AN EXPLORATION TO USER EXPERIENCE OF A MOBILE TABLET APPLICATION THROUGH PROTOTYPING

CHRYSOULA GATSOU*

School of Applied Arts, Hellenic Open University
Aristotelous 18, 26 335, Patra, Greece,
Technological Educational Institute of Athens
Ag. Spiridonos , 122 10, Athens, Greece
cgatsou@teiath.gr

ANASTASIOS POLITIS

Technological Educational Institute of Athens
Ag. Spiridonos , 122 10, Athens, Greece
politisresearch@techlink.gr

DIMITRIOS ZEVGOLIS

School of Applied Arts, Hellenic Open University
Aristotelous 18, 26 335, Patra, Greece
zevgolis@eap.gr
http://dzevgolis.eap.gr/

Abstract— One key factor in inexperienced users adopting mobile technology is the ease of use of mobile devices. The present paper investigates inexperienced user performance through a usability testing of three alternative prototypes of a mobile tablet application. For this purpose of the study, three prototypes have been investigated, built on the basis of previous research conducted in collaboration with users. The investigation takes place with a usability testing of these three alternative prototypes of a mobile tablet application. Our study involves five navigation tasks which novices users were asked to complete with each of the three prototypes. Our results showed that participants displayed better task performance with the interface structure and layout prototype F1, which was created in collaboration with participants, in contrast to prototypes interface F2 and F3 that have been designed without the collaboration of the users, which both caused navigation problems.

Keywords: mobile devices; mobile interaction; user experience; prototyping; usability testing; human factors.

1. Introduction

One main challenge regarding the various mobile devices, whose novice users are increasing exponentially, is the matter of interface design and layout. The challenge in designing proper interfaces based on human – interface interaction – involves examining

---

* cgatsou@teiath.gr, 25 Botasi, 185 37 Piraeus Greece

the experience of significant numbers of ordinary people who lack skills in computer science and who are drawn from a wide range of backgrounds. As has been noted by Hassenzahl (2002), there is no guarantee that users will actually perceive and appreciate the product in the way interface designers desire it to be perceived and appreciated. For example, a product with a specific screen layout intended to be clear and simple will not necessarily be perceived as such by novice users. Furthermore, if an interface is comprehensible within one culture, it may not be so in others. Despite the best efforts on the part of designers, new technologies often fail to meet basic human needs and desires [Norman (1999)]. The difficulties concerned in designing an interface that will deal effectively with individual preferences and experience, while minimizing frustration on the part of the user, transfer errors and   learning effort, is widely recognized as a persistent problem in human-computer interaction [Benyon and Murray (1993)]. While much has been written about what makes an interface usable, the interface itself may not always be understood, in particular by those lacking a background in information technology. Making things more usable and accessible is part of the larger discipline of user-centered design (UCD), which includes a number of methods and techniques [Rubin and Chisnell (2008)]. Usability testing is a method used to evaluate a product by testing it on representative users. Greenberg and Buxton (2008) point out that *"Usability evaluation is valuable for many situations, as it often helps validate both research ideas and products at varying stages in its lifecycle"* . The method we used to gather data in usability testing was the so-called 'think aloud protocol', in which the researcher observes, while the user completes, or attempts to complete, a defined task. Prototyping is an essential part of usability testing, as it confirms whether users can effectively complete tasks by means of the prototypes that are being tested and allows us to deal with various types of problems. Furthermore, prototypes can also be useful in dealing with the more subjective aspects of an interface. A previous study by the present authors has shown that inexperienced users structure content information in a mobile tablet application differently from experienced users, when the former interact with mobile devices [Gatsou *et al* (2012)]. Carroll argues that an effective way of dealing with system complexity for the inexperienced user is to provide a functionally simple system [Carroll (1999)]. In order to create more affordable mobile interactive artifacts for inexperienced users, we have focused on the interface design of a mobile tablet application and tested it on real users. The goal of this study is to investigate the effect of different interfaces in usability testing with regard to inexperienced user performance   and the perceived usability of a tablet mobile application. To present the results of our study, we start our paper with a review of the literature, which establishes the theoretical background for our study. We then describe the research methodology employed. We analyse the data and give our results, which we discuss, before offering some conclusions.

2. Background

2.1. *The think-aloud protocol*

The think-aloud protocol is frequently used as a method in testing usability. The researcher observes, while the user completes, or attempts to complete, a defined task, the observer intervening only to remind the user to talk aloud while solving the problem or performing the task in question. If the verbal protocol is recorded and then listened to after the experiment, this may help explain behaviour that cannot easily be interpreted on the basis of observation alone and help highlight cognitive activities that otherwise remain invisible [Hackos and Redish (1998)].

