

CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

CRITICAL EVALUATION OF CROSS-CULTURAL IMPACT ON USABILITY
STUDIES: ANALYSIS AND CASE STUDY DATA COLLECTION

A graduate project submitted in partial fulfillment of the requirements

For the degree of Master of Arts in Psychology

Human Factors and Applied Psychology

By

Zhimeng Li

April 2014

The graduate project of Zhimeng Li is approved:

Barry Berson M.A.

Date

Dr. Jill Quilici

Date

Dr. Tyler Blake, Chair

Date

Table of Content

Signature.....	ii
Table of Content.....	iii
Abstract.....	iv
Introduction.....	1
Usability Studies	1
Cultural Impact	2
Usability Studies and Culture	5
Theories in Culture Differences.....	5
Studies on cross-cultural usability	7
Think-Aloud Methods and Culture	12
Case Studies: Cross-Cultural Usability Studies for an Insulin Pen.....	17
The Case Studies.....	17
Participants.....	18
Procedures.....	18
Data Collection	20
Results.....	21
China Study.....	24
Germany Study	24
Discussion.....	25
Conclusions.....	30
Future Work.....	36
Recommendations.....	40
Conclusion	44
References.....	46

ABSTRACT

CRITICAL EVALUATION OF CROSS-CULTURAL IMPACT ON USABILITY
STUDIES: ANALYSIS AND CASE STUDY DATA COLLECTION

By

Zhimeng Li

Master of Arts in Psychology

Human Factors and Applied Psychology

Usability studies have been used to evaluate many kinds of products ranging from manual products such as ladders or hammers, to semi-automatic systems such as aviation operating systems. As corporations are gaining interest in the international market, usability studies have been conducted with participants from different

cultures. Those cross-cultural usability studies are often challenged by large cultural diversity. While many previous usability reports ignored the role that culture played in these studies, some researchers are becoming aware of the validity of cross-cultural usability studies. This paper focuses on two case studies that used *think-aloud* protocol. The *think-aloud* protocol requires participants to verbalize their actions and inner thoughts related to the testing product. One study was conducted in China, while the other took place in Germany. This paper compares the outcomes from these two studies; in particular, participants' behavior during the *think-aloud* session is observed and analyzed in the context of participants' cultural background. The results show that participants' attitudes toward thinking aloud vary from culture to culture, including what constitutes a comfortable environment for testing, and their overall understanding of the study they are participating in. At the conclusion of this paper, a few recommendations for new research areas are presented, possibly revealing the necessity of applying cultural variables in cross-cultural evaluations.

Introduction

Usability Studies

Usability studies are used to identify design flaws and potential problems that if remedied could eventually improve products. Instead of assuming what would potential users do, human factors analysts recruit potential customers and work as evaluators to assess how participants utilize products by implementing techniques such as *questionnaires*, *think-aloud* protocols, *probing questions* and *observation*. After running these studies, improvements are recommended based on their analysis of participants' performance. When usability studies are a part of the design process, any potential problems with the product being tested can be addressed.

Usability studies have been conducted to evaluate a wide range of products from manual products such as drywall sanding tools (Young-Corbett, Nussbaum, & Winchester, 2010) to semi-automatic systems such as pilot navigation systems (Beringer & Ball, 2002). Additionally, usability studies have been conducted in many countries. For instance, in 2006, a usability study of a hospital care cart was held in Canada with nurses and respiratory therapists to determine more desirable product features (Seto, Roach, Chagpar, & MacDonald, 2006). Moreover, online education programs were examined in the United States with different homepage layouts, aiming to figure out better strategies for this particular learning style as opposed to the traditional face-to-face learning style (Butler & Scherrer, 2009). Also in Sweden, a study of a multimedia program for school children was conducted, with a focus on identifying interface problems that would reduce the effectiveness of that program

(Berns, Klusell, & Pousette, 2000). Additionally, the usability of various e-commerce websites were tested in China, aiming to help *Business-to-Customer* electronic commerce designers improve their websites (Li & Li, 2011). Lately, A cross-cultural usability study was conducted in Namibia to test the *Information System for the Management of Rare Species in Namibia*, a hypermedia information system for biologists and wildlife managers (Paterson et al., 2011). In Japan, usability studies and related activities began to appear in recent years (Masaaki et al., 2004). Thus, although usability studies are not conducted on every product, it is now considered common practice for many companies to include usability research in their product development processes.

Cultural Impact

With commerce operating on an ever-increasing international market, corporations are no longer satisfied with promoting their products in a single environment. When large corporations first began launching their products in foreign countries, their success over local businesses was pre-determined by an advantage in resources and more advanced techniques. Recently, however, the competition has become more even as domestic companies are becoming aware of the situation and starting to obtain new techniques. Corporations are now facing challenges from domestic companies and other foreign corporations.

As the competition amongst corporations grows, the marketability of a company's products must increase in order to sustain business. The usability of

products is a major factor in attracting customers. However, local consumers have not always been involved in the product development. Hence, usability studies have now become a valuable source to understand local users' needs, preferences and behaviors regarding products.

Peterson and Parker (2005) discussed the challenges involved in cross-cultural usability studies, including the need for local resources, financial limitations, and time constraints, as well as listed some possible strategies for planning these kinds of studies. However, in their reports, the influence of cultural diversity was not taken into consideration. Moreover, few reports of cross-cultural usability studies have provided general strategies or guidelines for running studies with participants from any particular cultural background. Additionally, most evaluation reports did not describe using any culturally adaptable methods in their study.

On the other hand, cultural diversity has been studied for decades by specialists in other fields, such as psychologists and sociologists. Most usability analysts are familiar with various differences from culture to culture, yet there is a huge gap between being aware of the existence of cultural differences and realizing the need to eliminate or accommodate for its impact on cross-cultural usability studies.

Fortunately, some experimenters are starting to notice how various differences in culture can affect certain evaluation methods, such as the *think-aloud* protocol, and are now questioning the validity of these methods. This paper will review a few relevant cultural theories and recent articles about the validity of cross-cultural

usability studies. Additionally, two field studies of an insulin pen conducted in China and Germany will be analyzed and compared in order to investigate the impact cultural differences can have on usability studies. At the conclusion of this paper, a few recommendations for new research areas are presented that could help reveal the necessity of applying cultural variables in cross-cultural evaluations.

Usability Studies and Culture

Theories in Culture Differences

Although the cultures of participants or locations of cross-cultural usability studies were not commonly taken into consideration, many researchers were at least aware that these cultural differences exist. Some researchers have simply categorized cultures as either Western or Eastern. Others have viewed cultural diversity as a combination of different cultural factors, such as *power distance* and *uncertainty avoidance*. Among many theories in cultural diversity, a few relevant ones are discussed in the following paragraphs.

Richard Nisbett (2004) focused on the cultural differences between Westerners (primarily Europeans, Americans and citizens of the British Commonwealth) and Easterners (primarily the people of China, Korea and Japan). He noted several cognitive differences between Easterners and Westerners. In particular, Westerners focus directly on objects themselves, and are less sensitive to context, or the information associated with the objects. On the other hand, Easterners focus more on contextual information; they are more likely to reveal the relationship among events than Westerners, and they are more easily influenced by the environmental context of an object.

