CORE
Metadata, citation and similar papers at core.ac.uk
Provided by University of East Anglia digital repository

When to Ask Participants to Think Aloud: A Comparative Study of Concurrent and Retrospective Think-Aloud Methods

Thamer Alshammari *thamer.alshammari@uea.ac.uk*
School of Computing Sciences
University of East Anglia
Norwich, NR4 7TJ, UK

Obead Alhadreti *o.alhadreti@uea.ac.uk*
School of Computing Sciences
University of East Anglia
Norwich, NR4 7TJ, UK
College of Computer
Umm Al-Qura University

Pam J. Mayhew *p.mayhew@uea.ac.uk*
School of Computing Sciences
University of East Anglia
Norwich, NR4 7TJ, UK

Abstract

This paper presents the results of a study that compared two think-aloud usability testing methods: the concurrent think-aloud and the retrospective think-aloud methods. Data from task performance, testing experience, and usability problems were collected from 30 participants equally distributed between the two think-aloud conditions. The results suggest that while the thinking aloud method had no impact on task performance and testing experience, participants using the concurrent think-aloud method reported a larger number of problems with the test interface than participants using the retrospective think-aloud method. These findings suggest a reason for preferring the concurrent think-aloud method to the retrospective one.

Keywords: Usability Testing, Think-aloud Studies, Verbal Protocols.

1. INTRODUCTION

Usability is increasingly recognized as a key factor in the success of interactive software systems. Several studies have reported the benefits of more usable software interfaces, including greater user satisfaction, productivity and security [1]. As a result, a number of usability evaluation methods have been developed over the last four decades in attempts to improve the usability of software products. One of these evaluation methods is concurrent think-aloud (CTA) protocol, which has been used as the primary tool for understanding users' task-solving behaviours during usability testing [2]. This method involves test participants working on a set of predesigned tasks and talking aloud about their thoughts and task performance. This enables evaluators to identify the areas of a system that cause problems for the user and thus require further development. However, since the introduction of CTA into the usability field in 1982 cited in [3], questions have been raised regarding the validity of CTA data. Some scholars argue that thinking aloud while performing tasks may feel unnatural and thus may threaten test validity by altering what participants say and do [4]. Others contend that CTA protocol produces data mainly relating to descriptions of actions and procedures and does not yield more detailed explanatory data which usability evaluators often need to collect [5].

To counteract these issues, an alternative approach has been developed to gather verbal data which seeks to increase the utility and validity of the resulting verbal protocols. This method, retrospective think-aloud (RTA), addresses the above issues by allowing test participants to perform tasks silently and comment on their performance afterwards [4]. The retrospective method, therefore, does not interfere with participants' thought processes, and is deemed more capable to mirror real-world use. However, RTA also has its own limitations, chief amongst, which are its reliance on memory and the subsequent possibility of post-task rationalisations [4].

This paper presents the results of a study comparing the CTA and RTA methods. It is structured as follows: the next section discusses the existing literature focusing on the development of think-aloud methods and recent studies related to the evaluation of TA methods, and concludes by stating the aims and hypothesis of the current study. Further sections discuss the research method, the data analysis, and the results found. The papers concludes with a brief discussion and conclusion.

