Relationship Model in Cultural Usability Testing

Qingxin Shi and Torkil Clemmensen

Department of Informatics, Copenhagen Business School, Denmark
{qs.inf, tc.inf}@cbs.dk

Abstract. Culture plays an important role in the global market today. It not only affects products, but also impacts on usability evaluation methods. In this paper we first introduce culture theories and two kinds of relationships in thinking aloud usability testing and then review previous research. Based on the discussion, we extract the potential factors which may influence cross-culture usability testing and then propose a relationship model. Finally, we discuss how the two thinking aloud approaches may be used in cross-culture usability testing.

Keywords: Usability test, culture, thinking aloud theory, localization

1   Introduction

With the advent of globalization and IT revolution, we can no longer overlook the aspect of culture in the design of user interfaces and products. In order to capture global markets, the products and software must be tested in target cultures to make sure that they are acceptable and suitable for people's cultural characteristics. But some previous studies have found that culture not only influences the products or interface design, but also the design methods used in building interfaces [1]. Culture affects the usability evaluation methods (UEMs) of focus groups, questionnaires, structured interviews, and the understanding of metaphors and interface design [2]. Yeo's study [3] shows culture also impacts on usability testing.

Usability testing, also called the thinking aloud method, has been extensively applied in industry to evaluate a system's prototypes of different levels of fidelity [4]. The primary goal of a usability test is finding a list of usability problems from evaluators' observations and analysis of users' verbal and non-verbal behavior; thus, the relationship between the evaluator and user is very important for finding accurate usability problems. Tamler [5] suggested establishing a trusting and supportive relationship in order to make the users honestly disclose their thoughts and feelings.

During usability testing, representative users are required to complete pre-established tasks by using the system. This measurement is largely related to specific users and specific tasks. However, people differ across regional, linguistic and country boundaries; therefore, if the evaluator and user have different cultural backgrounds, they may be strongly influenced by their local cultural perspective, perception and cognition, so the interaction and communication between them may be different from those who are from the same culture. Since usability testing involves human-human interaction, the evaluator and user's cultural background must be considered, or else there may be a misunderstanding between them. Therefore, how to

build an effective relationship in the usability test has become a key issue in cross culture usability testing.

Although the thinking aloud usability test is generally thought to be an effective and successful technique [6], practitioners do not conform to the theoretical basis of the thinking aloud method in the industrial area which was described by Ericsson and Simon [7]. Therefore, Boren and Ramey [7] proposed speech communication theory as a theoretical basis for thinking aloud in usability testing, focusing on evaluator-test user communication. Later we will introduce two kinds of relationships in usability testing using the two thinking aloud theories.

Based on previous research, this paper analyzes the concept of cultural usability testing and brings forward a relationship model in cultural usability testing.

2   Two Kinds of Relationships in Usability Testing

Thinking aloud has been a widely used method to study people's thought processes and content. There are two different thinking aloud theories which can inform us about the evaluator and test user relationship in usability testing. One is the classical thinking aloud method which was put forward by Ericsson and Simon. This classical model emphasizes that during the usability test session, there should be little interaction and communication between the user and evaluator. The evaluator just tells the user to speak "as if alone in the room" [7, p.263]. The only interaction may be asking the user to keep thinking aloud. There is no tight relationship between the evaluator and test user.

The other thinking aloud theory in usability test was brought forward by Boren and Ramey [7] called speech communication theory. This theory focuses on the communication between the evaluator and test user. In the practice of usability testing, there is always a user and an evaluator. "Talk is not simply a form of action" performed by the user alone, "but a mode of interaction" between users and evaluators [7,p.267]. Relationship is much more important in this theory. The users cannot ignore the evaluators, even silent ones. They expect a response, agreement, sympathy, etc., from the evaluators. In the speech communication model, some key issues need to be clarified:

- The subject of the test is the interface, not the user.
- The test user is the expert, who is assumed to provide valuable information of the interface. The evaluator is the learner, whose main task is to get information from the user's speech and find usability problems.
- The evaluator should use undirected and undisturbed tokens to keep the users focused on the tasks and at the same time, verbalize their thoughts fluently.
- When encountering contingencies during the usability test, interaction between the evaluator and test user is required.
- In the practice area, it is okay to probe with questions to elicit more valuable information, which is not allowed in Ericsson and Simon's theory.