Furthermore, Westerners' cognitive style has been described as analytic, while Easterners' are described as being holistic (Kitayama, Duffy, Kawamura, & Larsen,

2003). For instance, Westerners are more likely to judge people by their group stereotypes; and Easterners tend to judge a person's behavior within a particular social situation. In other words, Westerners are more likely to fall for the *fundamental attribution error* than Easterners. They tend to overestimate the effect of a person's internal personality, while ignoring the external factors that could possibly affect a certain situation. In terms of usability studies, Western participants may be less influenced by evaluators due to their focus on the task at hand; Eastern participants may be more easily distracted than the Western participants since they afford more attention to the testing environment, such as the testing facility and the evaluators' reaction.

One of the most cited theories in cultural diversity literature is Geert Hofstede's *cultural dimensions*. According to his theory, culture should not be classified as Western or Eastern. Instead, he proposes six dimensions that can be used to distinguish one target group's culture from another (Hofstede). The following is the summarize of each dimensions.

Power distance is a dimension that refers to how much a member of lesser power within an organization accepts the inequality of his or her position in that organization. The higher the *power distance* is within one country, the more likely people with less power will accept society's inequality. Another dimension, *uncertainty avoidance* deals with the tolerance a member of an organization has for the ambiguity of the future. A society with strong uncertainty avoidance makes a

stronger effect to shape the future. *Individualism-collectivism* refers to how much a member of an organization integrates into the organization. *Masculinity-feminism* refers to the distribution of emotional roles between genders. *Long-short term orientation* refers to the degree that the society values either past and present, or future rewards. The final dimension, *indulgence-restraint*, stands for the degree to which the society allows or suppresses gratification of basic and natural human needs.

In usability studies, these dimensions can be used as criteria to create a comfortable environment for participants to engage in the study. In both Yeo's Malaysia study (2000) and Chavan's India study (2005), which I will return to later, participants were more critical of the testing products with evaluators that were of lower status than themselves, and more complimentary with those of higher status.

Studies on cross-cultural usability

Recent studies have suggested that culture can affect the outcome of a usability evaluation. As early as 1996, Herman indicated that cultural effects on usability evaluations not only exist, but also have a strong influence on the outcomes of the studies (Dunckley, Oyugi, & Smith, 2008).

Additionally, the role of evaluators is highly consequential in usability studies. Evaluators observe participants and guide them in disclosing their thoughts. Through observation, evaluators analyze participants' verbal and non-verbal behaviors. They summarize and report usability problems with the test product to the products'

manufacturer. In cross-cultural usability studies, evaluators and participants might not come from the same culture. For instance, cultural misconceptions can lead to evaluators' misunderstandings of their participants, which in turn can cause observation errors, and invalid data collection and results.

Communication Issue

Communication difficulties between evaluators and participants are one of the key catalysts to faulty usability evaluations. In most cross-cultural studies, local participants are allowed to use their native language. Moreover, differences in language fluency amongst participants can cause valuable data to be lost in translation.

The bottom line is that evaluators should be able to understand participants' language in order to analyze them accurately. There are many ways to address this language problem. When evaluators and participants do not speak the same language, a translator can be recruited in the study to translate between evaluators and participants. Evaluators can also work together with a local experienced evaluator and if necessary, bilingual evaluators can also be recruited.

Although the language barrier in cross-cultural usability studies is not difficult to remedy, previous reports showed that communication errors between evaluators and participants may be caused by more than language. The different cultural background between evaluators and participants is also a major factor in

communication error. The difference can affect evaluators' observation and analysis, and it cannot be easily solved by recruiting an interpreter. For example, a study in India had participants interviewed by either an Indian evaluator or an Anglo-American evaluator. The results showed that an evaluator who was a member of the same culture as the participants recorded more usability problems and participants' suggestions than an evaluator from a foreign country (Pérez-Quirónes & Vatrapu, 2006).

Different evaluators who actually share the same culture as participants can find different usability problems. This can be due to variations in evaluators' knowledge and experience as a usability expert. When evaluators and participants do not share the same culture, this effect can be even more significant. If evaluators are not familiar with participants' cultural background, cultural discrepancies in communication may arise. Evaluators may overlook or misinterpret some of the participants' facial expressions and nonverbal gestures, leading them to false impressions about the usability of the products.

According to Nisbett's theory, Easterners express less surprise than Westerners when things are evolving in inconsistent or unexpected ways (Clemmensen, Hertzum, Hornbak, Shi, & Yammiyavar, 2009). In regards to cross-cultural usability studies, evaluators may spend their time trying to record facial expressions that seldom happen or have little significance in the participants' culture, and misinterpret the lack of expressions as satisfaction with the test products.

Validity in Non-Western Countries

Some studies have raised concerns about whether the results from usability studies are reliable outside of Western countries, as this is where usability studies were developed. For instance, a study examined the efficacy of a Western designed program in the Malaysian market found that usability problems identified from a *think-aloud* session and those noted in the interview were inconsistent (Yeo, 2000). Some participants who avoid outward participation in the *think-aloud* session responded positively in the interview. An interesting possible explanation is that participants did not want to comment negatively about the product to avoid offending the evaluator.

Another study in Namibia of the *Information System for the Management of Rare Species in Namibia* also found contradicting results between participants' answers in their questionnaires and the evaluators' observations (Paterson et al., 2011). During the observation period, many participants were unable to solve tasks and struggled with some of the product's functions. In the questionnaire, however, participants' answers indicated the usability of the product needed no further improvement. Researchers suggested that one possible reason is that Namibians tend to answer questions according to what they think is socially acceptable. Additionally, they have a stronger desire to complete the questionnaire than to be truthful.

In the Malaysian study mentioned earlier, the experimenter found that most of the positive comments came from participants with low social status (Yeo, 2000). He

also found that when participants held a perceived higher rank than the evaluator, they are more likely to comment on the product negatively; in contrast, when they were of equal or lower rank than the evaluator, they tend to comment the product more positively. In accordance with Hofstede's *culture dimensions* theory, Malaysia is a country with a high *power distance*; in other words, people with lower status and power (i.e., employee) are more unlikely to contradict a person with a higher rank (i.e., employer) than vice versa. In their culture, going against a person of higher rank may jeopardize the harmonious relationship between the two people. However, a person of higher rank will be more likely to voice his or her feeling to a person of lower status.

Yeo suggested that differences in culture can offer an explanation of the inconsistent results from participants, and therefore these issues need to be taken into account when conducting usability studies in non-Western culture. Other usability studies carried out around would support this same conclusion. In India, which has cultural similarities to Malaysia, giving an obvious negative comment regarding a person or object is considered rude or strange. Participants were more willing to explore and critique the product when the evaluator has no affiliation with the product (Chavan, 2005).

Some researchers have argued that the usability studies themselves may not work as appropriately in certain cultures; therefore, new methodologies and techniques should be developed to work within local cultural contexts (Hsieh, 2011).