2. RELATED WORK

The development of think-aloud protocols is usually attributed to Ericsson and Simon [6]. These protocols were originally developed for the purpose of gaining insight into people's mental processes in cognitive psychology research. Later, they began to serve as a basis for understanding how people perform certain activities such as reading, writing and decision-making in different domains [7]. They are currently used in the context of usability testing to study human-computer interactions, and have become the methods of choice for many practitioners [8]. The primary benefit of think-aloud methods – unlike other forms of observation, which rely on many assumptions – is that think-aloud encourages the person being studied to verbalise their thoughts, thereby shedding light on the reasons for the behaviour being observed [4]. Nevertheless, the validity of think-aloud reports, particularly those generated from CTA, has been the subject of much heated debate within usability evaluation [9; 10]. For example, a study by [11] questioned the effect of concurrent verbalization on task behaviour and noted that participants who were encouraged to think aloud when performing tasks took longer than those who were not encouraged to think aloud to complete the same tasks. This change is often referred to as reactivity. This finding runs counter to Ericsson and Simon's [6] claim that thinking aloud does not affect task processing. For usability testing, reactivity can be problematic as it may lead to identifying and fixing false problems or failing to detect real issues. Other evidence suggests that verbal reports produced by the CTA protocol contain information largely related to the step taken to achieve a certain goal (such as reading or describing actions) which are less relevant to usability evaluators' interest [7]. By contrast, the less popular protocol method, RTA, appears to yields more elaborate and explanatory data as the RTA participant is not under strain and instead is free to think aloud naturally during the retrospective phase [7]. Moreover, since the participant is free to perform the tasks without the need to think aloud, the risk of reactivity is eliminated. As mentioned earlier, however, RTA reports are open to information loss and post-hoc fabrication.

There have been a small number of attempts to compare CTA with retrospective protocols [e.g. 12; 9; 23]. So far, most studies of CTA have compared its effectiveness to that of usability evaluation methods such as heuristics or walkthrough evaluation methods [13; 14]. The few studies that do evaluate CTA and RTA methods reveal little evidence of a difference between the two, describing them as comparable methods. However, these studies are limited, as is shown in a review by [15], by their failure to consider the nature of the problems detected; poor experimental design; and the omission of representative participants and realistic tasks.

2.1 The Present Study

The purpose of the present study was to investigate the utility and validity of two think-aloud usability testing protocols – CTA and RTA methods – taking into account the limitations of the above-mentioned studies. Based on previous research, the following hypotheses were proposed:

H1: RTA will outperform CTA in revealing more usability problems.

H2: There will be no difference in terms of task performance between the two conditions.

H3: Participants will be more satisfied with RTA.

3. METHODOLOGY

3.1 Study Design

This study involved a between-group design to prevent irreversible carry-over effects between the compared conditions. Participants were divided into two groups. The independent variable was the think-aloud method: one group of participants was allocated to CTA, and the other to RTA. The dependent variables were the usability problems that were discovered, task performance data, and participants' testing experiences. The effectiveness of the TA methods under study was evaluated based on the number and severity of problems discovered through each method. Task performance was measured by the time needed to complete a task and task success. Participants' testing experiences were evaluated via a post-test questionnaire which participants were asked to fill in; the questionnaire was developed based on previous research [7; 16] and consisted of rating scale questions. Participants were asked to rate on a five-point scale the level of ease they felt when participating in the testing session, their comfort, their convenience, the degree to which they felt at ease when concentrating on the tasks, the degree to which they felt at ease when remembering to think aloud, and their willingness to participate in a similar test in the future.

3.2 Selection of Test Object

The University of East Anglia (UEA) Ticket Bookings website was chosen to be the targeted object for the experiment in this study (www.ueaticketbookings.co.uk) (Figure 1). The UEA Ticket Bookings (UEA-TB) website is a commercial website overseen by the UEA Box Office. Its commercial services, which mainly target students but can also be accessed by the general public, include booking tickets and buying items such as UEA hoodies. The UEA-TB website was chosen for this study due to its interactive interface with multiple functions, processes, and features. The representative users of this website are easily accessible (as they are university students), which will assist in recruiting representative participants who actively use the targeted website. In addition, since the UEA-TB website is a significant element of UEA student life during the academic year, improving its usability may lead to increased student satisfaction.

FIGURE 1: Test Object.

3.3 Designing Tasks

It is vitally important that tasks in usability studies represent the activities that end users would complete when using the test website to accomplish specific goals [17]. To this end, the first author interviewed two UEA students who had previously used the UEA-TB website to gain insight into how users use the website in order to create representative tasks. The interviewer

