# Culture and Usability Evaluation: The Effects of Culture in Structured Interviews

**Ravi Vatrapu**

Communication and Information Sciences

University of Hawaii at Manoa

Honolulu, HI, 96826. USA

ravikira@hawaii,edu

**Manuel A. Pérez-Quiñones**

Department of Computer Science

Virginia Tech

Blacksburg, VA, 24061, USA

perez@vt.edu

## Abstract

A major impediment in global user interface development is that there is inadequate empirical evidence for the effects of culture in the usability engineering methods used for developing these global user interfaces. This paper presents a controlled study investigating the effects of culture on the effectiveness of structured interviews in international usability evaluation. The experiment consisted of a usability evaluation of a website with two independent groups of Indian participants. Each group had a different interviewer; one belonging to the Indian culture and the other to the Anglo-American culture. The results show that participants found more usability problems and made more suggestions to an interviewer who was a member of the same (Indian) culture than to the foreign (Anglo-American) interviewer. The results of the study empirically establish that culture significantly affects the efficacy of structured interviews during international user testing. The implications of this work for usability engineering are discussed.

## Keywords

culture and interviews, international usability evaluation, usability methods, cross-cultural HCI, human-computer interaction, global user interface design,

## Introduction

Culture, as it is often defined in Human-Computer Interaction, refers to the common values, attitudes, and behavioral patterns shared by a group of people. These common characteristics influence how people behave when interacting with computers. Thus, a product design intended for international audiences must consider the values, attitudes, and behavioral patterns that different cultures will bring to the product.

The influences of culture can be seen in products by the choice of language used, the use of colors, the use of symbols (for example, currency), and so on. Culture also influences higher level design issues, as discussed by Marcus and Gould (2000). To address all these cultural issues, revised development processes have been proposed, such as globalization and localization (Luong et al. 1995; Taylor, 1992; Uren et al. 1993). However, the impact of culture also affects the software development process itself (Smith and Yetim, 2004). For example, Smith, Dunckley, French, Minocha and Chang (2004) suggest that user-centered design might be influenced by the Western culture's "view that users as individuals have a democratic right to be involved in the development of software" that they will use. The authors state that these assumptions cannot automatically be transferred to other cultures.

Cultural background influences the design methods employed in building interfaces, in particular, usability methods. The cultural influence, if ignored, can compromise usability evaluation, thus providing inaccurate information when a localized product is tested using these techniques. When the usability methods involve human-human interaction, as is the case with structured interviews, the interaction of the cultures of the two participants must be considered. One example where two cultures interact occurs with usability evaluation during the internationalization of products. When differences in cultures exist between the usability expert and the local users, usability methods may mask the usability problems instead of discovering them. In particular, the use of structured interviews where the interviewer is from a different culture than the participants is strongly influenced by cultural factors, as we show in this paper.

In a recent article, Aaron Marcus (2006) has expressed how little we know or understand culture as it relates to user-interface usability and design. He uses several examples to highlight the importance of culture and points to recent developments as evidence that culture does matter for user interface design and usability. Our work is an echo of this point of view.

In this paper, we present a study designed to evaluate the cultural effect on structured interviews while evaluating the usability of a website. Two groups of participants, all graduate students from India, evaluated a website. One group had an interviewer from India; the other group had an interviewer from the US. The group with the Indian interviewer found more errors, provided more feedback about the website, and identified more culturally sensitive materials than the group with the US interviewer.

The next section discusses some of the literature on culture, cultural metrics, and how they were used in this study, as well as the literature of culture in human-computer interaction (HCI). This section is followed by the experimental design for the study, and a

presentation of the findings. The paper finishes with the implications of the results of this study for usability engineering.

## Related Work

*Culture*

Culture has been defined in many different ways by different researchers. A compiled list of over 200 different definitions of culture can be found in Kroeber and Kluckhohn's book (Kroeber & Kluckhohn, 1952). In a recent editorial to the *Special Issue of Interacting with Computers* titled 'Global human-computer systems: cultural determinants of usability', Smith and Yetin (2004) state that researchers often refer to culture as "patterns of values, attitudes, and behaviors, which are shared by two or more people". In our work, we use Geert Hofstede's definition of culture: "Culture is the collective programming of the mind which distinguishes the members of one group or category of people from another" (Hofstede, 1997). Hofstede's cultural model is well suited for empirical research because a score for each individual member of the culture can be computed unambiguously. Many other cultural models exist in the literature, but most of these models are typologies, which are problematic in empirical research because individuals rarely fall into an ideal type.