Think - aloud methods have long been used in psychology in order to study task-based cognitive processes [Ericsson and Simon (1984)]. Ericsson and Simon recognize two basic types of think –aloud protocols/methods, the concurrent think-aloud, in which users are asked to verbalize their thoughts while performing the tasks, and the retrospective think-aloud, where users express their thoughts after completing the task. A study by Van den Haak and De Jong (2003) point out that the two methods produce similar numbers and types of problems. Carroll et al. (1987) used thinking aloud studies to investigate how learners interacted with new software. Generally speaking, the method consists of collecting think-aloud protocols in a systematic way and analyzing the protocols to obtain the users' own model of their cognitive process. The advantage of using the think-aloud method is that the researcher can spot any usability problems that a user may face in interacting with a prototype and can analyze users' behaviour in dealing with the task set them. A representative sample of the tasks to be used in the think-aloud method is vital for revealing usability problems.

Boren and Ramey (2006) point out that most literature on the use of think-aloud in usability testing does not follow the model created by Ericsson and Simon, and that how think-aloud is actually practiced by investigators deviates significantly from Erisson and Simon's prototype. Boren and Ramey later investigated the think-aloud protocol in relation to usability studies and discovered that the methodology that Ericsson and Simon describe, although frequently referenced, was not actually used by the practitioners. They found that practitioners failed to give participants appropriate instructions and in an appropriate way. Then, having analyzed how usability researchers implement the think-aloud protocol, they developed a new methodology for employing usability testing. They argue that in some cases speech communication can provide more efficient rules for achieving the results that Ericsson and Simon were interested.

The method has been criticized for its limitations as regards detecting how useful the system itself really is [Norgaard and Hornbak (2006)], but the method remains a very practical and popular approach to recording feedback and assessing the effectiveness of the user interface. Of course, it is possible that, if the user is distracted by being required to speak at the same time as performing the task in question, s/he may not perform as well as s/he might otherwise. Yet, in general think-aloud remains a qualitative method appropriate for investigating the problems typical users will experience in their interactions with artifacts. As Nielsen (1993) point out   *"the method has a host of advantages. Most important, it serves as a window on the soul, letting you discover what users really think about your design"*.

2.2. *Prototyping*

Prototyping is an essential procedure that structures design innovation. The translation of user needs into a system specification was facilitated in our study by the use of iteratively refined prototypes validated by users. Beaudouin-Lafon and Mackay (2003) define any prototype as a concrete representation of part or all of an interactive system. A prototype is, in their view, a tangible artifact, rather than an abstract description that requires interpretation. In Moggridge's, view, a prototype is a representation of a design made before the final solution has been found [Moggridge (2007)]. By offering different prototypes of a mobile application to users and requesting feedback, we can be sure that we are designing for those who will actually use our designs. Prototyping serves various purposes in a human-centered design process. With a view to improving this process, we developed, as a continuation of two previous studies of card sorting and creative sessions, three interactive prototypes, in order to explore more intuitive navigation methods for inexperienced users who are to interact with mobile tablet devices. Our decision to construct three, rather than one prototype, rests on work by Tohidi et al. and Dow et al., who argue that multiple prototypes are more helpful than merely one in aiding users to formulate negative or positive comments [Tohidi *et al* (2006)],[Dow *et al* (2009)]. Moreover, we were able to verify whether or not the prototype that we created on the basis of user requirements has responded effectively to these. Houde & Hill (1997) classify the ways in which the prototypes can be of value to designers. Prototypes, in their view, include any representational design idea, regardless of the medium involved. Their model defines three types of issues that a prototype may affect, namely, the role of a product in the context in which it is used, the look and feel of the product and its technical implementation. Floyd, in an earlier laboratory prototyping of complex software systems, describes two primary objectives of prototypes, namely, 1) to act as a vehicle for learning and 2) to enhance communication between designers and users, as developer introspection of user needs often leads to inadequate products [Floyd (1984)]. Buchenau & Suri (2000) note that the prototypes perform an additional function. They strengthen empathy, "an original experience every kind of representation, in any medium, that is designed to understand, to explore or communicate what you could work with the product, space or system design". Lim *et al* (2008) stress the role of prototypes as a vehicle for learning, "prototypes are the means by which designers organically and evolutionary learn, discover, generate, and refine designs" .

2.3. *Usability testing*

The term "usability" is frequently employed in the field of human-computer interaction (HCI). For Nielsen (1994), usability is a matter related to the broader issue of acceptability. In his view, "Usability is a quality attribute that assesses how easy user interfaces are to use". Usability is a significant part of the user experience and therefore of user satisfaction. A formal definition of usability is given in the ISO standard 9241–11 : "…the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction, in a specified context of use". Effectiveness is defined as the accuracy and completeness with which users achieve specified goals and efficiency as the resources expended in relation to the accuracy and

completeness with which users achieve goals. Satisfaction is defined as the freedom from discomfort, and positive attitude to the use of the product, whilst the context of use is defined as users, tasks, equipment and the physical and social environments in which a product is used [ISO-9241-11(1998)].

