

A COMPARISON OF USABILITY TESTING METHODS FOR AN E-COMMERCE WEBSITE: A CASE STUDY ON A MALAYSIA ONLINE GIFT SHOP

K. N. Goh¹, Y.Y. Chen¹, F.W. Lai², S.C. Daud¹

¹Computer and Information Sciences Department

²Management and Humanities Department

Universiti Teknologi PETRONAS

Bandar Seri Iskandar, Tronoh,

31750, Perak, Malaysia

A. Sivaji, S.T.Soo

Product Quality and Reliability Engineering

MIMOS Berhad,

Technology Park Malaysia,

57000, Kuala Lumpur,

Malaysia

ABSTRACT

Customer's perceived trust towards an e-commerce website is crucial for the success of online business. Effective design of web interfaces increases perceived trust of customers. Given many associated usability issues when performing tasks on a website, it is important for technopreneurs embarking on online business to understand issues related to usability problems of an e-commerce website and the techniques to identify these issues. In this study, usability evaluation was performed on an online gift shop with a group of potential consumers with age range of 18-22. Four different evaluation methods were used: Feedback Capture after Task (FCAT), Retrospective Think Aloud (RTA), Retrospective Think Aloud with Eye Movement (RTE) and observation. This study reveals that the major defects found were language and content, user guidance and support, flexibility and control, and visual clarity. Therefore, this study suggest that applying good user interface design could provide better user experience and thus increase perceived trust and user satisfaction towards the website.

KEYWORDS

Evaluation method, retrospective think aloud, observation, usability testing, eye tracking

1. Introduction

According to PayPal's "Online and Mobile Shopping Insights" study [1], the size of Malaysia online shopping was RM1.8 billion in 2010, estimated to be RM5 billion by 2014. Out of the RM14.8 billion, about RM825 million (45%) was spent on local business websites. This shows strong development and opportunity of e-commerce market in Malaysia. Majority of the Malaysians who purchased online falls into the age group between 21-40. In terms of products bought online, travel products such as flight tickets has the highest percentage which constitutes 24% of the total product and services bought online, followed by bill payment (18%), entertainment (14%) and IT & Electronics (12%). The less favorable products to buy online were gifts and collectibles. Hence our study involves evaluating and proposing some guidelines for designing an online gift shop.

The importance of usability testing to evaluate e-commerce has been well defined [2, 3]. User interface of

computer applications affects how people interact with the website and also their perception towards the website. The overall goal of usability from a user perspective is to measure and improve effectiveness, efficiency and satisfaction [4]. The usability of an e-commerce website is to provide users with satisfactory transaction effectively and efficiently. It helps to obtain a complete understanding of user's needs and to improve product development in order to provide a better user experience. Indeed, the usability of an e-commerce website is of utmost importance as it will affect consumer's trust towards the website and in turn their purchase intention [3]. Hence, the usability of an e-commerce website will determine the success of the e-commerce website. The purpose of this study is to outline the user interface defects in an online gift shop that contribute to trust and purchase intention on consumers in Malaysia context. The study focuses on Malaysian age between 18-22, the group that has the most experience to shop online.

2. Related Works

Trust and satisfaction are the main determinants for successful business relationships in business to consumer electronic commerce [5]. Trust plays a vital role in any form of business that requires monetary transaction. Customer's satisfaction has a direct relationship to customer's purchase intention, repurchase intention and word of mouth marketing. Consumer satisfaction is a measure of how well the product or services provided by companies meeting its consumer's expectations. A satisfied customer can expect higher purchase intention, repurchase intention and in effect promote the company to other potential consumers. Hence, there is clearly a need to study how an e-commerce website will influence consumer's trust and satisfaction.

Usability of a website is critical in determining the success or failure of a company [6]. However, many e-commerce applications still do not meet customers' usability requirements [2, 7]. Customers judge the credibility of a website within the first 50 milliseconds. Therefore, the website needs to be design with customer's behavior in mind. In this line, usability is an essential component of e-commerce strategy [8, 9]. Considering the importance of usability on e-commerce website, there were many studies conducted to understand how every aspect of an e-commerce

website can influence customer's behavior. For example, Papadopoulou and Pelet [10] studied how colors in an e-commerce website can affect consumer's mood and purchase intention. Work by Kamoun and Halaweh [11] investigated HCI factors that contributes to customer's security perception towards an e-commerce website. Their work has also re-emphasized on the importance of a good user interface design as an effective technique for increasing customer's trust and purchase intention. Sivaji et al. [3] showed the importance of applying Gestalt Principle, Fitts' Law and affordance across e-commerce websites. Since these elements are frequently present in a conventional shopping mall, it has been recommended to be virtualized for online transactions. The study also showed that once fundamental usability principles have been applied to the website, other principles such as trust, social presence, online and offline communication elements needs to be incorporated.