*Hofstede's Cultural Model*

Hofstede's cultural model consists of five dimensions. Each dimension groups together phenomena in a society that were empirically found to occur in combination. Hofstede's seminal work on cultures in organizations formulated a framework of four dimensions of culture identified across nations. Michael Harris Bond (Hofstede & Bond, 1984; Hofstede & Bond,

1988) added the fifth dimension. The five dimensions in Hofstede's model are power-distance, collectivism-individualism, uncertainty avoidance, femininity-masculinity, and long/short-term orientation. Hofstede's model allows for independent evaluation along each of the five cultural dimensions. In our work, we used power distance because of its potential effect in structured interview use for the Indian population (see below). Although the interactional effects from the other dimensions cannot be absolutely ruled out, our controlled study was carefully designed to account for the dominant effect of the power distance dimension in the situation of structured interviews.

Hofstede defines power distance as "the extent to which the less powerful members of institutions and organizations within a country expect and accept the power that is distributed unequally" (Hofstede, 1997, p. 28). People in large power distance cultures are much more comfortable with a larger status differential than small power distance cultures. In large power distance cultures, subordinates display considerable dependence on bosses. In small power distance cultures, subordinates display limited dependence on bosses, and a preference for  mutual consultation and teamwork.

Huo and Randall (1991) have provided some criticism of Hofstede's dimensional model of culture. Hofstede formulated the dimensions from statistical analysis of the answers to questions provided by IBM employees in over 50 countries. It has been suggested that Hofstede's sample from a single multinational company might introduce some effects of organizational culture. However, there is ample empirical evidence to prove that Hofstede's dimensions are relevant and valid.

Many studies have revalidated Hofstede's dimensions (Bochner, 1994; Sondergaard, 1994; Robertson & Hoffman, 2000). For example, Sondergaard's paper contains a review of over 60 replications of Hofstede's cultural model and concludes that Hofstede's dimensions are "largely confirmed" (Sondergaard, 1994).

*Culture and Usability*
Prior research has found that culture affects the usability evaluation process and several of its well-known techniques (Beu, Honold, and Yuan, 2000; Day & Evers, 1997; Day & Evers, 1999; Evers, 1998; Evers, Kukulska-Hulme, and Jones, 1999; Honold, 2000; Yeo, 1998; Sears, Jacko, and Dubach, 2000; Yeo, 2001). Culture affects the functioning of focus groups (Beu et al., 2000), the think-aloud protocol (Yeo, 2001), questionnaires (Day & Evers, 1999), and the understanding of metaphors and interface design (Evers, 1998; Evers et al., 1999). Tractinsky (1997) found that culture affects the users' perception of aesthetics and apparent usability. However, the effects of culture in structured interviews when the interviewer and the participants are from different cultures has, until now, remained unexplored.

Marcus and Gould (Marcus & Gould, 2000) applied Hofstede's cultural dimensions to web and user-interface design. The authors mention each of Hofstede's five cultural dimensions and the aspects of design that can be influenced by that particular dimension. They present screen shots of different web sites developed in different nations and point out the cultural influences on design. The findings amplify the cultural differences, but lack empirical evidence.

Honold examined the notion of culture and its relevance to Human-Computer Interaction(HCI) and discusses the theories of culture in HCI (Honold, 2000). Honold found cultural influences when a washing machine developed in Germany was used in India. Honold identifies eight cultural factors that must be considered in any investigation of the context in which the product is used:

- objectives of the users
- characteristics of the users
- environment
- infrastructure
- division of labor
- organization of work
- mental modes based on previous experience and tools

Nielsen recommends traveling to the target country and conducting usability tests as the best choice in international usability testing. Another alternative suggested by Nielsen is to employ local staff to conduct the usability testing (Nielsen, 1996). Based on the results of our work, this might be required if culturally sensitive comments are sought. However, in general, usability assessment techniques have not been carefully studied in a cross-cultural context to evaluate cultural effects on their use.

*Interviews as a Usability Evaluation Method*
In Rubin's model (Rubin, 1994), interviews are used in the development stage and in the evaluation stage of usability testing. Interviews are used in the beginning of the development stage to design the questionnaire. They are used in the last stage of the evaluation to

clarify user responses and to collect additional information. Interviews are of two types: Structured and Open-ended (USINACTS Usability, 2004). Structured interviews have a predefined set of questions. A structured approach usually provides more reliable and quantifiable data than an open-ended interview and can be designed rigorously to avoid biases in the line of questioning.

This research explores the efficacy of structured interviews in the last stage of the usability evaluation in a cross-cultural context.

## Method

We conducted a two-phase experiment to explore the effects of culture in structured interviews when international usability evaluation involves participants from one culture and interviewers from the same and from different cultures.

