

Benefits of Complementing Eye-Tracking Analysis with Think-Aloud Protocol in a Multilingual Country with High Power Distance

Ashok Sivaji and Wan Fatimah Wan Ahmad

1 Introduction

Sherman (2009) reported the various user experience (UX) techniques practised across 34 countries. The 1,786 respondents comprise of UX practitioners, usability professionals, user researchers, user-centred design (UCD) practitioners, interface designers and others. Among the techniques reported in Fig. 1, this study includes eye tracking and lab-based usability testing (LBUT), as was reported by Sivaji et al. (2013).

2 Problem Statement

During the LBUT performed by Sivaji et al. (2011, 2013), Goh et al. (2013) and Abdollah et al. (2013), an eye tracker was used to capture the user's feedback such as audio, video and eye gaze data (fixation and saccades). The eye tracker keeps tracks of user's eye movement while they are performing the task. The LBUT also involves encouraging users to think aloud (TA) while they perform some task. However, from this, it has been found that the users have difficulty with the TA method. There are few reasons for this.

2.1 Hofstede's Power Distance

Hofstede (2005) and Yammiyavar et al. (2008) have shown that in countries like Malaysia and India, there exists high power distance as per the Hofstede's model. The power index score for Malaysia was recorded as 104 by Hofstede, being the highest

A. Sivaji (✉)
MIMOS Berhad, Kuala Lumpur, Malaysia
e-mail: ashok.sivaji@mimos.my

W. F. Wan Ahmad
Universiti Teknologi PETRONAS, Perak, Malaysia

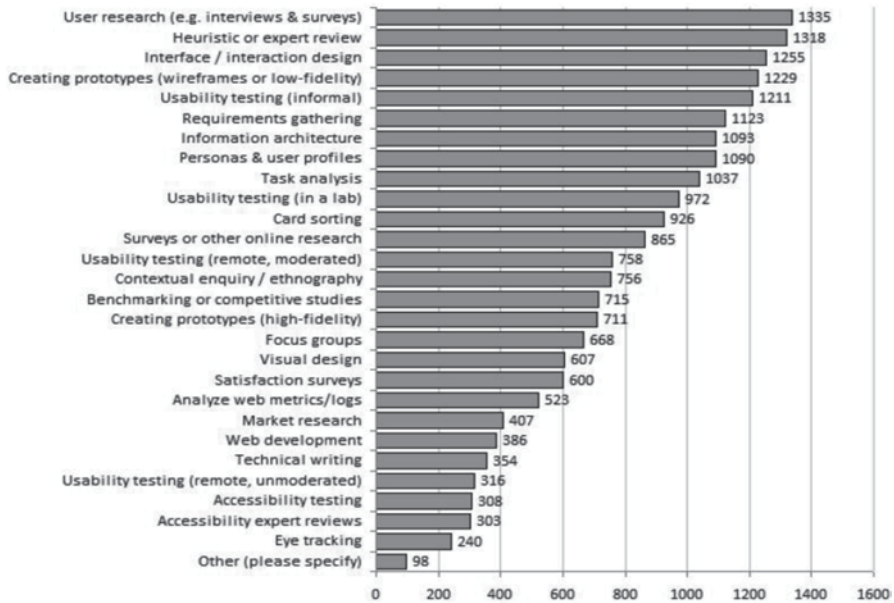


Fig. 1 User experience (UX) techniques employed by usability practitioners worldwide

around the world. In the context of usability testing, power distance could be defined as the relationship between the user who is being tested and the moderator who is facilitating the testing. In most cases, the moderator will recruit the users, after assessing his/her suitability to perform the testing based on a user screening questionnaire. During the testing, the user is expected to provide feedback based on the usability test session conducted by the moderator. Due to the power distance that is already present in the Malaysian culture, the user, during the TA process, sees the moderator as a supervisor and hence has a tendency to be afraid in disagreeing in the effectiveness, efficiency and satisfaction of degree of usability of a website under test. This is one reason why TA technique alone may not be suitable and reliable in usability studies in Malaysia. Yammiyavar et al. (2008) and Goh et al. (2013) also found that there is a rich amount of non-verbal behavioural data such as eye, hand and head gestures that are collected during the LBUT. It is almost impossible for the moderator to capture all these information during the LBUT. Sivaji et al. (2011) found that although the subjects have been encouraged to TA, some subjects are reluctant to do so, 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 LBUT is to assess the web interface and not the users themselves.

