

Cultural Usability in Computer Supported Collaboration

Ravi Vatrapu

Hawai'i Computer-Human Interaction Laboratory (hi'CHI)
Department of Information and Computer Sciences
University of Hawai'i at Mānoa
POST 317, 1680 East-West RD, Honolulu, HI, 96822 USA.
vatrapu@Hawaii.edu

ABSTRACT

This paper presents an empirical analysis of usability in the emerging field of computer supported intercultural collaboration (CSIC). The basic premise of the research project on which the empirical analysis is based is that social affordances of technologies vary across cultures. To empirically evaluate the premise, an experimental study was conducted to investigate how pairs of participants from similar and different cultures (American-American, American-Chinese, and Chinese-Chinese) appropriate affordances and relate to each other in a computer supported collaborative learning environment. Usability analysis consisted of evaluating objective performance measures and subjective user interface satisfaction measures. Statistical results show a systemic variation between cultures. We discuss the implications of these findings for the research and practice of cultural usability.

Categories and Subject Descriptors

H.5.3 Group and Organization Interfaces: *Theory and models, Asynchronous interaction Collaborative computing, Evaluation/methodology*; H.1.2 User/Machine Systems: *Software Psychology*.

General Terms

Affordances, appropriation, culture, cultural usability, technological intersubjectivity.

1. INTRODUCTION

Cultural aspects of usability have been a topic of study in the field of human-computer interaction (HCI). Early research focused on localization and internationalization user interface aesthetic issues of languages; colors; and conventions of data, time and currency [15, 23, 30]. Subsequent research investigated cultural considerations in the usability evaluation methods employed and usability problems detected. Prior empirical work has documented cultural differences in the *functioning of focus groups* [1], *think-aloud protocol* [7, 49], *questionnaires* [11], *understanding of metaphors and interface design* [11, 13], *non-verbal cues* [47], *web design* [24], *objective and subjective*

measures of usability [19], and *structured interviews* [42]. Currently, the CULTUSAB project [8] is conducting a systematic empirical research program of investigating cultural aspects of the practice of usability in general and usability evaluation methods in particular [e.g., 6, 7-10].

1.1 Culture, Collaboration, and Usability

Existing research in cultural usability research has largely focused on aesthetic issues, methodological aspects, and practitioner concerns. There is little empirical research on the cultural aspects of usability in computer supported collaboration environments. In this paper, we present usability analysis of an experimental study of intra- and inter- cultural computer supported collaborative learning of Chinese and Anglo-American participants.

1.2 Computer Supported Intercultural Collaboration

Computer supported intercultural collaboration (CSIC) is an emerging field of study centrally concerned with the iterative design, development, and evaluation of technologies that enhance and enrich effective intercultural communication and collaboration. There are two interrelated aspects of interaction design in developing CSIC systems: (i) interacting with computers and (ii) interacting with other persons. Both these aspects of interaction can be influenced strongly by culture, given the strong empirical evidence documenting cultural differences in *cognition* [26], *communication* [17], *behavior* [21], and *interacting with computers* [43]. In line with that research program articulated in [41], the research project discussed here originally focused on the cultural influence on (i) how participants from similar and different cultural backgrounds appropriate affordances in a CSIC environment [40, 41] and (ii) how participants relate to each other during and after computer supported collaborative interaction [40, 41, 44]. This paper focuses on usability aspects of the research project. The analytical objective is to investigate cultural aspects of usability in computer supported collaboration in general. Before discussing the methodological aspects of the research project, the key definitions of socio-technical affordance, appropriation of affordance, and technological intersubjectivity are provided below.

1.3 Definition of Socio-Technical Affordance

In computer-supported collaboration, each actor is both a user of the system as well as a resource for the other users. *Technology affordances* are action taking possibilities and meaning making opportunities in a user-technology system. Similarly, *social affordances* are action taking possibilities and meaning making opportunities in a social system. In socio-technical systems that

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '04, Month 1–2, 2004, City, State, Country.
Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

facilitate collaboration, these amalgamate into *socio-technical affordances*. There has been little work on the concept of socio-technical affordances. Drawing upon foundational work in ecological psychology on the formal definition of affordances [34, 39], the following definition is offered for *socio-technical affordance*.

Let $Wpqr$ (e.g., person-sending-email-to-another-person system) = (Tp, Sq, Or) be composed of different things T (e.g., email technology); S (e.g., email sender) and O (e.g., email receiver). Let p be a property of T ; q be a property of S and r be a property of O . The relation between p , q and r , $p/q/r$, defines a higher order property (i.e., a property of the socio-technical system), α . Then α is said to be a socio-technical affordance of $Wpqr$ if and only if:

- (i) $Wpqr = (Tp, Sq, Or)$ possesses α
- (ii) Neither $T, S, O, (T, S), (T, O), (S, O)$ possesses α

The formal definition of socio-technical affordance provided above reflects the duality of individuals' perception with respect to the technology as well as other persons. The formal definition informs the design of experimental studies of computer supported intra- and inter-cultural collaboration. Systematic variation of each of the three elements—technology (T), self (S), and other (O)—generates studies of the appropriation of affordances as higher-order relational properties.

