Copenhagen Business School

HANDELSHØJSKOLEN

A comparison of what is part of usability testing in three countries

Torkil Clemmensen

A comparison of what is part of usability testing in three countries

Torkil Clemmensen

Department of Informatics
Copenhagen Business School
Howitzvej 60
Frederiksberg, 2000, Denmark
tc.inf@cbs.dk

Abstract. The cultural diversity of users of technology challenges our methods for usability evaluation. In this paper we report and compare three ethnographic interview studies of what is a part of a standard (typical) usability test in a company in Mumbai, Beijing and Copenhagen. At each of these three locations, we use structural and contrast questions do a taxonomic and paradigm analysis of a how a company performs a usability test. We find similar parts across the three locations. We also find different results for each location. In Mumbai, most parts of the usability test are not related to the interactive application that is tested, but to differences in user characteristics, test preparation, method, and location. In Copenhagen, considerations about the client´s needs are part of a usability test. In Beijing, the only varying factor is the communication pattern and relation to the user. These results are then contrasted in a cross cultural matrix to identify cultural themes that can help interpret results from existing laboratory research in usability test methods.

Keywords: Usability test, cultural usability, ethnographic interviewing.

1. Introduction

Culture plays an increasing role in discussions of information and communication technology. As of today, we do not have any formal methods to guide us in evaluating a product to a certain standard while being sensitive to cultural issues. Cultural usability tests are not yet established methods. In this paper we look at the methods that companies already use.

In the past few years researchers have suggested paradigms for culture-specific Human-Computer interaction such as 'cultural computing' [14], 'culturally sensitive IT' [20], and 'cultural usability' [2]. Attempts have been made to include cultural knowledge such as cultural dimensions [10], cultural factors [16], cultural constraints [12], and cultural models [8], in research into HCI in general, and into cultural usability specifically, e.g. [19]. What is common in these newer approaches to cultural usability is a focus on the diversity of users and use of technology around the globe, social-cognitive approaches to usability (as opposed to psycho-physiological approaches), and the utility of HCI.

A major finding from the existing literature on cultural models in HCI is that there are differences in cultural models in the East (Asia) and in the West (USA, Europe). These differences imply the need for localized designs [10] and for local adaptations of usability evaluation procedures [17]. Specifically, empirical studies show that Chinese users adapt a more holistic approach to using software compared to European users [16]. This resembles the general finding from cultural psychology on human cognition that easterners are context focused, while westerners are object focused [11]. An example of this is that asked to report what is on a scene, easterners mention the background, while westerners report the focal objects. The cross cultural differences in cognition lead us to expect cross cultural differences in usability evaluation.

In this paper we study the following research questions. How does the practice of usability testing address cultural diversity for both the evaluator and the user? What is part of a standard usability test in India? What is part of a standard usability test in Denmark? What is part of a standard usability test in China? What is part of a standard usability test across all three countries?

2. Method

The study that we report in this paper is part of a multi-site, cross-cultural, grounded theory field study of think aloud (TA) usability testing in eight companies in three countries (Denmark, China and India). Our field studies of how these companies ran a typical, or standard, think aloud usability test, had given us initial ethnographic recordings consisting of 52 hours of videotaped usability tests, interviews with usability managers, evaluators, notetakers, and test users, and notes from confronting the employees of the company with our observations and interpretations. All of this we analyzed with grounded theory, see [5]. In the present paper, however, we report on a follow-up study which we did one year after the original field study.

The method used in the study was a classical ethnographic interview [18], which was done as a follow up study in one of the companies one year after the first round of observation. The three companies selected for this follow-up study were a) a Mumbai based company with more than 200 usability and user centered design specialists that is an Indian branch of an international usability consulting company, b) a Copenhagen based usability vendor with 12 employees, and c) a Beijing based branch of a major telecommunications international company with an in-house group of usability specialists.

Our initial ethnographic record from the original large scale field study that we did the year before, was the basis for a taxonomic and a paradigm analysis [18] of what is a part of a usability test in the companies in question. The next step was, in each of the three companies, to do two days of ethnographic follow-up interviewing with our key informant from the year before. In all three cases the informant was a usability evaluator with senior responsibility. We followed a classical ethnographical interview procedure suggested by Spradley [18]:

1. Create network/set of codes related to the code 'Usability test' by an is-a-part-of relation.
2. Print a code hierarchy (a specific procedure in the software used in the analysis).
3. Ask the informant questions about each term (code) in the hierarchy: name, other of same kind, difference to others etc.
4. Do it for one sub domain at a time.
5. Enter all the responses in the code hierarchy
6. Go back, change the network of codes accordingly
7. Iterate the process, if necessary

In the day 1 interview, we created, adjusted and verified the taxonomy by asking the informant structural and contrast questions such as [18]:

- Is <x> a term (code) you would use?
- Would most people here at <this company> usually use this <X> term?
- Is <Y> a part of <X>? Are there different parts of X? What other parts of <X> are there?
- Do you see any differences between <X .1> and <X.2>? and <x.1 >and <x.3>? (and so on)

The <X> term could for example be on the highest level of the taxonomy 'usability test' or on a lower level for example 'inform participant'.