2.2 Persuasive Power

During the LBUT, when the moderator transcribes the TA feedback and raises the problem faced by the users as a defect, it is common for the design and development

team to be sceptical and defensive. They even go to the extent of requesting for further evidence. This shows that LBUT and TA are not sufficient to convince the design and development team of the validity of the defects. In this case, it is important to complement the findings from LBUT and TA with a more visual method. Eye-tracking analysis has high persuasive power. Blandford et al. (2008) has identified persuasive power as one of the important criteria of assessment from multiple UX methods. This study aims to show how eye tracking can increase persuasive power of highlighting usability defects to the stakeholders.

2.3 Multilingual Society

Malaysia is a multicultural, multiracial and multilingual country. According to the 2010 Population and Housing Census of Malaysia by the Department of Statistics Malaysia, the citizens comprise of 67.4% Bumiputera, 24.6% Chinese, 7.3% Indians and 0.7% others. The official language of Malaysia is Malay or also known as Bahasa Malaysia. English remains as a second language and is taught in school. Malaysian English sees wide use in business, along with Manglish, which is a colloquial form of English with heavy Malay, Chinese and Tamil influences. It is common that during the usability moderation, some participants may tend to highlight some words in Malay, Chinese or Tamil. The moderator should translate this into English for the benefit of the international audience. However, there are chances for misinterpretation or miscoding of the TA feedback between the moderator and the users.

3 Objectives

The objectives of this study are as follows:

- Assess the level of the power distance index (PDI) in Malaysia
- Propose a new LBUT methodology that will incorporate both TA and eye tracking
- Perform an LBUT case study on a website based on the Malaysian demography
- Analyse and assess impact of the newly derived LBUT methodology

4 Current Study

4.1 Power Distance Assessment

Previous studies (Hofstede 2005 and Oshlyansky 2007) have performed surveys among Malaysians to gauge PDI. In order to validate these findings in the current context, Values Survey Module (VSM) 1994 questionnaire was distributed among Malaysians. The study revealed a PDI score of 36.8 from the 44 respondents. This shows that our findings (PDI of 36.8) is significantly lower than Hofstede (2005) hav-

ing a PDI of 104. Our result, however, is closer to that reported by Oshlyansky (2007) being 23.47. One significant difference between the subjects for this study is that the sample was more targeted to working adults (working experience ranging from 1 to 35 years) as compared to postgraduates and undergraduates Oshlyansky (2007).

Although we found a lower PDI score of 36.8, we still believe there existed a significant power distance, as a closer look at Question 14 of the VSM revealed that significant power distance still existed as the average rating is “Sometimes”; for the response to “How frequently, in your experience, are subordinates afraid to express disagreement with their superiors?” This has a significant impact when users perform usability testing as there could be some reluctance in providing an honest opinion or feedback, especially negative feedback. Additionally, the moderators from MIMOS UX Lab who were involved in previous UX studies, such as Sivaji et al. (2011, 2013), Goh et al. (2013) and Abdollah et al. (2013), have also observed this high power distance response whereby subjects feel that their responses are being recorded and could be used against them when an issue is reported although anonymity in recruitment and reporting is practised.

4.2 *LBUT with Eye Tracker*

The process flow for the LBUT methodology employed in this study is shown in Fig. 2. The users are recruited as per the Malaysian demography who have participated in the VSM survey. Out of the six users, four were Malay, one was Chinese and another Indian. The seventh user was a native English speaker who was chosen as a control sample for this study. The websites chosen for this study was the Nielsen Norman Group (<http://www.nngroup.com/>) whereby the users were tasked to determine the registration details for a particular event.