employed a semi-structured interview technique by directing the interview to a general discussion about users' activity on the UEA-TB website, thus allowing the interviewees some freedom to express their opinions and giving the interviewer a deeper understanding of their responses. The interviews took place in a meeting room at the UEA and lasted for about 15 minutes each. The interviewees allowed the interviewer to take notes and record the conversation, which helped the interviewer to summarise the interviews. Based on the information acquired, six tasks were created: logging in, updating an address, adding items to the basket, editing basket, simple searching for events, and advanced searching (see Appendix B). All tasks were designed to be carried out independently from one another, meaning that even if a task was not completed successfully, participants could still carry out the other tasks. The tasks were also subjected to four pilot tests prior to final testing to ensure that they were free from bias and ambiguity. During the experiment, the six tasks were presented in ascending order of difficulty. The researchers were aware that participants might initially have felt nervous, and therefore ensured that the first task was the easiest in an attempt to help participants to overcome the pressure and cope with the tasks.

3.4 Participants

The number and backgrounds of the participants who used the TA methods under study to evaluate the UEA-TB website were carefully considered in order not to influence the results obtained. The ideal number of participants in usability testing has been the subject of much debate. Some researchers state that three participants are sufficient to identify about 50% of the most important usability issues [18], while others suggest that five participants can find 80% of all usability issues [19]. However, it is arguable that these numbers are not applicable to the current study, as its aim is to compare the effectiveness of two usability testing methods rather than to detect usability issues using a single method. Therefore, with consideration to time constraints and to enhance validity, 30 participants were engaged in this study, with 15 participants assigned to each method. The researchers ensured that the participants were a representative sample of university students and were interested in the test website via an online pre-test questionnaire (Appendix A) which gathered information about their background and web usage. Participants were divided as evenly as possible in terms of gender and online usage experience, and received a £5 shopping voucher as a reward for their participation.

3.5 Resources and Equipment

All testing sessions were conducted in the same laboratory at the University of East Anglia. Participants used the same lab computer which had a high-speed connection to the Google Chrome browser, an external microphone, and a large screen to ensure that even sight-restricted participants could see the website content. ScreenPresso software was used to record the participants' screen footage and their verbal comments. Other resources included a consent form, tasks sheet, an observation sheet (Appendix C) on which the test facilitator (first author) took notes, and a timer.

3.6 Experimental Procedure

This study was approved by the UEA Ethics committee. Each testing session began with the test facilitator welcoming the participant to the lab and then asking him/her to read and sign the consent form before starting the test (Figure 2). The consent form explained the aim of the study and informed the participant that s/he would be observed and his/her voice and screen actions would be recorded during the session. Next, the participant was asked to fill in a background questionnaire. The participant was then given two minutes to familiarise himself or herself with the lab computer. On completion of this step, the participant was given a task list and was asked to perform each task in sequence on the test website. CTA participants were asked to think aloud while performing the tasks, whereas RTA participants were required to perform all tasks silently and, once they finished, were invited to verbalise their thoughts on a video recording their performance. During the test, the facilitator strictly followed Ericsson and Simon's [6] guidance, and only issued a think-aloud reminder if the participants fell silent for 15 seconds; there were no other interactions. The observation sheet was used to record the following information: time spent

on tasks (in seconds), task completion rate (successful or unsuccessful), usability problems discovered, type of test, participant's identification number. At the end of the session, participants in both conditions were requested to fill in the post-test questionnaire and were then thanked and permitted to leave.

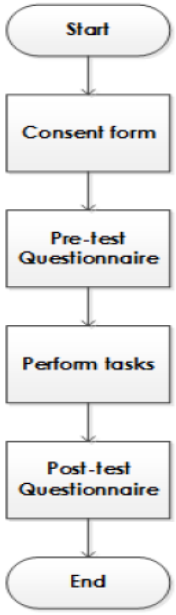

FIGURE 2: Experimental Procedure.

3.7 Piloting and Correction

Prior to the actual experiments, a pilot study was crucial for testing and refining the study procedure and its instruments. As suggested by [20], four participants were recruited to conduct the pilot test. These participants had similar characteristics to the participants in the main experiment, but did not participate in the main study. The pilot test revealed that some tasks were ambiguous and were not clear enough for the piloting subjects. As a result, these tasks were reworded and clarifying information was added. The pilot study also helped the researchers to enhance the content of the consent form. All other aspects of the pilot test went smoothly and remained part of the actual experimental procedure.