Usability testing is a method employed in user-centered design to evaluate product design by testing it on representative users. Such users thus yield quantitative and qualitative data in that they are real users performing real tasks. In the view of Rubin and Chisnell (2008), usability testing allows the researcher to collect empirical data, while observing representative end-users using the product to perform realistic tasks. Usability testing requires an artifact that is fairly complete and rationally designed, which means that the appropriate place for usability testing is at a stage quite late in the design cycle [Cooper (2007)].

Dumas and Redish (1999) argue that usability testing is a "a systematic way of observing actual users trying out a product and collecting information about the specific ways in which the product is easy or difficult for them". They also recommend that usability test possess the following five features:
1. The primary goal is to improve the usability of a product. For each test, you also have more specific goals and concerns that you articulate when planning the test.
2. The participants represent real users.
3. The participants do real tasks.
4. The researcher observes and records what participants do and say.
5. The researcher analyzes the data, diagnoses any problems, and recommends solutions to such problems.

There are normally three participants involved in a usability test, namely a test person, a test moderator and an observer. The test can be carried out in a professional test lab, but frequently an ordinary meeting room or the place where the test person might possibly use the product in reality is equally adequate.

3. Research methodology.

To examine how novice users conceptualize a mobile tablet application, we created a user test involving three prototypes of a mobile tablet application themed around the topic of "first aid" (Fig.1). All three interfaces had the same "look and feel", so as to standardize the visual appeal and the emotional impact made by the various alternative versions employed in the test. These versions vary in terms of conceptual models and menu navigation, one of them, F1, having been created on the basis of the participant collaboration in previous studies by the present authors [Gatsou *et al* (2012a)], [Gatsou *et al* (2012b)].

Fig. 1. Participant during the usability testing

3.1. *Participants*

The literature gives no clear optimum number of participants to be employed in usability testing. Nielsen (2000) argues that five participants will discover 80% of the problems in a system. In any case, a small amount of users, that is, generally fewer than 10 subjects, is sufficient for any formative evaluation of usability [Petrie and Bevan (2009)]. On the other hand, Spool and Schroeder (2001) state that five users identified only about 35% of the problems in a website. The research by Turner *et al* (2006) implies that a group size of seven may be optimal, even when the study is fairly complex.

 In the view of Sauro and Lewis (2012) "the most important thing in user research, whether the data are qualitative or quantitative, is that the sample of users you measure represents the population about which you intend to make statements" . Our session was designed specifically to include a pool representative of potential users of the mobile application that we were testing. Twelve participants (N=12) ranged from 18 to 79 (mean age = 41,6, SD = 20.9, years), seven of whom were men and five women, all of whom had participated in one or more previous studies. All participants were novices in terms of computing. They had no visual or cognitive impairment and their education was of at least high school level. Given the evidence from our previous studies, the number of people in this experiment was sufficient to provide satisfactory evidence and depth for us to study. The age and gender of the participants is shown in Table 1.

Table 1. Age, gender and number of Participants.

ID	P1	P2	P3	P4	P5	P6
Age	50	38	26	27	57	79
Gender	M	F	M	F	M	M
ID	P7	P8	P9	P10	P11	P12
Age	52	31	18	45	65	67
gender	M	M	F	F	F	M

3.2. *Materials*

Our usability testing was performed on a Dell Inspiron Duo, 10.1tablet computer with a touch screen. A Panasonic HDC-SD40 digital camera was used to create a complete record of all user interactions with the interface. Furthermore, Camtasia Studio software was used to record a video of the movements made by the user on the interface during the test. Camtasia studio software captures the action and the sound from any part of the desktop. A digital tape recorder was also used.

3.3. *The three prototypes*

To reproduce a realistic software environment, over a period of three months three prototypes were developed in Adobe Flash, which were used as a tool for recording user behaviour during interaction. Prototypes help designers to balance and resolve problems that occur in different dimensions of design. Each prototype allowed the user to interact with mobile application and to carry out various tasks.

