ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

Affect of Context on Think-aloud During Usability Evaluation

Mohammadi Akheela Khanum[1], Munesh C. Trivedi[2], Sajjad Anees Nagrami[3]

Associate Professor, Integral University, Lucknow, India [1]

Professor, Department of CSE, DIT, Greater Noida, India [2]

Project Engineer, Al-Sharq Office Co. KSA[3]

Abstract: This paper investigates the effect of physical context on the outcomes of usability evaluation when think-aloud usability testing protocol is applied on children. Usability evaluation is performed with 18 children by assigning them randomly to two different physical settings: a lab and a field. Traditional think–aloud protocol is applied to elicit verbal comments from the children while solving tasks on the given system. The amount of verbalization, impact of test monitor on solving tasks, amount of prompting by the test monitor is quantitatively measured during the test sessions. The results indicates that context plays important role in influencing the results of usability evaluation and that think-aloud when applied in different physical contexts gives different results.

Keywords: Think-aloud, physical context, usability evaluation, child computer interaction, ICDL, verbalisation

I. INTRODUCTION

Child computer interaction (CCI) is a sub-field of HCI which deals with how children use interactive products. CCI spans multiple scientific disciplines informing and supporting an area of research and industrial practice that is concerned with the design of interactive systems for children. CCI is still finding its way in research. *"Relating to sociology, education and educational technology, connected to art and design, and with links to storytelling and literature, as well as psychology and computing, CCI borrows methods of inquiry from many different disciplines"* [1] .

With the growing demand of technologies for children, user interfaces are becoming a much more important part of the computer. The personal computers are now available to ever broader groups of users, and users are using these computers for a variety of tasks. The emergence of video games and advanced computer software has made it possible to produce pleasant and familiar interfaces. Therefore, the users are less willing to compromise with the usability of the interfaces.

Usability evaluation* is an important part of today's software development process as it can help to improve the usability of system under development. Usability evaluation can save money, time, and effort, if introduced in the software development process correctly and at the right time [2] .

Usability evaluation involves number of activities, such as, designing tasks that reflect the future use of the system [3,4] , deciding on a method or protocol to be used for the evaluation [5,6], deciding what data to collect and how to collect it [5,7] , the activity of recruiting participants that are representative of the end user group [5,7] and deciding if the evaluation is best done in a usability laboratory or as a field study [5,3].

One of the predominant discussions during usability evaluation is about the choice of context. Brown et al.[8] define context as location, identities of the people around the user, time of the day, season, and temperature. Ryan et al., [9] define context as the user's location, environment, identity, and time. The choice of location as context during usability evaluation is considered an important topic of discussion in research. The focus on location for usability evaluation indicates that this aspect of context is considered important, when choosing which aspects of the use context to recreate in usability evaluation. Context plays an important role in usability evaluation. Studies have shown that children are mostly affected by the context than adults. Children show varying behaviour when they are tested in the laboratory environment and when they are tested in the field environment. The importance of the physical context has been explored and studied by several usability researchers. The research on context is scattered and scarce, and it lacks a unifying overview. Therefore, an understanding of the influence of context and how it impacts the process of usability evaluation is required.

II. LITERATURE REVIEW

The importance of physical context in usability evaluation has been researched for a long. Out of the many factors that can effect usability evaluation, physical context is considered to directly influence the behaviour of the people taking part in the usability evaluation. Tullis et al. [10] compared usability test performance between remote and lab settings, with a focus on task times and issues discovered. Study involved a prototype of a Web site for providing the employees of a company with access to information about their own benefits, including retirement savings information, pension information, medical and dental coverage, payroll deductions and direct deposit, and financial planning. The study found no significant difference between remote and traditional task times. Both remote and traditional lab testing revealed usability issues on existing websites. However, Tullis's participants scored the subjective tasks and interface

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

differently between the different testing locations. It was theorized that the remote condition would incite participants to be more honest regarding the test. Tullis offered no explanation for the difference in these scores, other than small sample size.

Tsiaousis & Giaglis [11] examined the effects of environmental distractions on mobile website usability. They hypothesize that environmental distractions can decrease user performance levels. They proposed a model hypothesizing on the effects of environmental distractions on the usability of mobile websites. They categorized the environmental distractions into auditory, visual and social. A preliminary test on 30 users was conducted to investigate the effect of environmental distractions on mobile website usability. Results indicate that the environmental distractions have direct effect on mobile website usability.