4. RESULTS

4.1 Usability Problems

Overall, 53 problems were extracted from the test sessions files of both think-aloud (TA) conditions (Figure 3). The CTA condition generated 45 problems, 33 of which were unique to that condition, while the RTA condition yielded 20 problems, 8 of which were unique to that condition. Thus, both groups commonly identified 12 of the total number of problems (see Appendix D).

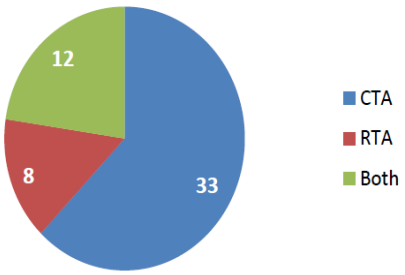

FIGURE 3: Unique and Shared Usability Problems.

As shown in Table 1, t-test suggests that the difference between the two methods' performance in terms of finding usability problems is significant, CTA appeared to be more effective than RTA in detecting problems.

CTA		RTA		Value
Mean	SD	Mean	SD	
16.8	6	9	3.3	(p= 0.01)

TABLE 1: Number of problems detected per task in both testing methods.

According to [21], usability issues can fall into one of the following categories of severity: minor, major, and critical. To ensure an objective assessment of the problems discovered in this study, the researchers asked a usability expert with a PhD degree in usability evaluation to rate the final set of usability problems discovered by the two groups. The expert was asked to rate severity of the problems based on their frequency, impact and persistence, as suggested by [22]. The 45 usability problems discovered by the CTA group were classified by the usability expert into 17 minor problems, 20 major problems and eight critical problems. The 20 problems discovered by the RTA group were classified into five minor problems, eight major problems and seven critical problems. Interestingly, there were no significant statistical differences found between the groups' performance in terms of the types of problems found.

A number of researchers claim that a group of five participants is able to find 80% of usability issues [19]. To further examine this controversial issue, a detailed analysis was undertaken to find out the relationship between the number of issues found and the number of participants in each TA group. It was found that the first five participants in the CTA indeed found just under 80% of the total number of usability problems identified by their group, and that the first five participants in the RTA found 85% of the total number of usability problems identified by their group (Figure 4). The researchers also found that the last three CTA participants and the last four RTA participants could not identify any new usability problems whatsoever, as they only discovered what had already been identified by their predecessors. In total, 12 CTA and 11 RTA participants were able to discover all reported problems.

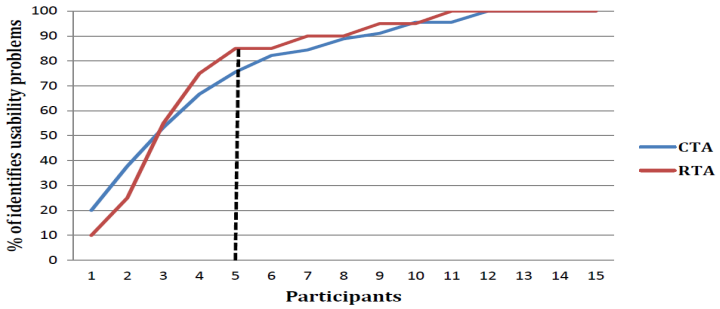

FIGURE 4: Relationship between usability problems found and number of participants.

4.2 Task Performance

Each participant was asked to perform six tasks on the UEA-TB website, meaning that a total of 90 tasks were performed by each group. Participants in the RTA group successfully completed 86 out of 90 tasks, whereas participants in the CTA group were able to successfully complete 83 tasks (Table 2). On average, 93% of the tasks were completed successfully. Participants in the CTA condition completed 5.53 out of the six tasks, in contrast to 5.73 by participants in the RTA condition. The RTA group spent a total of 6206 seconds on the tasks, while the CTA group spent a total of 7175 seconds on the tasks (Table 3). Nevertheless, an independent t-test found no significance differences between the two TA groups on any of the performance measures (Table 4).