In the day 2 interview, we created the paradigm by this procedure:

1. Place the first level of the taxonomy in a column in a worksheet.
2. Inventory all other codes related to "usability test" by other relations than is-a-part-of relations and place them as the top-row in the worksheet.
3. Prepare contrast questions such as "is moderating dependent on the test user's age or gender?"
4. Conduct an interview with the informant to elicit needed data.

The final step was to use the analysis to discover general cultural themes. This was done by careful analysis and comparison of the interview data (see the next sections in this paper).

To give an example of the procedure, in the interviews with the Indian informant the final taxonomy showed that 182 concepts are part of a standard usability test in the studied company. Of these, 23 were main concepts that had up to three sublevels.
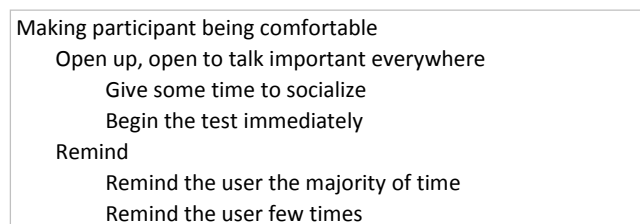
```
Making participant being comfortable
    Open up, open to talk important everywhere
        Give some time to socialize
        Begin the test immediately
    Remind
        Remind the user the majority of time
        Remind the user few times
```

Figure 1. Excerpt from taxonomy of usability testing, showing an entry: 'making participant being comfortable' and sublevels.

Figure 1 gives illustrates how a part of a standard usability test in the company in Mumbai is to make the participant (the test user) comfortable. The purpose is to get

the user to 'open up' and varies in duration, depending on the user. It is also done to get users to think out loud in the proper way.

The full paradigm for the usability test in the India company was then made by setting up a matrix in which the rows illustrate the different parts of the usability test taken from the main elements in the taxonomy illustrated in Figure 1, and the columns illustrate the context-factors. For example, how exactly important parts of the usability test in Mumbai, such as greeting, compensation and thanks to the participant, will be carried out, depend in the Mumbai company on the context-factor that is the user's gender: *"…I may not have a handshake with a lady..."*. This and other findings are presented in more detail in the next sections of this paper.

Comparing the taxonomic and paradigm analysis across the three countries/companies/informants was done in two steps. First we described what were the common parts (taxonomy analysis) and common context (paradigm analysis) of a usability test across the three countries. The result was a cross cultural paradigm for a usability test with cross cultural parts (those parts all mentioned) and cross cultural context (those contexts which all the informants said had influence on one or more cross cultural parts). Second we described for each country, in addition to the common cross cultural elements, what were specific parts and specific contexts for a usability test in each country. The results were three country specific paradigms for a usability test process.

3. Cross cultural findings

This section presents what parts of and what context-factors for a usability test are common across the three countries.

From the taxonomic analysis based on the interviews, and inspired by the usability literature [4], we learned that cross culturally there are four main parts of a usability test, which each has a number of sub-parts, see Table 1.

Table 1. **The interview-based, cross cultural taxonomy for a think aloud usability test**

Main part of usability test	Sub part of usability test
Instruction and tasks	Introduce user to think aloud
	Introduce user to technology to be tested
	Introduce user to test task
Verbalization	Probe for specific information
	Remind the user to think aloud
	Communicate with the user
Reading the user	Observation room, one way mirror
	Video of user and screen
	Expectations to users task performance
Overall user-evaluator relationship	Explain user not tested, design tested

The influence of context-factors on one or more parts of the usability test is seen in the paradigm in Table 2. Cross culturally there are eight context-factors that in some way or another influence the main parts of a usability test.

Table 2. The paradigmatic context-factors in a cross cultural a usability test.