Hummel, Hess & Grill [12] studied the effect of environmental conditions such as light, acceleration, sound, temperature, and humidity, on the user's action during usability testing. They developed a mobile context-framework based on a small wireless sensor network, to monitor the effect. User experiments were conducted in a laboratory with seven test participants where the environmental conditions were altered. The test users have to complete predefined tasks on an application running on a mobile smart phone. The results proved that under varying environmental conditions the user's performance level on an average was reduced in terms of higher error rates and delays.

Andreasen et al. [13] compared four usability testing methods:synchronous remote usability testing, asynchronous remote expert testing, asynchronous remote user testing and conventional lab-based think-aloud method. The test was performed on 24 subjects, 14 male and 10 female. All the participants were the students at Aalborg University and aged between 19 to 30 years. The tested system was Mozilla Thunderbird 1.5. user had to perform 9 tasks which included creating an account, checking new e-mails, creating a folder and mail filter, run the mail filter to find the number of mails in the folder, creating a contact, to activate spam filter, find and mark spam mails, labelling mails in the inbox. The results show that remote testing reveal interface issues. Asynchronous methods required more user time to complete, and revealed fewer issues. However, asynchronous methods can be distributed to larger groups, and the authors cite this as a benefit trade-off for asynchronous decreased performance. These findings were further supported by Bruun et al. [14], who found that remote, asynchronous testing identifies about half of the problems found by traditional usability testing, and their study concludes that the time savings introduced by the remote asynchronous method make them viable for software usability testing.

Kaikkonen et al. [15] carried out usability testing of mobile consumer application in two environments: in a laboratory and in a field. The test was carried out with 20 users in each environment. Think-aloud protocol was used to elicit verbal comments from the users. The tested application was known as Mobile Wire, which transfer files between the computers and mobile handset. Results indicate that field testing is more time consuming the lab testing. They found that field testing is worthwhile when combining usability tests with a field pilot or contextual study where user behaviour could be investigated in a natural context.

Razak et al. [16] conducted usability testing with children in both laboratory and field. Drawing applications were tested in their preschool and an educational game was tested in the usability laboratory. The test involved 4 children of five years age. The results indicate that field study is viable solution for understanding children experience with technology than it is with testing for usability problems and laboratory study is more suitable for evaluating user interfaces and interaction with the application than it is with understanding children's experience.

Andrrzejczak & Liu [17] conducted a study to evaluate the effect of testing location on usability test elements such as stress levels and user experience. They compared traditional lab testing with synchronous remote testing. They investigated two groups of users in remote and traditional settings. Each group participants have to complete two tasks, a simple task and a complex task. The dependent measures were task time taken, number of critical incidents reported, and user-reported anxiety score. Task times differed significantly between the physical location conditions; this difference was not meaningful for real world application, and likely introduced byoverhead regarding synchronous remote testing methods. Critical incident reporting counts were similar in all conditions. No significant differences were found in user reported stress levels. Subjective assessments of the study and interface also did not differ significantly. Study findings suggest a similar user testing experience exists for remote and traditional laboratory usability testing.

III. METHOD

A. Participants

The purpose of the experiment was to explore the effect of physical context on children's verbalization during think-aloud sessions. According to Piaget theory of cognitive growth [18], children in the age ranging between 11-15 years have their cognitive functions-formal thinking, negatious, logic as well as social skills well developed. Keeping in mind the children's cognitive growth and several researches in literature, we chose to experiment with the children in the age group of 11-15 years old. To select the most appropriate children for the experiment we obtain some preliminary information from them. The information consists of their age, family income, academic grades, experience with computers/Internet, and weekly usage of computer/ Internet. Since most of the children could not provide information on their family income, we excluded it while selecting the children for the

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

test. We ended up selecting 18 children (8 girls and 10 boys) at the age ranging from 10 years to 13 years as test subjects in the experiment. We did not receive responses from equal number of boys and girls. Therefore, the division was slightly uneven. All these children were the 6th and 7th grade pupils from two different English medium schools in the Lucknow area of India. The mean age of the selected children was 11.56 years. Academic grades of the students were very high, with a mean of 3.84 points on a scale of 4.00. The exposure to computer/Internet was found to be less with a mean experience of 1.69 years. The amount of time that they spent each week on using computer/Internet was about 2 hours, which could be an indication that the selected children were less involved in the online activities. We also found that only 24% children were involved in online reading activities.

