Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 557–564

Exploring inexperienced user performance of a mobile tablet application through usability testing.

Chrysoula Gatsou
School of Applied Arts
Hellenic Open University,
Patra , Greece
Email: cgatsou@teiath.gr

Anastasios Politis
Graphic Arts Technology
Faculty of Fine Arts and Design,
TEI of Athens
Athens, Greece
Email: politisresearch@techlink.gr

Dimitrios Zevgolis
School of Applied Arts
Hellenic Open University,
Patra, Greece
Email: zevgolis@eap.gr

Abstract—This paper explores inexperienced user performance through a usability testing of three alternative prototypes of a mobile tablet application. One key factor in inexperienced users adopting mobile technology is the ease of use of mobile devices. The interface layout one of the three prototypes was built on the basis of previous research conducted in collaboration with users. More specifically, our study involves five navigation tasks which novice users were required to complete with each of the three prototypes. Our results showed that participants displayed better task performance with the prototype F1, which was created in collaboration with participants, in contrast to prototypes F2 and F3, which both caused navigation problems.

I. Introduction

THE rapid growth of mobile tablet technologies has lead to exponential growth in numbers of novice users, that is, in ordinary people who lack skills in computer science and who are drawn from a wide range of backgrounds. According to Hassenzahl [1], there is no guarantee that users will actually perceive and appreciate the product in the way 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. Despite the best efforts on the part of designers, new technologies often fail to meet basic human needs and desires [2]. 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 of Human Computer Interaction [3]. 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 [4]. Usability testing is a method used to evaluate a product by testing it on representative users. Greenberg and Buxton 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"* [5].

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 [6]. Carroll argues that an effective way of dealing with system complexity for the novice user is to provide a functionally simple system [7]. 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.

II. Background

A. 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 iteratively refined prototypes validated by users. Beaudouin-Lafon and Mackay define any prototype as a concrete representation of part or all of an interactive system [8]. 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 exists [9]. 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 session , 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 [10],[11]. 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 classify the ways in which the prototypes can be of value to designers [12]. 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 [13]. Buchenou & Suri 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"*[14]. Lim et al., 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*" [15].

B. Usability testing

The term *"usability"* is frequently employed in the field of human-computer interaction (HCI). Nielsen describes usability as an issue related to the broader issue of acceptability [16]. 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 [17].

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. 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 [18].

Dumas & Redish 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"* [19]. 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. You observe and record what participants do and say.
5. You analyze the data, diagnose the real problems, and recommend changes to fix those problems.

III. 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).

Fig. 1 Participant during the usability testing.

All three interfaces had the same look and feel, in order 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 [6], [20].

A. Participants

The literature gives no clear optimum number of participants to be employed in usability testing.

Nielsen [21] 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 [22]. On the other hand, Spool and Schroeder [23] state that five users identified only about 35% of the problems in a website. The research by Turner et al. implies that a group size of seven may be optimal, even when the study is fairly complex [24].

According to Sauro & Lewis *"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"* [25]. Our session was designed specifically to include a pool representative of potential users of the mobile application being tested. 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 I.

TABLE I.
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

B. Material

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. Digital tape recorder was also used.

C. The three prototypes

To reproduce a realistic software environment, for a period of three months, three prototypes were developed in Adobe Flash and we used them as a tool for recording user behavior 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 some tasks.

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 [6]. 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

Interface F2

The colors remain the same in prototype F2, but the main menu has 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

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.

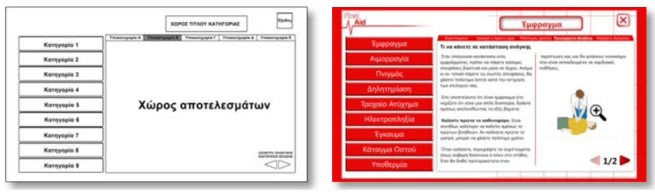


Fig. 4 The Interface F3

D. User Tasks

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

TABLE II
PARTICIPANTS 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

E. 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.

F. 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 System Usability Scale (SUS) developed by Brooke [26], since this is the most precise type of questionnaire for a small number of participants, as is shown by Tullis and Stetson's study [27]. 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 [27],[28],[29]. In addition to these advantages over other systems, the SUS is a powerful and multifunctional instrument [30].

G. Test protocol

Participation in the study lasted approximately one hour and 20 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.