Ericsson and Simon's theory is primarily focused on cognitive processes, like problem solving. However, in the usability test, the main purpose is not only to get the user's thoughts, but more importantly, to get the user's expectations, feelings,

design ideas, etc., of the interface/software. So as long as the evaluator does not force his/her own opinion on the user, it is okay to build a proper interactive relationship between the evaluator and test user in the usability test. We did some field studies in Denmark, India and China. We observed how the usability practitioners do the usability test in the industrial area. Actually in all three countries the evaluators did not listen passively, but actively interacted with the users on necessary occasions to get more valuable information about the interfaces.

We can say that having a good communication and interaction is very necessary for usability testing. In order to get the fluent communication and suitable interaction, a warm, supportive and trusting relationship cannot be ignored.

3   Culture and Culture Theory

Culture has been defined in many different ways by different researchers. With regards to the usability test, we need a more specific definition of culture. Thus we introduce Honold and Nisbett's conception of culture. Honold [8] defines culture for the purposes of human computer interaction. One of her definitions is worthy of mentioning:"Culture does not determine the behavior of individuals but it does point to probable modes of perception, thought and action. Culture is therefore both a structure and a process" [8,p.329]. From Nisbett's research, we understand that people in different cultures perceive the world differently, which means that people's cognition and perception are different in different cultures. "Cultural practices and cognitive processes constitute one another. Cultural practices encourage and sustain certain kinds of cognitive processes, which then perpetuate the cultural practices" [9, p.3]. Usability testing is a cognitive activity [10] which, from the evaluator's eye, sees the user's behavior and comments. If they are from the same culture, it may be much easier for the evaluator to get the user's real meaning. If using a foreign evaluator, it will require extra effort to understand the user's real meaning. Hence, the effective communication and interaction is much more important in a cross cultural usability test.

Regarding culture theory, considerable usability research cites Hofstede [2, 3, 11-13] who has proposed five culture dimensions: power distance (PD), collectivism/individualism (IC), femininity/masculinity (MF), uncertainty avoidance (UA), long-term Confucian orientation. Marcus [11] has investigated how culture dimensions might affect user-interface designs. His research seeks to help user-interface designers cope with global product and service development. Although it is hard to design a universally usable interface, it is possible to provide guidelines for UEMs applied in different cultures.

This paper elaborates on Nisbett's culture theory [9, 14]. His theory focuses on the cognition and perception differences; for example, people from western countries and eastern countries will be different in causal attribution, categorization, and attention to the context vs. salient object [15]. This theory is more relevant to usability testing because thinking aloud usability evaluation methodology asks users to work on typical tasks and to verbalize their task performance and thought process [16]. The whole process involves users' cognition and perception characteristics. The results of

the usability test, i.e., usability problems, which are found by the evaluators, are also involved in the evaluators' cognition and perception of the whole test process. When cultural differences exist between the evaluator and test user, some usability problems might be masked, instead of being uncovered. If the cultural influence is ignored, the usability test UEM methodology may be inefficient to provide accurate information about the localized product.

From Nisbett's culture theory, there are two kinds of orientation [17]: task-focus orientation and socio-emotional relational orientation. Task-focus orientation means people's effort is directed towards task-related goals, and attention is focused on monitoring the extent to which these goals are being accomplished. Socio-emotional relational orientation means people's effort and attention are directed towards the interpersonal climate of the situation, and they strive to maintain social harmony. Users from different cultures may be affected by the foreign evaluator/interviewer in quite different degrees. Users, from socio-emotional relational orientation cultures, may be influenced more by the perception of a foreign evaluator. On the other hand, users from a typical task-focus culture may not be influenced by the foreign evaluator, since they focus only on their task and do not care much about the evaluator's status.