Most previous study uses existing HCI evaluation method such as user testing, heuristic evaluation, inquiry methods and simulation methods to evaluate web applications. Among all, user testing and heuristic evaluation method are most commonly used for e-commerce website [3, 12]. To perform a more complete evaluation, most previous studies uses more than one evaluation method in evaluating their web application. For example, a combination of user testing (think aloud protocol) and inquiry methods (interview, questionnaire). These methods help in generating a list of usability problems; however it does not provide guidelines on how to resolve these problems. Therefore, usability evaluation must take into account discovering the usability problems and repairing them [12]. With regards to the sample size for performing lab based usability testing, [13, 14] found that based on a binomial probability, six users will be needed to discover between 85% - 90% of the problems, given that the occurrence of the problem is 30%. Although this sample size does not represent the Malaysian demography, it is sufficient to detect significant usability problems. Another study by [15], also suggested 6 users if the research method involves qualitative eye tracking such as analyzing the gaze replays as opposed to solely depending on heat maps, which would require larger sample size.

3. Methodology

Six users were purposefully sampled from among the targeted web visitors to participate in the first user-based web usability testing at MIMOS Berhad UX Lab using the Tobii T60 Eye Tracker, Tobii Studio and URANUS [16]. All of them were male; with age range from 18-22. They were chosen as users as the website we are testing is mainly targeted for male users who are keen on purchasing gifts online. The details of the procedure have been published in previous studies [3, 17]. The six users took turns to enter the UX Lab, guided by a moderator, to complete a list of five outlined test cases (tasks). Each session lasted for the duration of one hour. The entire tests for all the six users are observed and the entire conversation between the users and the moderator were recorded. Collected observations and recorded audio files were used as inputs to further enhance

the online gift shop website interface to make it more user-friendly.

4. Results and Discussion

Table 1 list the task that needs to be completed by each user.

Table 1
Tasks and its description

Tasks	Description
1	Next Friday is your partner's birthday; you wish to buy him/her a small gift. You have a budget of RM30 to buy a gift for your partner. What are the 2 gifts that you think is most suitable for your partner?
2	Read the description of each product.
3	Find the product comparison tool and compare the price of 'Love Letter Keychain' and '3D Character Keychain'.
4	You are a new user to this website. Create a new account for this website.
5	You do not wish to buy anything today. You want to log off the account.

From the usability testing conducted, four sets of qualitative data was obtained, namely textual feedback or also known as feedback capture after task (FCAT), verbalization or also known as retrospective think aloud (RTA), retrospective think aloud with eye movement (RTE) and observation. In the traditional think aloud method, also known as concurrent think aloud (CTA) [18, 17], a participant would verbalize each action that they are performing at the point of time. However, it was found that cultural barriers make CTA method less suitable in countries with high power distances per the Hofstede's model [19] such as Malaysia. Sivaji and Ahmad [19] found that although the subjects have been encouraged to think aloud, they are reluctant as they are afraid that a failure of completion of a given task would reflect poorly on their performance. This is despite the moderator briefing the user at the beginning of the task that the purpose of the usability testing is to assess the web interface and not the user themselves. A detailed thorough study in regards to culture and think aloud method was done by Clemmensen et. al. [20]. Since the website was in English, language barrier also exist [19]. Thus, the traditional think aloud method will not provide us with the necessary feedback, which is useful for us to capture usability issues. Hence, [19] has proposed to complement retrospective think aloud (RTA) methods with eye tracking analysis as it is able to reveal key biometric information regardless of the cultural and language barriers. In this situation, a combination of 3 methods has been applied to capture defects or difficulties faced by the subjects to complete the tasks mentioned.

4.1 Feedback Capture after Task (FCAT)

This method is suitable, as it has been found that Malaysian users prefer to type their thoughts instead of verbalizing them due to the language and cultural barrier of the users [19]. The URANUS system [16] used in this study facilitates the end-to-end usability testing and FCAT. This is

achieved by prompting the user after the completion of each task to provide feedback on their experiences. FCAT however does not involve playback of any videos but instead rely purely on their short-term memory of the experience. There are high chances of users forgetting some of the issues they have faced. In this circumstance, the role of the moderator is to remind them. However, the advantage of this method is that users will remember the issues that impacted them the most, hence it is expected that most of the important issues will be fed back. Since no video playback is done, this method is also the fastest method to capture defects. The details of the defects captured using FCAT is further described in Section 5.2.