In Chavan's study, she used the *Bollywood method* to make Indian participants disclose their thoughts and feelings to the evaluators (Chavan, 2005). Bollywood films and film reviews are a major part of the Indian culture. With the *Bollywood method*, the movie review format is used as an evaluating method. The experimenters prompt the participants to mimic the critiquing behavior of film reviewers, and to "review" the product being tested in the study. Interestingly, the participants were more critical towards the product when they used the *Bollywood method*.

Think-Aloud Methods and Culture

The *think-aloud* protocol is one of the most valuable and widely used methods in usability studies. It requires little training of participants and does not require expensive equipment. It can also be used at any stage of the design process. During a *think-aloud* session, participants are instructed by evaluators to vocalize their thoughts out loud. By doing this, evaluators can observe participants' behavior and try to access their thinking process.

Ericsson and Simon (1984) defined two types of commonly used *think-aloud* protocols. In the *concurrent think-aloud* method, participants verbalize their thinking process while performing tasks. The *retrospective think-aloud* method begins after participants have completed these tasks. Also, according to their *think-aloud* theory, before the think-aloud session begins, evaluators are supposed to instruct participants with "neutral" words and allow them to "warm up" in a trial study prior to the data collection. Evaluators sometimes remind participants to "keep talking" if they stop

thinking aloud. During the entire *think-aloud* process, evaluators are expected to keep the interaction between them and participants to a minimum.

There are ongoing debates on how evaluators should utilize the *think-aloud* method in usability studies. Some recommend using the instructions provided by Ericsson and Simon (1984). Others, however, have suggested that evaluators should instruct participants from the perspective of user experience, such as participants' preference, affective or emotional response during interaction. Still, others have recommended the evaluator perform a think-aloud demonstration and practice ahead of a study, while others do not (McDonald, Edwards, & Zhao, 2012). Although the best way to implement the *think-aloud* method is undetermined, cultural factors should still be taken into account when performing cross-cultural usability evaluations.

Participants' Verbalization

The amount and quality of information evaluators can glean from participants in usability studies largely depends on participants' verbalization, the description of their behaviors and inner thoughts. However, a person's verbalization is constrained by his or her cultural customs. For instance, participants' comfort levels with talking while performing tasks varies from culture to culture. In the Western societies, talking and thinking is considered interdependent, and it is an essential ability to express one's thoughts. Conversely, in Eastern tradition, there is little connection between

talking and thinking. Silence and introspection are considered to be beneficial for higher levels of thinking (Kim, 2002).

In Kim's study (2002), Asian American participants and European American participants were asked to solve the same set of problems while either thinking aloud or repeating letters from A-Z. As predicted, European American participants performed better when they were thinking aloud than when they were distracted by the alphabet. In contrast, Asian American participants performed worse when they were asked to think aloud than when they were distracted by the alphabet. This supported the hypothesis that European American participants' problem solving process is more verbal than Asian Americans participants. Thus, the *think-aloud* protocol may not be sufficient when doing usability evaluations with Eastern participants.

Shi (2008) investigated the interaction and communication between evaluators and participants in China. During the *think-aloud* session, he noticed that participants generally kept quiet and did not speak out actively. Furthermore, they often forgot to speak out when they were working on tasks. He suggested that in order to get Chinese participants to be more forthcoming in their opinions, evaluators should interact with them more often by reminding them to talk out loud and by raising questions.

In addition to the already mentioned missteps involved with the extraction of information from cross-cultural participants, the *think-aloud* method itself may influence participants' performance of their assigned tasks. Kim (2002) had

participants perform the first half of the assigned problems in silence and then prompted them to think aloud while solving the other half of the problems. The results showed that not only did European American participants tend to perform better while they were thinking aloud, but that thinking aloud impaired Asian American participants' performance.

Relationship between Evaluators and Participants

The interpersonal relationship that is established between an evaluator and a participant can be a major factor in the outcome of a usability evaluation, because the final reports largely depend on evaluators' personal interpretation. As mentioned above, evaluators' judgments can be influenced by their cultural misunderstandings.

Many studies have suggested that various cognitive processes, such as *attribution of causes* (Nisbett, 2004), *decision making* (Briley, Morris, & Simonson, 2000), *conversational indirectness* (Lee, Sanchez-Burks, Choi, Nisbett, Zhao, & Koo, 2003), and *object perceiving* (Kitayama, Duffy, Kawamura, & Larsen, 2003) are not performed the same way in different cultures. In order to help participants honestly disclose their thoughts and feelings, the relationship between evaluators and participants must be appropriate and comfortable for participants; what is appropriate and comfortable, however, depends upon the cultural background of the participant. For example, for participants to be critical of a product in countries with a high *power distance*, evaluators should have what is considered in that culture as a lower rank (i.e., younger age) than participants.

The participant-evaluator relationship can have a larger influence in the outcome of a usability study that involves participants from Eastern cultures. Since Westerners are more task-oriented and Easterners are more context-oriented, Westerners may focus more on completing tasks while Easterners may put more effort and attention on the environment. Since the evaluators performing the study are part of this environment, participants from Eastern cultures might therefore be more vulnerable to their behavior. Yeo's (2000) study supported this theory as Malaysian participants' comments during the *think-aloud* session were also influenced by their relationship with the evaluator. For instance, participants tended to comment on the product more negatively when they were of a higher rank than the evaluator, and conversely, those who were of equal to lower rank tended to comment more positively.

Case Studies: Cross-Cultural Usability Studies for an Insulin Pen

The Case Studies

An insulin pen with disposable needles and cartridges was developed for the international market in 2011. This product is used for diabetic patients to self-inject insulin on a daily basis. Patients can adjust the dosage with a "dose selector" and inject with a "dose button". The insulin pen was designed with a digital display showing important user information, including the dosage selected, the present action such as an injection confirmation, warnings such as battery expiration, and the recent injection time.

Usability studies of the insulin pen were conducted in both China and Germany in 2011. The main purpose of these usability studies was to assess the functionality of the insulin pen, ensuring the pen was safe to use. The studies hoped to achieve, after launching the pen in both locations, that future users in both China and Germany would understand the interface of the insulin pen, and would be able to inject the right amount of the medication at the right time. In both usability studies, local participants and native speaking evaluators were recruited. Both participants and evaluators spoke their native language during studies.

This paper compares the *think-aloud* session in the study that took place in China and the study located in Germany. Video tapes from the German study were reviewed as part of my research for this paper. Since the participants were tested in

German, the video tapes were translated into English. For the Chinese study, I was working as a member of the evaluation team. In order to write this paper, I reviewed video tapes recorded from the Chinese study as well. The observations made in this paper are primarily based on participants' performance during the *think-aloud* session. To be more specific, it is primarily based on participants' behavior during the session where they talked out aloud while working on the given tasks. The purpose of this paper is to investigate and discuss whether cultural differences affect participants' performance during *think-aloud* sessions, how the differences influenced participants' task performance and how to avoid uncontrolled variance in future studies.

Participants

Thirty-five local participants (23 females and 12 males) were recruited in the Chinese study, between the ages of 11 and 67 (Mean = 42.22, SD = 18.09). Twenty-three local participants (15 females and 8 males) were recruited in the German study, between the ages of 12 and 81 (Mean = 41.14, SD = 20.57). All participants were familiar with at least one type of injecting method, but none were familiar with this particular insulin pen.