3.1.1. *Interface F1*

The first screen of the interface consists of icons that offer easy accessibility to the topic. We settled on this layout after a participatory session with users involved in our previous study [Gatsou *et al* (2012b)]. There we concluded that users preferred icons for main menu selection, rather than a representation of options in words arranged hierarchically.
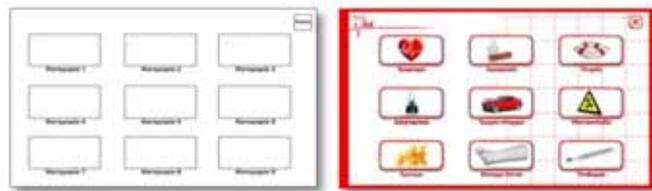
Fig. 2. The Interface F1

3.1.1. *Interface F2*

The colours remain the same in prototype F2, but the main menu has now been moved to the left of the screen and now employs words, instead of icons. The options are the same in number as in the prototype F1. The   subcategories are now placed in the middle of the screen. The aim of this layout was to explore whether a larger amount of text helps or hinders the inexperienced user to interact with a mobile application.

Fig. 3. The Interface F2

3.1.1. *Interface F3*

Prototype F3 is identical in basic design to prototype F2, except for a horizontal bar at the top of the screen, which enables the user to select subcategories. This layout resembles that of a website. The aim of this arrangement, which simulates the web environment, was to test the familiarity of users with little experience of surfing.

Fig. 4. The Interface F3

3.4. *User Tasks*

For the usability test, the participants were required to complete the five tasks given in Table 2. The tasks were chosen as being representative and covered as many as possible of the features of the application.

Table 2. Participant's tasks.

Task 1	Turn on the mobile tablet device and select the icon "first aid".
Task 2	Find the information on Cardiopulmonary resuscitation (CPR.)
Task 3	Enlarge the image in order to see details.
Task 4	Select information on heart attacks.
Task 5	Find the information on symptoms of broken bones. Turn off the mobile device

3.5. *User Performance*

User performance was recorded in terms of the effectiveness, efficiency and ease of use of prototypes. In order to evaluate task effectiveness, we measured the percentage of tasks successfully completed within the set time limit. Task completion time refers to the time needed to accomplish the task. To evaluate efficiency, we recorded the time needed

to process a task. To measure user satisfaction, we asked users to complete a post-test questionnaire.

3.6. *Post- test Questionnaire*

The main aim of administering written questionnaire after the test (post-test questionnaire) is to record participants' preference, in order to identify potential problems with the product. Information collected usually includes opinions and feelings regarding any difficulties encountered in using the product. Our questionnaire was based on the System Usability Scale (SUS) developed by Brooke (1996), since this is the most precise type of questionnaire for a small number of participants, as is shown by Tullis and Stetson's (2004) study. SUS employs a "quick and dirty" approach in evaluating the overall subjective usability of a system (Appendix A). While SUS was originally intended to be used for measuring perceived usability, i.e. measuring a single dimension, recent research shows that this provides an overall measure of satisfaction of the   system [Tullis and Stetson (2004)],[Lewis and Sauro (2009)],[Sauro (2011)]. In addition to these advantages over other systems, the SUS is a powerful and multifunctional instrument [Bangor *et al* (2008)].

3.7. *Test Protocol*

Participation in the study lasted approximately one hour and twenty minutes and was conducted in an isolated room in our department. It consisted of the series of tasks that we mention above. All participants were tested individually.

After being welcomed by the experimenter, participants were told that they were to take part in a usability test and were to work with a prototype of a mobile tablet application. All participants gave their permission to be recorded on video. Subsequently participants completed the five tasks. The process of user testing is illustrated in Fig. 5. To minimize the potential for learning bias, the presentation order of the prototypes was counterbalanced.
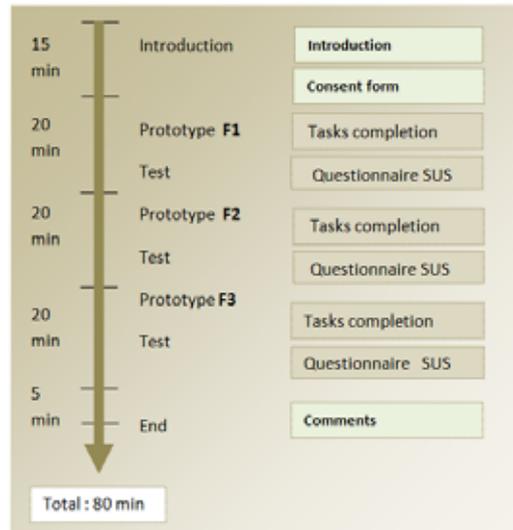
Fig. 5. User test process.

4. Results analysis.

The main factors to be examined when testing usability are effectiveness, efficiency and user satisfaction. Effectiveness refers to how "well" a system does what it supposed to do. In order to evaluate task effectiveness, we measured the percentage of steps successfully solved within the time limit (7min). Efficiency refers to how quickly a system supports the user in what he wants to do. To evaluate efficiency, we recorded the time needed to process the task. Satisfaction refers to the subjective view of the system on the part of the user [Rubin and Chisnell (2008)]. Qualitative and quantitative data were collected from each participant. Qualitative data included the participants' verbal protocol in video recording.