4.2 Retrospective Think Aloud (RTA)

The second method, known as RTA, includes video playback of the task performed by the user. Using this method, users are prompted by the moderator to talk about what he or she has just done to complete the tasks, which will include any additional comments the users might have. This method is particularly useful for us as it has proven to detect usability defects [17, 19] and it enables the users and moderator to observe and verbalize the activities that they have performed. Similar to FCAT, during RTA users have a tendency to forget their train of thought during performing the tasks, which may lead to fewer, incomplete, or reconstructed verbalizations of their thoughts [21].

4.3 Retrospective Think Aloud with Eye Movements (RTE)

The third method is similar to RTA but additionally during playback, the various eye tracking features such as gaze overlay, gaze plot and heat map is used to provide more insights visually. It is a combination of RTA with eye tracking. Also known as RTE or retrospective think aloud with eye movements, a term coined by Elling et al. [21], it is expected that there are a considerable increase in visual cues that may help the user's memory, hence on the number and types of observation and verbalization. RTE involves watching what each individual user does as they perform task and also listening to what they are saying as they think aloud. RTE provides the ability to follow user's gaze (gaze plot) around the screen and to review the recording in slow - motion to watch exactly where the user was looking [15].

4.4 Observation

The fourth method is a commonly used data collection method by observing the user without probing or intervening the situation. The user completes the usability testing with the moderator in a room which has one-way mirror. The observer sat in the observation room while observing, behaviour, actions of users and verbal comments were noted down as complimentary information to the other forms of qualitative data collected.

5. Analysis and Discussion

5.1 Comparison of defect count by method

A comparison on the number (Table 2) and types of defect detected (Table 3) for each method was carried out. From Table 2, it could be seen that using the RTE, 18 defects were observed and/or verbalized as compared to 10 and 6 defects

for the RTA and FCAT defects respectively. The findings from this study are similar to Ball et al. [22] whereby RTE offers promising results and more insightful information. This however is in contrary to Elling et al. [21] findings whereby they found no additional value of showing users eye movements. Our findings however shed some light over Elling et al.'s [21] expectation of RTE whereby it improves the interaction between subjects and moderator in terms of difficulties.

Table 2
Defect count for each method

FCAT	RTA	RTE
6	10	18

Now, we describe each defect in detail. Table 3 shows the defect detection capability of each method. A 'Yes' indicates the method was able to detect the defects while a 'No' means otherwise.

5.2 Defects Detection by RTE Analysis

Figures 1 to 12 shows the various visual cues obtained using the eye tracker such as a gaze plot (Fig 1,2,4,5,7,9a,9b,10,11) , gaze overlay (Fig 8a, 8b,12) and heat map (Fig 3, 6) . RTE reveals difficulty in information extraction by users as shown in Fig 9a, fixation 22 which lasted for a duration of 1.366 seconds. This is consistent with previous studies [23] whereby long fixation duration last more than 320ms. In this study, the long fixation is due to the usage of highly saturated colors that increase cognitive load. RTE also reveals user interface that has less efficient search as shown by the high number of saccades and fixation pairs in Fig 1, 2, 4, 5 and 7, which is consistent with studies [23, 24, 25]. The major defects experienced by users are described in detailed in Table 3 (defect details column) and correspondingly the RTE column make reference to visual cues in Fig 1 to 12.

Table 3
Defect Detection Capability of FCAT, RTA and RTE

User	Defect Details	FCAT	RTA	RTE
6	It takes a user a significant amount of time (23 seconds) to spot the "Gift" label	No	No	Yes
6	During the 23 seconds of searching for the "Gift" label, there was a significant amount of cognitive workload resulting in 108 saccades	No	No	Yes (Fig. 1)
6	User unable to find pricing information from the "Gift" link. After searching for 45 seconds, user gave up	No	Yes	Yes (Fig. 2)