1.4 Appropriation of Affordances

The concept of affordance simultaneously specifies the two concurrent levels of meaning and action. Although the perception of objects and other persons can be accounted by a Gibsonian ecological approach, the perception of events cannot be accounted on strictly ecological grounds [33]. Gibson's rejection of a role for higher order cognitive processes in perception-action is problematic, as the perception of events has interactional consequences in computer-supported collaboration. Interactions in socio-technical environments are a dynamic interplay between ecological information as embodied in artifacts and individual actions grounded in cultural schemas. The essential mediation of all interaction is the central insight of socio-cultural theories of the mind [46]. The conceptualization of interaction as being mutually "accountable" (observable and reportable) is the critical insight of ethnomethodology [16]. Following these two schools of thought, interactions in socio-technical systems are conceptualized as appropriation of socio-technical affordances. Adapting Stoffregen's discussion of behavior [34, p.125], appropriation is defined as "*what happens at the conjunction of complementary affordances and intentions or goals*". Research into social aspects of HCI [28] has shown that even computer-literate users tend to use social rules and display social behavior in routine interactions with computers. Social interaction is grounded strongly in culture as every person carries within patterns of thinking, feeling, behaving and potential interacting. Thus, participants in computer supported collaboration make culturally appropriate and sensitive choices and decisions in their actual appropriation of affordances.

1.5 Technological Intersubjectivity

Information and Commutation Technologies (ICT) and the Internet have changed our social relations with others and objects in fundamental ways that sometimes transcend their very technology mediation. Our interactions with others and objects

are increasingly informed by the operational logic of technology, hence technological intersubjectivity. For example, technology lets us assign distinct ring tones, images, priorities to our significant others. Our psychological perception and phenomenal relation with social others is being increasingly transformed by the advances in information and communication technologies and social software. Human beings are not only functional communicators but also hermeneutic actors. In technological intersubjectivity, technological mediation can sometimes (but not necessarily always) disappear like in Clarke's third law of technology [5]. Technological intersubjectivity is an emergent resulting from psychological-phenomenological nexus of the electronic self-other social relationship.

1.5.1 Definition of Technological Intersubjectivity
Technological intersubjectivity (TI) refers to a technology supported interactional relationship between two or more actors. TI emerges from a dynamic interplay between the technological relationship of actors with artifacts and their social relationship with other actors.

From a functional perspective, psychological intersubjectivity doesn't require two or more persons to have the same or similar subjective experience. Put differently, having a collective phenomenal experience is not a necessary condition for psychological intersubjectivity. In psychological intersubjectivity, the other human being is always an object of our attention and an object in our awareness. We observe the other person for communicative cues and informational structures relevant to the ongoing interaction. Unlike in phenomenological intersubjectivity there is no requirement for an emphatic relationship with the other person. However, in the emergent technological case, there is a dynamic interplay between these psychological and phenomenological aspects. In technological intersubjectivity, information processing entailed by computational support can enhance and enrich the communicative possibilities and communion potentials of two or more human beings. Socio-technical systems and online communities have potentials for both psychological and phenomenological intersubjective experiences without the requirement that interacting persons be co-present in the same place and interact at the same time.

2. METHODOLOGY

The formal definition of socio-technical affordance α in $Wpqr = (Tp, Sq, Or)$ has two important elements: technology T and individual actors S, O [40, 41, 44]. The definition of appropriation of affordances has two important elements: affordances and intentions [41]. Based on these two definitions, an experimental study was designed that introduced a variation in the cultural background of individuals (by selecting participants from a nation-state based ethnically stratified random sampling frame) but kept invariant the technological interface T and interactional setting. Briefly, the experimental study investigated how pairs of participants from similar and different cultures (American-American, American-Chinese, and Chinese-Chinese) appropriated affordances in a quasi-asynchronous computer supported collaborative learning environment with external representations in order to collaboratively solve a public health science problem. Usability analysis reported in this paper was conducted on the empirical data generated by this experimental study.

2.1 Research Questions

Four separate lines of empirical research have demonstrated that culture influences: (1) *social behavior*, [21], (2) *communication* [17], (3) *cognitive processes* [26], and (4) *interacting with computers* [43]. These four lines of empirical research were integrated into a methodological framework (Figure 1).

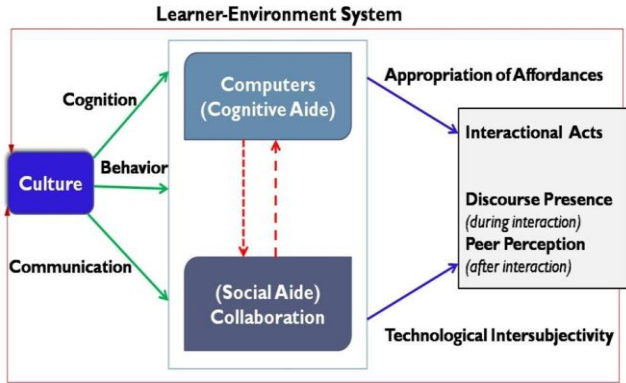


Figure 1: Methodological Framework

The primary purpose of the study was to answer two basic research questions [see 40, 41]. The first research question asked “to what extent does culture influence the appropriation of socio-technical affordances?” The second research question asked “to what extent does culture influence technological intersubjectivity?”