Our hypothesis was that that having an interviewer of the same culture might mediate the effect of power distance. The cultural common ground between people of the same culture (interviewer and participant) will help effective communication and in identifying culturally related usability problems and/or design issues.

*Measurements of Culture in Cross-Cultural Research*
Any research dealing with culture must carefully measure culture in order to study its effect. In this study, we measured the power-distance for all Indian participants using the Early/Erez power differential scale (Earley & Erez 1997). The goal of using such a metric in our study was to show there was no

difference between the two experimental groups (of Indian participants) with respect to power-distance.

A second concern in cross-cultural research is the issue of acculturation. Acculturation is a process that occurs when two or more cultures interact. Acculturation occurs as the dominant host culture absorbs, to a certain extent, the minority immigrant culture (Suinn, Ahuna, and Khoo, 1992). In cross-cultural research, the user's perception of his/her identity is important because it is a subjective statement of cultural character. Individuals from the minority immigrant culture with high acculturation may behave like the individuals from the dominant host culture. This behavior becomes an external variable in cross-cultural research. We control for effects from this external variable by measuring the acculturation level of the participants belonging to the minority immigrant culture (Triandis, Kashima, Shimada, and Villareal, 1986). Participants with high level of acculturation can best be used as members of the dominant host culture, or not included in the study (Triandis, et al., 1986).

In this research, we used the Suinn-Lew Asian Self Identity Acculturation (SL-ASIA) scale (Suinn et al., 1992) to measure the acculturation levels of the Indian participants. We chose this scale because it was specifically designed for Asians. The SL-ASIA is a multiple-choice test with 21 questions and five choices per question. A score is obtained by adding across the answers for all 21 questions and dividing the total value by 21. The scores range from 1.00 (Low Acculturation) to 5.00 (High Acculturation).

The independent variable of the experiment was the cultural profile of the interviewer. The dependent

variables in the experiment were number of usability problems found, suggestions made, positive comments made, negative comments made, culturally related comments made, and the rating given to the interface.

The hypothesis for the experiment can be summarized as follows. We expected that participants with an Indian interviewer would provide more comments, identify more usability problems, make more negative comments, and make more culturally sensitive comments than those with an Anglo-American interviewer. Similarly, we expected that participants would make more positive comments to the Anglo-American participant. In general, we expected participants to give a higher rating of the interface to the Anglo-American interviewer than to the Indian interviewer.

Phase two consisted of the usability evaluation of a website followed by a structured interview. All the interviews were recorded with both audio and video. We selected 16 participants from phase one based on the SL-ASIA acculturation scores. The participants were divided into two groups with each group having eight participants: seven male and one female. Participants were assigned to the two groups in such a manner that the two groups matched with respect to acculturation. One group had an Indian interviewer and the other an Anglo-American interviewer. See the results section for details on participant demographics, their power distance, and acculturation scores.

The interviewers were not affiliated with the study; they were recruited from the Virginia Tech student population. The primary consideration in selecting the interviewers was to minimize their familiarity with the participants. Both interviewers selected had prior usability experience as both evaluators and as participants. The authors did not pay the interviewers and the interviewers were not informed about the experimental design and hypothesis. The interviewers were instructed to treat the session as a regular usability evaluation. The interview guidelines (USINACTS Usability, 2004) were adhered to in conducting the interviews.

*Association for India's Development Website*
For the experiment, we used a website intended primarily for Indian students at Virginia Tech. The Association for India's Development (AID) is a voluntary, not-for-profit organization that supports a wide variety of social service and development projects, such as literacy, health care, vocational training, women's empowerment, and children's welfare.

For the experiment, we introduced usability problems in a local copy of the AID website. Using Nielsen's Ten Usability Heuristics (Nielsen, 1994), we introduced some problems into the website that matched the heuristics. The content of the AID web pages was not altered, modified, or enhanced. No new web pages were created, and none of the existing ones were removed.

For each of the five tasks listed in the procedure section, non-culturally specific usability problems were introduced. Introducing non-cultural usability problems allowed us to determine if the cultural background of the interviewer had an impact on finding culturally sensitive usability issues, as well as on finding non-cultural ones. We believe that culture strongly affects all the results obtained from a usability assessment

technique, because the effects are manifested in the communication between the participant and the interviewer. Furthermore, we introduced these problems to control for the floor and ceiling effects in the study. A short list of the non-culturally specific usability problems introduced is given below:

- Clicking on the About AID link leads to a page without headings or links.
- Organizational language and jargon were used more than necessary.
- Some pages do not have a 'Home' link, and other links are missing as well.
- Links are arranged inconsistently and standards are not followed.
- Some links, options, and objects are made less visible than the others.