6	User misses gazing on pricing information while zooming into the image - Pricing information was not prominent as there was no fixation recorded both from heat map and gaze plot	No	No	Yes (Fig. 3, 4)
6	When the zoom window is open, the price information is clipped off	No	Yes	Yes (Fig. 3)
6	Insufficient product choices	Yes	Yes	Yes (Fig. 1)
1	User was scanning for more products	No	Yes	Yes
1	User was scanning within the same page for a period of 1 minute and 21 seconds with 250 fixation and saccades. User was attempting to look for more product options, gave up at the end and settled for 2 products within the budget as shown in the main page	No	No	Yes (Fig. 5, 6)
2	Insufficient product choices	Yes	Yes	Yes
2	User hopes to see all products at once	Yes	Yes	Yes
2	User states inability to sort product by price range	Yes	Yes	Yes
2	User states that there is no advance search option for user to type budget and display products according to budget	Yes	Yes	Yes
2	User states that there is no information that list the most purchased items within the users budgets	Yes	Yes	Yes
2	User faces difficulty performing task as it requires scanning for 1 minutes and 42 seconds with over 336 gazes	No	No	Yes (Fig. 7)
2	User's cognitive load increases when	No	No	Yes (Fig. 8a,

	trying to understand the items in the sub-menu. The high red saturation used for the menus slows down reading as the white fonts used are not visually clear.			8b, 9a, 9b)
4	While searching for product, fixation 8 shows that user was trying to perform some action on the image for about 3 seconds. (2:17:894 – 2:25:309). The user tried to click on the View “Love Letter Keychain” but there was no response. Two reasons for this: 1) The target size of the “View” label is too small (Fitt’s Law) 2) The performance of the system was slow Hence, the user starts to right click instead to try performing some action. When this did not work, the user went away to fixation 9, 10 and in fixation 11, the user came back to the target (“View”) to try again for 1.016seconds and only at 2:28.551, the system responded, so there was a delay of about 7 seconds and the user had moved on to fixation 38	No	No	Yes (Fig. 10, 11,12) Yes (Fig. 10, 11,12)
4	User is unsure the difference between selecting the “View” hyperlink and clicking on the product label. Poor visual clarity on the purpose of each function. Visual for product label is black in color as compared to the view that is in yellow. There exist consistency issues.	No	Yes	Yes



Figure 1. Cognitive workload – 108 fixation-saccades pair



Figure 4. Gaze plot of user misses gazing on pricing



Figure 2. Missing pricing information in gift link

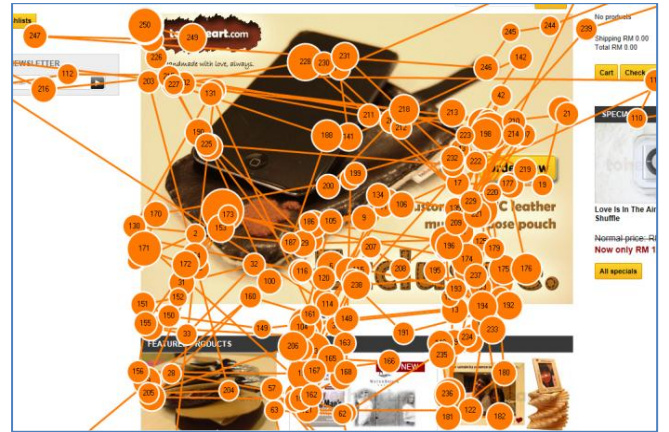


Figure 5. Searching task producing 250 fixation within a period of 1 minute and 21 seconds