Method	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Total
CTA	15	15	15	14	13	11	83
RTA	15	15	15	14	15	12	86

TABLE 2: Number of tasks completed successfully.

Method	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Total
CTA	1106	1231	764	742	1897	1435	7175
RTA	720	967	696	540	1671	1612	6206

TABLE 3: Time spent on tasks in seconds.

	CTA		RTA		P Value
	Mean	SD	Mean	SD	
Completion rate	5.53	0.74	5.73	0.59	ns
Time on tasks	478	87	414	159	ns

TABLE 4: Task performance measures.

4.3 Participants' Experiences

Figure 5 shows participants' ratings of their experiences. A Mann-Whitney test found no differences in testing experience between the two conditions. The CTA and RTA groups gave similar ratings for the ease of the experiment, the ease of concentrating on the tasks, the ease of thinking aloud, the level of distractions caused by the evaluator, and their willingness to participate in similar experiments in the future.

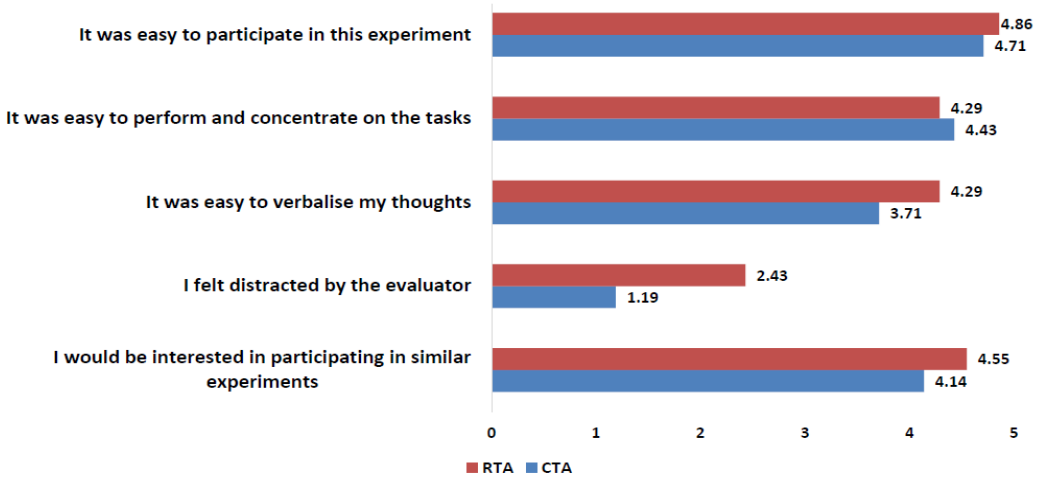

FIGURE 5: Average subjective ratings given by participants.

5. DISCUSSION AND COMPARATIVE EVALUATION

For maximum reliability, this study followed a systematic approach recommended for use in usability studies. The number and characteristics of test participants, number of tasks, targeted object and evaluation measures all were taken into consideration in order to eliminate any factors that might have affected the results. The following provides a discussion of the study results and a comparative evaluation of the findings with other empirical studies in the field.

The first hypothesis of this study was rejected, as the CTA method was found to be more effective than the RTA method at detecting usability problems. A possible explanation for this finding may be that asking the CTA participants to think aloud while performing the test tasks afforded them the opportunity to report usability problems as soon as they arose, whereas the

RTA participants were more likely to forget usability problems during the retrospective phase even though they may have noticed them during the performance of the tasks. This result is in line with the findings of [23], who concluded that CTA testing outperformed RTA testing in identifying usability problems. This finding lends support to Ericsson and Simon's [6] argument that vital information may be lost in the case of retrospective research. This would cast doubt on using the outcome of a RTA evaluation as an overall indication of the usability of the test object that is being assessed.