Usability test parts	Context for usability test							
	Recruitment of participants	User's age	User's personality	User's cultural back-ground	Design test/formative evaluation	The final report	Consolidate the data	Design recommendation
Instruction and tasks	Yes	Yes	Yes	No	Yes	Yes	No	No
Verbalization	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reading the user	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Overall user-evaluator relationship	No	No	No	No	Yes	No	No	No

Note: 'Yes' means that the part of the usability test is influence by the context-factor, 'No' means that it is not influence by the context-factor.

In the following, the interview data on the four parts and the eight contextual factors are discussed one by one.

First, already during the recruitment of test participants it is important to consider how to introduce the to-be-tested technology solution to the future test user. The Indian informant pointed out that *"the client may define a target audience that require that…[we consider issues of]…computer literacy or application awareness"*. Furthermore, the think aloud technique may sometimes requires what the Danish informant described as *"smooth talkers"*, and she also said that it may be important during recruitment to tell the future test users that there will be an observation room with people from the client.

Second, the users' diversity influences the parts of a usability test. The instruction and task part is related to age; the Indian informant described the situation like this: *"introducing that to older people requires a bit more explanation sometimes, this is how it is going to work, they are not very tech savvy as such…"*. Age does also play a role for verbalization, as the Chinese informant said: *"older persons need more encouragement"*, and for expected task performance when observing the user: *"young or very old people are not expected to be able to solve the same tasks as a standard adult"* (Danish informant). Interestingly, despite the common sense nature of the statements there are only few studies that deal with usability across the life span (see e.g. [3], and the few findings on specific age groups like the elderly, shows that they are just as technological savvy as the other population [15]. I have not been able to find studies that compare usability testing with different age groups, which could corrode or confirm my informants' views.

Third, the user's personality is a contextual factor that is relevant for giving 'Instructions and tasks', 'Verbalization' and the 'overall user –evaluator relationship. For 'Instructions and tasks' *"you would like to give the same [*instructions*], but in a different language"* (Danish informant). *F*or 'Verbalization' *"it depends on the extrovert- or introvert-ness of a person, if he or she feels comfortable with verbalizing…"* (Danish informant). Both the Danish and Chinese informant felt that when and how to encourage and stop a test user from speaking depend on the test users' personality. Finally, *"if he is an introvert, your body language and everything will move towards making him comfortable …"* (Indian informant). These statements can be compared to experiences from other fields that apply evaluation methods. For example, in the field of administering psychological personality tests, the tester is expected to adapt the communication with the test taker according to his or her personality. However, so far few studies of usability has included personality or briefly state that they consider personality unimportant, e.g. [6]. One exception is a study of interactive television prototypes that were designed in color and shape to show different degrees of extrovertness, and evaluated in reference to the users' measured personality traits [1].

Fourth, the user's cultural background influences the 'Verbalization' and 'Reading the user' parts. Usability evaluators may experience that *"…in Singapore the users are more shy than in the US…[we do not] give as many reminders [to think aloud] in US as in Singapore and India…"* (Indian informant). A difference in cultural background is something the evaluator can use *"…you may use the difference, you can take the role as the stranger entering from outside…"* or something to be learned *"…culture can be a professional qualification that you do not have…"* (Danish informant). Cultural background can also be viewed as a matter of differences in education *"…[yes, cultural background matters] if we talk education, people with low education need more encouragement…"* (Chinese informant). There seems to be three aspects of the concept of cultural background: national/ethnic culture (Indian informant), professional culture (Danish informant) and educational culture (Chinese informant). Despite the multiple meanings of the concept, we kept the cultural background as a context in the cross cultural paradigm for pragmatic reasons, see [9].

Fifth, the test methodology acts as a context for all the parts of a usability test in the sense that all four parts of a usability test are fixed in a summative evaluation/beta test, but in formative evaluation or design evaluation the properties of the test parts varies. The variation is related to how to instruct the user in thinking aloud, how to tell the user about paper prototype, if there are any real test tasks or only interview questions, and how much the user should think aloud and how much qualitative data is needed. In a formative test, the expectations to task performance are plastic: *"…yes, if they say something outside the expectation, you should reverse the expectation, if all the results go far away…you can stop and reverse it"* (Chinese informant). Finally, in a formative test, the necessity of making the user comfortable varies. The existence of such variation in formative usability testing, which is not the case in summative usability testing (where all parts of the test is performed in a fixed way), indicates a need for investigating more how this variation in the context of a usability test influence the test procedures. This aligns with recent suggestions to limit research on criteria for evaluating usability evaluation methods to formative usability evaluation [7].