4   Previous Work on Relationship in Cultural Usability

Yeo [3] examined cultural factors that may affect the results of usability evaluation techniques. The aim of his study was to identify, examine and reduce the effect of cultural factors that influence usability testing. Initial results showed that an important possible cultural factor is power distance: a test user who was of higher rank than the experimenter gave more negative comments about the product than the one who was of lower rank than the experimenter. Nisbett's culture theory suggests that Malaysian culture is socio-emotional relational orientation culture. In the usability test the users hope to establish a harmonic relationship with the evaluators, so they do not want to give too many negative comments during the usability test, even if it is very hard for them to complete the task using the system. If the user thinks the evaluator has a higher rank, they may be more reluctant to provide negative comments [3] since they do not have a task-focus orientation; rather, they hope to build a good relationship with the higher ranking evaluator. So in Malaysian culture, in order to get honest results from usability testing, the experimenter should be of the same rank or of lower rank than the test subjects.

Yeo [13] explored the efficacy of the global-software development lifecycle (global-SDLC), which includes the design, implementation and usability evaluation phase. He found that adapting software from a source culture to a target culture, the design and implementation phase is efficacious, but the evaluation phase is not. He employed three Usability Assessment Techniques (UATs): Thinking-aloud Technique (objective measure), System Usability Scale (subjective measure) and Interviews. The results of the Usability evaluations were found to be inconsistent. He found that for the less experienced computer users, or for the users who were not familiar with the evaluators, the objective measure and subjective measure were not matched.   Even though these users performed poorly on the task, they still provided positive

comments of the software in the interview. According to Yeo, the cause of these inconsistencies was the users' reluctance to provide critical negative comments. Malaysia is a collectivistic country where users want to 'preserve the face' of the designer. If Malaysian users were familiar with the evaluator, they would not be concerned about making negative comments [13] since they would understand the evaluator's role in the usability test, and know that their negative comments would not destroy the good relationship with the evaluator.

Vatrapu and Pérez-Quiñones [2] investigated the effects of culture on structured interviews in the usability test. They carried out controlled experiments using two independent groups of Indian participants by two interviewers. One interviewer was from Indian culture and the other from Anglo-American culture. The results showed that the culture of the interviewer had an effect on the number of usability problems found, on the number of suggestions made, and on the number of positive and negative comments made. They found that the participants who were from the same culture as that of the interviewer (Indian culture) brought more usability problems and made more suggestions than participants who were interviewed by the interviewer who was not of the same culture (Anglo-American). From their study, we can see that when using a foreign evaluator, users may not be willing to talk as freely and accurately as when using a local evaluator. Language may not be the key issue, since in this research both interviewers and users could speak English fluently. We will analyze the potential factors that may influence cross-cultural usability testing.

5   Main Factors in Cultural Usability Testing

From the above discussion related to thinking aloud theories, culture theories, and previous research, we have extracted the basic factors that may influence cross-cultural usability testing. We will briefly discuss these factors now.

5.1   Evaluator and User's Cultural Background

Culture background needs to be considered since users from different cultures may not be influenced to the same degree when they are with a foreign evaluator. Sanchez-Burks's study [17] found that Northern European culture is a typical task-focus culture, which means that users in those countries may not be influenced when the evaluator is from another country since they pay more attention to the task, not   the evaluator. While East Asian culture and Indian culture are socio-emotional relational orientation cultures, users in these countries may be influenced more when they are with a foreign evaluator. For example, the study done by Vatrapu and Pérez-Quiñones [2] shows that Indian users who were with a foreign evaluator did not like to talk as freely as those who were with a local evaluator. But this may not be the case for Danish users. In our future study, we will use a foreign evaluator in India, China and Denmark to see whether the effect degree is the same in different kinds of cultures.

5.2   The Application/Software/Interface Being Tested

The requirement of an evaluator's cultural background is also related to the application or product which is tested in the target culture. There are two approaches to designing products for international markets: globalization and localization [18]. "Globalization seeks to make products general enough to work everywhere and localization seeks to create custom versions for each locale" [18,p.158]. If testing a localized application which adapts specific cultural elements for a specific target culture [19], the results of the usability test may be more related to the evaluator and user's cultural background.   Usability testing will not provide accurate information when a localized product is tested without considering cultural issues.