B. System

The selected system for our experiment was International Children's Digital Library (ICDL). This particular website was selected because digital libraries are becoming a common place for children and many researches are now focusing on how the children are using these new learning tools. During the children's demographic data collection we also found that none of the children had ever used ICDL for reading books online. Figure 1 is the screenshot of ICDL homepage.

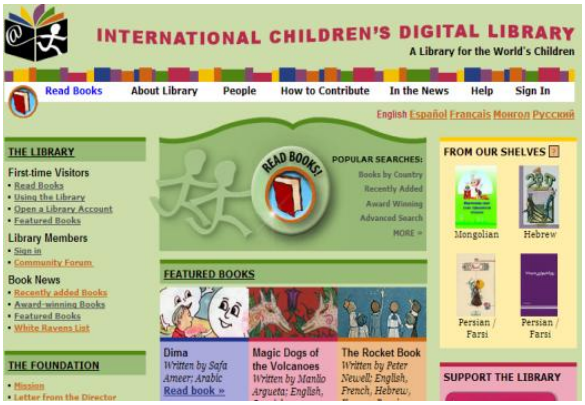

Fig.1 Screenshot of ICDL homepage

International Children's Digital Library is an online collection of books which has in store various books for children in age group ranging from 3-13 years. It is a multilingual library offering books for a wide range of languages and children. ICDL has four search tools for accessing the current collection of books: Simple search, Advanced search, Location search, and Keyword search.

C. Test Sessions

The children were assigned as test subjects to one of the two setups: as individual testers in lab and in field for think-aloud sessions. Each individual setup had 9 individual testers (4 girls and 5 boys). Children were randomly assigned to each of the two test setups. Each of the test sessions was configured to have same gender. The test setups and assignment of children to each setup is given in table 1.

TABLE I

TEST SETUPS AND CHILDREN ASSIGNMENT

Participants	Think-aloud	
	Lab	Field
Boys	5	5
Girls	4	4
Total	9	9

D. Test Tasks

The children were asked to solve five tasks. The tasks involved the use of different search options in ICDL. This included searching books by country, searching books by title, searching books by language, searching award winning books in English and reading a specified book in the language of their preference. We did not specify any time limits for the tasks, but required the participants to try to solve all tasks.

E. Procedure

We received consent from 18 children. At the beginning of the test session children were introduced to the experiment by two of the participating researchers. We explained the children about their roles in the experiment and how they would contribute to our research. Participation in the experiment was voluntarily and interested children got an information sheet describing the experiment in detail and a consent form that had to be signed by a parent or a guardian.

The sessions were held at the school's campus itself due to not getting permission to commute to other place where the usability laboratory was set. We created two labs, one for field testing sessions, and one for laboratory testing sessions. For the field testing, we used the school's computer lab which with the students was familiar and we tried to keep it as it was used by the children. No restrictions were imposed on the people to move in the lab during the test session. This created a perfect field environment for the children. For testing in lab environment, we setup a usability laboratory in one part of the school.

The lab environment was different as compared with the field. Lab was located in a quiet place where people not related with the test sessions were not allowed. The lab only had the test monitors and the test participants at any given time during the test sessions. Outside disturbances were also kept minimal. Hanna et al. (1997) guidelines for usability testing with children were followed. We greeted and children and introduced ourselves. Particularly, we focused on stressing the importance of the participation, and stressing that they were not the object of the test. The purpose of the usability test was explained to the children in detail. The children received questionnaires on which they had to provide answers to such as age, name, school, computer/internet experience, number of hours of usage, online reading experience. The usability test sessions were conducted in two labs, one a specialized usability laboratory setup in the school and the other was the school's computer lab.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

During the test sessions, all the screen activities and children's interaction with ICDL were recorded using CamStudio for later analyses. CamStudio is an open source screen recorder. The children were asked to solve the five tasks as described above. We did not specify any time limits for the tasks, but required the participants to try to solve all tasks. All children were able to solve all specified tasks. Think-aloud was explained to the individual testers in terms of the descriptions in (Nielsen, 1993).