IV. Results And Discussion

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 [4]. Qualitative and quantitative data were collected from each participant. Qualitative data included the participants' verbal protocol as recorded in videotapes.

Problems of usability were identified and categorized. We also collected comments on the prototypes and preference

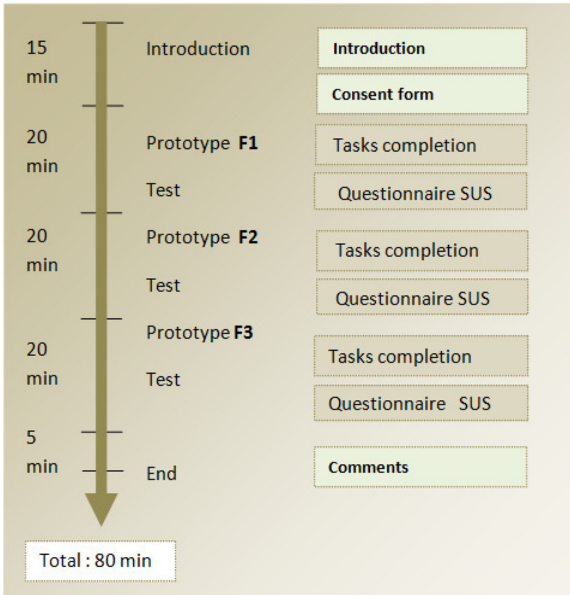


Fig. 5 User test process.

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.

A. 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 [4]. 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 III, IV show the user tasks and the error rate.

TABLE III
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%**

Errors were classified into two main categories, navigation errors and comprehension errors. Navigation errors occurred when partcipants didn't move as expected. Comprehension errors occured when participants didn't understand the design of the interface.

B. .Efficiency - Task Completion Time

We recorded the total amount of time required to complete each task in prototypes F1,F2 and F3, starting from turning

TABLE IV
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

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 V 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.

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.

TABLE V
TASKS 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

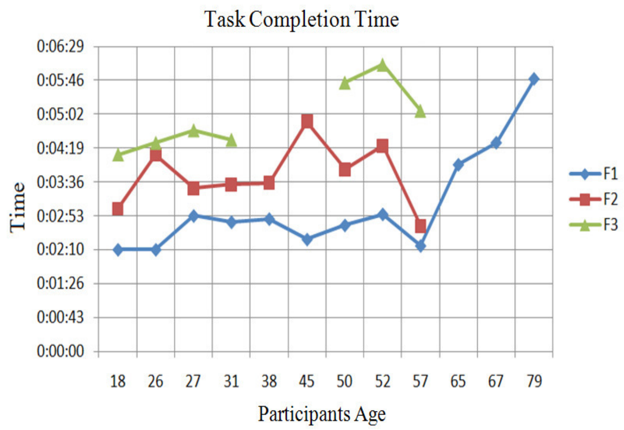


Fig.7 Task completion time per interface type.

C. 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 asks 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 VI. 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.

The survey results showed the overall satisfaction. Sauro [31] 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.

TABLE VI
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**

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.

V. CONCLUSION

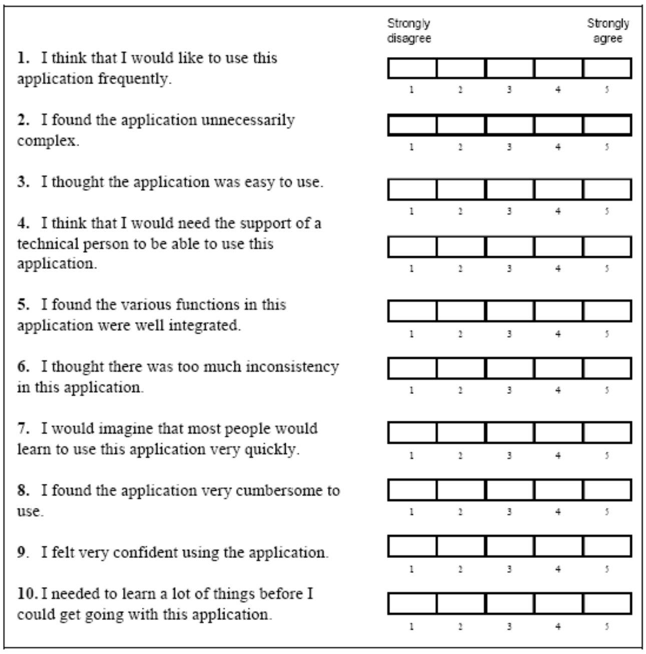
The aim of our study was to examine whether an interface design approach could improve performance of tasks by inexperienced users during interaction. To do this, we employed three different prototypes of the same application. We tested our empirical methodology on twelve individuals, all of them novices 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 [2].