In Vatrapu and Pérez-Quiñones's study, the website which was tested was a culturally localized website, which means people in other cultures might not understand the background, purpose and other detailed issues of it. It is not easy for a foreign interviewer to find the culturally sensitive usability problems. On the other hand, the users also did not discuss too much with the foreign interviewer since they thought the foreign interviewer did not understand it. The users with the foreign interviewer just gave their opinions with little communication and interaction with the interviewer which, in turn, influenced the usability problems that the foreign interviewer would find. This implies that when testing a culturally neutral application, the influence of the difference in cultures between interviewer and user may not be as big as a culturally localized application. In our future study, if we want to see bigger cultural influences, maybe we should still use culturally localized application/software. Of course, we can also compare the difference of testing culturally localized applications and culturally neutral applications to see whether cultural issues have the same effect.

5.3   Evaluator Effect

The influence of culture on usability testing may also be derived from another factor called the Evaluator Effect: the total number of usability problems found will depend upon the knowledge and experience of the evaluator and the number of evaluators [6].

Hertzum and Jacobsen [10] examined three of the most widely used usability evaluation methods, cognitive walkthrough, heuristic evaluation, and thinking aloud, and found that all of them suffer from a substantial evaluator effect. No two evaluators evaluating the same interface and using the same usability evaluation method found the same set of problems. The evaluator effect exists "for both novice and experienced evaluators, for both cosmetic and severe problems, for both problem detection and severity assessment, and for evaluations of both simple and complex systems" [10,p.421].

The evaluator effect indicates that even in one culture, evaluators with different experience will find different usability problems. The effect may be much more significant when the evaluators are from two different cultures, since they do not even have the same cultural background. Even though they are both very qualified and professional, their cognitive process and knowledge cannot be the same, which may be a strong impact factor on cross cultural usability testing.

In a cross-cultural usability test, how can we minimize the evaluator effect which is derived from culture? It is very hard to change the foreign evaluator's cognitive process, but it may be much easier to increase his/her knowledge related to the culturally localized application. The foreign evaluator does not need to master all the target culture, because it is impossible. But he can get some important information just related to this application. Maybe he/she needs to know the background, using habits and some related culture features of the application in the target culture, which will be very helpful for them to understand and communicate with the users in the usability test.

6   Relationship Model in Culture Usability Testing

Based on the above discussion, a cultural usability testing model was brought forward (see Figure 1).
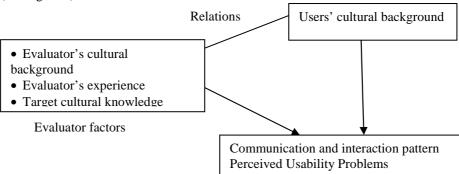

Figure 1: Relationship Model in Cross-Cultural Usability Testing

This model considers the evaluator's cultural background (task-focus versus socio-emotional relational orientation), experience and knowledge related to the localized application. There are four basic relationships between the evaluator and test user:
- 1. Foreign experienced evaluator with little target culture knowledge about the localized application----local users.

In this model, all the participants speak English in the target country. The foreign evaluator just gets the instructions of the tasks and the procedures of the test, but does not have any training for the localized application. The foreign evaluator does not have much knowledge about the usage of the culturally localized application in the target culture. But the application also exists in the evaluator's culture and the evaluator is familiar with such application in his/her own culture; thus, the only knowledge the foreign evaluator would need to master is the related cultural issues. If the foreign evaluator does not know the application at all, for example, chopsticks are seldom used by Danish people, then it would not be necessary to ask a Danish evaluator to do the usability test with Chinese users in China.

In our pilot study [20], we used Microsoft Clipart as the application, since regardless of which culture the evaluators came from, they would know what Clipart

is and how to use it. But the Clipart which was tested is a culturally localized one. We added a collection of culturally specific images and icons and a text document with preformatted invitation text called "cultural clipart" to My Collections in Microsoft Word's clipart organizer. The usability test is to see whether the "cultural clipart" is good enough for the user to make a traditional wedding invitation in the target culture.