IV. DATA ANALYSIS

After recording the 36 test sessions they were then analysed in detail. The sessions were analysed based on how well children verbalized during think-aloud sessions and collaborated during constructive interaction sessions. First, we analysed each session for assessing the children behaviour in general. Then we analysed each session based on gender. The different aspects of our analysis were (i) degree of verbalization and collaboration, (ii) quality of verbalization , (iii) impact of test monitor on solving the tasks, (iv) communication between the test monitor and the user, and (v) prompting by the test monitor during the test. Quantitative values were assigned to each of these parameters on a scale of 1 to 5. A score of 1 means the lowest and 5 means the highest. For instance, a score of 5 assigned to verbalization means that the children verbalized their thoughts to the maximum during think-aloud sessions.

As part of our assessment of the two setups, we applied five different aspects of verbalization in usability evaluation. Theses aspects are illustrated in table 2.

TABLE II
ASSESSMENT OF VERBALIZATION IN TWO SETTINGS FOR ALL CHILDREN

Assessment Parameters	**Think-aloud**	
	Lab	Field
Degree of verbalization	M=1.89 SD=0.74	M=2.67 SD=0.67
Quality of verbalization	M=1.67 SD=0.67	M=2.44 SD=0.68
Impact of test monitor on solving the tasks	M=2.56 SD=0.88	M=2.56 SD=0.53
Communication between test monitor and tester	M=2.44 SD=0.88	M=2.56 SD=0.53
Prompting by the test monitor	M=3.11 SD=0.33	M=3.00 SD=0.71

M refers to the mean and SD indicates the standard deviation. Verbalization refers to the verbal comments during think-aloud sessions which would facilitate identification of what the tester is feeling about the interface under test.

Interestingly, we found that the degree of verbalization and quality of verbalization was different in different settings. The analysis of variance shows significance difference for the degree of verbalization F (1, 16) =4.9, p=0.042. Analysis of variance also show significance difference for the quality of verbalization with F (1, 16) =5.297, p= 0.035.

The test monitor plays an important role during usability evaluations. Test monitor's role has been emphasized in many usability studies. Test monitor is a person who closely monitors the usability test activities and notes the tester's behaviour, verbalization and other such things which may be of interest for the usability test under consideration. We analysed the impact of test monitor on solving the usability tasks. Even though the test monitor had to intervene more with the children during the field sessions, however, the impact of test monitor did not show significance difference between the setups.

Communication between the test monitor and testers was reported higher in the field. But the difference was not statistically significant F (1, 16) =0.11, p=0.75.

We also assessed the level of prompting that was required to make the testers verbalize their actions during the test sessions. Higher level of prompting was required in the lab. However, this difference was not statistically significant.

V. RESULTS AND DESCUSSION

Lab characterizes the real world context in a simulated controlled environment. It allows collection of high quality of data by minimizing the unwanted distractions and disturbances. However, it may lack ecological validity. The behaviour of the children in the lab was different from the field. Even though the lab which we had setup in this research was a known place to the children but the settings was new and formal. The lab had only the usability test monitors and the children participating in the test. Children behaved more disciplined and formal in the lab.

Field represents a place of natural use context of an application. Field comprises of the workplace such as an office, school, or a market, where the usability test participants can be tested by observing them use the application naturally. The data collected during field observation holds ecological validity as it has disturbances, noise and other obstructions. Children showed a slightly different behaviour in the field than in the lab. Field is a place where they felt more casual. They expressed a free attitude when they were tested in the field.

Results are an indication that physical context can affect the verbalization of children during usability evaluation. Our findings show that the amount of verbalization by children while completing the tasks in the lab was lesser compared to the field testing sessions. Also, the quality of verbalization that resulted during the field sessions was better than the lab sessions. This may be due to the fact that field gives children a natural environment to work and express themselves freely as they would do normally.

Test monitors role was more important in the field than in the lab. This could be due to the fact that field is an

uncontrolled environment with lot of disturbances and obstructions. Even though the test monitor was important in the field but lesser prompting by the test monitor was required to verbalize thoughts and actions especially when children were tested in pairs. This could be due to the fact that field offers natural environment to verbalize and also having a partner while solving tasks could make the children feel relaxed and confident.