The steps that are shaded in grey (Fig. 2) involve using an eye tracker. The eye tracker that was used in this study is the Tobii T60 with Tobii Studio 2. This version enables retrospective think aloud (RTA) where the moderator and users can play back the video to view the session. The test environment setup involves arranging a logical sequence of task, selecting a website URL to be tested, designing of the subjective ratings questionnaire. These information are then configured into URANUS, that would automatically link to the Tobii Studio. URANUS is an open source software that is developed by Sivaji et al. (2012) to facilitate usability testing of websites and any user interface. It is developed in such a way that practitioners are able to integrate with Tobii Studio. After the briefing session, subject will start to perform the first task. This will involve the subjects concurrently thinking aloud (CTA) as they perform the task. These will be recorded by the eye tracker. The subjects gaze patterns and mouse clicks are recorded for further analysis. The moderators would rate the effectiveness and efficiency of the task performed by the subjects. The subjects would also be able to provide some feedback based on a questionnaire. After all tasks are completed, the moderator will play back the recordings to confirm with the users on why they have reacted in a particular manner. For instance, if they really liked an interface, more details could be asked on what elements of the interface that attracted them. Conversely, if they had problems

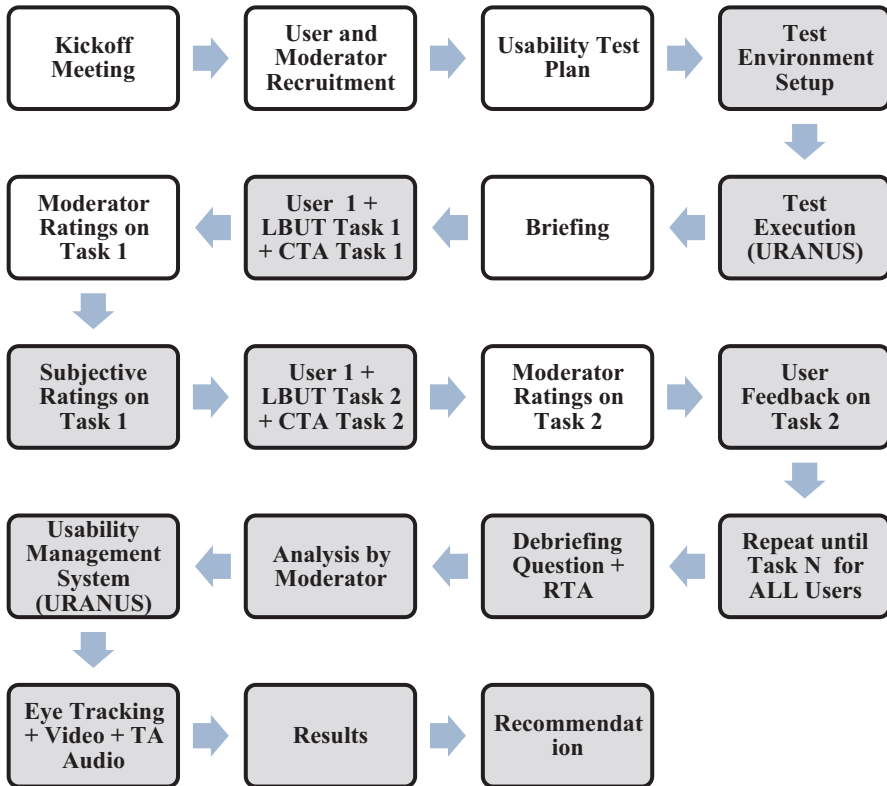


Fig. 2 Process flow for lab-based usability testing (LBUT) with eye tracker

with the interface, it will be interesting to determine the reason for that. In addition to that, the moderators would also gather the feedback from the observers to validate certain segments of the task with the users to gain a thorough understanding of the subject reaction towards the interface. This is called RTA. The moderators would also use the eye tracker to perform analysis, and obtain recommendation for fixes.