Sixth, the final three contextual factors all relates to considerations about how to communicate the results, i.e. what to write in the final report, how to consolidate the data, and how to present the design recommendations, all influence the parts of a usability test. In some cases the instructions and tasks are written in the final report. The final report usually also contain information about the verbalization: *"... usually we write if we had to lead them [the test users] a lot, if they were helped or not, if they acted spontaneously or not..."* (Danish informant), *"...if probing happened [we write it]... especially if there is a common response from that kind of user, this kind of user needs more encouragement..."* (Chinese informant). Also information about how the user was read/understood is entered into the final report *"...if the client has certain expectations we show them the graphics [on user performance]"* (Indian informant). Expectations to typical ways of consolidating the data, e.g. reports typically provide excerpts of user verbalization, and ways of giving design recommendations, e.g. results are often presenting using power points with wireframe examples, influence how the verbalization occurs and how the user is read. This influence, from consideration about how to communicate the results, to the main parts of a usability test (and reverse), indicates that usability testing is carried out with the philosophy of iterative testing. It is congruent with recommendations from research on communicating the results of usability tests to designers, which says that evaluators should be explicit about the data behind their claims, but not overwhelm the designers with information, and rather involve them in a learning process [13].

3.1. A cross-cultural template for the usability test process

The information from the interviews analyzed above indicates that a usability test is a complex affair; a simple count of the cells in Table 2 gives four times eight = 32 pos-
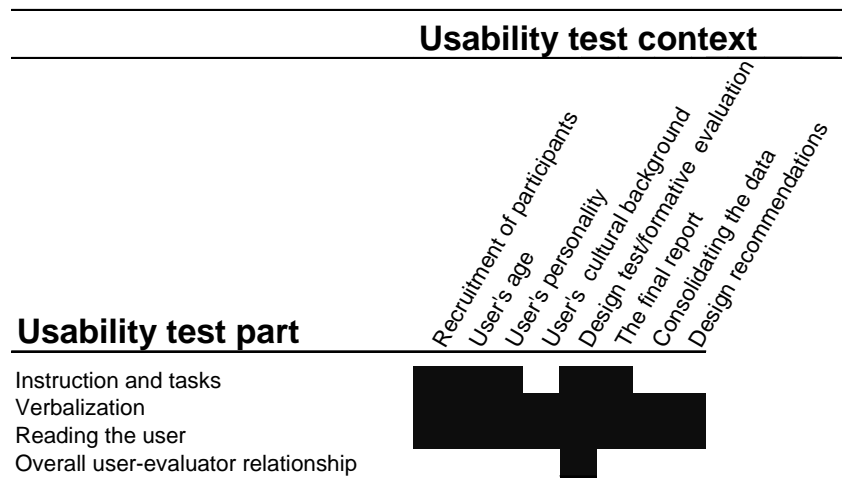

Figure 2. The interview-based, cross cultural paradigm/template of a usability test (Graphic view of the information presented in table 2).

sible aspects for consideration by usability vendors when they carry out a usability test. The interviews and the above analysis indicate however also that not all of these 32 aspects are equally important. Figure 2 depicts graphically the 21 important aspects of the four parts of a usability test. The depiction can be regarded as a paradigm or a *template* for a cross-cultural usability test process.

The template for a cross-cultural usability test process that is shown in Figure 2 can be interpreted as follows. If a practitioner (or a researcher) is going to perform a cross-cultural usability test, as a minimum he or she has to consider the aspects that are colored black in Figure 2. For example, he or she should ask the question: "how will the think aloud verbalization that I require from the user, support the design recommendations that I will give?" The gain from having a graphical depiction of the cross-cultural usability test process is that it can be used as a "must-have" checklist in the preparation of tests, but also that a representation of the type illustrated in Figure 2 is useful, when looking for variations in how the usability test should be carried out in different cultures and countries.

4. Culturally specific parts of a usability test

In each country, the informant mentioned parts and context that were not shared by the informants in the other countries. While this should not be generalized to believe that all usability tests in a given country are performed in a similar way, the interview data indicates three distinctly different ways of performing a usability test. For example, in a usability test in Beijing it is important, the Beijing informant told me, during 'Instructions and tasks' to *"explain test to give user an overview* [of the whole test]", and, in order to facilitate 'Verbalization', to use a *"direct and encouraging communication style"* when talking to the user. The first of these statements pertain to the holistic cognitive style identified by e.g. [11], and the second statement indicate that verbalization of your thoughts is an unusual activity that needs to be directly addressed in a Chinese situation. In Danish usability tests, the focus on the client (customer) needs, which we will discuss more in the country specific sections below, could explain why the Copenhagen interviewee did see not see important distinctions between those who 'Observe the user', but simply talked about varying groupings of people could observe from the *"observation room"*. For the last part: 'Overall user-evaluator relationship', the informant in Mumbai insisted that it was most important to *"make the user feel comfortable"* ; the validity of this statement it is supported by interviews, observations and group discussions with fifty plus employees of a major Indian usability vendor that we did previously [5]. Thus the interviews indicated that the four parts of a usability test should be interpreted differently in the three countries. Further analysis of the data suggested that it make sense to distinguish between an "evaluator-centered", a "user-centered" and a "client-centered" approach to the usability testing process. These three approaches are described in the following.