Procedures

The processes of the Chinese study and the German study were comparable in all major aspects. First, participants in both studies were asked to complete a survey

where they provided their medical history. Then they were introduced to the beginning section of study, which prompted them to think aloud while they performed some tasks. In the warm-up task, participants were asked to demonstrate a simulated injection with their own insulin pens. During the actual study, they were asked to perform several tasks with the provided new insulin pen. For example, one task involved a simulated injection which included assembling the insulin pen, injecting the required amount into a cushion, and disassembling the pen. It also included checking the display for details of previous injections and reading symbols on the display. The entire study was recorded and each study took roughly 60 minutes. There were two evaluators assigned to each study. One evaluator interacted with participants and helped them to complete the study. The other one recorded verbal and non-verbal behavior of each participant.



Figure 1: Screen Shot from Germany Study



Figure 2: Screen Shot from China Study

Data Collection

Each task was divided into several subtasks in order to facilitate data analysis. For instance, the first task, which was to simulate an injection followed by confirming that injection, was divided into several subtasks (e.g., assembling the pen, preparing for the injection, injecting into a cushion and so on). In each subtask, participants were assessed as whether they were thinking aloud, whether they described their thoughts when having trouble completing the subtask, and whether they were interacting with evaluators.

In each subtask, participants were considered as "*think-aloud*" when they described their behaviors or thoughts for more than one sentence during the subtask. Participants were identified as struggling with the task when they either indicated that

they are having a specific problem, or when they are not making any progress with the task. In that situation, they were graded as "*think-aloud*" only when they pointed out the problem. Participants were considered interacting with the evaluator when they talked to the evaluator, turned to the evaluator for help, or showed their performance to the evaluator.

After scoring all subtasks, participants fell into the category of "able to *think-aloud*" when they completed over 66.67% of the subtasks while talking out aloud. They were considered as "not able to *think-aloud*" when they completed less than 33.33% of the subtasks talking out loud. They were considered as "somewhat *think-aloud*" when they were in between the other two categories. If the number of subtasks participants described out loud exceeds the number of subtasks they did not describe out loud when they were having trouble completing tasks, they were considered as "*think-aloud*" when they were struggling with the task. Participants were considered as interacting with evaluator if they talked to the evaluator during more than one subtask.

Results

For the most part, German participants were more engaged in the *think-aloud* session than were the Chinese participants. German participants spent more time talking aloud, and they provided more verbal information to the evaluator by thinking aloud in a more complete and comprehensive way than Chinese participants. In contrast, more Chinese participants performed their tasks relatively quietly. In fact,

many of them performed their tasks in silence. This resulted in less verbal information being provided to the evaluator during the tasks. In both studies, participants tended to talk out loud less often if not reminded regularly by the evaluator. After being reminded, Chinese participants tended to speak out for only the next few steps and then quickly became quiet again. There were some Chinese participants that did not talk out loud even after being asked to do so. One participant pointed out that she felt strange vocalizing her thoughts and outright refused. German participants, on the other hand, were able to think aloud for many tasks, particularly after being reminded.

Regardless of whether being reminded to talk out loud, 20 out of 23 German participants (87.0 %) were able to talk aloud throughout the entire study. This means they were able to describe their own behavior in each task and continued to do so as they progressed through the study. There was, however, one German participant (4.3%) who gave almost no verbal information. Eleven out of 35 Chinese participants (31.4%) talked out loud throughout the study. There were 14 Chinese participants (40.0%) who seldom spoke aloud, and were silent for most of the tasks. Ten participants (28.6%) gave virtually no verbal information at all. Table 2 shows the number of participants and the percentage of subtasks participants spoke out loud in the study.

When participants were struggling with tasks, 12 out of 23 German participants (52.2%) described their thoughts. In contrast, 6 out of 35 Chinese participants

(17.1%) verbalized their thoughts actively. They described their thoughts such as “What is going on here?” “I’m doing this because I want to check ...” and “This is not what I meant to do.” These participants were able to express what was troubling them at the time, and explained what they were doing to solve the problem. Many of the Chinese participants started to focus only on solving the problem, and when the tasks became more difficult they stopped thinking aloud completely.

Instead of talking out loud to themselves, 18 Chinese participants talked to the evaluator in the *think-aloud* session. This was not found in the German study. Only three German participants were found talking to the evaluator in the *think-aloud* session. When these participants were talking out loud, they were often building a relationship with the evaluator. They asked the evaluator questions, and demonstrated the tasks directly to the evaluator. They showed the insulin pen to the evaluator, and explained what they were doing with the pen. Additionally, there were two Chinese participants who kept turning to the evaluator for help even after they were instructed to complete the tasks by themselves.

Main Findings		China Study (N=35)		Germany Study (N=23)	
		Number of participants	Percentage	Number of participants	Percentage
Engage in Think-Aloud	Able to Think-aloud	11	31.4%	20	87.0%
	Somewhat think-aloud	14	40.0%	2	8.7%
	Not Think-aloud	10	28.6%	1	4.3%
Struggle with tasks	Think-aloud	6	17.1%	12	52.2%
	Not Think-aloud	10	28.6%	3	13.0%
Interact with evaluator		18	51.4	3	13.0%

Table 1, Study Results: Number of participants

	China Study	Germany Study
N	35	23
Mean	54.27%	79.31%
Median	54.50%	81.80%
Mode	61.5%	100%
Minimum	0.00%	1.54%
Maximum	100%	100%

Table 2, Study Results: Percentage of tasks participants spoke aloud.