Two culturally sensitive usability problems introduced. These consisted in changing the color background of two pages to culturally sensitive colors. We changed the background color of the "Home Page" from white to saffron (orange). We changed the background of the "Join Us" page to a black background, which is considered inauspicious in Indian culture.

## Procedure

In phase one, we collected demographic data, the power distance score as measured by the power differential scale (Earley & Erez 1997), and the acculturation score as measured by the SL-ASIA scale (Suinn, Ahuna, and Khoo, 1992). Based on the scores of these tests, and to balance the two experimental groups for acculturation effects, we selected 16 participants by removing participants who were outliers from the phase one sample of 25. The 16 participants

were asked to return for the evaluation of the AIDS website in phase two. We contacted these 16 participants using email and telephone correspondence.

We recruited the two interviewers, as indicated above. The interviewers were given a copy of the guidelines, and were asked to read them at least twice to understand and follow the instructions. The interviewer greeted the participant and welcomed the participant into the usability test room. Each interviewer was given a predetermined introduction script and was advised to use it. This was done to control for any interviewer-specific characteristics that might affect the participants' first impressions. The two introduction scripts were identical in all aspects, except for the three sentences marked in bold. The Indian interviewer leveraged the cultural background by referring to the home state and an Indian cultural event at Blacksburg, Virginia Tech. The introduction script is given below, with the three extra sentences used by the Indian interviewer marked in bold.

Good Evening. I am *<<name withheld>>*. How are you doing today? Which part of India are you from? **I am from Andhra Pradesh. Did you attend the Indian Film festival held recently? I watched three movies**. We are testing the design of the AID website and want to get feedback from the Indian students here in [Virginia Tech]. It's so nice that you are willing to participate. Shall we get started?

We gave the participants a written description of the five usability tasks. We asked them to read each task description carefully and then perform the task. We designed the five tasks to give participants experience using the site and to expose them to most of the errors

we introduced on the site. They performed the tasks on the redesigned AID website using a Windows 2000 computer and Internet Explorer 6.0. The five tasks are listed below:

- Task #1: Become a member of AID and find out the time and location of the next community service hour for AID.

- Task #2: Find out the information about the current executive committee members, and find out the web coordinator for the AID Virginia Tech (VT) chapter.

- Task #3: Learn about the project coordinated by *<<name withheld>>*. Note down his contact info.

- Task #4: Learn about the home schools project (SSGS) and contact the coordinator for the project.

- Task #5: Learn about the grocery certificates. Get involved in the program.

After completing all the tasks, the interviewer conducted the structured interview. The structured interview questions were centered on the five test tasks of the study. The interviewer took notes on the interview, and a meta-evaluator (one of the co-authors) took notes in the adjoining observation room of the usability lab. One of the co-authors made transcripts from the interviews using the audio-video recordings and by cross-checking the notes. We removed all pauses like "um" and "ahh". "Yeah" was interpreted as "yes".

We scored the transcripts by recruiting four graduate students who had taken at least the CS 5714 Usability Engineering course at Virginia Tech. None of the four coders were informed about the specific hypothesis of the study or the general premise of the research project. The Indian interviewer group transcripts were scored by two evaluators and the other two evaluators scored the Anglo-American group transcripts. All four coders were given the corresponding transcripts and asked to count the total number of replies, the usability problems identified, suggestion given, positive and negative comments, and any comments related to cultural issues in the interface. The following list provides definitions of each of these categories:

- **Usability Problem:** interaction design flaw or a user difficulty that is directly associated with an interaction design flaw.
- **Suggestion:** subjective preference of the participant to the implemented design choice or tradeoff.
- **Positive comment:** participant's subjective approval of a design choice or tradeoff.
- **Negative comment:** participant's subjective disapproval of a design choice or tradeoff.
- **Cultural comment:** participant's reference to his/her native culture, country, customs, symbols, rituals, and tradition.

We opted for inter-rater consensus rather than a quantitative measure of reliability like Cronbach's Alpha. This was possible as we had only 16 relatively short transcripts to code, and we thought consensus rather than a quantitative measure of reliability is desired given the nature of our data. There was some confusion among the coders about how to count

negative comments because they could count toward both usability problems identified and negative comments. The rule applied was that any negative comment made about the interaction design and/or interface design element was also a usability problem found. Thus, there was some overlap between Negative Comments and Usability Problems found. Apart from this, all the scores are unique. Once this clarification was made, the scores from all the coders agreed in the numbers. The scoring variables and their relationship is given by the following formula:

Total Replies = Usability Problems (U) + Suggestions (S) + Negative Comments (N) + Positive Comments (P) + Cultural Comments (C) + Reply to the Website Rating- that are also Usability Problems Negative Comments.