Problems of usability were identified and categorized. We also collected comments on the prototypes and preference data and evaluations in the form of the SUS data questionnaire completed by the users after the test. Any user action that did not lead to the successful completion of a task we defined as error.

4.1. *Effectiveness*

The percentage of users that manage to complete a task successfully thus becomes a measure of the effectiveness of the design. The number of errors made on the way to completing a task is an example of a performance measure [Rubin and Chisnell (2008)]. An interaction effect is noticeable in the results, suggesting that the approach employed in the interface F1 may well have a marked impact on reducing the number of errors made. Tables 3, 4 show the user tasks and the error rate.

Table 3. Tasks completion rates.

	Prototype F1	Prototype F2	Prototype F3
Task1	12/12	11/12	12/12
	100%	91%	100%
Task2	11/12	9/12	6/12
	91%	75%	50%
Task3	9/12	8/12	7/12
	75%	66%	58%
Task4	9/12	7/12	4/12
	75%	58%	33%
Task5	11/12	7/12	6/12
	91%	58%	33%

4.2. *Qualitative results*

Qualitative data includes the participants' verbal protocols, as recorded during the videoing of the session. Qualitative research ″can be used to reveal and interpret whatever lies behind any phenomenon of which we know little. It can be used in the search for new ideas on matters that are already fairly well known…and it can yield complex details of phenomena that can tied to quantative methods only with difficulty ″ [Strauss and Corbin (1990)]. Participants' verbal reports were transcribed and analyzed by means of qualitative content analysis. In order to understand the meaning of these qualitative results, errors gathered from the verbal data given by participants was analyzed further. The analysis of the text of the recording was performed in successive steps, by means of the open technique of codification of Strauss and Corbin (1990). These oral phrases were decoded and then analysed in combination with the appropriate screen of the application. The content analysis tends to be inductive, as it goes through the data and looks for dominant patterns.

The basic idea is to locate sections from the transcripts that offer the most guidance and to derive from them the important messages hidden in the verbal expressions used by the participant. Having listened repeatedly, we classified the still codified sections in broad categories. During the initial processing of the recorded data, it became clear some phrases were errors. During second processing, the codified sections were assigned a descriptive label (e.g. 'wrong key stroke', 'navigation failure', 'failure to recognize graphics'). During the third processing, we defined the broader categories that arose from grouping together all labels of the same type. A word-processing programme and Microsoft Excel was used to record and examine this data. Errors were assigned to one of two categories, navigation and comprehension, depending on the activities demanded by each task.

Questions of navigation arose when the user did not proceed as the application expected him/her to or when s/he was unable to choose the piece of information required. Errors of comprehension arose when the users failed to understand the significance of the graphics. Tables 4 to 8 show the frequency with which such problems appeared in each category.

Table 4. Types of errors by prototype.

Type of error	Prototype F1	Prototype F1	Prototype F1
Navigation	3	7	12
Comprehension	3	6	6
Total	6	13	18

4.3. *Efficiency - Task Completion Time*

We recorded the total amount of time required to complete each task in prototypes F1,F2 and F3 (Fig.6), starting from turning the device on to turning it off. The mean amount of time required by participants in each age group is shown in Fig.7. Participants P6, P11, P12 failed to complete their tasks in prototype F2 within the time set (7min). In prototype F3, participants P2, P6, P11, P10, P12 failed to complete their tasks. Table 5 shows the results of the mean completion time and standard deviation for the participants for prototypes F1, F2 and F3. Data regarding time taken by each participant for each task is given in Appendix B.

Participant	Age	F1	F2	F3
P1	50	0:02:41	0:03:52	0:05:43
P2	38	0:02:49	0:03:35	
P3	26	0:02:10	0:04:11	0:04:26
P4	27	0:02:53	0:03:28	0:04:42
P5	57	0:02:15	0:02:40	0:05:07
P6	79	0:05:47		
P7	52	0:02:55	0:04:22	0:06:06
P8	31	0:02:45	0:03:33	0:04:30
P9	18	0:02:10	0:03:02	0:04:11
P10	45	0:02:23	0:04:53	
P11	65	0:03:58		
P12	67	0:04:26		

Fig. 6. The tasks completion mean time (seconds).