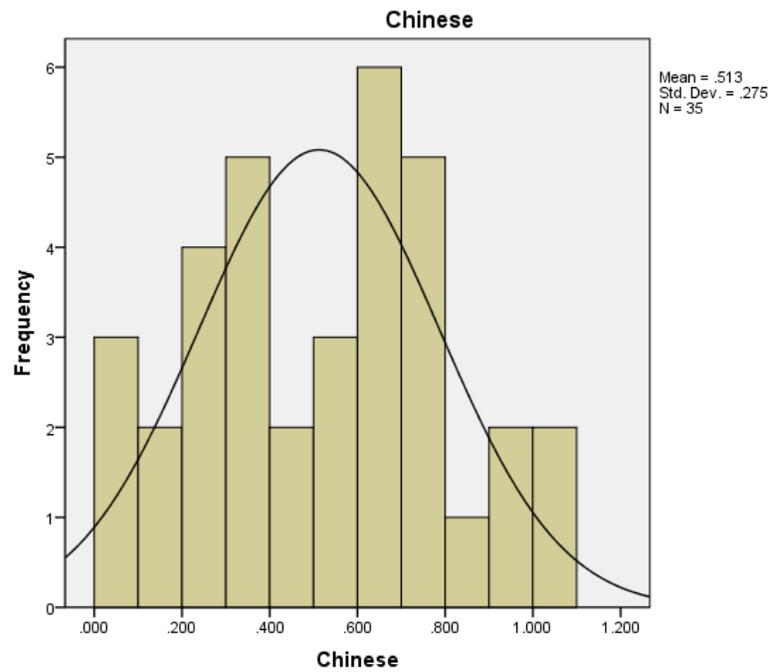


Figure 3, Frequency of Chinese participants talking out loud

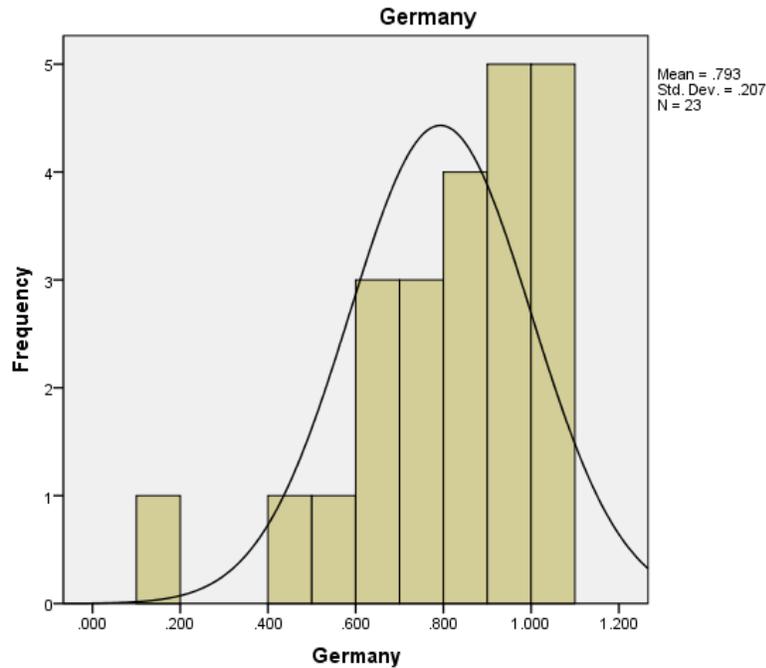


Figure 4, Frequency of Germany participants talking out loud

Discussion

This paper investigated how different cultural factors may influence the results derived from the *think-aloud* protocol. Overall, German participants were more proactive in the *think-aloud* session than were the Chinese participants. Compared to Chinese participants, German participants talked out loud more actively throughout the entire study. Most of the German participants were able to describe both their behavior and their thoughts while less than half of the Chinese participants described both. Being reminded to think out loud was more useful to German participants than

to Chinese participants, and German participants were more independent when thinking out loud than were the Chinese participants.

Verbalization

From the analysis of the video tapes, less than half of the Chinese participants described their behavior and thoughts throughout the entire study. Repetitive the majority either provided no verbal information at all by thinking aloud, or tended to stop thinking aloud when they were struggling with tasks. Most German participants seemed to have little trouble thinking aloud when they were trying to solve problems. The explanation of the inconsistency between Chinese participants and German participants' performance in the *think-aloud* session could possibly be found in the cultural theories expounded by Nisbett; theories that differentiated between Eastern and Western communication customs.

One idea is that Eastern and Western cultures value self-expression differently. In Western culture, expressing thoughts, preferences and feelings is a sign of individualism; where in Eastern culture, expressing one's thoughts is neither commonly practiced nor viewed positively (Kim & Sherman, 2007). In the Kim and Sherman's study, which used open-ended questions as prompts, European American participants placed a higher value in self-expression than Asian American participants.

Another theory expounded upon is that Easterners may not habitually talk and think simultaneously as Westerners do. As described in Kim's (2002) study

mentioned before, Asian American participants' problem solving process is less verbal than European American participants. This suggested a closer connection between talking and thinking for European Americans than for Asian Americans.

In the think-aloud session, evaluators reminded participants to “keep talking” when participants stopped thinking aloud. In the current study, reminding participants to think aloud was not as effective for Chinese participants as it was for German participants. Reminding Chinese participants to talk out aloud did not make them open up more to the evaluator. Similar results were found in Clemmensen's (2011) cross-cultural field study. Clemmensen used a sample of both Danish and Chinese participants. In the Danish sample, evaluators did not need to give many reminders. With Chinese participants, he noted, retrospective reminders appeared to be a necessary way to acquire information.

From the observation of the current study, asking Chinese participants probing questions turned out to be more efficient than reminding them to think aloud. When evaluators asked them questions right after a subtask, such as “Why were you doing that?” or “What was bothering you?” most Chinese participants had no trouble answering questions. In fact, they appeared to be quite talkative. They not only explained their behaviors and disclosed their thoughts from the previous tasks, but also actively informed evaluators of their daily experience with insulin products, providing recommendations for the test product.

This revealing bit of information leads to another possible solution when conducting a usability study with Chinese participants. If reminding participants to speak out loud does not encourage them to reveal their thoughts, using *probing questions* or *retrospective think-aloud* protocol can serve as alternative methods to elicit the opinions of Chinese participants. However, this technique requires that the evaluators have more advanced observation and communication skills. Evaluators need extensive knowledge of participants' cultural behavior, habits and facial expressions. Such cultural knowledge could potentially reveal wealth of information hidden beneath the veil of silence manifested by some participants.

Moderator-Participant Relationships

The relationship between evaluators and participants in the Chinese study was different from the relationship in the German study. The majority of German participants did not actively interact with the evaluator during the *think-aloud* session. In contrast, one-third of the Chinese participants tried to interact with the evaluator when they were thinking aloud. This supports the assumption that there is a culture difference between Easterners and Westerners in cognitive style that Easterners pay more attention to the overall testing environment and Westerners focus more on the research tasks. In the case study, the presence of the evaluators had a greater effect on Chinese participants than German participants. Similar results were discovered in Shi's study in China (2008) where the relationship between evaluators and participants was described as a "student-tutor" relationship. Instead of verbalizing

their thoughts, participants considered themselves as teachers. They "taught" the evaluators the correct application of the product.

The role of evaluators in the *think-aloud* protocol is to observe participants and guide them in disclosing their thoughts. Meanwhile, evaluators are required to minimize their interaction with participants so that they can study participants' performances without affecting the testing results. In this situation, it was difficult to determine an appropriate relationship between evaluators and participants that is close enough to gain participants' trust, yet distant enough to elicit neutral testing results.

Few studies have discussed the boundary between maintaining limited interaction with participants and building a supportive relationship when participants are constantly seeking interactions. In cross-cultural usability studies, the idea of a proper relationship between participants and evaluators may be different from the traditional concept. There is no guideline so far that evaluators can follow when conducting usability studies with participants in non-Western cultures.

Interpretation of Tasks

In certain instances, some Chinese participants misunderstood the relationship between tasks. More specifically, four Chinese participants assumed that all tasks were sequential, basing every task on the previous task. For example, several different insulin pens were used at different points in the case study. Four out of 35 Chinese participants showed obvious confusion when the pen was changed. They either did

not notice the pen was switched or assumed that all test pens were the same. They tried to find evidence from the previous tasks to solve the current task. Because participants did not prefer to talk out loud, it took evaluators some time to identify the problems and correct them. This could be a common problem with usability studies in general, and one that could go without being noticed by evaluators. It is possible that evaluators may have attributed participants' frustrations to the usability of the test product and not the cultural miscommunication in the testing itself.