An experimental study was originally designed and conducted to answer these two research questions. Several theoretical predictions were generated from prior empirical evidence [17, 21, 26] warranting the claim that both the perception and appropriation of affordances might vary across cultures and that interpersonal perceptions and relations will vary across cultures [see 40, 41]. As mentioned earlier, this paper presents usability analysis of the empirical data generated by the experimental study. Specifically this paper seeks to answer the following two research questions on cultural usability in computer supported collaboration settings:

1. To what extent does culture influence *objective usability measures of usability* in computer supported collaboration?
2. To what extent does culture influence *subjective usability measures of usability* in computer supported collaboration?

In the next five subsections experimental design, materials, research hypotheses, sampling and procedure are discussed briefly.

2.2 Experimental Design

The experimental study design consisted of three independent groups of dyads from similar and different cultures (Anglo-American, Chinese) doing collaborative problem solving in a knowledge mapping learning environment. The three experimental conditions were the Chinese-Chinese *intra*-cultural condition, the American-American, *intra*-cultural condition and the Chinese-American *inter*-cultural condition.

In all three experimental conditions, the collaborative dyads were given the same experimental task. All the collaborative

dyads interacted in the same computer supported collaborative learning environment after reading the same instructions, software tutorial and demonstration. The same instruments were administered to all participants. Internal validity and external validity were actively considered when designing and conducting the experiment. Construct validity was addressed by using existing instruments with high validity and reliability [2, 21, 31, 35]. Brief descriptions of the experimental study’s software and topics are provided next.

2.3 Materials

2.3.1 Software

The computer supported collaborative learning environment used in this experimental study has an “information viewer” on the left in which materials relevant to the problem are displayed. This information viewer functions as a simple web browser, but the presentation of materials is constrained as discussed in the next section. The environment has a shared workspace or “information organizer” on the right hand side in which participants can share and organize information they gather from the problem materials as well as their own interpretations and other ideas. The “discussion” tool below the “information viewer” on the left enables participants to discuss their ideas in a threaded discussion format. Figure 2 below displays a captioned screenshot of the environment used in the experimental study.

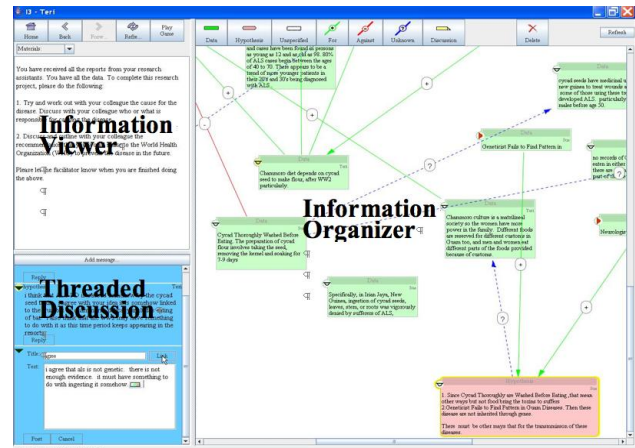


Figure 2: Screenshot from I3P1's session

The “information organizer” workspace includes tools derived from Belvedere [36] for constructing knowledge objects under a simple typology relevant to the experimental task of identifying the cause of a phenomenon (e.g. a disease), including data (green rectangles, for empirical information) and hypotheses (pink rectangles, for postulated causes or other ideas). There are also linking tools for constructing consistency (“for”) and inconsistency (“against”) relations between other objects, visualized as green links labeled “+” and red links labeled “-”, respectively. “Unspecified” objects and “unknown” links are also provided for flexibility. Finally, an embedded note object supports a simple linear (unthreaded) discussion that appears similar to a chat tool; except that a note is interactionally asynchronous and one can embed multiple notes in the knowledge map and link them like any other object. In the “threaded discussion” section of the environment (see bottom-left of Figure 2 above) participants can embed references to knowledge map objects in the threaded discussion messages by

selecting the relevant one or more graph object while composing the message. The references show up as small icons in the message. When the reader selects the icon, the corresponding object in the knowledge map is highlighted, indicating the intended referent.

Mutual awareness of participants' *artifacts* is supported in the software environment as follows: all knowledge map nodes and threaded discussion messages carry the name of the participant who first created it. The mutual awareness features of *artifacts* and of *activity* are shown in Figure 2, a screenshot taken from I3P1 (I stands for Anglo-American--Chinese inter-cultural session, 3 stands for the number of experimental session in this condition, P1 stands for Participant 1). In Figure 2, the I3P1's screen name of "Teri" (screen name selected by participant) appears on the title bar of the application window and on knowledge map nodes and message created by I3P1. Similarly, I3P2's screen name of "Sue" appears on artifacts created by him. Artifacts marked with a solid red triangle in the top right corner are from I3P2 and are yet to be opened by I3P1. The yellow circle on the threaded discussion message of I3P1 in the lower left region of Figure 2 indicates artifacts created by "Teri" (I3P1) but not yet read by the study partner, "Sue" (I3P2). Thus each participant is aware of the new artifacts from the study partner as well as the artifacts not yet read by their study partner. In the lower-left corner of Figure 2 shows I3P1 appropriating the affordances for referencing knowledge map artifact (yellow outlined hypothesis node in the bottom-right of Figure 2).