For users testing the prototype F1, the time needed to complete tasks ranged between 2:10 min and 2:53 min up to the age of 57. For participants aged 57 years or older, task completion time increased. This affected mean task completion time and standard deviation. For prototype F2 tasks, completion times were clearly higher. Participants older than 57 failed to complete their tasks within the specified time. The mean completion time of those who did finish their tasks was 20.43%   greater than the corresponding figure in prototype F1. For prototype F3, the mean completion time for those who succeeded in finishing was 33.04% greater than the corresponding figure in prototype F1.

Table 5. Task completion time (mean, SD)

	Prototype F1	Prototype F2	Prototype F3
Task completion time (mean)	03:06	03:44	04:58
Standard Deviation (SD)	01:06	00:41	00:43

Elderly users were thus not able to complete all the tasks in prototype F2 and F3 and specifically in prototype F3, where the layout of the prototype was slightly different, in that it resembled a web site. They had more information to process located on the left and at the top of the screen. These users found the interaction difficult to understand and to activate. On the whole,   all users were more comfortable when interacting with prototype F1.

Fig. 7. Task completion time per interface type.

4.4. *Post test Questionnaire*

We realised that time-on-task measures can be useful for collecting data on the efficiency of a system. On the other hand, such data does not give any information on overall

satisfaction on the part of the user. User satisfaction may be an important factor in motivating people to use a product and may affect user performance. So, as a final point we decided participants were to complete an SUS questionnaire, so as to explore their experiences when interacting with the prototypes. A crucial feature of the SUS lies in the fact that it requires the user to evaluate the system as a whole, rather than specific aspects.

All 10 questionnaire statements having been processed, the overall SUS score for each prototype turned out to be that given in Table 6. To calculate the SUS score, first we summed the score contributions of the items 1, 3, 5, 7 and 9 (Appendix A). The score contribution of these items are their scale position minus one. We then summed the score contributions of the other items: five minus their scale position. Finally, we multiplied the sum of the scores by 2.5, to obtain the overall score with a range between 0 to 100.

Table 6. Overall SUS Score

Participants	F1	F2	F3
P1	80.0	70.0	70.0
P2	82.5	70.0	25.0
P3	90.0	82.5	60.0
P4	95.0	82.5	72.5
P5	87.5	80.0	72.5
P6	65.0	25.0	25.0
P7	77.5	70.0	27.5
P8	75.0	75.0	75.0
P9	92.5	82.5	75.0
P10	82.5	72.5	25.0
P11	70.0	25.0	25.0
P12	70.0	25.0	25.0
Mean	**80.6**	**63.3**	**48.1**

The survey results showed the overall satisfaction. Sauro (2011) reports that a mean value over 74 is level B, value above 80.3 is level A. An average value of below 51 is level F (fail). The prototype F1 with an average value of 80.6 passes the threshold of 80.3 and are to be placed on level A, F2, with an average value of 63.3, belong to level B and F3, with a value of 48.1, is to be placed at Level F, which is regarded as failure.

However, with respect to F1, nearly all participants preferred the interface with the icons over the other two interfaces (F2, F3), in which there was a large amount of text. Some of the participants simply misunderstood the graphics keys that depicted a lens and whose purpose was to increase the photographs on the screen and the arrows that represented the act of selecting the next screen. If perhaps users had understood the graphics more fully,

the error rate for prototype F1 may perhaps have been as low as zero. Overall users liked the process and regarded their interaction with the prototypes positively. Nevertheless, in some cases, the participants were apprehensive. Uncertain in their selections, they demanded greater confirmation and reassurance about the actions they were to take. In such cases, it is important for the researcher to motivate participants, encouraging them discreetly to investigate alternative directions, while simultaneously recording any mistakes made. As for individual prototypes, participants prefered the design of the first interface, which contained icons (F1). This was to be expected and users commented positively on its simplicity, ease of use and intuitiveness.

5.   Conclusion

The aim of our study was to examine whether an interface design approach could improve performance of tasks by inexperienced users during interaction. To do this, we employed three different prototypes of the same application. We tested our empirical methodology on twelve individuals, all of them novices in terms of computer use.

One of the most remarkable discoveries we made is the large degree of difference in performance among the three different prototypes with regard to user effectiveness and the number of errors. The effectiveness and efficiency of the F1 prototype is evident in the fact that users made fewer errors and took less time to complete their tasks. Participants reported that the icon menu of the F1 prototype facilitated the execution of their tasks, as did the absence of text in menu selections. This confirms what emerged from a previous study by the present authors. Our findings imply that the users did not understand the basic conceptual models informing prototypes F2 and F3   [Norman (1999)].
The usability test performed on each of the prototypes showed that most users considered the prototype easy to use and intuitive. When evaluated by SUS, the same prototype received an overall score which placed it on level A. The test also helped in locating various issues regarding the other two prototypes F2 and F3 and, in particular, regarding what is to be corrected, so as to improve its usability for the elderly. The need for usability testing is rapidly increasing with each new artifact. However, we believe that our paper, which focuses more on the users and their cognitive abilities, offers a new insight into how inexperienced users perform tasks on mobile tablets.