Above all, in this model, the foreign evaluator should be usability professional. The application which is tested should be a common application but culturally localized. The evaluator does not have such knowledge about the localized application related to the cultural issue. Our aim is to see what is the communication and interaction pattern in this situation, and how many and what kind of usability problems the foreign evaluator would find.

- 2. Foreign experienced evaluator with more target culture knowledge about the localized application ----local users.

This model is similar to the first relationship. The only difference is that the foreign evaluator will be trained with important related cultural information about the localized application.

- 3. Local experienced evaluator----local users.

In this model, all the participants speak English. If the communication and interaction is better than the first two models, then we can safely infer that the language does not have a great influence that the difference may be from the cultural background. Of course, in all the four models, participants are chosen who are good at English.

- 4. Local novice evaluator----local users.

This model is similar to the third, except that the local evaluator is not experienced.

In the four relationship models we can compare Model 1 and Model 2 to see the influence of knowledge on the results of the usability test (relationship built in the test; communication and interaction pattern; perceived usability problems). Comparing Model 2 and Model 3, what kind of knowledge does the foreign evaluator have to get in order to do the usability test as efficaciously as the local evaluator? Suppose in Model 2 the foreign evaluator mastered all the related information, and then compared it to Model 3, would they get the same result? If not, what are the other main factors that influence the cross-culture usability test? Comparing Model 3 and Model 4, what is the influence of the evaluator effect? Comparing Model 1 and Model 4, which factor is more important, knowledge related to the culture or the skill of doing a usability test. By making these four groups of comparisons, we hope to gain a clearer understanding of cross-cultural usability testing.

6.1   Two Thinking Aloud Theories in Cross-Culture Usability Testing

As introduced above, there are two different thinking aloud approaches in a usability test. The usability practitioners usually do not follow the rigid thinking aloud approach which was proposed by Ericsson and Simon. Tamler [5] claims that thinking aloud data which is generated by the users themselves is often inadequate. The evaluator needs to probe questions which are important for the interface but not noticed by the user, and he/she also needs to share his/her understanding of the user's speech and behavior and get feedback from the user in order to get the user's real idea and experience to the interface[5]. Therefore, the communication and interaction is

very important for a usability test. A fluent and successful communication and interaction also relies on a supportive relationship. Actually, it is much harder for foreign evaluators to establish such supportive relationships with native users in the target country, even for experienced usability professionals.

When professional evaluators conduct a usability test with foreign users in the target culture, they may follow the traditional way (Ericsson and Simon's approach) to do the thinking aloud, which means that there might be less interaction and communication. Since they may not be familiar with the culturally localized application, they may not be certain what the critical issue is that needs to be probed. In order not to disturb and influence the users and get more accurate information, the better way is to interrupt them less and avoid false leading.

When native professional evaluators conduct a usability test with native users, they may be following the communication theory proposed by Boren and Ramey [7]. Previous studies [2, 20] show that, compared to foreign evaluators, local evaluators had more interrogative reminders, affirmative reminders, and help out behaviors.

Krahmer and Ummelen conducted two variants of usability tests under controlled circumstances. One condition was based on Ericsson and Simon's protocol, and the other on Boren and Ramey's proposal [21]. From their research, they found that although the evaluators used different approaches, the process of thinking aloud while carrying out tasks is not affected by the type of approach that was used. The task performance does differ. More tasks were completed in the Boren & Ramey condition, and subjects were less lost. But the number of different navigation problems that were detected and users' evaluations of the website quality were similar.

From this study, we can see that no matter which thinking aloud theory was followed, experienced usability professionals will find similar usability problems. In a specific culture with a local evaluator, the users' evaluation of the application/software will not be influenced by the thinking aloud approach that is used in the usability test. If foreign and local evaluators find quite different usability problems, this may not be because they are following different protocols with local and foreign users, but may be due to their varied cultural background.

7   Conclusion

This paper has discussed the effects of culture on thinking aloud usability testing and the application of two thinking aloud theories in cross-cultural usability testing. As the usability test needs test organizers, evaluators and users who may be from different cultures, it is becoming increasingly important to avoid the effect brought about by cultural differences. In this paper we have discussed the cultural influence on usability testing from only a theoretical viewpoint. In future studies we intend to investigate from an empirical viewpoint what kind of relations and communications between evaluators and test users are the most effective for finding usability problems of a culturally localized application during the usability test. We will design experiments to validate the models proposed above.