5 Results and Discussion

During the LBUT, an eye tracker was used to capture the user's feedback such as audio, video and eye gaze data (fixation and saccades). The eye tracker keeps track of the user's eye movement while they are performing the task. Although it is not mandatory to use an eye tracker in an LBUT study, there are some benefits to using one. If a study is carried out without an eye tracker, the moderator only has to rely on the feedback obtained from TA. With the eye tracker, the moderator can now support the TA feedback with one of the human biometric feature, which in this case is the user's eye. This increases the data integrity obtained from all users.

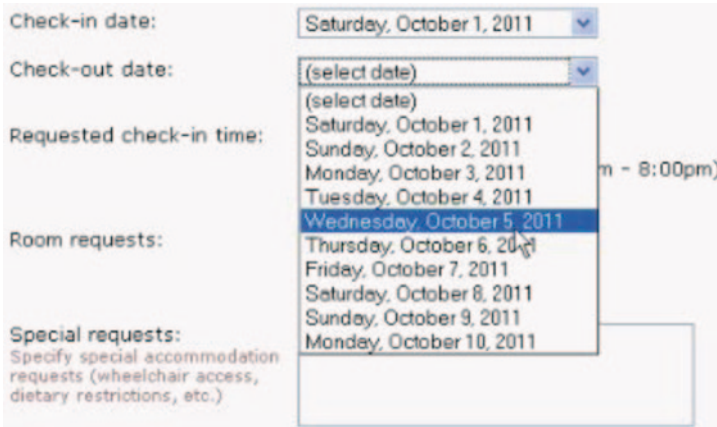


Fig. 3 Original image to be analysed

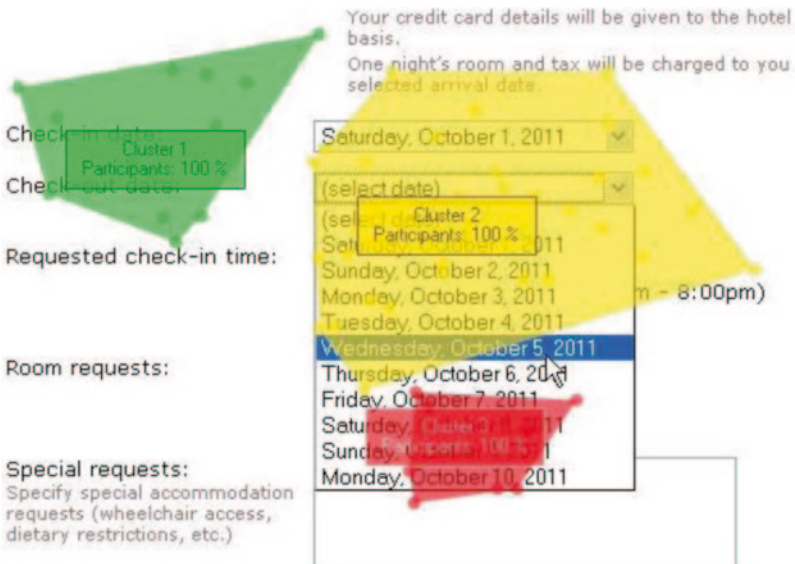


Fig. 4 Cluster plot

Figures 3–7 illustrate the various eye-tracking analysis features that are used in this study. Figure 3 has been chosen to be analysed, as it has some insightful TA feedback. Using the Tobii Studio’s visualisation tools, the following observations could be made:

- The cluster plot (Fig. 4) shows the areas of high concentration of gaze data points when this task was performed. Based on this, the moderator could mark certain areas of interest (AOI) for further analysis.