The second hypothesis of this study was accepted, as an independent t-test found no differences between the CTA and RTA conditions, neither in terms of successfully completed tasks nor in terms of the time it took the participants to complete the tasks. This suggests that the task performance of the participants in the CTA condition was not affected by the double workload of having to think aloud and carrying out the tasks at the same time. This finding supports Ericsson and Simon's argument that thinking aloud does not have an effect on task performance [6], and in agreement with [12]. Task performance outcome in CTA testing can, therefore, be regarded as a valid representation of the behaviour of real-life users. With regard to the third hypothesis, the hypothesis was rejected, as CTA participants appeared to have the same experience as those in the RTA condition. For the CTA condition, this means that participants were not affected by the dual-task. Nevertheless, this result should be accepted with some caution, as it is primarily based on participants' subjective rating and may be biased due to factors like social desirability [7]. This finding echoes that of [12] but conflicts with [9] who found that RTA participants reported more satisfaction than CTA participants. One possible justification for this difference may be the latter study did not take steps to control the participants' individual differences by matching them as closely as possible between conditions as in the case of this study.

5.1 Limitations and Future Work

As with any research, this study has its inevitable limitations. However, these limitations may represent opportunities for further work. First, this study used a between-group design which did not allow for a rigorous control for individual differences and their possible effects on think-aloud performance. Second, this study only focused on CTA and RTA methods. For future expansion, the researchers are currently working on a study combining both methods into a single method and comparing that combined method with both CTA and RTA.

6. CONCLUSION

The aim of this study was to provide usability evaluators with a better understanding of the utility and validity of the CTA and RTA testing methods in collecting usability data, with a view to contributing to existing knowledge on TA approaches and helping evaluators to make more informed decisions about which TA method to use in particular contexts. The results of this study indicate that thinking aloud did not lead to reactivity under CTA conditions; however, CTA significantly outperformed RTA in terms of the number of usability problems discovered, although no statistically significant differences were found in the types of problems detected. The participants in the two groups were equally satisfied with the methods under study. The CTA method would, therefore, seem an appropriate method for usability testers who are interested in detecting as many usability problems as possible, regardless of the types of the problems found. If usability testers are interested in portraying user performance as it might occur in the real context of use, they have the choice between using the CTA method or the RTA method. Overall, taking the results obtained and practical considerations into account such as time taken to complete study, a case can be made for preferring the CTA protocol to the RTA protocol in usability evaluation studies.

7. ACKNOWLEDGMENTS

The authors would like to thank all those people who took time to partake in the experiments. The authors also thank Sidney Brouet, who helped them greatly in recruiting the study participants. Thanks also to Kelly Kanayama for her assistance with proofreading and to the anonymous reviewers for their helpful comments.

8. REFERENCES

[1] Seffah, A., Donyaee, M., Kline, R., and Padda, H. 'Usability measurement and metrics: A consolidated model.' Software Quality Journal, 14(2): 2006, 159–178.

[2] McDonald, Sharon, and Helen Petrie (2013). 'The effect of global instructions on think-aloud testing.' Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM.

[3] Lewis, C. & Rieman, J. 'Task-Centered User Interface Design: a Practical Introduction'. 1993. [Online). Available from: http://hcibib.orgltcuidl. [Accessed: 22 November 2014).

[4] Gray M., Wardle H. 'Observing gambling behaviour using think aloud and video technology: a methodological review'. NatCen Social Research. Available at: www.natcen.ac.uk. 2013, [Accessed: 22 January 2015).

[5] Cotton, D. and Gresty, K. 'Reflecting on the think-aloud method for evaluating elearning'. British Journal of Educational Technology, 37 (1), 2006, pp. 45-54.

[6] Ericsson, K. A. and Simon, H.A., (1993) Protocol Analysis: Verbal Reports as Data. Revised ed. Cambridge: MIT Press.

[7] Haak, V, (2008). 'A penny for your thoughts – investigating the validity and reliability of think-aloud protocols for usability testing'(PhD dissertation).

[8] McDonald, S., Edwards, H. and Zhao, T. 'Exploring think-alouds in usability testing: an international survey'. IEEE Transactions on Professional Communication, 55(1), 2012, pp.1-17.

[9] Haak V, Menno D, and Peter J. 'Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue.' Behaviour & Information Technology 22.5, 2003, 339-351.

[10] Hertzum, M., Hansen, K.D. and Andersen, H.H.K. 'Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload?'. Behaviour & Information Technology, 28 (2). 2009, pp. 165-181.