4.1. "Evaluator-centered" as the approach to the usability test process

The Beijing informant told us about several parts of a usability test that we could not recognize from our studies in the other countries, see Figure 3. These standard parts of a usability test in Beijing were 'translation' (you always consider the need for translation for client's, moderator's or test user's sake), 'Problem fixing' (there are always some problems with the test that you have to fix and continue), 'Usability problem description' and 'User experience description' (you always focus on describing the user's interaction with the technology solution), 'Combined moderator, observer and note taker' (you are always in all three roles during a test), and 'Professional notetaker' (it is a standard thing to have a professional that makes a full transcription). However, besides the four standard parts of a usability test, only two more parts varied with the different contexts. Firstly, the 'Choice of usability evaluation method' could vary for intranet, web search portals and government web pages *"…a few methods like expert review can be used…"*, but also for cross cultural contexts. For example, 'Choice of usability evaluation method' was about fitting the method to user's personality *"…this kind of user is more talkative, maybe more useful in some kind of tests…"*, or to user's cultural background *"…to only country or religion, but education, different people can give different feedback, like the IT person gives totally different feedback from the general user…"* and to the users' lifestyle and family background *"…products like sports, select the natural observation method…"*. Secondly, the 'Choice of task scenario or task list' varied both with technology solutions as a matter of fitting topic with methodology, and would accordingly as part of data consolidation and design recommendations be written in the final report.

Usability test context

Usability test part	Recruitment of participants	User's age	User's personality	User's cultural background	Design test/formative evaluation	The final report	Consolidating the data	Design recommendations	Company intranet web page	Goverment website	Mobile phone interface	Newspapers travel website	Telecommunication web site	Online computer game	Industrial proces control room	Web search portal
Instruction and tasks									-	-	-	-	-	-	-	-
Verbalization									y	y	y	y	y	y	y	y
Reading the user									y	y	y	y	y	y	y	y
Overall user-evaluator relationship									-	-	-	-	-	-	-	-
Choice of usability evaluation method	w	-	w	w	y	-	y	-	y	y	-	-	-	-	-	y
Choice of task scenario or task list	-	-	-	-	y	y	y	y	y	y	y	y	y	y	y	y

Figure 3. Culturally specific paradigm for tests in Beijing.

The technology solutions to be tested were important contexts for the standard parts of a usability test. 'Verbalization' was related to all kinds of technology solution *"...but only through the methodology, sometimes you need more qualitative data...".* 'Reading the user' was related to who would be the observers of that technology *"...if hot topic, the marketing should hear directly from the users...[if intranet or other internet technology] the designer has no need to hear it directly from user...[if mobile phone interfaces] usually the designer will join...".* When asked about who chose the methodology and why the methodology was most important, the informant explained that the evaluator chooses the methodology in each case of a usability test, and that this was possible, because the evaluators usually were highly qualified human factors professionals. Together with the findings from the year before [5], this suggested, for the template shown in Figure 3, the label "evaluator-centered" usability test process.

4.2. "User-centered" as the approach to the usability test process

In Mumbai, in addition to the cross cultural usability test parts the informant explained about 10 other parts of a usability test, see Figure 4.

Figure 4. Culturally specific paradigm for usability tests in Mumbai.

The task scenario part of a usability test is influenced by whether summative or formative tests are done, and it is usually described into the final report. The task scenario is also influenced by where geographically the test takes place, which was a context-factor that was only mentioned by the Indian informant.

There are four distinct parts of a usability test that are related to dealing with the user as a person. The conduct of these parts varies in different contexts. For example, when recruiting the user *"...we think about what kind of compensation to be given...to a certain extent we ask them if they are willing to come here...nothing in writing but we have the consent...we are asking the permission [from family]..."*, and when greeting the user *"...in US, for example, even if he is an older or younger person, the greeting will be the same. Here the elderly will feel good if I bow..."*. Specific contexts for the Mumbai usability test include gender *"...I may not have a handshake with a lady..."*, government websites *"...in US if your client is government agency you cannot give them compensation...[actually] I don't know if it is the case in India..."*.