## Data Analysis and Results
*Demographic & Cultural Metrics of Participants*
For the Indian interviewer group, the average age of the participants is 23 years. Of the eight participants, seven were male (87.5 %). Their average stay in US was 17.75 months, ranging from a high of 24 months to a low of 8 months. The average stay in India was 21.25 years. The participants belonged to four different states in India and spoke three different first languages. The power distance score of the participants averaged 19.375, ranging from a lowest score of 14 and a highest score of 23 (standard deviation was 3.66). Acculturation was low, with an average score of 2.08. The lowest acculturation score was 1.80 and the highest was 2.47, with a 0.2 standard deviation.

For the Anglo-American interviewer group, the average age of the participants was 24 years. Of the eight

participants, seven were male (87.5 %). Their average stay in US was 18.25 months, ranging from a high of 24 months to a low of eight months. The average stay in India was 22.75 years. The participants belonged to four different states in India, and spoke five different first languages. The power distance score of the participants averaged 19.375, ranging from a lowest score of 11 to a highest score of 27 (standard deviation of 6.39). Acculturation was low, with an average score of 2.19. The lowest acculturation score was 1.71 and the highest was 2.57, with 0.28 in standard deviation.

None of the participants in either of the groups had the maximum possible power distance score of 40 or the minimum possible score of 5. Furthermore, none of the participants in either treatment group could be classified as bicultural or Anglo-American acculturated, according to the rules of interpretation of scores given in the SL-ASIA scale (Suinn, Ahuna, and Khoo, 1992). The low acculturation scores mean that the influence of the majority host culture (USA) is not significant, and the participants are reasonable representatives of the Indian culture. An ANOVA analysis with respect to power distance scores found no significant difference between the two groups. Similarly, an ANOVA analysis of acculturation scores found no significant difference between the two groups.

*Quantitative Results*
The dependent variables produced statistically significant results for all the analyses done. The results are shown in Table 1. The participants found more usability problems and made more suggestions with the Indian interviewer than with the Anglo-American interviewer. More positive comments and fewer negative comments were made by the participants to

the Anglo-American interviewer, possibly leading to a false picture of subjective preferences. More importantly, the participants were reluctant to make culturally related comments to the Anglo-American interviewer. The whole purpose of finding culture-related data from the structured interviews can be lost if an interviewer from a foreign culture is used. On the other hand, with the interviewer from the same culture, participants were more forthcoming.

*Qualitative Results*
We also report here some observations based on the comments provided by the participants in the structured interviews. We looked at the comments that were identified as culturally sensitive, to identify how these comments differed with the two groups. We found that the comments to the Indian interviewer were more candid in expressing the problem being identified. The comments on the same questions to the Anglo-American were more general. It was evident that the Indian participants shared more common ground with the Indian interviewer and that was markedly affecting the culture-sensitive comments made by participants related to color, identity, and nationality.

*Colors*
We found interesting the expressiveness of the comments made that were related to color. For example, question 2 asked the participants "What do you think about the colors used?" The responses were markedly influenced by the interviewer. Remember that we changed the background color of the home page to use saffron, one of the colors in the Indian flag. Some comments to this question from participants with the Indian interviewer include the following:

- #2: Indian flag colors or something like that are good.
- #3: Ok. Since I am an Indian, I know that the saffron is a color of the flag.
- #4: I am not sure if they used saffron for the Indian flag or as a Virginia Tech color.

The responses to the Anglo-American interviewer were more generic, and certainly lacked references to the cultural meaning of the colors. For example:

- #1: ... But orange color is not good.
- #4: Colors are good.
- #5: Orange color's purpose is not evident.
- #6: Saffron and white are OK
- #8: Colors are looking good.

Notice how the "feelings" about the saffron color come through in both groups, but only in a culturally neutral way with the Anglo-American interviewer. In this group, notice that the comments do not identify why the color might or might not be appropriate. Even the name of the color used is different; "saffron" was more common in the Indian-interviewer group, while "orange" the type of responses they provided. In general, the comments made by the participants with the Indian interviewer show a clear connection to the color of the Indian flag.

*Identity*
Question 18 asked about the appeal of the AID site design to the Indian students at the university. In general, we noticed that the participants used "we" to refer to the cultural group when they had an Indian interviewer. One participant referred to the collective design of the site by indicating "…we want it to be

better than any site". It is interesting to contrast this statement with a comment made by one participant to the Anglo-American. The participant referred to the designers of the site in third person ("They should focus more on Indian students").