[11] Griffiths, M.D. 'The role of cognitive bias and skill in fruit machine gambling'. British Journal of Psychology, 1994, 85: 351-369.

[12] Haak V, Maaike J., Menno DT de Jong, and Peter Jan Schellens. 'Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: a methodological comparison.' Interacting with computers 16.6: 2004, 1153-1170.

[13] Khajouei, Reza, Arie Hasman, and Monique WM Jaspers. 'Determination of the effectiveness of two methods for usability evaluation using a CPOE medication ordering system.' international journal of medical informatics 80.5, 2011, 341-350.

[14] Molich, R., Ede, M. R., 'Kaasgaard, K., & Karyukin, B. Comparative usability evaluation' Behaviour & Information Technology, 23(1), 2004, 65-74.

[15] Gray, W. D., & Salzman, M. C. Damaged merchandise? A review of experiments that compare usability evaluation methods. Human-Computer Interaction, 13, 1998 203-261.

[16] Eger, N., Ball, L. J., Stevens, R., & Dodd, 'Cueing retrospective verbal reports in usability testing through eye-movement replay'. In Proceedings of the 21st British HCI Group Annual

Conference on People and Computers: HCI... but not as we know it-Volume 1 (pp. 129-137). British Computer Society, 2007.

[17] Dumas, Joseph S., and Janice Redish. A practical guide to usability testing. Intellect Books, 1999.

[18] Virzi, R.. Refining the test phase of usability evaluation: How many subjects is enough? Human Factors, 1992, 34(4):457–468.

[19] Nielsen, J. (2000). Why you only need to test with 5 users. Nielsen Norman Group. Department of Computer Science. Machine Learning. Available at: bit.ly/1gpks7w [Accessed 25-04-2014].

[20] Fagan, J. C. Usability studies of faceted browsing: A literature review. Information Technology and Libraries, 2013, 29(2):58–66.

[21] Lindgaard, G. and Chattratichart, J. 'Usability testing: what have we overlooked?' In Proceedings of the SIGCHI conference on Human factors in computing systems, 2007, pages 1415–1424. ACM.

[22] Nielsen, J. Usability engineering. 1994, Elsevier.

[23] Peute, L. W., de Keizer, N. F., & Jaspers, M. W. 'Effectiveness of Retrospective and Concurrent Think Aloud in Formative Usability Testing; Which Method Performs Best?'. Human factors methods in health information systems' design and, 2013, 65.

APPENDIX A: Background Questionnaire

ID ……..

Please circle/underline your answer of following questions:

1. Age:

18-21 22-25 26-29 30+

2. Gender:

Male Female

3. What is your first language?

……

4. What is your nationality?

……

5. What is your current education level?

Undergraduate student

Master student

PhD student

Other (please specify)

……

6. How long have you been using the Internet?

Less than a year 1-2 years

3-5 years More than 5 years

7. On average, how many hours a day do you use your WWW browser?

Less than an hour 1-2 hours 3-4 hours 5+ hours

8. Have you ever visited UEA Ticket Booking website?

Yes No

9. Have you ever participated in a usability testing experiment?

Yes No

APPENDIX B: List of Tasks

1. Login in UEA ticket booking website by using email address: thmmr@hotmail.com and password: a2014a.

2. Update your address to House Number: 37 City: Norwich, Street: Caddow Road County: Norfolk, Postcode: NR5 9PQ Country: United Kingdom.

3. Find the artist "Motorhead" and add it to your favourite.

4. Find the venue "The Talk" and view its contact details.

5. Find "Unisex Grey T-Shirt", add 3 size medium (include postage and packaging inside UK) and 2 size XL (include postage and packaging inside UK) to your basket.

6. Find all events that will take place in venue: the Aquarium/The Zoo, Lowestoft, during: August.

APPENDIX C: Usability Observation Sheet

Usability Test Observation Sheet

Participant's ID No.: ..... TA Method Type: ..... Date: .....

Session Starts at: ..... Session Ends at: .....

Task #: ...1....