References

Bangor, A.; Kortum, P.; Miller, J. (2008): An empirical evaluation of the system usability scale, International Journal of Human-Computer Interaction, **24**(6), pp. 574–594.

Beaudouin-Lafon , M. ; Mackay, W. (2003). *Prototyping Tools And Techniques*. In: J. A. Jacko and A. Sears (Eds) The Human-Computer Interaction Handbook. Lawrence Erlbaum Associates.

Benyon, D.; Murray, D. (1993): Applying user modeling to human-computer interaction design. Artificial Intelligence Review, 7(3-4): pp.199-225.

Brooke, J. (1996): SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), Usability Evaluation in Industry (S. 189 -194). London: Taylor and Francis.

Buchenau, M.; Fulton Suri, J. (2000): Experience prototyping. In Proceedings of Design of Interactive Systems New York: ACM Press, pp. 424-433.

Boren, M. T. ; Ramey, J. (2000): Thinking Aloud: Reconciling Theory and Practice, IEEE Transactions on Professional Communication, **43**(3), 261- 277.

Carroll, J. (1999): *The Nurnberg Funnel*, Cambridge, Mass.: MIT Press.

Carroll, J. ; Smith-Kerker, P.; Ford, J. ; Mazur-Rimetz , S. (1987): The minimal manual, Human Computer Interaction, **3**, pp. 123-153,.

Cooper, A.; Reimann,R.; Cronin, D. (2007): *About Face 3: The Essentials of User Interface Design.* John Wiley & Sons, Inc.

Dow, S. ; Glassco, A.; Kass, J.;Schwarz, M.; Klemmer, S. (2009): The effect of parallel prototyping on design performance, learning, and self-efficacy. Tech. Rep. CSTR-02, Stanford University.

Dumas, J.; and Redish, J. (1999): *A Practical Guide to Usability* Testing (revised Ed.). Portland, Oregon: Intellect Books.

Ericsson, K. A.; Simon, H. A. (1984): Protocol analysis. Verbal reports as data. Cambridge, MA: MIT Press.

Floyd, C. (1984): Systematic Look at Prototyping. In Budde, ed., Approaches to Prototyping. Springer Verlag, pp. 105-122.

 Gatsou ,C ; A, Politis, A; Zevgolis, D. (2012a): Text vs visual metaphor in mobile interfaces for novice user interaction. In Proceedings of the 16th International Conference on Electronic Publishing, ElPub, pp.125-135.

Gatsou ,C. ; A, Politis, A. ; Zevgolis, D. (2012b)Novice User involvement in information architecture for a mobile tablet application through card sorting. In proceedings FEDSIS-MMAP, Wroclaw, pp. 711–718.

Greenberg, S. ; Buxton, B. (2008): Usability evaluation considered harmful (some of the time). Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human factors in Computing Systems, Florence, pp.111-120.

Hackos, J. ; Redish, J.(1998): *User and task analysis for interface design.* New York: John Wiley & Sons, Inc.

Hassenzahl, M. (2002):The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction, **13**(4), pp. 479-497

Houde, S.; Hill, C. (1997): What do prototypes prototype? In M. Helander, T. Landauer, & P. Prabhu (Eds.), Handbook of human-computer interaction ,2nd ed., Amsterdam: Elsevier Science, pp. 367-381.

ISO 9241-11 (1998): Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11, Guidance on usability, ISO.

Lim, Y.; Stolterman , E.; Tenenberg, J . (2008): The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. ACM Trans. Comput.-Hum. Interact. **15**, 2 , pp1-27.

Moggridge , B. (2007): *Designing Interactions.* The MIT Press.

Nielsen, J. (1993): *Usability Engineering*. Boston, MA, Academic Press.

Nielsen, J.(1994): Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. In R. G. Bias & D. J. Mayher (Eds.), Cost-Justifying Usability Boston, MA: Academic Press. pp. 242-272.

Nielsen, J. (2000): Why You Only Need to Test With 5 Users. Jakob Nielsen's Alertbox

Norgaard, M.; Hornbaek, K. (2006): What do usability evaluators do in practice?: an explorative study of think-aloud testing. In: Proceedings of DIS06: Designing Interactive Systems: Processes, Practices, Methods, & Techniques pp. 209-218.

Norman, D. (1999): *The Invisible Computer:* Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution. MIT Press, Cambridge, MA, USA, 1999.

Petrie, H.; Bevan, N. (2009): The evaluation of accessibility , usability and user experience In: The Universal Access Handbook, C Stepanidis (ed), CRC Press, pp. 299–315.