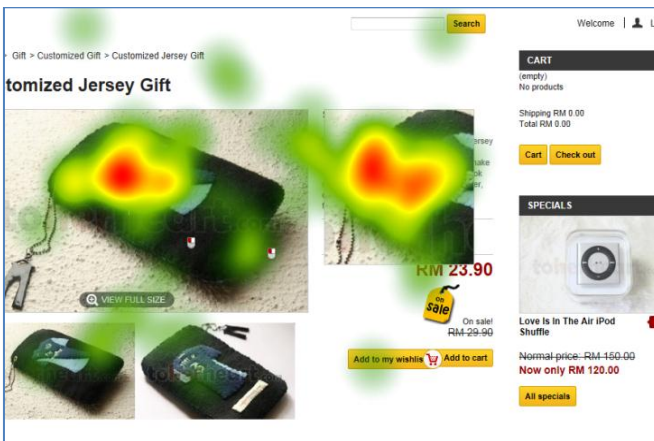


Figure 3. Heat map of user misses gazing on pricing information

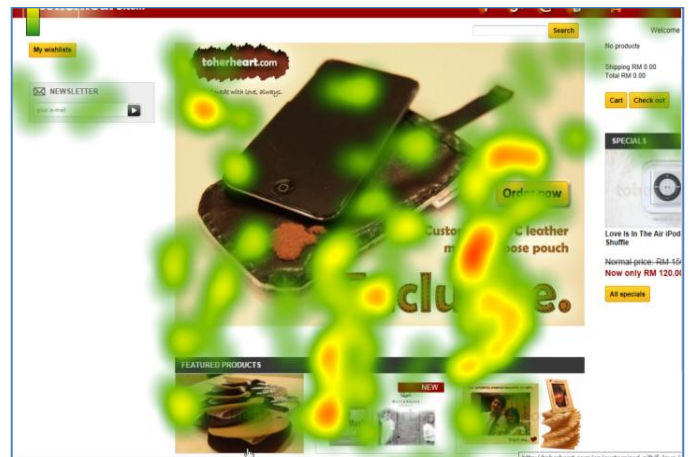


Figure 6. Searching task within a period of 1 minute and 21 seconds. The areas of high interest in the heat map could have been filled up with more product choices

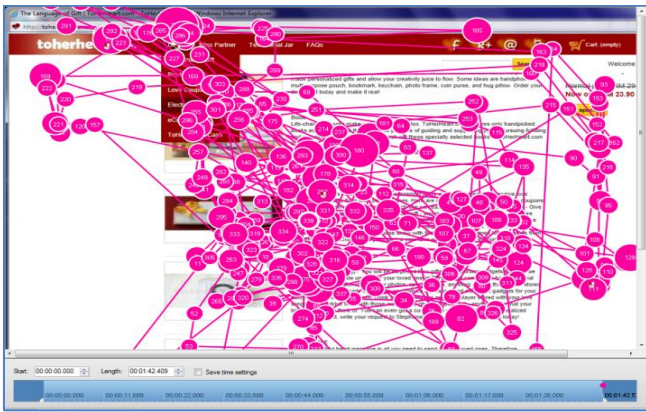


Figure 7. User faced difficulty performing task as requires scanning for 1 minute and 42 seconds with over 336 gazes