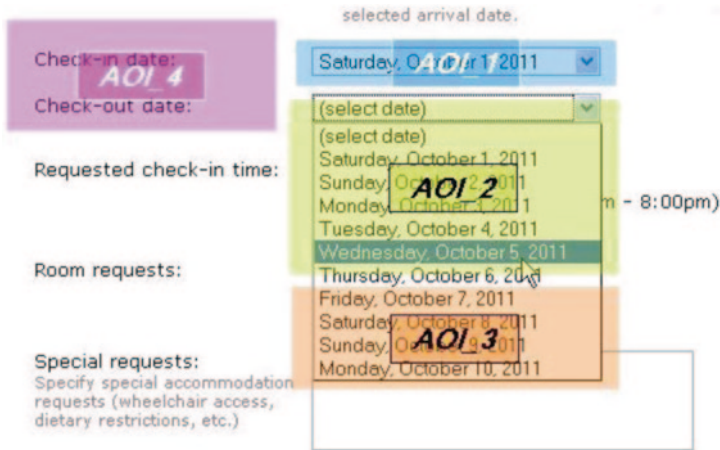


Fig. 5 Areas of interest



Fig. 6 Heat map

- Figure 5 shows the four AOIs that have been marked with AOI 1–4. AOI 4 corresponds to cluster 1, while AOI 3 corresponds to cluster 3. Since cluster 2 spans a larger space, it is divided down to 2 AOIs, namely AOI 1 and AOI 2.
- The heat map as shown in Fig. 6 highlights areas based on the fixation duration and fixation count. Areas that receive more fixation concentration is shown as

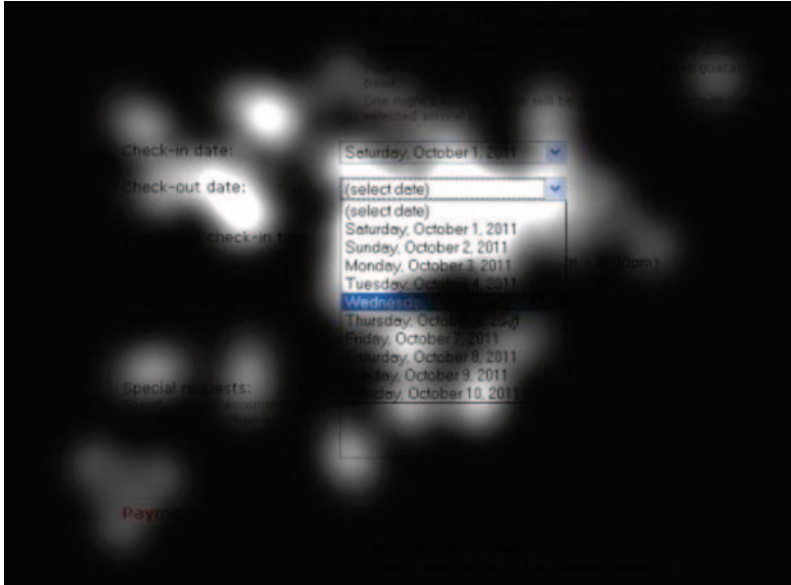


Fig. 7 Gaze opacity

red, while areas that received lesser fixation are marked as green. It is a quick way to show how a group of users have focussed on some areas of the web page that is of high interest.

- The inverse plot of the heat map is the gaze opacity plot as shown in Fig. 7. This plot hides out areas that have received fewer fixations. This way the areas with the most fixations are highlighted.

6 Analysis

Some of the other descriptive statistics that are generated from the eye tracker, based on these AOIs, are fixation count, fixation duration, time to first fixation and mouse click-related data. It is useful to correlate the TA feedback obtained from the users with the descriptive statistics obtained from the eye tracker.

6.1 Correlation of TA with Eye Tracker

When subjects (user 1–7) were performing the registration task on one of the eight websites, the TA feedback as shown in Table 1 was recorded from the task being carried out.

Table 1 Concurrent think aloud (CTA) feedback for the interface used to display the date to register for the conference

User	Concurrent think aloud
1	“Wow, I like this feature, it is different from other sites, others sites would have used a calendar, and expected me to scroll up and down to the landing page to look for the conference dates to register, ...I could have even selected the wrong conference date”
2	“Date is Sunday October 2011”
3	“Found the date”
4	“Sunday October 2011”
5	“”
6	“tak dapat jumpa tarikh”
7	“”

6.2 TA and Eye Tracker’s Visualisation

Figure 7 clearly shows that users were attracted to the drop-down menu, which actually displayed the relevant content that is the “Check Out Date” data of the conference. Even the heat map (Fig. 6) highlighted the drop-down menu with red, indicating areas of high fixation. This shows that the AOI 2 was the areas of high interest.