Even though the effect of physical context on the outcomes of usability evaluation is clear from the test experiments we conducted, however, this research also poses some limitations such as: The children involved in our experiments were chosen from a particular geographical area of India having distinct culture and characteristics. Therefore, the results may not be generalizable to other user groups with different culture and characteristics than the user groups involved in our experiments.

REFERENCES

[1] Markopoulos,P., Read, J.C., Hoysniemi,J.& MacFarlane, S. (2008). Child Computer Interaction:Advances in Methodological Research. Cognition, Technology & Work, 10(2), 79-81.

[2] Neilsen, J. (1993). Usability Engineering. San Francisco: Morgan Kaufmann.

[3] Heim, S. (2008). The Resonant Interface-HCI Foundations for Interaction Design. USA: Pearson Addison Wesley.

[4] Rubin, J. (1994). Handbook of Usability Testing-How to Plan, Design, and Conduct Effective Tests. USA: John wiley & Sons.

[5] Dix,A., Finlay, J., Abowd, G.D. & Beale, R. (2004). Human Computer Interaction (3 ed.). Essex, England: Pearson Education.

[6] Schneiderman, B. (1998). Designing the User Interface-Strategies for Effective Human-Computer Interaction (Vol. 3rd.ed). Addison Wesley.

[7] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S.& Carey, T. (1994). Human-Computer Interaction. Addison Wesley.

[8] Brown, P.J., Bovey,J.D & Chen, X. (1997). Context-Aware Applications: From the Laboratory to the Marketplace. IEEE Personal Communications, 4(5), 58-64.

[9] Ryan, N., Pascoe, J. & Morse, D. (1997). Enhanced Reality Fieldwork: The Context-Aware Archaeological Assistant. In V. Gaffney, M. van Leusen, & S. Exxon, Computer Applications in Archaeology.

[10] Tullis, T., Fleischman, S., McNulty, M., Cianchette,C.& Bargel, M. (2002). An Empirical Comparison of Lab and Remote Usability Testing of Websites. Presented at the Usability Professionals Association Conference. Orlando, Fla.

[11] Tsiaousis, A.S.& Giaglis,G.M. . (2008). Evaluating the Effects of the Environmental Context-of-Use on Mobile Website Usability. Mobile Business, ICMB '08, 7th International Conference on, (pp. 314-322).

[12] Hummel, K.A., Hess, A.& Grill, T. (2008). Environmental Context Sensing for Usability Evaluation in Mobile HCI by Means of Small Wireless Sensor Networks. In MoMM '08: Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia, (pp. 302-306).

[13] Andreasen, M.S., Nielsen, H., Schoder, S.& Stage, J. . (2007). What Happened to Remote Usability Testing? An Empirical Study of Three Methods. Conference on Human Factors in Computing Systems, (pp. 1405-1414).

[14] Bruun, A., Gull,P., Hofmeister, L. & Stage,J. (2009). Let Your Users do the Testing: A Comparison of Three Remote Asynchronous Usability Testing Methods. CHI'09: Proceedings of the 27th International Conference on Human Factors in Computing Systems (pp. 1619-1628). Boston, MA, USA: ACM.

[15] Kaikkonen, A., Kallio,T., Kekäläinen, A., Kankainen, A. & Cankar, A. (2005). Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. Journal of Usbaility Studies, 1(1), 4-16.

[16] Razak, F.H.A., Hafit, H., Sedi, N., Zubaidi, N.A.& Haron, H. (2010). Usability Testing with Children: Laboratory vs Field Studies. User Proc. 2010 International Conference on Science and Engineering (i-USEr), (pp. 104-109).

[17] Andrrzejczak , C. & Liu,D. (2010). The Effect of Testing Location on Usability Testing Performance, Participant Stress Levels, and Subjective Testing Experience. Journal of Systems and Software, 83(7).

[18] Simatwa, E. (2010). Piaget's Theory of Intellectual Development and its Implication for Instructional Management at Presecondary School Level. Educational Research and Reviews, 5(7), 366-371. Retrieved October 4, 2011, from www.academicjournals.org/err/PDF/Pdf% 202010/July/Simatwa.pdf