References

1. Vatrapu, R., *Culture and International Usability Testing: The effects of Culture in Structured Interviews. Master thesis.*, in *Virginia Polytechnic Institute and State University*. 2001.

2. Vatrapu, R. and M.A. Pérez-Quiñones, *Culture and Usability Evaluation: The Effects of Culture in Structured Interviews.* Journal of Usability Studies, 2006, 2006. **1**(4): p. 156-170.

3. Yeo, A.W. *Cultural Effects in Usability Assessment*. in *CHI 98, Doctoral Consortium*. 1998.

4. Law, E.L.-C. and E.T. Hvanneberg. *Analysis of Combinatorial User Effects in International Usability Tests*. in *CHI 2004*. 2004. Vienna, Austria.

5. Tamler, H., *High-tech versus high-touch: The limits of automation in diagnostic usability testing.* http://www.htamler.com/papers/techtouch/.

6. Clemmensen, T. and S. Goyal, *Cross cultural usability testing.* Working paper, Copenhagen Business School, Department of Informatics, HCI research group, 2005. **2005-006**: p. 20.

7. Boren, M.T. and J. Ramey, *Thinking aloud: Reconciling theory and practice.* IEEE Transactions on Professional Communication, 2000. **43**(3): p. 261-278.

8. Honold, P., *Cultural and context: an empirical study for the development of a framework for the elicitation of cultural influence in product usage.* International Journal of Human-Computer Interaction, 2000. **12**(3&4): p. 327-345.

9. Nisbett, R.E. and A. Norenzayan, *Cultural and Cognition*, in *Stevens' Handbook fo Experimental Psychology, third edition*, D.L. Medin, Editor. 2002.

10. Hertzum, M. and N.E. Jacobsen, *The evaluator effect: A chilling fact about usability evaluation methods.* International Journal of Human-Computer Interaction, 2001. **13**(4): p. 421-443.

11. Marcus, A., *User Interface Design and Culture*, in *Usabillity and Internationalization of Information Technology*, N. Aykin, Editor. 2005. p. 51-78.

12. Vöhringer-Kuhnt, T., *The influence of culture on Usability. Master thesis.*, in *Dept. of Educational Sciences and PsychologyFreie Universität Berlin, http://userpage.fu-berlin.de/~kuhnt/thesis/results.pdf, July 2004.* 2002: Berlin, Germany.

13. Yeo, A.W. *Global-software Development Lifecycle: An Exploratory Study*. in *CHI 2001*.

14. Nisbett, R.E. *Cognition and Perception East and West*. in *28th International Congress of Psychology in Beijing*. 2004.

15. Nisbett, R.E. and T. Masuda, *Cultural and point of view.* PNAS, 2003. **100**(19): p. 11163-11170.

16. Ramey, J., et al. *Does Think Aloud Work? How Do We Know?* in *CHI 2006, April 22-27*. 2006.

17. Sanchez-Burks, J., R.E. Nisbett, and O. Ybarra, *Cultural Styles, Relational Schemas and Prejudice Against Outgroups*. 2000, University of Michigan.

18. Horton, W., *Graphics: The not quite universal language*, in *Usabillity and Internationalization of Information Technology*, N. Aykin, Editor. 2005, Lawrence Erlbaum. p. 157-188.

19. Bourges-Waldegg, P. and S.A.R. Scrivener, *Meaning, the central issue in cross-cultural HCI design.* Interacting with Computers, 1998. **9**(3): p. 287-309.

20. Clemmensen, T. *Cultural models in psychological usabililty evaluation methods (UEM)*. in *Indo-Danish HCI Research Symposium*. 2006.

21. Krahmer, E. and N. Ummelen, *Thinking About Thinking Aloud-A comparison of two verbal protocols for usability testing.* IEEE Transactions on Professional Communication, 2004. **47**(2): p. 105-117.