2.3.2 Protocol for Workspace Updates

To simulate asynchronous online interactions, the actions of each participant in the shared workspace were not displayed immediately in the other participant's workspace. As a person worked, the actions of that person were sent to the other participant's client application, but were queued rather than displayed. Participants were given a new report after playing the game of Tetris™. Tetris™ was chosen as it presents a different sensory-motor perceptual task than the primary experimental study task of collaborative knowledge map co-construction and simulates taking a break from the studies in real-worlds asynchronous learning settings [36]. After the game of Tetris™, all of the currently queued actions on that client were displayed. Conflicts that might arise when both participants edited the same object were resolved through operational transformations [36]. The delayed updating protocol simulates one aspect of the experience of asynchronous collaboration: a participant sees what one's partner has done upon returning to a workspace after a period of time. It excludes the possibility of synchronous conversation in which one participant posts a message in the workspace and receives an immediate reply. The "refresh" feature of the software enables one to get all updates to that point in time.

2.3.3 Alternates for Action

The software environment provides multiple alternatives for appropriation of affordances. For example, participants can discuss with each other using the threaded discussion tool or the embedded notes tool. Participants can also use the knowledge-map objects to discuss the task at hand or any other topic of interest. Participants can refer to artifacts by deictic referencing (this, that, etc...) or use the cross-referencing feature of the threaded discussion. Participants can externalize the perceived relations between their concepts by creating external evidential

relations between objects in the knowledge-map, by spatial arrangement, or by mentioning them in discussion. Participants have multiple ways of sharing the information presented to them (threaded discussion, embedded notes, and knowledge-map).

The research strategy was to provide participants with a feature rich collaborative environment with multiple alternates for action. By incorporating systematic variation in the assignment of participants to the collaborative dyad based on their cultural background and gender, the experimental design measured and observed systemic differences in how participants used the tools and resources of the technology (research question 1, appropriation of affordances) and related to each other during and after their interaction (research question 2, technological intersubjectivity).

2.3.4 Topics

The study presented participants with a "science challenge" problem that requires participants to identify the cause of a disease known as ALS-PD on the island of Guam. This disease has been under investigation for over 60 years, in part because it shares symptoms with Alzheimer's and Parkinson's diseases. Only recently have investigators converged on both a plausible disease agent (a neurotoxic amino acid in the seed of the Cycad tree) and the vector for introduction of that agent into people (native Guamians' consumption of fruit bats that eat the seed). Over the years numerous diverse hypotheses have been proposed and an even greater diversity of evidence of varying types and quality explored. These facts along with the relative obscurity, multiple plausible hypotheses, contradicting information, ambiguous data and high interpretation make this a good experimental study task for measuring cultural effects on appropriation of affordances and on technological intersubjectivity.

All experimental study materials were in English. All participants began with a mission statement that provided the problem description and task information. Four mission statements corresponding to the four participant assignment configurations (Chinese vs. Anglo-American x P1 vs. P2) were administered (<http://lilt.ics.Hawaii.edu/culturalreps/materials/>). Due to the distribution of conflicting evidence, sharing of information across participants and study sessions is needed to expose the weakness of genetics as well as to construct the more complex explanation involving bats and cycad seeds. Given the nature of the information distribution between the two participants, working out the consumption of bats as an optimal hypothesis involves making these cross-report collaborative connections and also considering and rejecting other probable factors. The study task and task materials are designed to highlight "social division of cognitive labor". The experimental study encouraged participants to interact with each other by including the following reinforcing task instruction on each report (set of 4 articles): "Please share and discuss this information with you colleague. Please play the game to receive the next report from your research assistant." The next section discusses several research hypotheses generated from the culture theory and empirical findings in cross-cultural psychology.

2.4 Participants

Participants were recruited from the graduate student community at the University of Hawai'i at Mānoa. Each participant was offered a payment of US\$75 for participating in

the study. Participant selection and treatment assignment are discussed next.

2.4.1 Sampling

There is a tendency in cross-cultural computer mediated communication research to use cultural models bounded by modern nation-states. Although nationality based stratified sampling frames remain a methodologically convenient way to select participants provided, cultural homogeneity of the participants is not to be assumed but empirically measured. We used the PVQ individual values survey [31] and the GLOBE [21] instrument to empirically assess differences in the two participant groups at the individual and group levels respectively.

2.4.2 Assignment

Participants were randomly assigned to either the intra- or the inter-cultural profiles and the same or different gender profiles. Excluding 6 pilot studies, a total of 33 experimental sessions involving 66 pairs of participants were conducted. Data from 3 experimental sessions was discarded due to issues of a missing screen recording, a software crash, and a disqualification. There were 10 pairs of participants for each of the three treatment groups: Chinese-Chinese intracultural; Anglo-American-Anglo-American intracultural, and Anglo-American-Chinese intercultural groups. All the three conditions were gender-balanced because gender can substantially influence social interaction [37]. Each treatment group included 3 female-female, 3 male-male and 4 female-male dyads.

2.5 Instruments

2.5.1 Demographic Questionnaire

A demographic questionnaire [41, pp.275-276] was administered to collect participants' familiarity with each other, with online learning environments, with usability evaluation studies as well as data about age, gender, ethnic background, duration of stay in the USA, duration of stay in the state of Hawai'i. All participants were requested to make a self-report of their CGPA and also assign a release form for obtaining official records of their CGPA, graduate record examination scores (GRE), and test of English as a foreign language (TOEFL) (Chinese participants only).