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. 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.

APPENDIX

Appendix **A** *System Usability Scale*



ACKNOWLEDGMENT

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund. We would also like to thank the participants in our study.

REFERENCES

[1] M. Hassenzahl, The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction, 13(4), 2002, pp. 479-497
[2] D. Norman, *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.
[3] D. Benyon and D. Murray: Applying user modeling to human-computer interaction design. *Artificial Intelligence Review,* 7(3-4): pp.199-225, 1993.
[4] J.Rubin, and D. Chisnell, *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests* (2nd Ed.). Indianapolis, IN: Wiley Publishing, 2008.
[5] S. Greenberg and B. Buxton, Usability evaluation considered harmful (some of the time). Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human factors in Computing Systems, Florence, 2008, 111-120.
[6] C. Gatsou, A, Politis and D. Zevgolis, Novice User involvement in information architecture for a mobile tablet application through card

sorting. In proceedings *FEDSIS-MMAP* Wroclaw, 2012, pp. 711–718.

[7] J.M.Carroll, *The Nurnberg Funnel,* Cambridge, Mass.: MIT Press), 1999.

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

[9] B. Moggridge, *Designing Interactions*. The MIT Press, 2007.

[10] Tohidi, M., Buxton, W., Baecker, R., and 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.

[11] S.P. Dow, S. P, A.Glassco, J. Kass, M. Schwarz, and S.R Klemmer,. The effect of parallel prototyping on design performance, learning, and self-efficacy. Tech. Rep. CSTR-2009-02, Stanford University.

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

[13] C.A.Floyd, Systematic Look at Prototyping. In *Budde, ed., Approaches to Prototyping.* Springer Verlag, pp. 105-122, 1984.

[14] M. Buchenau, & J. Fulton Suri, “Experience prototyping”. In Proceedings *of Design of Interactive Systems* New York: ACM Press, 2000, pp. 424-433.

[15] Y.Lim, E. Stolterman and J.Tenenberg, “The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas”. *ACM Trans. Comput.-Hum. Interact.* 15, 2 , pp1-27, 2008.

[16] J. Nielsen, 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.1994 pp. 242-272.

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

[18] A.Cooper, R. Reimann, and D. Cronin, *About Face 3: The Essentials of User Interface Design.* John Wiley & Sons, Inc. 2007.

[19] J.S Dumas and J.C Redish, *A Practical Guide to Usability Testing* (revised Ed.). Portland, Oregon: Intellect Books,1999.

[20] C. Gatsou, A. Politis and D, Zevgolis, “Text vs visual metaphor in mobile interfaces for novice user interaction” *In Proc. of 16th International Conference on Electronic Publishing,* 2012 pp.271-279

[21] J.Nielsen, “Why You Only Need to Test With 5 Users”. Jakob Nielsen's Alertbox, March 19, 2000.

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

[23] J. Spool and W. Schroeder “ 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; 2001.

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

[25] J. Sauro and J.R. Lewis, *Quantifying the user experience: Practical statistics for user research*. Burlington, MA: Morgan Kaufmann, 2012.

[26] J. Brooke, 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,1996.

[27] T. Tullis, and J. Stetson, “A comparison of questionnaires for assessing website usability,” In *Proc. of the Usability Professionals Association (UPA)* 2004, pp. 7–11.

[28] J. R. Lewis, and J. Sauro, “The factor structure of the system usability scale”, *Proc. Human Computer Interaction International Conference (HCII 2009),* San Diego, CA, 2009, pp. 94–103.

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

[30] A. Bangor, P. T. Kortum, and J. T. Miller, “An empirical evaluation of the system usability scale,” *International Journal of Human-Computer Interaction,* vol. 24, issue 6, 2008, pp. 574–594.

[31] J. Sauro, *A practical guide to the System Usability Scale (SUS): Background, benchmarks & best practices*. Denver, CO: Measuring Usability LLC.2011

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	