It is interesting to point out that the AID site is designed and maintained by Indian members of the community. Yet, having an interviewer of a different culture conduct the usability presumes that the developers of the site might be from a different culture. This particular comment shows certain disconnect between the participants and the developers of the site.

| Usability Problems(U) | n | Mean | SD | SE | Source of variation | SSq | DF | MSq | F | p | Contrast Difference | Bonferroni 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Anglo-American Interviewer Group | 8 | 5.000 | 1.512 | 0.5345 | Interviewer | 60.063 | 1 | 60.063 | 19.61 | 0.0006 | -3.875 | -5.752 to -1.998 |
| Indian Interviewer Group | 8 | 8.875 | 1.959 | 0.6928 | Within cells | 42.875 | 14 | 3.063 | | | | |
| **Suggestions(S)** | n | Mean | SD | SE | Source of variation | SSq | DF | MSq | F | p | Contrast Difference | Bonferroni 95% CI |
| Anglo-American Interviewer Group | 8 | 5.500 | 1.069 | 0.3780 | Interviewer | 14.063 | 1 | 14.063 | 5.20 | 0.0388 | -1.875 | -3.639 to -0.111 |
| Indian Interviewer Group | 8 | 7.375 | 2.066 | 0.7304 | Within cells | 37.875 | 14 | 2.705 | | | | |
| **Positive Comments(P)** | n | Mean | SD | SE | Source of variation | SSq | DF | MSq | F | p | Contrast Difference | Bonferroni 95% CI |
| Anglo-American Interviewer Group | 8 | 6.875 | 2.588 | 0.9149 | Interviewer | 25.000 | 1 | 25.000 | 4.68 | 0.0483 | 2.500 | 0.022 to 4.978 |
| Indian Interviewer Group | 8 | 4.375 | 1.996 | 0.7055 | Within cells | 74.750 | 14 | 5.339 | | | | |
| **Negative Comments(N)** | n | Mean | SD | SE | Source of variation | SSq | DF | MSq | F | p | Contrast Difference | Bonferroni 95% CI |
| Anglo-American Interviewer Group | 8 | 3.875 | 2.167 | 0.7662 | Interviewer | 156.250 | 1 | 156.250 | 17.68 | 0.0009 | 6.250 | 3.062 to 9.438 |
| Indian Interviewer Group | 8 | 10.125 | 3.603 | 1.2739 | Within cells | 123.750 | 14 | 8.839 | | | | |
| **Cultural Comments(C)** | n | Mean | SD | SE | Source of variation | SSq | DF | MSq | F | p | Contrast Difference | Bonferroni 95% CI |
| Anglo-American Interviewer Group | 8 | 0.500 | 0.756 | 0.2673 | Interviewer | 25.000 | 1 | 25.000 | 7.95 | 0.0136 | 2.500 | 0.599 to 4.401 |
| Indian Interviewer Group | 8 | 3.000 | 2.390 | 0.8452 | Within cells | 44.000 | 14 | 3.143 | | | | |
| **Website Rating(R)** | n | Mean | SD | SE | Source of variation | SSq | DF | MSq | F | p | Contrast Difference | Bonferroni 95% CI |
| Anglo-American Interviewer Group | 8 | 3.063 | 0.496 | 0.1752 | Interviewer | 3.516 | 1 | 3.516 | 7.46 | 0.0162 | -0.938 | -1.673 to -0.202 |
| Indian Interviewer Group | 8 | 2.125 | 0.835 | 0.2950 | Within cells | 6.594 | 14 | 0.471 | | | | |

**Table 1: ANOVA results from the structured interview data.**

*Nationality*

One final set of comments is worth pointing out. The original site of the AID organization has a logo that includes a drawing of Mahatma Gandhi. No participants with the Anglo-American interviewer made a reference to this image. Two participants in the other group made references to this image ("Father of Nation's image is good" and "I like Gandhi's image"). It is this type of culturally specific comments that international usability evaluation is intended to obtain. Yet, by ignoring the influence of culture in the usability evaluation methods, usability engineers are bound to miss identifying these culture-specific usability issues.

## Discussion

The results shown in this paper have a serious impact on the process of developing international user interfaces. In particular, the results have implications to international usability evaluation with users from one culture and interviewers from a different culture. The results empirically establish that culture affects the type of responses participants provided in a structured interview. Participants responded more freely and accurately to the interviewer from the same culture than to the interviewer from a different culture.
The findings are one more example of the "evaluator effect" that has been reported in the literature. However, it is not, we believe, an example of the same type of "chilling effect" described by (Hertzum & Jacobsen, 2001) because, in our study, the evaluators did not assess the problems found, nor did they make any judgment decisions about what was or was not a usability error.

