "fake" elder-friendly.

To tackle the problems, corresponding design improvement suggestions were proposed. First, the systematic effects of color and physical size on the visual focus of the elderly should be considered, rather than simply magnifying size or enhancing color to attract the elderly' attention. Second, simplify the multiple choices of the same function, to reduce the screening time and the selection burden for the elderly. Third, multi-level tasks are not desirable on the same page, and even if they are, the order of operations within the same page needs to be perceivable. For complex tasks involving multiple operations, the visual clues to the purpose of the current action and the progress of the task should be presented. Finally, the icons can be designed with symbols and imagery metaphors familiar to the elderly.

Admittedly, this study also has certain limitations. In terms of sample selection, our study is based on a small sample size, and the data used may be biased, which may have limited generalizability. In terms of data collection instruments, due to the blurred vision of the elderly, the data captured by the eye tracker (Tobii Pro Glasses 2) may still have some deviations after subsequent calibration. Therefore, considerably more work will need to be done to implement usability evaluation and further research on elder-friendly design.

5. Conclusion

Generally, this study is intended to research and evaluate the usability of the elder-friendly design, and further contribute design improvement suggestions. For this purpose, the specific usability test on Alipay's elder-friendly mode was presented.

In this study, 16 middle-aged and elderly people were recruited and randomly divided into two groups for several experiments. By comparing the normal mode and elder-friendly mode, we evaluated the usability and learnability problems of the elder-friendly design in the Alipay app. It was found that the elder-friendly mode had no significant improvement. For deeper analysis, we carried out a detailed comparison of information architecture, interface design, and task flow of the two modes and identified three types of usability problems. Finally, some suggestions for elder-friendly design principles are proposed.

The contribution of this work may lie in the design strategies proposed for elder-friendly design. For example, relevance should be considered in information architecture design, which means that the content above and below the hierarchy must be relevant to improve the usability of the elderly. Consistency should be embodied in function portal design to ensure that the elderly can recognize the portals at any interface. Moreover, the task flow should be straightforward and as visual as possible on the interface. Furthermore, various new interactive methods and technical means such as voice, vision, gestures, etc, can be tried to further enhance the elder-friendly design. With the interdisciplinary design improvement suggestions, it is expected to solve some of the usability difficulties of the current elder-friendly design, which could also offer some important insights for further study.

References

[1] H.M. Salman, W.F.W. Ahmad, and S. Sulaiman, Usability evaluation of the smartphone user interface in supporting elderly users from experts' perspective, *IEEE Access,* Vol. 6, 2018, pp. 22578–22591.

[2] J. Tavares, G. Santinha, N.P. Rocha, Age-friendly health care: a systematic review, *Healthcare*, 2021, Vol. 9(1), 83.

[3] J. J. Chang, N. S. H. binti Zahari, Y. H. Chew, The Design of Social Media Mobile Application Interface for the Elderly, In: *2018 IEEE Conference on Open Systems (ICOS),* 2018, IEEE, pp. 104-108.

[4] ISO, W, 9241-11, Ergonomic requirements for office work with visual display terminals (VDTs), *The international organization for standardization,* Vol. 45, 1998.

[5] J. Nielsen, *Usability engineering,* Academic Press, San Diego, 1994.

[6] World Health Organization, *World report on ageing and health,* World Health Organization, 2015.

[7] H. Manaf, Barriers to participation in physical activity and exercise among middle-aged and elderly individuals, *Singapore Med J,* Vol. 54, 2013, pp. 581–586.

[8] A. Bangor, P. T. Kortum, and J.T. Miller, An empirical evaluation of the system usability scale, *Intl. Journal of Human-Computer Interaction,* 2008, Vol. 24, pp. 574–594.

[9] R.J. Holden, N.L. Campbell, E. Abebe, D.O. Clark, D. Ferguson, K. Bodke, M.A. Boustani, C.M. Callahan, et al, Usability and feasibility of consumer-facing technology to reduce unsafe medication use by older adults, *Research in Social and Administrative Pharmacy,* 2020, Vol. 16, pp. 54–61.

[10] J. Fletcher and R. Jensen, Mobile health: barriers to mobile phone use in the aging population, *On-Line Journal of Nursing Informatics,* Vol. 19(3), 2015, http://ojni.org/issues/?p=3364, accessed June 20, 22.

[11] E. Paez, Lesly, and C. Zapata, Elderly users and their main challenges usability with mobile applications: A systematic review, In: *International Conference on Human-Computer Interaction,* 2019, pp. 423–438.

[12] E. Goodman and M. Kuniavsky, *Observing the user experience: A practitioner's guide to user research,* Elsevier, Amsterdam. 2012.

[13] J. Sauro and J.R. Lewis, *Quantifying the user experience: Practical statistics for user research,* Morgan Kaufmann, Cambridge, 2016.