2.5.2 Self-Perception: Portrait Value Questionnaire (PVQ)

The 40 item version of the PVQ instrument [41, pp.277-279] recommended for intercultural contexts (Schwartz, S. H, personal communication) was used in the study. The PVQ scale measured cultural values at the individual level. Cronbach's "alpha measures of internal consistency range from .37 (tradition) to .79 (hedonism) for the PVQ (median, .55)" [31, p.532]. Gender specific versions of the self perception PVQ scale were administered.

2.5.3 GLOBE Cultural Dimensions Instrument

The GLOBE instrument [21] was used to measure cultural values at the group level [41, pp.280-293]. Section 1 and Section 3 of the original GLOBE instrument were used in this study. Section 1 of the GLOBE instrument measures a responder's perceptions of their society "Section 1 — The way things are in your society". Section 3 of the GLOBE instrument measures a responder's preferences for their society "Section 3 — The way

things generally should be in your society". According to the "Guidelines for the Use of GLOBE Culture and Leadership Scales," "the construct validity of the culture scales was confirmed by examining the correlations between the GLOBE scales with independent sources (e.g., Hofstede's culture dimensions, Schwartz's value scales, World Values Survey, and unobtrusive measures)" [21]. Phrasing of "this country" has been changed to "my home society" to remove possible ambiguity for Chinese graduate students who might rate Hawai'i, USA instead of the society they grew up in.

2.5.4 Individual Essays

At the end of their collaborative science problem solving, the immediate post-test consisted of each participant individually writing an essay. Identical essay writing instructions were provided to all participants. The instructions asked the participants to (a) state the hypotheses they considered, (b) whether and how their hypotheses differed from those of their study partners', and (c) their final conclusion.

2.5.5 Peer-Perception: Portrait Value Questionnaire (PVQ)

Technological intersubjectivity after interaction was measured by the second immediate post-investigative-test. This was the administration of the Portrait Value Questionnaire (PVQ) [31] instrument with a reversal of the direction of assessment [41, pp.304-306]. This time instead of assessing themselves, participants assessed their collaborative partners. Based on their collaborative interactions, each participant rated his/her impressions of the study partner on the Portrait Value Questionnaire (PVQ).

2.5.6 Acculturation: SL-ASIA Questionnaire

Acculturation is a process that occurs when two or more cultures interact together. This becomes an external variable in cross-cultural research conducted with participants from an immigrant culture in a host culture (in our case, Chinese participants in Hawai'i, USA. Although one could argue that Hawai'i is different culture for Anglo-American participants who grew up on mainland USA). This external variable can be controlled by measuring the acculturation level of the participants belonging to the minority immigrant culture [38]. Participants with high level of acculturation can be best used as members of the majority host culture or not included in the study [38]. This research project used the Suinn-Lew Asian Self Identity Acculturation (SL- ASIA) scale [35] to measure the acculturation levels of the Chinese participants [41, pp.307-311]. This scale was chosen as it is specifically designed for Asians. Suinn et al. [35] reported an internal-consistency estimate of .91 for the SL-ASIA instrument.

2.5.7 Intercultural Sensitivity: Intercultural Sensitivity Instrument

Intercultural sensitivity is a vital skill for intercultural collaborations [2]. The SL-ASIA scale provided a measure of Chinese participants' assimilation to USA. The intercultural sensitivity instrument (ICSI) [2] was used to measure Anglo-American participants' self-assessment of intercultural sensitivity [41, pp.312-315]. Bhawuk and Brislin (1992) report that "the ICSI was validated in conjunction with intercultural experts at the East-West Center with an international sample (n=93)" (p. 423). The word "Japan" in the original ICSI scale

was changed to “China” to fit the context of Chinese-American collaboration setting of the experiment. Part three of the original ICSI instrument was not used, as pilot studies indicated that it was irrelevant to the purposes of this experimental study.

2.5.8 User Satisfaction: QUIS Questionnaire

The QUIS 7.0 questionnaire [18] was administered to collect the participants subjective perceptions and preferences of the learning environment [41, pp.316-321]. The QUIS has high reliability (Cronbach’s alpha = 0.95 and high construct validity (alpha = 0.86) [18].

2.6 Procedure

Two students participated in each session. Experimental sessions lasted about 3.5 hours on average. Informed consent was obtained from all participants for both the pilot studies and the experimental studies. After signing the informed consent forms, participants completed a demographic survey. They were then given CGPA/GRE/TOEFL score release form, Self-Perception PVQ [31], and the GLOBE instrument [21]. After completing these three forms, participants were brought into a common room. Participants were then introduced to the software and the structure of the experimental study through an identical set of instructions and demonstrations across all three conditions.

After the software demonstration, the two participants were led back to their respective workstations in two different rooms. They were then instructed to begin work on the study task. Participants had up to 90 minutes to work on the information available for this problem. The update protocol described in [36] was used to synchronize the workspaces of the two participants. At the conclusion of the investigative session, each participant was given up to 30 minutes to write an individual essay. The CSIC environment remained available to each participant during the essay writing, but the participants were requested not to engage in any further communication. After each participant had finished writing the individual essay, the other-perception PVQ instrument [31] and the QUIS instrument [18] were administered. This concluded the experimental session. Participants then completed the payment forms and were debriefed.