We found a strong effect caused by the common ground between the Indian participants and the AID website. Based on a qualitative assessment of the structured interview data, in our opinion, the Indian participants felt a cultural connection, an affinity, to the site and the goals of the site. This has, we believe, a strong impact on the results of the evaluation. However, the findings are not limited to the culturally sensitive comments alone. The number of usability problems reported by the participants to their interviewer was affected, as were the numbers of positive and negative comments. Therefore, the impact goes beyond identifying culturally sensitive usability problems. We can only speculate as to what would happen if we replicate the study using a culturally neutral site. Would the impact of the difference in cultures between interviewer and participants have such a marked effect? More research is needed to explore these issues.

In our particular study, the users evaluating the interface were from a large-power distance culture. We believe that this might have had a lot to do with the results obtained. However, it is not clear if the results are more influenced by the difference in culture between the participants and the interviewer, or the fact that the participants were from a large-power distance culture. It is difficult, however, to evaluate all possible combinations of two cultures to effectively determine which factor might have had a stronger impact. A question that should be asked next is: what happens if we repeat the experiment with Anglo-American users and with two interviewers from two different cultures (for example, Indian and Anglo-American). Would the results be the same? From a research point of view, it might be worth considering the question and its possible implications. Also, the Indian interviewer's referral to the Indian cultural

events in the structured interview introduction script might have primed the participants. While this does not confound the results, it can be investigated further to determine the effect of pre-interview banter in cross-cultural usability evaluation settings.

A more relevant question is the effect that differences in cultures might have in other usability evaluation techniques. We have shown one case where difference in culture affects the results of the evaluation. In our particular instance, the usability method, structured interview, depends heavily on human-human interaction. This gives the opportunity for social and cultural norms and practices to come to the front. However, would the same be true in other usability evaluation methods? For example, what impact might culture have in a remote usability (Hartson, Castillo, Kelso, Kamler & Neale, 1996) evaluation method? In that method, the interaction between the evaluator and the users is minimized or completely removed. It is possible that such a method might not be influenced as strongly by social and cultural factors. Further research is needed.

Finally, the research methodology followed in our work can serve as an example of how culture and acculturation can be considered in HCI research. Successful integration of Hofstede's cultural model and acculturation models with usability engineering will result in exciting and useful results in cross-cultural HCI research.

## Practioner's Take away

- Cultural mismatch between the interviewer and the users affects the usability assessment technique of structured interviews.
- Interviewers from the same culture might be more effective in eliciting usability problems in when users come from hierarchical cultures.
- Hofstede's cultural dimensions model can be used to inform the selection of usability assessment techniques cross-cultural user testing.
- Culture might influence the efficacy of a usability method that involves high degree of social interaction.
- Considering culture in the usability evaluation process will lead to better informed products and services.

## Acknowledgements

## References
Beu, A., Honold, P. & Yuan, X. (2000). How to Build Up an Infrastructure for Intercultural Usability Engineering. The International Journal of Human-Computer Interaction, 12 (3&4), 347-358.

Bochner, S. (1994). Cross-Cultural Differences in the Self Concept: A Test of Hofstede's Individualism/Collectivism Distinction. Journal of Cross-Cultural Psychology, 25(2), 273-283.

Day, D. & Evers, V. (1997). The Role of Culture in Interface Acceptance. Proceedings of Human-Computer Interaction, INTERACT '97, IFIP TC13 International Conference on Human-Computer Interaction, pp. 260-267.

Day, V. & Evers, V. (1999). Questionnaire Development for Multicultural Data Collection. In E. del Galdo & G. Prahbu (Eds.), Proceedings of the International Workshop on Internationalization of Products and Systems, Rochester, 20-22 May 1999.

Earley, P.C. & Erez, M. (1997). The Transplanted Executive: Why you need to understand how workers in other countries see the world differently. New York: Oxford University Press.

Evers, V. (1998). Cross-cultural Understanding of Metaphors in Interface Design. In C. Ess & F. Sudweeks (Eds.), Attitudes toward Technology and Communication , London, 1-3 Aug. 1998.

Evers, V., Kukulska-Hulme, A. & Jones, A. (1999). Cross-Cultural Understanding of Interface Design: A Cross-Cultural Analysis of Icon Recognition. In E. del Galdo and G. Prahbu (Eds.), Proceedings of the International Workshop on Internationalization of Products and Systems, Rochester, 20-22 May 1999.

Fernandes T. (1995). Global Interface Design. Academic Press.

Hartson, H.R., Castillo, J.C., Kelso, J., Kamler, J. & Neale, W.C. (1996). Remote evaluation: the network as an extension of the usability laboratory. Proceedings of the Conference on Human Factors in Computing Systems, pp. 228-235.

Hertzum, M. & Jacobsen, N.E. (2001). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. International Journal of Human-Computer Interactions, 13(4), pp. 421-443.