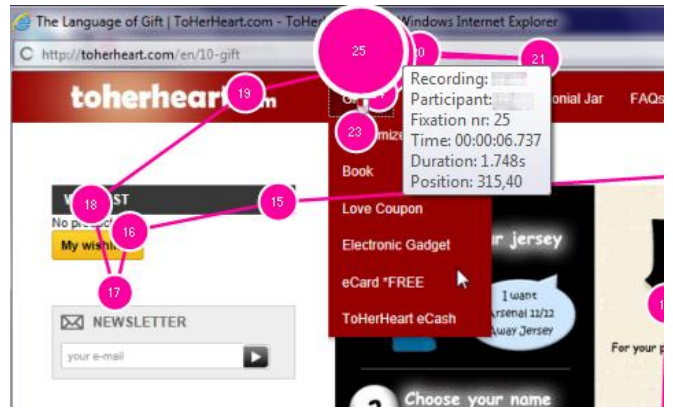


Figure 9b. Fixation 25 at 1.748 second duration



Figure 8a. Large gaze overlay on 'Gift' menu



Figure 10. Fixation 8 and 11 with 3 mouse clicks

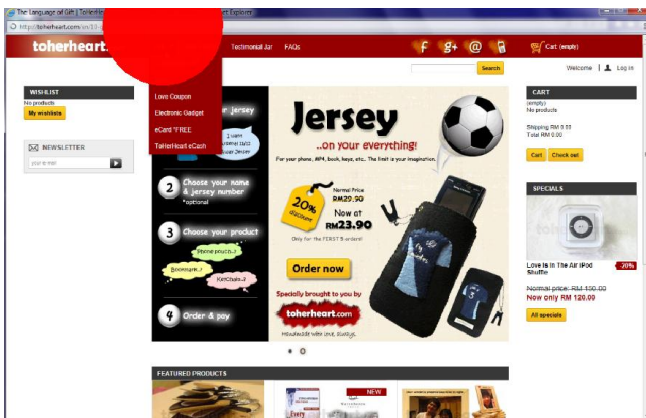


Figure 8b. Larger gaze overlay on 'Gift' menu



Figure 11. Fixation 8 duration 3.265 seconds

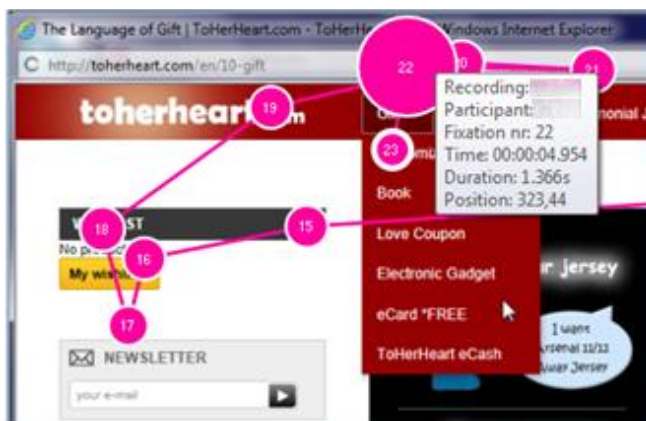


Figure 9a. Fixation 22 at 1.366 second duration

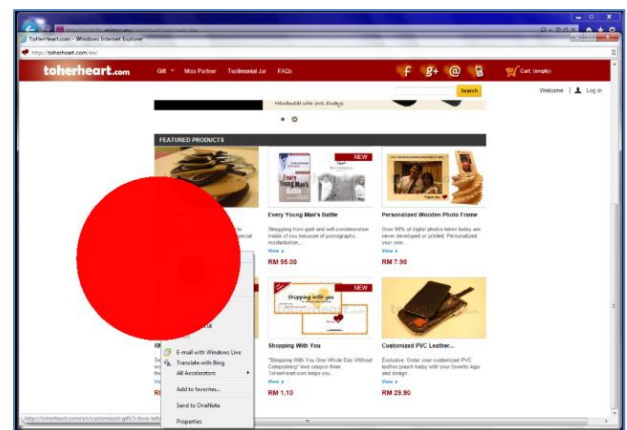


Figure 12. Gaze overlay with and right hand click action

5.3 Defect Categorization for FCAT

Based on FCAT method, we extracted the problems mentioned by users into categories. The qualitative feedback gathered from all six users broadly falls into eight categories: 1. Language and content, 2. User guidance and support, 3. Flexibility and control, 4. Visual clarity, 5. Consistency and standard, 6. Navigation, 7. Functionality, 8. Informative feedback. This paper would only highlight major problems faced in some of these categories.

1) *Language and content*: 33% of the users raised the issue that there were insufficient product choices. As the items sold on this website were handmade and personalized, mass production and variation of products is not feasible in such a small-scale business. However, more products were eventually added into the website to allow variations to potential buyers.

2) *User guidance and support*: It was raised by 33% of the participants that they found it confusing trying to sign up for the website. In task 4, users were asked to create a new account. Users did not expect this feature to be found in login (Figure 13). This comes to show that separate term must be explicitly mentioned to show separate process to sign up and login as shown in Figure 14.

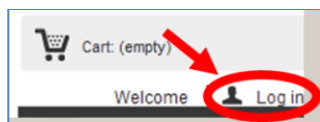


Figure 13. Login

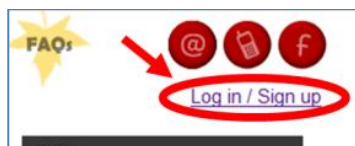


Figure 14. Login and Sign up

3) *Flexibility and control*: In task 3, users were asked to find the tool to compare the price between two products. 50% of the users mentioned that they were not able to compare more than two products at a time. And another 50% were not able to find the comparison tool, thus were unable to complete the task. This comparison feature was taken off because users mentioned the function is not too crucial and is more useful in comparing products like mobile phones.

4) *Visual clarity*: Each product displayed on the website has a description of the product and its features. 33% mentioned that the font size used for its product description and features were not legible enough to read. Font size was increased to ensure legible reading for users. As mentioned, it was also noted by 50% of the users that it was very difficult to find the comparison tool, thus they were not able to complete task 3. 50% of the users also mentioned that some products were not visible due to overlapping product description while doing task 3 as seen in Figure 15. This was a technical issue that was not expected prior to the test.