6.3 TA and Eye Tracker’s Descriptive Statistics

The eye tracker can also support descriptive statistics as shown in the following tables. From Table 2, it could be seen that AOI 2 had 25 fixation counts. As pointed out by Ehmke and Wilson (2007), longer fixation indicates that the elements are more noticeable and/or more important. It also had the longest fixation duration of 9.22 s as compared to the remaining AOIs. Table 2 also shows that the visit duration for AOI 2 is 10.45 s, which is much higher than the remaining AOIs. This is consistent with the TA feedback from the users “...Wow, I like this feature...”.

The correlation and consistency between the TA and eye tracking enable the moderator to conclude that this site has a better navigation strategy than other sites. Just like the above segment of TA, many other parts of the user’s recording could be further analysed with an eye tracker to gain insightful data of the web page design. From the results shown in Table 2, the moderator can highlight this feature as desirable and present it to the project team with visual evidence. Since the project team has the visual evidence to support their design, they could reuse this interface and incorporate it as a design best practice for their organization. Eventually, other designers will be able to adopt this practice, since it has been validated. Without the eye tracker, the moderator will not be able to recall visually what was happening, just based on the moderator notes which were gathered from the audio recordings from the TA. Now they have video and audio evidence with user’s eye as a biometric to proof a point. This enables the usability analyst (moderator) to present the findings with visual, audio and statistical evidence,

Table 2 Descriptive statistics from eye tracker for user 1

User	Fixation and visit duration											
	AOI_1			AOI_2			AOI_3			AOI_4		
	N (C)	Mean (s)	Sum (s)	N (C)	Mean (s)	Sum (s)	N (C)	Mean (s)	Sum (s)	N (C)	Mean (s)	Sum (s)
Fixation	4	0.16	0.65	25	0.37	9.22	8	0.31	2.46	6	0.22	1.33
Visit	4	0.16	0.65	10	1.05	10.45	6	0.50	2.98	5	0.30	1.52

eventually increasing the persuasive power towards the design and development team to implement changes.

6.4 Hofstede’s Power Distance and RTA

Among the seven users, it could be seen that user 1 has low power distance while the remaining users have high power distance. User 1 was from a native English-speaking country, while the remaining users were from Malaysia. Among the Malaysian users, two out of the six were quiet throughout the entire moderation. Although they were successful at completing the task (finding the date), they were not comfortable to TA because they had problems verbalising in English. This would make it difficult for the moderators to provide ratings on the interface without justification if there were only depending on RTA. But with the eye tracker, the moderators are able to provide the ratings with evidence as shown in the gaze opacity plot (Fig. 7) and with the time to first fixation metric.

Although users 2–4 managed to provide feedback, it was very succinct. This would not help the designers and developers to understand the degree of affective elements used in the design and how a user may react to it. The author finds this a common problem in usability studies conducted in Malaysia, whereby users are so focussed on completing the task and moving on to the next task. When the users are probed by the moderator on their feedback on the interface, they would provide a short or succinct answer, instead of providing a general comment on their UX and satisfaction level. There is still a strong belief that it is their fault when they are not able to complete a task, instead of criticising or praising a design of the interface.

User 6 managed to find the date. However, the CTA provided was in Malay language. This will require the moderator to translate to English. The eye tracker also enables the user and the moderator to perform an RTA. With the RTA, two interesting observations were made:

1. Even though user 6 mentioned during the RTA that the date was not found (“tak dapat jumpa tarikh”), during the playback, the moderator and user realized that the user was indeed staring at the date entry, but has mistakenly mentioned that the date was not found. The RTA feedback was corrected to “date was found” instead. In other words, eye tracking was able to reduce miscoding of information.