Rubin, J. ; Chisnell, D. (2008): Handbook of Usability Testing. (2nd Ed.). Indianapolis Wiley Publishing.

Sauro, J.; Lewis, J. (2009): The factor structure of the system usability scale, Proc. Human Computer Interaction International Conference , San Diego, CA, pp. 94–103.

Sauro, J. (2011): Does prior experience affect perceptions of usability? Available at http://www.measuringusability.com/blog/prior-exposure.php, January 19, 2011 [Nov.15, 2012]

Sauro,J.; Lewis, J. (2012): Quantifying the user experience: Practical statistics for user research. Burlington, MA: Morgan Kaufmann.

Spool, J.; Schroeder, W. (2001): Testing web sites: Five users is nowhere near enough", In: Proceedings of the Conference extended abstracts on Human Factors in Computing Systems, CHI'2001. New York: ACM Press.

Strauss, A.L., Corbin, J.M., (1998). *Basics of Qualitative Research*, second ed. Sage Publications, Thousand Oaks, CA.

Tohidi, M.; Buxton, W.; Baecker, R.; Sellen, A. (2006): Getting the right design and the design right. In proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '06. ACM, New York, NY, pp. 1243-1252.

Tullis, T.; Stetson, J. (2004): A comparison of questionnaires for assessing website usability," In Proc. of the Usability Professionals Association (UPA), pp. 7–11.

Turner, C.; Lewis, J.; Nielsen, J. (2006): Determining usability test sample size. In W. Karwowski (ed.), International Encyclopedia of Ergonomics and Human Factors Boca Raton, FL: CRC Press, pp. 3084-3088.

Van den Haak, M.J. ; De Jong, M.D.T. (2003): Exploring Two Methods of Usability Testing: Concurrent versus Retrospective Think-Aloud Protocols. IEEE International Professional Communication Conference Proceedings Piscataway, New Jersey.

Appendix A. System Usability Scale

	Strongly disagree				Strongly agree
1. I think that I would like to use this application frequently.					
	1	2	3	4	5
2. I found the application unnecessarily complex.					
	1	2	3	4	5
3. I thought the application was easy to use.					
	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this application.					
	1	2	3	4	5
5. I found the various functions in this application were well integrated.					
	1	2	3	4	5
6. I thought there was too much inconsistency in this application.					
	1	2	3	4	5
7. I would imagine that most people would learn to use this application very quickly.					
	1	2	3	4	5
8. I found the application very cumbersome to use.					
	1	2	3	4	5
9. I felt very confident using the application.					
	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this application.					
	1	2	3	4	5

Appendix B. Data regarding time taken by each participant for each task.

"F1" Task	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	mean time	st.dev.
E1	20	25	19	18	22	33	24	25	17	18	24	27	23	5
E2	25	31	37	42	29	110	55	35	29	20	49	35	41	24
E3	60	56	28	41	35	122	38	52	39	25	55	60	51	25
E4	35	30	24	43	25	46	21	29	21	38	71	111	41	26
E5	21	27	22	29	24	36	37	24	24	42	39	33	30	7
SUM (sec)	161	169	130	173	135	347	175	165	130	143	238	266	186	
SUM (min:sec)	0:02:41	0:02:49	0:02:10	0:02:53	0:02:15	0:05:47	0:02:55	0:02:45	0:02:10	0:02:23	0:03:58	0:04:26	0:03:06	

"F2" Task	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	mean time	st.dev.
E1	33	34	38	32	27		30	33	25	42			33	5
E2	40	39	60	38	34		53	40	37	50			43	9
E3	73	67	54	65	40		59	65	56	67			61	10
E4	47	40	53	39	30		77	39	34	86			49	19
E5	39	35	46	34	29		43	36	30	48			38	7
SUM (sec)	232	215	251	208	160	0	262	213	182	293	0	0	224	
SUM (min:sec)	0:03:52	0:03:35	0:04:11	0:03:28	0:02:40	0:00:00	0:04:22	0:03:33	0:03:02	0:04:53	0:00:00	0:00:00	0:03:44	

"F3" Task	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	mean time	st.dev.
E1	52		37	41	43		54	39	34				43	8
E2	60		45	52	57		73	46	42				54	11
E3	92		60	62	65		89	63	67				71	13
E4	81		81	79	85		92	80	55				79	11
E5	58		43	48	57		58	42	53				51	7
SUM (sec)	343	0	266	282	307	0	366	270	251	0	0	0	298	
SUM (min:sec)	0:05:43	0:00:00	0:04:26	0:04:42	0:05:07	0:00:00	0:06:06	0:04:30	0:04:11	0:00:00	0:00:00	0:00:00	0:04:57	