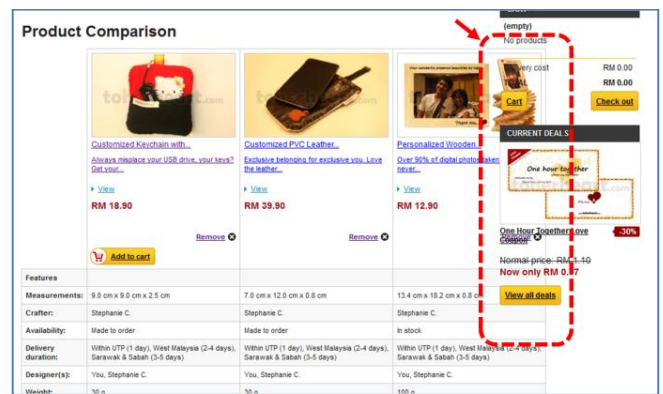


Figure 15. Overlapping problem faced by user

For other categories not mentioned specifically, problems raised by the users were minor and only affects a small percentage (16.67% or one user). Some examples would be no welcoming note when the user signs into the website or was not able to zoom into the product picture. Minor issues were easily fixed based on participants' comments and recommendation.

After users completed each task, they were prompted to type in some comments. Comments typed were found to compliment the retrospective think aloud conducted. However, prompting by the moderator was necessary to encourage them to comment in detail. Reminders of what they said during the think aloud session helped them to remember the problems encountered earlier and that will trigger them to explain more.

5.3 Observations

1) *Task 1*: Most of the users did not realize or took some time to realize the existence of the "Gift" menu in the top navigation bar to start their product browsing. Four out of six users suggested to have an additional product category that holds all the products sold in order to allow users to choose gift from all the products that are available, instead of having to browse product category after product category to see what gifts are available. Other suggestions included enabling product filtering function and eliminating malfunctioning links.

2) *Task 2*: Almost all users requested the fonts used to be larger. Two users commented that the product descriptions could come with more practical details, like how long would it take for the customers to receive the ordered products from the time the order is placed, as well practicality of some of the products that might be new and unfamiliar to users. Contrarily, other four users commented that the existing product descriptions were sufficient to help them make purchase decision. One user mentioned that he would initiate to contact the web owner to clarify any doubt before placing order.

3) *Task 3*: Three users skipped this task case due to the difficulty in locating the product compare button, while the last user did not participate in accomplishing test cases 3 through 5 due to the unexpected downtime of the web hosting server. For the two users who managed to accomplish this task, they both commented that the

overlapping interface, which occurred when comparing three products, distracted their browsing experience. Another suggestion was to allow users to add the product into comparison basket right from the product page itself instead of only enabling the comparison button in product category page.

4) *Task 4*: Among the five users, all of them were able to locate the log in button very quickly but two were reluctant to create an account as requested by the task due to privacy concern of personal information. However, all five users created their respective accounts in order to accomplish this task. Suggestions received included adding tooltips on required fields on the account registration form which are unfamiliar to users, eliminating unnecessary fields required in order to make purchase, displaying welcoming message to explicitly inform users that they have successfully logged in and allowing registered users to personalize the website theme color according to their preference. Besides, the position of the displayed user's name after logging in overlapped other web content due to long user name.

5) *Task 5*: All five users who did this test case managed to locate the log out button very quickly and they could log out successfully. One user suggested adding a confirmation dialogue box to make sure that the user really wants to log out every time after the log out button is clicked.

From RTA and observational method, it was found that there were no significant differences between the two methods used, even as reported by [26].

6. Conclusion

From this study, it was found that by using FCAT, RTE and RTA method, a significant number of defects were detected through this test. The methods used complemented each other in finding defects, while some methods were easier to identify than others or revealed more. However, for RTE method, we manage to find 3 times more defects as compared to RTA. RTE reveals more data than we expected with the usage of eye-tracking device. Since data through this method were abundant, proper analysis is necessary to identify and extract only relevant data. Eye tracking (cognitive load) method reveals lots of information, thus further understanding is necessary to comprehend cognitive processing and how this method could be better exploited to reveal identification of more defects and its accuracy.