Conclusions

As usability studies are proliferating across various markets around the world, the cultural backgrounds of participants should be fully considered in study design and researcher training. Previous studies have also supported questioning the validity of certain testing methods in non-Western countries (Hsieh, 2011), and suggested new methods need to be developed to better fit the culture and environmental context (Chavan, 2005).

The *think-aloud* protocol is one of the most valuable and widely used methods in usability studies. This paper analyzed the *think-aloud* session with both Chinese and German participants in hoping to raise awareness of cultural misconceptions in usability researchers.

The outcomes of the case studies showed that German participants were generally more adaptive to the thinking-aloud protocol than Chinese participants.

This paper makes the suggestion that in order to encourage participants to speak out, using effective communication methods and building a more supportive relationship might be beneficial when working with non-Western participants. Future studies need to be conducted to develop practical and efficient approaches for evaluators in cross-cultural usability studies. The following are summarized from the case studies.

Attitude towards Think-Aloud

First, culture could impact participants' attitudes toward thinking aloud. In the Chinese study, most participants did not actively think aloud. In order to acquire valuable information, it is especially important for evaluators to encourage participants to speak out. One way to achieve this is to make it clear that the usability study does not test participants themselves, or judge them personally in their ability to do anything. This requires evaluators to explain the purpose of the study – to test the product, not the participants. This explanation should occur before the testing starts and be consistently repeated throughout. Although this is done in some studies, it appears to be more critical when dealing with Chinese and similar Asian cultures.

Although similar explanation was included in the briefing session of the China study, it could be more effective if evaluators emphasize this subject rather than just mentioning it among many other announcements. It could also be helpful if evaluators do not judge participants during the study. This is important to avoid even if the judgments are positive. Certain utterances that were used in the Chinese study,

such as “You are doing great” or “You are doing better than others,” should be avoided so as not to mislead participants of the study's purpose.

When participants are determined to not open up to evaluators, reminding them to talk out loud might not work as expected. Instead, use alternative methods, such as *retrospective think-aloud* protocol, and *probing questions*. In the Chinese study, participants either forgot to speak out or were not willing to do so, and they did not change their behavior after being reminded to think aloud. *Probing questions* became a necessity to address important issues in the case study.

Honestly Disclose Comments

Secondly, culture could impact participants’ ideas of a comfortable testing environment. As in Hofstede’s cultural dimensions theory, one nation is distinguished from another in dimensions such as the *power distance*, *uncertainty avoidance*, *individualism-collectivism*, as well as other aspects. These cultural dimensions can affect participants’ behavior in usability studies. In Yeo’s (2000) study, Malaysian participants’ comments varied based on their societal rank with their evaluators. In Chavan’s (2005) study, Indian participants were more skeptical when evaluators pretended to be participants. Consistent in the findings of both studies, when the study did not provide participants an equal or higher societal position with an evaluator, participants did not feel comfortable disclosing their thoughts or opinions.

In the case study, Chinese participants were trying to interact with evaluators. The Chinese participants preferred a closer relationship with the evaluators as opposed to an isolated relationship. It is unclear whether such a close relationship between participants and evaluators in the Chinese study had a positive or negative impact on the testing results. Further studies are necessary to investigate a proper relationship between participants and evaluators.

Understanding the Purpose of the Study

Previous research has suggested that culture affects cognitive styles. Westerners are usually described as analytical and task-oriented; Easterners are described as holistic and relationship-oriented. In usability studies, participants from different cultures might have different understandings of the study itself.

In the current study, a few Chinese participants misunderstood the relationship between tasks. They automatically viewed tasks in a sequential order and were confused when the following tasks did not match or were not directly related to previous ones. To prevent such confusions, evaluators in future studies should aim to be more specific with the task descriptions and instructions. For instance, instead of saying “Now, move on to the *next/second* task,” say “Now, move onto *another task*” or “*a different task*”. Evaluators should remind participants that the testing device is changed. For example, an evaluator could say “Now, you are going to use *this* pen...”

Evaluator's Effect

Not only can participants' culture impact the way they performed in the study, evaluators' behaviors and judgments can also be affected by their own cultures. In cross-cultural usability studies, evaluators may be recruited from a foreign country or from the country under study. They may or may not come from the same culture as participants, and they may lean towards either in Western or non-Western practices. Their knowledge and habits are influenced by their background.

Evaluators with different styles of conducting studies may lead to different study results. Shi (2008) described Chinese evaluators as "always thoughtful" in his field study. He pointed out that Chinese evaluators adjusted their behavior according to participants' personality and background. In the case studies, for instance, the German evaluator placed the testing sheet in front of participants, asked participants to read the task out aloud and they would point to areas on the testing sheet if participants miss any subtasks. The Chinese evaluators sometimes placed the testing sheet between them and participants, and they read the task requirements to participants if a subtask was not performed. This may influence the results of the study, as this variation it would give participants a different sense of the presence of evaluators which in turn could affect participants ability to perform tasks on their own.

Limitations

Since this paper is based on the comparison of two existing usability studies, some factors could not be controlled during the analysis. The results of this paper are further limited by the conditions of the usability studies.

First, the purpose of the case studies was to address the usability issues of the insulin pen. All tasks were designed based on the functions of the product, rather than participants' reaction to the *think-aloud* protocol. Therefore, the amount of think aloud required to finish the task was not controlled. It appears in the studies, when the task was too easy, some participants were able to complete the steps in a very short period of time. This gave participants less opportunity to think aloud. Some participants stayed quiet during the performance, and only informed the evaluator that they have completed the task. On the other hand, when the task was too difficult, some participants focused only on the tasks, and their ability to think out loud was decreased.

During the observation, it was not possible to eliminate participants' behavior caused by the difficulty levels of some tasks. For example, if one participant did not speak out loud in a task because it was too easy for him/her to finish, but not too easy for other participants, this participant would be considered as "not able to *think-aloud*" in this particular task for that reason. In order for a more subjective observation in the future, tasks could be designed in various difficulty levels. Tasks that are moderately difficult to one participant can be selected and compared with other participants. In this way, participants could have an equal opportunity to think aloud. Moreover, with the difficulty level added to the observation, more variables can be analyzed, such as whether participants' ability to think aloud from this culture is affected in tasks with a certain difficulty level.

Second, not only were different evaluators recruited for the China study and the Germany study, the evaluators were not trained to study participants' performance during the *think-aloud* protocol. The evaluators' influence on the usability studies, for one thing, was not controlled. Even with the same evaluator, different strategies were applied with different participants. To minimize evaluators' influence on the observation result, standard strategies can be developed before conducting the usability studies (e.g., the evaluator should always remind participants to talk out loud, the evaluator should only remind participants to talk aloud under certain situations, or the evaluator should never remind participants to speak out.)

Furthermore, some information was not recorded in the video tapes that were used in the observation. In the video tapes from the German study, the original soundtrack was translated to English. Information might have been skewed according to the translator's personal interpretations. Meanwhile, because of the restrictions of the recording equipment, the recorder was set up either behind or above participants. This made it impossible to observe participant's facial expression in some of the video tapes. More information could be analyzed in a follow up study, if the observer understand the languages used in both locations, and if the observer is able to study the facial expression of participants.