3. RESULTS

Results are grouped under the five subsections of *demographics*, *culture measures*, *objective usability measures*, and *subjective usability measures*. The empirical data generated by the experimental study were analyzed at four levels: culture (Anglo-American, Chinese), gender (female, male), dyadic culture (American-American, American-Chinese, Chinese-Chinese), and dyadic gender (female-female, female-male, male-female). Unless otherwise indicated, the statistical summaries refer to two-way analysis of variance with respect to culture (Anglo-American, Chinese) and gender (female, male).

3.1 Demographics

There was no age difference at any of the four levels of analysis (culture, gender, dyadic culture, dyadic gender). As expected, Anglo-American participants reported to have spent significantly more time in the United States of America than the Chinese participants. On the other hand, the time spent by the participants in Hawai’i with respect to culture and gender was not statistically significant. There were no significant differences at any of the four levels of analysis for prior experience with

experimental studies, prior knowledge about the experimental task, and partner familiarity.

3.2 Culture Measures

As mentioned before, Portrait Values Questionnaire (PVQ) [31] was used to measure culture at the individual level. The GLOBE instrument [21] was used to measure culture at the group level.

Ten individual values are measured by the Portrait Values Questionnaire (PVQ) [31]. Ten individual values are measured by the Portrait Values Questionnaire (PVQ) [31]. Statistical analysis showed that at the level of culture, the PVQ values of Conformity ($F(1,56)=7.71, p=0.008$), Benevolence ($F(1,56)=5.60, p=0.02$), Universalism ($F(1,56)=6.66, p=0.01$), Self-Direction ($F(1,56)=7.48, p=0.01$), Stimulation ($F(1,56)=10.02, p=0.003$) and Security.

Significant differences were observed on both sections of the GLOBE instrument. For the “AS IS” section, significant differences between the American and Chinese groups were observed for Institutional Collectivism ($F(1,56)=43.55, p<0.01$), In-Group Collectivism ($F(1,56)=102.43, p<0.01$), and Assertiveness ($F(1,56)=28.57, p<0.01$). For the “SHOULD BE” section of the GLOBE instrument, statistically significant differences were found for Uncertainty Avoidance ($F(1,56)=49.65, p<0.01$), Assertiveness ($F(1,56)=4.20, p=0.04$), Future Orientation ($F(1,56)=14.23, p=0.01$), Humane Orientation ($F(1,56)=7.90, p=0.007$), and Gender Egalitarianism ($F(1,56)=4.89, p=0.03$).

In summary, there is necessary and sufficient evidence to conclude that Chinese and Anglo-American participants significantly differ on specific PVQ individual values as well as GLOBE cultural dimensions. Even though a nation state based stratified random sampling frame was utilized; systemic variation between the two participant groups is thus empirically documented and not stereotypically assumed or dogmatically asserted.

3.3 Objective Usability Measures

Objective usability measures consisted of the *efficiency* (total task time in minutes), and *effectiveness* (usage of certain features of interest). Each of these is discussed below.

3.3.1 Efficiency

On average, task time was greater for Chinese participants ($M=156.07$ minutes, $SD=19.22$) than the Anglo-American participants ($M=144.96$, $SD=25.14$). On average, female participants’ task time ($M=155.58$, $SD=20.88$) was greater than the male participants ($M=145.44$, $SD=24.00$) in the study. A two-way ANOVA showed marginal main effects for culture ($F(1,56)=3.77, p=0.06$) and gender ($F(1,56)=3.14, p=0.08$). On the other hand, total task time varied significantly between the intra- and inter-cultural conditions of the experimental study ($F(2,51)=5.17, p=0.009$). A Bonferroni post-hoc comparison showed that the Anglo-American intra-cultural group had significantly lower task time than the Chinese intra-cultural group and the American-Chinese inter-cultural group. No significant differences were observed at the dyadic gender level.

3.3.2 Effectiveness

Effectiveness measures consisted of the software features of structural and functional significance to computer supported

collaborative learning. Each measure is introduced, briefly discussed, and then empirical results are presented.

3.3.2.1 Cross-referencing

Video analysis of the screen recordings of participant sessions was done to obtain the counts for the referencing. No statistically significant differences were found at any of the four levels of analysis (culture, gender, dyadic culture, dyadic gender).

3.3.2.2 Shared workspace refresh

As discussed under software in the methodology section, the shared workspace (information organizer +discussion) could be refreshed (a) automatically after returning from game or (b) on demand when the participant clicks on the “Refresh” button (see top-right in Figure 2). Recall that there were four reports and a final page. Participants had to play and quit the game in order to receive the next report. All the participants played and quit the game at least four times and therefore, received all the four reports. However, the refresh count varied due to the differences in the number of on-demand refreshes of the shared work-space. There was no significant main effect for refresh count with respect to culture. However, the refresh count was lower for the female participants with a marginal main effect for gender ($F(1,56)=3.50, p=0.067$). At the dyadic gender level of analysis, a marginally significant effect was found ($F(2,51)=2.632, p=0.82$). A post-hoc Bonferroni comparison showed a marginally significant difference between the male-male and female-female collaborative dyads.