References

- [1] *PayPal: Online Shopping Hits 1.8 Billion*. (2011). Retrieved from <http://marketing-interactive.com/news/25946>
- [2] Li, F. and Li, Y. (2011). Usability evaluation of e-commerce on B2C websites in China, *Procedia Engineering*, 15, 5299-5304.
- [3] Sivaji, A., Mazlan, M. F., Soo, S. T., Abdullah, A., & Downe, A. G. (2011). Importance of Incorporating Fundamental Usability with Social, Trust & Persuasive Elements for E-Commerce Website, *IEEEExplore Business, Engineering and Industrial Applications (ICBEIA)*, pp. 221-226.
- [4] Bevan, N. (2008). Classifying and selecting UX and usability measures, *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, 18th June 2008, Reykjavik, Iceland
- [5] Kim, D.J., Ferrin, D. L. and Rao, H. R. (2009). Trust and Satisfaction, Two Stepping Stones for Successful E-Commerce Relationships: A Longitudinal Exploration, *Journal Information Systems Research*, 20(2), pp. 237-257.
- [6] Tezza, R., Borgia, A.C. and Andrade, D. F. (2011). Measuring web usability using item response theory: Principles, features and opportunities, *Journal Interacting with Computers*, 23(2), pp. 167-175.
- [7] Offutt, J. (2002). Quality attributes of Web software applications, in: *IEEE Software: Special Issue on Software Engineering of Internet Software*, pp. 25-32.
- [8] Oneupweb. (2010). *Revolutionizing Website Design*. Retrieved from: <http://www.oneupweb.com>
- [9] Shi, M. (2011). Website Characteristics And Their Influences: A Review On Web Design, Association of Business Information Systems (ABIS) Refereed Proceedings. Retrieved from http://www.abis-fbd.org/wp-content/uploads/2011/03/ABIS_2011_Proceedings.pdf#page=147
- [10] Papadopoulou, P. and Pelet, J. (2009). The effect of colors of e-commerce websites on consumer mood, memorization and buying intention, *MCIS 2009 Proceedings*. Paper 116.
- [11] Kamoun, F. and Halaweh, M. (2012). User Interface Design and E-Commerce Security Perception: An Empirical Study, *E-business Research*, 8(2).
- [12] Fernandez, A., Insfran, E. and Abrahao, S. (2011). Usability evaluation methods for the web: A systematic mapping study, *Information and Software Technology*, 53, pp. 789-81.
- [13] Sauro, J. Why you only need to test with five users (explained) (2010) Retrieved from <http://www.measuringusability.com/five-users.php>
- [14] Sivaji, A., Mohd Nor, Z. and Downe, A. G. (2011). Comparison of the Lab Based Usability Testing and Exploratory Heuristic Evaluation: A Case Study on Human Factor Conference Website Usability. *Malaysian Journal of Ergonomics*, pp. 57-71.
- [15] Pernice, K. and Nielsen, J. (2009). How to conduct eye tracking studies, 20, 51 Retrieved from http://nngroup.com/s3.amazonaws.com/media/reports/free/How_to_Conduct_Eyetracking_Studies.pdf
- [16] Sivaji, A. and Soo, S. T. (2012). Website User Experience (UX) Testing Tool Development using Open Source Software (OSS). In press *IEEEExplore, 2nd South East Asian Network of Ergonomics Societies (SEANES)*, Malaysia.
- [17] van den Haak, M. J., DeJong, M. D. T. and Schellens, P. J. (2010). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22 (5), pp. 339-351.
- [18] Yu, B. and Voll, K. (2011). Probing student problem solving skills in mathematical induction using a scenario based think aloud protocol. *Proceedings of the 16th Western Canadian Conference on Computing Education (WCCCE '11)*, pp. 33-37.
- [19] Sivaji, A. and Wan Ahmad, W. F. (2012). Benefits of complementing Eye Tracking analysis with Think Aloud protocol in a multilingual country with high power distance (Hofstede's model) for website user experience testing. In press *Current Themes in Eye Tracking Research: Method and Application*.
- [20] Clemmensen, T., Hertzum, M., Hornbaek, K., Shi, Q. and Yammiyavar, P. (2008). Cultural Cognition in the Thinking-Aloud Method for Usability Evaluation. *Journal Interacting with Computers*, 21 (3), pp. 212-22
- [21] Elling, S., Lentz, L. and de Jong, M. (2011). Retrospective think-aloud method: using eye movements as an extra cue for participants' verbalizations. *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, pp. 1161-1170.
- [22] Ball, L. J., Eger, N., Stevens, R. and Dodd, J. (2006). Applying the PEEP method in usability testing. *Interfaces*, 67, pp. 15-19
- [23] Ehmke, C. and Wilson, S. (2007). Identifying web usability problems from eye-tracking data. *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it*, 1, pp. 119-128.
- [24] Goldberg, J. H. and Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs, *International Journal of Industrial Ergonomics*, 24(6), pp. 631-645.
- [25] Poole, A. and Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research. *Encyclopedia of human computer interaction*. Idea Group, Pennsylvania, pp. 211-219
- [26] Baauw, E. and Markopoulous, P. (2004). A comparison of think aloud and post-task interview for usability testing with children. *Proceedings of the 2004 conference on Interaction design and children: building a community (IDC '04)*, pp. 115-116.