Future Work

Participants' behavior in usability studies can be influenced by a lot of factors, such as the nature of the test product, the evaluation methods, the physical

environments, and the interaction between all of these factors. Additional studies are necessary to explore more effective evaluation methods in cross-cultural usability studies. Based on the observation of the case studies, there are a few interesting phenomena, or anomalies, that warrant in-depth experiments in the future.

In the Chinese study, some participants were extremely talkative, while some were very quiet. Although the talkative participants did not always perform according to the *think-aloud* instructions (i.e., pretending that they were operating the product alone, describing every action and thought simultaneously), they had no trouble expressing themselves. They described detailed information of their own products, shared their daily experiences, and made positive and negative opinions on the test product. On the other hand, some participants were very quiet. They did not think aloud, and they would only whisper short answers to evaluators' questions. This variation in behavior was not evident in the German study.

One possible explanation is that Chinese participants are more influenced by their self-concept. The talkative participants might have had a more extroverted personality than the quiet participants. Although there are always different kinds of participants on the extroversion-introversion scale in usability studies, when they are task-oriented, they can suppress their social interaction habits and remain focused on the task at hand. Since participants from Eastern cultures are considered to be more context-oriented than Western cultures, it was more challenging for Chinese participants to shut down their self-awareness when completing tasks. In other words,

they seemed very conscious that they were in a study, observed by evaluators and there was a social interaction taking place. Therefore, extroverted participants appeared to be talkative and they started many conversations with evaluators, while introverted participants remained shy and quiet.

Adult participants were more talkative than teenage participants in the Chinese study. There are some possible explanations that future studies should investigate. In the case study, parents or guardians of teenage participants were allowed to observe the study. Teenage participants' behavior was seemingly influenced by the presence of their parents. When teenage participants and their parents were sitting in the same room, they sometimes turned to their parents for answers or parents spoke for them. Although parents have the right to observe the study, instructions could be made prior to the study to prevent parents from helping out the tasks or answering questions for their children. Even when the parents are the caregivers, meaning they help participants with their daily insulin injection, the study would still benefit from removing the influence of the parents from the study.

Secondly, adult participants are usually more skilled communicators than teenage participants. In China, teenagers usually spend most of their time in school, and they do not start working until after they graduate high-school or college. They have very limited social experience compared to adults. In the study, adult participants might have felt more comfortable communicating with another adult stranger than the teenage participants did.

Thirdly, the evaluators in the China study were around 30 years old. According to Shi's description, the relationship between Chinese participants and evaluators are like a "student-tutor" relationship where participants teach evaluators how to operate the product. Based on the *culture dimension* theory, China is a country with high *power distance*. It is more acceptable for participants with high status, older in this case, to lecture evaluators, who were younger. On the other hand, it was socially awkward for teenage participants, who were younger than the evaluators, to either teach evaluators how to operate the product or comment on the product.

The *culture dimension* theory could also explain why the professional healthcare participants showed stronger confidence than other participants in the Chinese study. In Chinese hospitals, the *power distance* between healthcare professionals and patients are large. Professionals, such as doctors and nurses, are aware of their skills and patients' dependency on them, which leads to the development of an unequal relationship.

In the Chinese study, professionals were confident with their task performance. They spoke fast and got to the point with few words. They also showed impatience with evaluators' questions, and ignored evaluators' requests to speak slower or to talk out loud. Besides professional participants' familiarity with insulin products, they expressed a superior sense of themselves over the evaluators which could impact the evaluators' role in the study. *Power distance* does not only occur in age and working position, but also between careers. Some careers are considered to have higher status

than other careers. Future studies in countries with high *power distance* might also need to consider the careers of potential participants prior to testing.

One of the things that was not addressed in the paper is whether the different behavior towards the *think-aloud* protocol affected the data collection of usability studies. Again, the purpose of many usability studies is to estimate product functions. *Think-aloud* protocol is one of the most widely used methods in usability studies, because it is an effective tool to collect important information. Cultural customs were not involved when the standard strategies for the *think-aloud* protocol were developed.

For example, during the *think-aloud* session, evaluators are suppose to keep their influence to a minimum. According to the results from the Chinese study, it might be more difficult to collect data by using *think-aloud* protocol only. Many participants provided more verbal information when talking to the evaluator. Since there are not enough supporting theories or existing studies for reference, it is not clear the results are affected by the interaction, and what alternative strategies could be used in such situations. More studies are needed to understand whether culture is affecting the data collection process, and, if so, what is a more proper way to minimize or avoid the impact.

Recommendations

This paper has discussed cultural issues that may develop in cross-cultural usability studies. Although previous studies have shown that some researchers are

becoming aware of the impact that culture can have on usability studies, few reports provide guidelines or strategies to rectify or probe this issue. After analyzing two case studies in this paper, here are some recommendations that future usability teams could consider before running.

Pre-Study Research

Conducting cross-cultural usability studies requires a great deal of planning, especially for usability teams that have never run studies in that particular country. There are some ways that usability teams can improve their planning process. Research on the cultural backgrounds of the country being tested would help usability teams to better understand participants' behaviors and habits.

Studying a culture can be a vast and limitless endeavor. Instead of investigating unspecified literature of the testing country, usability teams could spend time on a few relevant works. If available, research teams should focus on previous usability studies published about the country being tested. Previous usability studies will provide usability teams with more concrete information such as the study planning, study process and the challenges. Usability teams could learn from their mistakes and prepare for possible obstacles.

Consulting with local usability professionals is another way to obtain quality and focused information. Local professionals are usually more experienced in the cultural

region than foreign usability teams. Additionally, consulting with the local professionals can also be helpful to the planning process.

Evaluators

When foreign evaluators are involved in the observation, whether having recruited a translator or not, the team should make sure that evaluators are familiar with the culture. This includes understanding participants' reaction when they are struggling with tasks, noticing when participants are hesitate to comment, and recognizing if participants are indeed satisfied with the test product. Otherwise, the evaluators might fail to catch some expression and will be unable to guide participants to complete the study correctly.

Usually in the planning process, usability teams need to coordinate with the manufacturer on subjects such as study budget, testing purpose (i.e., improve product functions, ensure product safety), and evaluation focuses. If local evaluators are recruited, usability teams should make sure that local evaluators and team members are on the same page. Since there is a good chance that local evaluators and team members do not acquire their knowledge and practice their expertise under the same guidelines, it is important to come to an agreement on aspects of the study such as communication styles, the types of questions to be used and evaluation emphases, prior to the study.

Evaluation Methods

As mentioned above, the validity of certain evaluation methods were questioned by some researchers. In previous studies, certain methods were not as effective as expected. Questionnaires did not receive honest answers in Namibia, and Malaysian participants responded positively about the test product in the interview portion of the usability test but performed poorly in the observation portion. In the case study in China, participants did not always respond well to the *concurrent think-aloud* protocol even though they were instructed to think aloud and were reminded to do that.