2. Unlike other users, user 6 was staring at the date and also staring away from the date. The user admitted that he was not in the right frame of mind to perform the test because he was rushing for another appointment after that. But the fact remained that the date entry was clearly visible by the user as shown by the fixations.

Despite advising the users that their feedback will remain anonymous, some users (user 5) were more comfortable on reporting some problems of the interface after the LBUT session. In these situations, the moderator was able to correlate the feedback provided with the eye-tracking visual cues to persuade the design and development team to improve the user interface.

7 Conclusion

From the PDI assessment performed, it was found that the existence of power distance in Malaysian is significant and this would make traditional usability testing problematic. Hence, the proposed LBUT is derived as shown in Fig. 2. To overcome the barrier, eye tracking is used in addition to TA during the LBUT. And the results from the case study do support the argument that despite the power distance and language barrier, the eye tracker is able to reveal key biometrics information. This is so because the LBUT method proposed with TA and eye tracking (Fig. 1) is able to reflect cognitive behaviour supported with visual cues to increase the persuasive power of the findings from the usability testing. In the future, it is recommended to use LBUT to enrich the UX of web and standalone interfaces, especially when power distance and language barrier is a constraint.

Acknowledgements The author would like to thank Mohamed Redzuan Abdullah from MIMOS Berhad who sponsored the MIMOS Usability/UX Lab that is equipped with the eye tracker. Many thanks go to Shi Tzuaan, Fahmi Mazlan and Zulkifle for facilitating the LBUT and Sujitra Balakrishnan for reviewing this chapter.

References

- Abdollah, N., Soo, S.T., Sivaji, A., (2013). Eyes on OSH – Usability Testing with Eye-Tracking and User Preference Rating. Springer International Publishing. Book Section Advances in Visual Informatics. Lecture Notes in Computer Science Volume 8237, 2013, (pp 323–334).
- Blandford, A., E, A., Hyde, J. K., Green, T. R., & Connell, I. (2008). Scoping Analytical Usability Evaluation Methods: A Case Study. *Human-Computer Interaction* (pp. 278–327).
- Claudia, E., & Wilson, S. (2007). Identifying web usability problems from eye-tracking data. *British Computer Society* (pp. 119–128.)
- Goh, K.N., Chen, Y.Y., Lai, F.W., Daud, S.C., Sivaji, A., Soo, S.T. (2013). A Comparison of Usability Testing Methods for an E-Commerce Website: A Case Study on a Malaysia Online Gift Shop. Information Technology: New Generations (ITNG), 2013 Tenth International Conference on , vol., no., pp.143,150, 15–17 April 2013

- Hofstede, G., & Hofstede, G. J. (2005). *Cultures and organizations: software of the mind* (Revised and expanded 2nd ed.). New York: McGraw-Hill. ISBN 978-0-07-143959-6. OCLC 57069196.
- Oshlyansky, L. (2007). Cultural Models in HCI: Hofstede, affordance and technology acceptance, Swansea University.
- Sherman, P (2009), *UPA 2009 Salary Survey*. Usability Professional's Association. http://usabilityprofessionals.org/usability_resources/surveys/2009salarysurvey_PUBLIC.pdf. Accessed 15 Feb 2012.s
- Sivaji, A., Mazlan, M. F., Shi-Tzuaan, S., 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).
- Sivaji, A., Shi-Tzuaan, S. (2012). Website user experience (UX) testing tool development using open source software (OSS). Network of Ergonomics Societies Conference (SEANES), 2012 Southeast Asian, vol., no., pp. 1, 6, 9–12 July 2012.
- Sivaji, A., Abdullah, M.R., Downe, A.G., Ahmad, W.F.W. (2013) Hybrid Usability Methodology: Integrating Heuristic Evaluation with Laboratory Testing across the Software Development Lifecycle. *Information Technology: New Generations (ITNG)*, 2013 Tenth International Conference on , vol., no., pp.375,383, 15–17 April 2013.
- Yammiyavar, P., Clemmensen, T., & Kumar, J. (2008). Influence of cultural background on non-verbal communication in a usability testing situation. *International Journal of Design*, 31–40.