3.3.2.3 Threaded discussion messages

Counts for *discourse usage* were obtained from the software logs of participant sessions. For threaded discussion messages, Anglo-American participants created more threaded discussion messages than the Chinese participants and the difference was statistically significant ($F(1,56)=8.88, p=0.004$).

3.3.2.4 Embedded discussion notes

For the embedded discussion notes, no statistically significant difference was found. However, the observed empirical trend was that Chinese participants created more embedded discussion notes than the Anglo-American participants.

3.3.2.4 Evidential relation links

Counts for *evidential relation links* were obtained from the software logs of participant sessions. The observed empirical evidence was exactly opposite to that of the theoretical prediction at the level of culture ($F(1,56)=5.54, p=0.02$) for the number of evidential relation links created. On average, Anglo-American participants created more evidential relation links compared to the Chinese participants of the experimental study.

3.3.2.5 Knowledge-map nodes

No significant differences were observed between the Chinese and Anglo-American participants in the number of “data” and “hypotheses” nodes created. However, Chinese participants created less number of “unspecified” nodes than the Anglo-American participants ($F(1, 56)=5.76, p=0.02$). At the gender level of analysis, female participants created significantly more “hypothesis” nodes than the male participants ($F(1, 56)=4.68, p=0.035$).

3.3.2.6 Verbosity

Language remains a potential mediating variable in the performance of Chinese participants. To empirically evaluate this mediating variable, session verbosity (total words individually produced by a participant in the collaborative session) and essay verbosity (total words produced by a participant in the individually written essay) were calculated. A two way analysis of variance for session verbosity showed significant main effects for both culture ($F(1,56)=4.46, p=0.39$) and gender ($F(1,56)=6.70, p=0.012$). On average, Anglo-American participants produced more words in the collaborative session than the Chinese participants. Female participants produced more words in the collaborative session than the male participants. Similar results were obtained for the essay verbosity measure.

3.4 Subjective Usability Measures

As mentioned earlier, the validated usability instrument, QUIS questionnaire [4] was administered to collect the participants subjective perceptions and preferences of the learning environment. The QUIS 7.0 instrument also measured participants’ subjective satisfaction with the instructions and the software tutorial besides various systems measures. The coding key for the QUIS instrument was used for the quantitative analysis of the data (<http://lap.umd.edu/QUIS/QuantQUIS.htm>). Table 2 below presents a summary of the QUIS results with respect to culture.

Table 1: Summary of QUIS Results with respect to Culture

| QUIS Construct | Anglo-American Mean (SD) | Chinese Mean (SD) | ANOVA F-Statistic (p-value) |
|----------------------------------|--------------------------|-------------------|-----------------------------|
| Overall Reaction | 6.18(1.35) | 6.49(1.22) | 0.84(0.36) |
| Screen | 6.95(1.06) | 6.15(1.19) | 8.00(0.01)* |
| Terminology & System Information | 7.46(1.02) | 6.81(1.22) | 4.84(0.03)* |
| Learning | 7.78(1.09) | 7.22(1.32) | 3.17(0.08) |
| System Capabilities | 7.17(1.13) | 6.85(1.43) | 0.92(0.34) |
| Tutorial | 7.90(1.03) | 7.71(1.03) | 0.49(0.48) |

* $p < 0.05$

3.4.1 Overall user satisfaction

On average, the overall user reaction for the Chinese participants was higher than Anglo-American (see table 1). However, no significant differences were found at any of the four levels of analysis (culture, gender, dyadic culture, dyadic gender).

3.4.2 Perception of information display

Significant differences were observed between the Chinese and Anglo-American participants on the QUIS section for information display on the screen (see table 1). Chinese participants’ subjective satisfaction scores for the screen information display were lower than the Anglo-American participants.

3.4.3 Perception of system terminology

System Terminology and System Information section of the QUIS instrument received significantly lower ratings from the Chinese participants.

3.4.4 Perception of system capabilities

No significant differences were observed despite lower scores by Chinese participants compared to the Anglo-American participants of the experimental study.

3.4.5 Perception of ease of learning

Results for the Learning section of the QUIS instrument showed a marginally significant difference on the ease of learning measure at the level of culture.

3.4.6 Perception of tutorial

Results for the Tutorial section of the QUIS instrument showed no significant difference for participants' subjective evaluation of the software demo and experimental instructions at any of the four levels of analysis. Therefore, experimenter bias and "demand characteristics" [27] are ruled out as confounding variables in the study.

In summary, there is a discrepancy between Chinese participants' higher overall satisfaction ratings and their lower satisfaction ratings for the specific components of the system (**Screen and Terminology & System Information**). Similarly, Anglo-American participants' reported lower overall reaction ratings but higher satisfaction ratings to the specific components of the system.

3.4.7 Analysis of comments

The QUIS instrument includes an open-ended comments solicitation at the end of each of the six sections. The user comments were transcribed. A few illustrative user comments are included below:

C2P2: "The link function is very helpful and I'll expect a drag and drop from the organizer to the message panel."

C6P2: "Having a zoom in/out feature might help."

I3P1: "I like the idea and that can link data boxes. However, a function that would put everything into a condensed list or a short outline to see all data at once should be helpful b/c sometimes too much information is displayed at once to work with in a rational manner."