The moderator skill level is related to user's personality *"...if he [the user] is an introvert, you [the moderator] may need to..."*, user's cultural background *"...You may need to know the nuances of the culture and understanding of that..."*, test methodology *"...for formative tests it would be good to have an experienced moderator...summative is fairly straight forward...."*, writing the final report *"...it is part of the contract that we have experienced moderators...also an experienced moderator is involved in creating the final report...very closely...various models here, he could be writing it, he could be overseeing [parts of] of the final report.."*, consolidating the data *"...the notetaker in consolidation with moderator, usually the moderator is much senior than the notetaker..."*. For the moderator skill level there are also specific contexts such as user's motivation *"...in the sense that you have to realize that when a person is not motivated, he is probably not giving you the real feedback...the moderator has to realize that ..and he has to do some twists..."*, the kind of technology solution *"...if it is a complex application we would need experienced moderators..."*, and remote testing *"...it is good to have someone who has done some remote testing, because the technology issues...making a phone, call, the supporting things, you should be aware of the things that can go wrong during the test, ...it be phone line, internet connection, web example, accent..."*.

The use of test documents such as formal test protocols and notes about usability problems are also important parts of the usability test in the Mumbai company. These have to be visible in the final report and data consolidation *"...if you have not been able to conduct all the tasks as per the protocol, you have missed out something...you do mention what data will be captured...in remote testing you will not be able to capture body language and facial expressions..."*. Getting other insight about the user interacting with the technology solution is also a standard part of a usability test, in particular in formative tests *"[the data are] much richer here than in summative, things can strike you here..."*. The time of test session is a standard part of the test that is considered during recruitment of test users *"I just mention to the user that it will take one hour..."*.

Eight Mumbai company specific context-factors were mentioned by the informant, see the right-most eight columns of Figure 4. These eight context-factors influenced both the 10 culture-specific parts of the test, and the four cross cultural parts of a

usability tests. For example, 'Verbalization' depends on user's motivation *"…[if low] very strongly, lot more probing would be required, give more reminders, also correlation to assists..",* and is considered during test protocol development, when interacting with the client *"…sometimes the clients are saying that we really need you to get more information ….then we identify were more probing is required..".* 'Overall user-evaluator relationship' is related to the user's motivation and considered during test protocol development *"You have to spend extra efforts if that person is not motivated… one of the reasons for using [specific kind of] scenarios is to make the user comfortable…".*

The heavy focus on the test parts related to greeting and informing the user and the test context-factors significant influence on the overall relationship between evaluator and test user, inspired to label the template illustrated in Figure 4 a "User-centered" approach to usability test process.

4.3. "Client-centered" as the approach to the usability test process

Figure 5 shows the findings from Copenhagen. Besides the moderator's experience with clients, the Danish informant mentions using "clickable prototypes" as an important part of usability testing that is always there, no matter the different contexts.

Usability test context

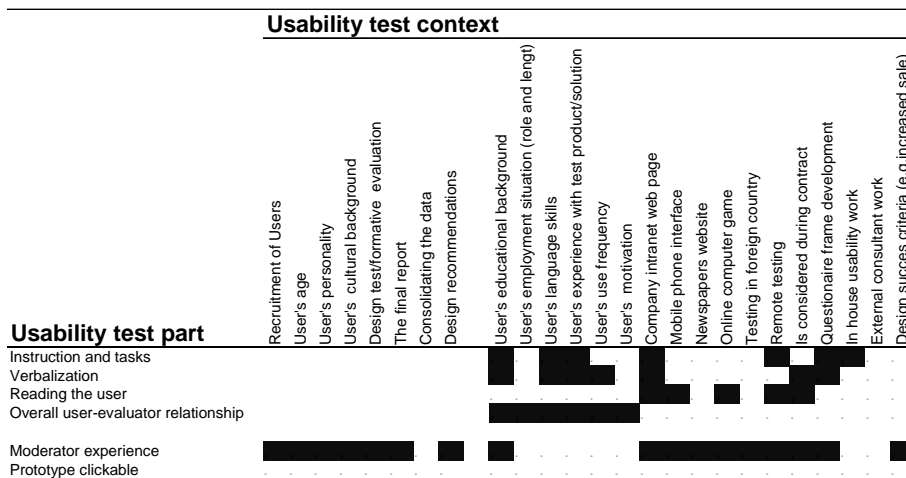| Usability test part |
| --- |
| Instruction and tasks |
| Verbalization |
| Reading the user |
| Overall user-evaluator relationship |
| |
| Moderator experience |
| Prototype clickable |

Figure 5. Culturally specific paradigm for usability tests in Copenhagen.