Hofstede, G. & Bond, M.H. (1984). Hofstede's Cultural Dimensions: An Independent Validation Using Rokeach's Value Survey. Journal of Cross-Cultural Psychology, 15(4), pp. 417-433.

Hofstede, G. & Bond, M.H. (1988). The Confucius Connection: From Cultural Roots to Economic Growth. Organizational Dynamics, 16(4), pp. 4-21.

Hofstede, G. (1997). Cultures and Organizations: Software of the Mind, Intercultural Cooperation and its Importance for Survival. New York: McGraw-Hill.

Honold, P. (2000). Culture and Context: An Empirical Study for the Development of a Framework for the Elicitation of Cultural Influence in Product Usage. International Journal of Human-Computer Interaction, 12(3 & 4), pp. 327-345.

Huo, Y.R. & Randall, D. (1991). Exploring Subcultural Differences in Hofstede's Value Survey: The Case of the Chinese. Asia Pacific Journal of Management, 8(2), pp. 159-173.

Kroeber, A.L. & Kluckhohn, C. (1952). Culture: A Critical Review of Concepts and Definitions. Harvard University Peabody Museum of American Archeology and Ethnology, 47.

Luong, T., Lok, J., Lok, S., & Driscoll, K. (1995). Internationalization: Developing Software for Global Markets. New York: John Wiley & Sons.

Marcus, A. & Gould, E.M. (2000, July/August). Cultural Dimensions and Global Web User-Interface Design. Interactions, pp. 32-46.

Marcus, A. (2006). Culture: Wanted? Alive or Dead? Journal of Usability Studies, 1(2), pp. 62-63.

Nielsen, J. (1994). Heuristic Evaluation. In Nielsen, J., & Mack, R.L. (Eds.), Usability Inspection Methods, New York: John Wiley & Sons.

Nielsen, J. (1996). International Usability Engineering. In Del Galdo, E. and Nielsen, J. (Eds.), International User Interfaces, New York: John Wiley & Sons.

Robertson, C.J. & Hoffman, J.J. (2000). How Different Are We? An Investigation of Conucian Values in the US. Journal of Managerial Issues, 12(1), pp. 34-47.

Rubin, J. (1994). Handbook of Usability Testing. New York: John Wiley & Sons.

Smith, A. & Yetim, F. (2004) Editorial: Global human-computer systems: cultural determinants of usability. Interacting with Computers, 16, pp. 1-5.

Smith, A., Dunckley, L., French, T., Minocha, S., & Chang, Y. (2004) Process model for developing usable cross-cultural websites. Interacting with Computers, 16, pp. 63-91.

Sondergaard, M. (1994). Hofstede's Consequences: A Study of Reviews, Citations and Replications. Organization Studies, 15(3), pp. 447-456.

Suinn, R.M., Ahuna, C. & Khoo, G. (1992). The Suinn-Lew Asian Self-Identity Acculturation Scale: Concurrent and Factorial Validation. Educational & Psychological Measurement 52(4), pp. 1041-1046.

Taylor, D. (1992). Global Software: Developing Applications for the International Market. New York: Springer-Verlag.

Tractinsky, N. (1997). Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. Proceedings of the Conference on Human Factors in Computing Systems, pp. 115-122.

Triandis, H.C., Kashima, Y., Shimada, E., & Villareal, M. (1986). Acculturation Indices as a Means of Confirming Cultural Differences. International Journal of Psychology, 21, pp. 43-70.

Uren, E., Howard, R., & Preinotti, T. (1993). Software Internationalization and Localization: An Introduction. Van Nostrand Reinhold.

USINACTS Usability Tutorial Internet. Retrieved 20 June 20 2004, from http://www.hft.org/HFT99/paper99/Poster/30_99.pdf

Yeo, W.A. (1998). Cultural Effects in Usability Assessment. Proceedings of the Conference on Human Factors in Computing Systems, pp. 74-75.

Yeo, W.A. (2001). Global-Software Development Life Cycle: An Exploratory Study. Proceedings of the Conference on Human Factors in Computing Systems, pp. 104-111.

Ravi Vatrapu is a PhD student in Communication and Information Sciences at the University of Hawai'i at Manoa. He works at the Laboratory for Interactive Learning Sciences (LILT). His research interests are in multicultural online learning, computer-supported collaborative learning, cultural studies, and consciousness studies. He is generally considered to be a jolly good fellow.



Manuel A. Pérez-Quiñones is an Associate Professor in Computer Science and a member of the Center for Human Computer Interaction at Virginia Tech. His research interests are in personal information management, multi-platform user interfaces, uses of computers in education, and cultural issues in interface design. He is a Multicutural Fellow of the Office of Multicultural Affairs at Virgina Tech.