When usability teams are not sure if the evaluation method will work as they expect, they should prepare other methods as back-ups in case the primary method does not attain the information they are seeking. For instance, when *concurrent think-aloud* did not work in the Chinese study, *probing questions* or *retrospective think-aloud* were used to elicit important information.

Participants

Participants are influenced by their cultural norms. Their habits or reactions may differ from what the usability team experienced before. Additionally, because of participants' differing cultural backgrounds, some participants from certain countries may have trouble vocalizing their thoughts. Moreover, some participants might be nervous talking to strangers, and some are not used to being critical in front of others.

There are many strategies usability teams could consider to create an appropriate testing environment for participants based on their cultural customs. For example, when conducting studies in countries with high *power distance*, evaluators' appearance will at some level affect the amount of information they can receive from participants. Teenage participants felt more comfortable disclosing their opinions to evaluators who were closer in age or were more personable. On the other hand, when the participants are professionals, they might consider evaluators who look more professional as an equal in social interaction, and may be willing to share their opinions more frequently.

Pilot Study

Usually pilot studies were conducted ahead of data collection to ensure the evaluation serves the study's purposes. In cross-cultural usability studies, usability teams may have limited knowledge or experience of the testing country. Conducting a pilot study is a practical way to gather information on the effectiveness of evaluation methods. If usability teams are not experienced in the testing country, the pilot study can provide instant feedback on the efficiency of the study. Moreover, usability teams need to be prepared to adjust their strategies at any time during the pilot study.

Conclusion

Usability studies have been used to evaluate many kinds of products. As corporations are gaining interest in the international market, usability studies have

been conducted with participants from different cultures. Those cross-cultural usability studies are often challenged by large cultural diversity. In this paper, two case studies that used the *think-aloud* protocol were observed, analyzed and compared. One study was conducted in China while the other took place in Germany.

The results showed that German participants were more engaged in the *think-aloud* session than were the Chinese participants. They spent more time talking aloud, and provided more verbal information to the evaluator by thinking aloud than Chinese participants. It is possible that culture can affect participants in their attitude towards *thinking-aloud*, their ideas of a comfortable testing environment, and their cognitive styles. Moreover, culture can also affect evaluators' behaviors and judgments. Since there are limitations in the analysis of the case studies, future studies need to be conducted to develop practical and efficient approaches for evaluators in cross-cultural usability studies.

References

- Beringer, D., & Ball, J. (2002). A flight simulator usability assessment of a multi-function flight-planning and navigation display for general aviation. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, 46, 6.
- Berns, T., Klusell, L., & Pousette, S. (2000). Usability evaluation of a multimedia program for school children, grade K-6. *Proceedings of the IEA 2000/HFES 2000 Congress*, 44, 685.
- Butler, R., & Scherrer, C. (2009). Developing a standard student interface for online courses through usability studies. *American Society for Engineering Education. AC 2009-2195*.
- Briley, D. A., Morris, M. W., & Simonson, I. (2000). Reasons as carriers of culture: dynamic versus dispositional models of cultural influence on decision making. *Journal of Consumer Research*, 27 (2), 157-178.
- Chavan, A. L. (2005). Another culture, another method. *HCI 2005 Conference Proceedings*.
- Clemmensen, T. (2011). Templates for cross-cultural and culturally specific usability testing: results from field studies and ethnographic interviewing in three countries. *International Journal of Human-Computer Interaction*, 27 (7), 634-669.
- Clemmensen, T., Hertzum, M., Hornbak, K., Shi, Q., & Yammiyavar, P. (2009). Cultural cognition in usability evaluation. *Interacting with Computers*, 21, 212-220.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis Verbal Reports as Data*. Cambridge: The MIT Press.

- Dunckley, L., Oyugi, C., & Smith, A. (2008). Evaluation methods and cultural differences: studies across three continents. *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges*, 318-325.
- Herman, L. (1996). Towards effective usability evaluation in Asia: Cross-cultural differences. *Proceedings of the 6th Australian Conference on Computer-Human Interaction*, 135-136.
- Hofstede, G. Retrieved 11 20, 2012, from Greet Hofstede & Gert Jan Hofstede: <http://geerthofstede.nl/dimensions-of-national-cultures>
- Hsieh, H. C. (2011). Exploring the impact of cultures on web usability test. *Lecture Notes in Computer Science*, 6776, 47-54.
- Kim, H. S. (2002). We talk, therefore we think? A cultural analysis of the effect of talking on thinking. *Journal of Personality and Social Psychology*, 83(4), 828-842.
- Kim, H. S., & Sherman, D. K. (2007). "Express Yourself": Culture and the effect of self-expression on choice. In press, *Journal of Personality and Social Psychology*, 92 (1): 1-11.
- Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. T. (2003). Perceiving an object and its context in different cultures: A cultural look at new look. *Psychological Science*, 14(3), 201-2003.
- Lee, F., Sanchez-Burks, J., Choi, I., Nisbett, R., Zhao, S., & Koo, J. (2003). Conversing across cultures: East-west communication styles in work and nonwork contexts. *Journal of Personality and Social Psychology*, 85 (2), 363-372.
- Li, F., & Li, Y. (2011). Usability evaluation of e-commerce on B2C websites in China. *Procedia Engineering*, 15, 5299.

- Masaaki, K., Tadashi, K., Ryoji, Y., Hideaki, T., Haruhiko, U. Trends in usability research and activities in Japan. *International Journal of Human-Computer Interaction*, 17(1), 103-124.
- McDonald, S., Edwards, H. M., & Zhao, T. (2012). Exploring think-alouds in usability testing-an international survey. *55(1)*, 2-19.
- Nisbett, R. E. (2004). *The Geography of Thought: How Asians and Westerners Think Differently...and Why*. Free Press, New York, United States.
- Paterson, B., Winschiers-Theophilus, H., Dunne, T. T., Schinzel, B., & Underhill, L. G. (2011). Interpretation of a cross-cultural usability evaluation: A case study based on a hypermedia system for rare species management in namibia. *Interacting with Computers*, 23, 239-246.
- Peterson, M., Parker, L. (2005). Surviving the international usability test: Tools and strategies for a successful outcome. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 49th Annual Meeting*. 49, 2080-2084.
- Pérez-Quiñones, M., & Vatrappu, R. (2006). Culture and usability evaluation: The effects of culture in structured interviews. *Journal of Usability Studies*, 1, 156-170.
- Seto, E., Roach, C., Chagpar, A., & MacDonald, A. (2006). Usability Evaluations as Part of the Procurement Process: Case study of hospital point of care carts. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 50, 1095.
- Shi, Q. (2008). A field study of the relationship and communication between Chinese evaluators and users in think aloud usability tests. *Proceedings: NordiCHI*, 344-352.
- Yeo, A. W. (2000). Are usability assessment techniques reliable in Non-Western cultures? *The Electronic Journal of Information Systems in Developing Countries*, 3, 1-21.

Young-Corbett, D. E., Nussbaum, M., & Winchester, W. W. (2010). Usability evaluation of drywall sanding tools. *International Journal of Industrial Ergonomics*, 40, 112.