Among the 17 different contextual influences on a usability test mentioned by the Danish informant are the user's gender: *"…females over 40 years are less talkative…"* ('Verbalization'), and *"…usually we select both male and female users because their context of use can be different"* ('Reading the user'); educational background *"…you would like to express yourself slightly different…"* ('Instructions and tasks'), *"…higher education gives more verbalization and better verbalization,*

people with low education has a tendency to apologize not being able to do the test correctly...” ('Verbalization'), and different expectations to task performance ('Reading the user'); user's employment situation (role and length) *“...there will be different tasks to managers and employees, there will be something that employees cannot answer, it is not part of their job...”* ('Reading the user'); users' language skills *“the problem can be that you cannot be sure to understand what they are saying...”* ('Instructions and tasks'), and *“...if their English is so bad that they cannot read the task instructions...”* ('Reading the user'); users' experience with the technology to be tested, use frequency and motivation *“...if it is a super user, I must say 'tell me about you knowledge'...”* ('Verbalization') and expectation to task performance ('Reading the user'); intranet webpages *“...in a test of intranet you have to make a point to the user that he or she is anonymous, try to explain to them that in the report it will not be written who they are...in such a test you often speak about use and frequency of use...the boss might not like to hear that...”* ('Instructions and tasks'), *“...[in intranet tests] you are the stranger coming from outside and has to try to figure out how much you know the concept and tools that are there...can you be part of that organizational culture...”* ('Verbalization'), and *“...there is no observation room because the test [of an intranet] will be done in the field, not in the lab...”* ('Reading the user'); mobile technology *“sometimes there are no observation room, because we move around [during test of mobile units] in town, in traffic...”* ('Reading the user'); online computer gaming *“...you may go to the user's home to watch them gaming in their environment...”* ('Reading the user'); remote usability testing *“...you have to explain a bit technical how the test will proceed, how you as a user get access, that you at the end of the test will be disconnected...”* ('Instructions and tasks'), and *“usually there are no observation room, if there are one, the observer has also logged on with a separate screen...”* ('Reading the user'); contract negotiation *“...we write about method in the contract...”* ('Verbalization'), *“...we always write [in the contract] that we have an observation room customers like that we are open and trust them, that they may come and learn about their users by observing them...”* ('Reading the user'); questionnaire guide *“...then there will be a script that tells us how to talk to these users...”* ('Verbalization'); and, furthermore, in-house usability work *“clearly explain the technology solution is not my design, so I will not be offended [by the user's critique]”* ('Instructions and tasks').

Finally, the context of success criteria (increased sale, interested community users, number of users seeing the advice given, etc.) is relevant for a part of usability test that only the Danish informant mentioned, namely the moderator's experience with working with clients: *“...there is not sufficient focus on that [client's success criteria] if you have taken the way through usability in your education – instead you need to have experience working with clients, you need to know what makes your clients pay attention to this and to be persuaded that this is necessary...it is important for your design recommendations that you reflect on what are the business goals for the use of this technology solution...”.* This focus on business goals, together with the view of the legal contract with the client as an important context for carrying out the usability test, suggested the use of the label "Client-centered" approach to usability test process.

5. Conclusion

The ethnographic interviews with the taxonomic and paradigm analysis indicate that a typical or standard usability test across countries has some clear similarities, with four main parts and eight important contexts to consider when doing the test.

In each country there are specific parts of a usability test and specific contexts for a usability test, which are not found everywhere. In Mumbai, most parts of the usability test are not related to the interactive application that is tested, but to differences in user characteristics, test preparation, method, and location. In Copenhagen, considerations about the client´s needs are part of a usability test. In Beijing, the only varying factor is the communication pattern and relation to the user.

The significance for academic researchers is that this kind of knowledge from the field can help provide a conceptual basis for interpreting the results of studies of usability test methods in laboratories. In particular, the use of paradigms as templates for usability evaluation process in this study can be seen as a simple and practical way to compare and improve how usability evaluation methods are carried out in multiple different cultures and countries.

The take-away for practitioners from this study is that the taxonomies and paradigms indicates areas of concern and help interpret results for practitioners who wants to execute and compare results from usability tests done in multiple countries. In particular, the templates can be used as "must-have" checklists in the preparation stage of multi-country usability tests.