Time on tasks: ..... No. of clicks: .....

Task Completion Rate:

❑ Successful ❑ Unsuccessful

No.	Usability Problems Discovered
	.....

Notes: .....

APPENDIX D: Usability Problems List

Method: C=CTA; R=RTA; B=Both
Severity: 1=Minor, 2=Major; 3=Critical

#	Description	Method	Severity
1	Too many links on home page.	C	1
2	Log in link is not on the top.	B	2
3	Log in link is not clear (small).	C	1
4	No welcome message nor user's name shown as a proof of a successful logging in.	C	2
5	Too many and too small pictures on home page.	C	2
6	Too much content on home page (difficult to scan).	C	2
7	Some of the text on the address page looks like clickable links.	C	1
8	Confusing labels/names for address fields	C	1
9	On the address page there is a "View" current address link and an "Add new address" link, but there are no direct links for updating or deleting a current address.	B	2
10	The link "View" next to the user address is confusing. Users think clicking this link will open a map.	C	1
11	Font size is too small on the address page.	B	1
12	County field should be optional (not required). House no., street name, city and country fields are enough.	C	1
13	Should allow searching for address through postcode and house number.	R	1
14	Main menu font is too small.	R	2
15	Search option on home page is not comprehensive.	B	3
16	No comprehensive search tool.	C	3

17	There is no "Add to basket" button on the item pages.	B	1
18	The button "Buy now" does not comply with what it says (it only adds items to basket).	C	2
19	When items are added to basket, an ambiguous message appears "5 minutes left before your reserved tickets ...etc."	C	2
20	On the basket page, there is a "Back to Events" button. The basket page should include a "Back to Shop" button.	C	1
21	Confusing instructions on product pages: "Choose delivery and then at next stage choose (collection) - to avoid a double delivery charge".	C	3
22	Quantity field on item page should be changed to a drop-down list.	B	1
23	There is not enough of a breakdown list for the total cost on the basket page (should present delivery cost).	C	1
24	Too many options on item page (a list of 16 options to choose from).	R	3
25	Too much text on artist page.	C	1
26	Small font size on artist page.	C	1
27	Favourite tab on customer zone menu is too close to "Add to favourites" button on artist page, which causes some confusion.	R	2
28	Adding the artist's logo as well as the artist's name would simplify finding an artist from the artist list.	C	1
29	No contact details on Venue Contact Details page.	B	3
30	Low contrast between main menu font and background colour (sidebar links stand out more than main menu).	C	3

31	On the Venues page, a disturbing message appears every time users hover over the venue links.	C	2
32	No search tool to search for venues by name.	B	2
33	Events page is not well organised.	C	2
34	Too many pictures on the Events page (difficult to access via slow internet connection).	C	2
35	Too much scrolling on the Events page.	C	2
36	No "back to top" button after scrolling down a long way.	C	2
37	Too many options on advanced search options.	C	2
38	Inconsistency on options layout (difficult to scan).	B	2
39	No enough spaces between search options.	C	1
40	Some tick-box labels are partially overlapping.	C	3
41	Venue names on advanced search are not sorted in alphabetical order.	C	1
42	No validation on data entry on advanced search ("from" and "to" date fields).	R	3
43	Confusing to have two buttons next to each other: "View all" and "A-Z"; they both have the same functionality.	C	2
44	On the Venue Address page, the text is duplicated.	B	2
45	Some pages are broken and a message is shown: "the code, which causes an error".	B	3
46	Once advanced search is opened, there is no button to close it.	C	2
47	Duplicate map on one page (one is too small).	B	3

48	No confirmation messages after completing certain tasks (e.g. updating address).	C	2
49	In main menu "Customer Service" is not properly worded. It should be changed to "Contact us".	C	2
50	No "FAQ" link in the main menu (nor in footer).	R	2
51	Low contrast between the colours of the website elements (header, body, and footer).	R	1
52	Some icons such as "Add to favourites" are static. The "Add to favourites" icon does not change even if the artist is added to favourites.	C	1
53	Even though the website contains too much content, about 40% of the screen width is wasted on the margin	R	3