The methodology insights from this study was that, compared to the current popular grounded theory approach, the strength of the ethnographic approach is that the terms revealed are the informant's own terms. Thus idiosyncrasies and tacit knowledge developed in the particular company can be revealed by our study. On the other hand, the informant may do his best to conform to some international standard or what believes that he has learned during his formal education, i.e. he violate the interviewer's need for a "non-analytic informant" [18, p52]. Another methodological challenge, which has been met with some success in this study, is to compare ethnographic studies from three sites distributed globally.

The major limitations of this follow-up study is that only one informant in one company was interviewed in each country, though these interviews were in-depth interviews done across several days, and based on previous extensive field observations. However, the results of this study suggest the need for more empirical evaluation of the depth and significance of apparent cross cultural similarities and subtle differences in what are parts of a usability test, how the test is practiced in different contexts, and what are the implications of such findings. This kind of knowledge from the field may provide a conceptual basis for interpreting the results of studies of usability test methods in laboratories.

Acknowledgements

6. References

1. Arvid, K. Personality preferences in graphical interface design *Proceedings of the second Nordic conference on Human-computer interaction*, ACM, Aarhus, Denmark, 2002.
2. Barber, W. and Badre, A., Culturability: The Merging of Culture and Usability in *4th conference on Human Factors and the web*, (1998).
3. Brouwer-Janse, M., D. . From our past to our future: user interfaces over the lifespan (panel session) *Conference companion on Human factors in computing systems*, ACM, Denver, Colorado, United States, 1995.
4. Clemmensen, T., Hertzum, M., Hornbæk, K., Shi, Q. and Yammiyavar, P. Cultural cognition in usability evaluation. *Interacting with Computers*, *21* (3). 212-220.
5. Clemmensen, T., Shi, Q., Kumar, J., Li, H., Sun, X. and Yammiyavar, P. Cultural Usability Tests -- How Usability Tests Are Not the Same All over the World. in Aykin, N. ed. *Usability and Internationalization - Second International Conference on Usability and Internationalization, UI-HCII 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27, 2007, Proceedings, Part I* Springer, Heidelberg, 2007, 281-290.
6. De Angeli, A., Sutcliffe, A. and Hartmann, J. Interaction, usability and aesthetics: what influences users' preferences? *Proceedings of the 6th conference on Designing Interactive systems*, ACM, University Park, PA, USA, 2006.
7. Hartson, H.R., Andre, T.S. and Williges, R.C. Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, *13* (4). 373-410.
8. Jagne, J. and Smith-Atakan, A.S.G. Cross-cultural interface design strategy. *Universal Access in the Information Society*, *5* (3). 299-305.
9. Marcus, A. Culture: Wanted? Alive or Dead. *Journal of Usability studies*, *1* (2). 62-63.
10. Marcus, A. and Gould, E.W., Cultural dimensions and global web user-interface design: What? So what? Now what. in *6th Conference on Human Factors and the Web*, (Austin, Texas, 2000), 1–15.
11. Nisbett, R.E., Peng, K.P., Choi, I. and Norenzayan, A. Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108* (2). 291-310.
12. Norman *The design of everyday things*. Doubleday, 1988.
13. Nørgaard, M. and Høegh, R.T. Evaluating usability: using models of argumentation to improve persuasiveness of usability feedback *Proceedings*

   of the 7th ACM conference on Designing interactive systems, ACM, Cape Town, South Africa, 2008.

14. Rauterberg, M. From Personal to Cultural Computing: how to assess a cultural experience. *Proceedings of the 4 thUsability Day (9. Juni 2006), Applied University Vorarlberg, Dornbirn, Austria. Pabst Science Publisher.*

15. Rousseau, G.K. and Rogers, W.A. Computer usage patterns of university faculty members across the life span. *Computers in Human Behavior*, *14* (3). 417-428.

16. Smith, A., Dunckley, L., French, T., Minocha, S. and Chang, Y. A process model for developing usable cross-cultural websites. *Interacting with Computers*, *16* (1). 63.

17. Smith, A. and Yetim, F. Global human-computer systems: Cultural determinants of usability. Editorial. *Interacting with Computers*, *16*.

18. Spradley, J.P. *The ethnographic interview*. Wadsworth Pub Co, Holt, Rinehart and Winston, 1979.

19. Sun, H. Expanding the Scope of Localization:A Cultural Usability Perspective on Mobile Text Messaging Use in American and Chinese Contexts *Rensselaer Polytechnic Institute Troy*, New York, 2004.

20. Zakaria, N., Stanton, J.M. and Sarkar-Barney, S.T.M. Designing and implementing culturally-sensitive IT applications. *Information Technology & People*, *16* (1). 49-75.