I8P2: "In general the system speed is satisfactory, and it's reliable, but it needs more on other functions such as undo, correcting typo, etc."

I11P2: "It is good if there is a "undo" and "redo" (ctrl+z or ctrl+y)."

A3P1: "When new text boxes appear from the partner, they should be in a separate section so it is easy to see them and sort them out from mine and older ones. They should be color coded differently until read. The size of screens should be adjustable to allow more."

A7P2: "Instructions were well laid out and easy to use. Messages sometimes appear overlapping, difficult to see everything that way."

A8P1: "Overall fairly clear & easy to navigate."

A9P2: "In the boxes, the word "text" should be eliminated in a click. It shouldn't need deleting."

Qualitative analysis of the comments shows that Undo, Copy+Paste, Zooming and Color Coding of Contributions are the most frequent usability suggestions. Usability problems mentioned included scrolling issues. Negative comments were mainly about screen clutter.

The coding scheme developed in [42] was modified and used for the content analysis of the comments. The modified coding scheme is described below:

- **Usability Problem (U):** interaction design flaw or a user difficulty that is directly associated with an interface/interaction design flaw.
- **Suggestion (S):** subjective preference of the participant to the implemented design choice or tradeoff.
- **Positive Comment (P):** participant's subjective approval of a design choice or tradeoff.
- **Negative Comment (N):** participant's subjective disapproval of a design choice or tradeoff.
- **Other Comment (O):** User comment that couldn't be categorized under one of the above categories.

Total Comments = Usability Problems (U) + Suggestions (S) + Negative Comments (N) + Positive Comments (P). Figure 3 presents aggregate comment counts with respect to culture.

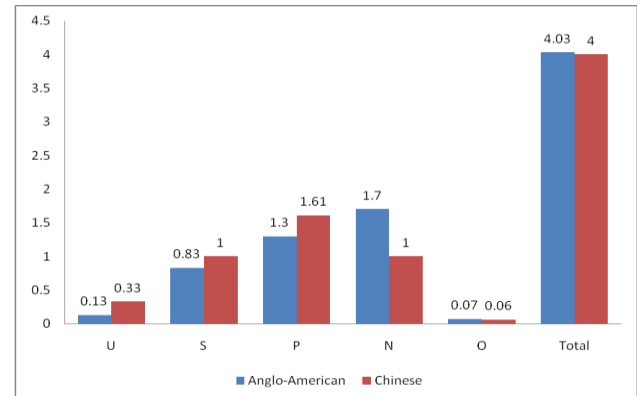


Figure 3: Aggregate Comments across Cultures

Even though Chinese participants made more usability suggestions, more positive comment, and less negative comments than the Anglo-American participants, no significant differences were observed at any of the four levels of analysis.

4. DISCUSSION

The empirical findings reported above can be summarized as follows: Culturally different participants were found to be engaged in different collaborative activities and created different quantities of artifacts. Further, participants from the two cultural groups had differential user interface satisfaction scores along with different quality and quantity of comments. The empirical findings discussed add to the literature on the cultural effects on usability evaluation process [e.g., 3, 7, 8, 11-14, 19, 25, 32, 47-49]. Specifically, Chinese participants reported higher overall user interface satisfaction scores but

gave significantly lower ratings for information display and terminology aspects of the system. One candidate explanation for this empirical finding is the cultural-cognitive difference in holistic vs. analytical reasoning [26]. Participants with an analytical thinking style might offer more specific user interface satisfaction ratings of the individual components of the system compared to participants with a holistic thinking style. This could account for the greater specificity of Anglo-American participants with individual aspects of the socio-technical system compared to Chinese participants' overall assessment of the socio-technical system. Cultural concerns with "face saving" might be another explanation.

4.1 Implications for Research

Honold [20] identified eight factors for cultural usability: *objectives of the users; characteristics of the users; environment; infrastructure; division of labor; organization of work; and mental modes based on previous experience and tools.* Objectives of users have been the focus of recent work in cultural usability that focuses on evaluator-subject interpersonal relations [32, 42]. Cultural usability should consider the debate around "demand characteristics" [27] in experimental psychology. A better understanding of demand characteristics is of benefit to the field of usability in general. Characteristics of user are often ignored the concept of culture is rarely operationalized. Cultural characteristics are attributed to individuals by virtue of ethnic affiliation or nation-state membership. Given the "fading quality of culture" the homogenous assumptions of culture are unwarranted [29]. As stated above under culture measures in the results section, cultural characteristics of participants in this experimental study are assessed, evaluated, and documented at level of individual values as well as at the level of group dimensions. In computer supported collaboration, *environment; infrastructure; division of labor; organization of work; and mental modes based on previous experience and tools* all become crucial factors of concern. For example, despite the differences in cultural characteristics of the users, their activities, their artifacts, and their subjective interface perceptions, as reported in [see 40, 41], no significant differences were found in individual learning outcomes. In a prior contribution, we have proposed [45] a comprehensive design evaluation framework of usability, sociability, and learnability for computer supported collaborative learning (CSCL) systems.

In my opinion, cultural usability needs to move beyond the comparative method approach to usability evaluation. Computer supported collaboration in intra- and inter- cultural contexts presents some unique challenges which do not entirely fall under the purview of usability typically conceived of as efficiency, effectiveness and satisfaction. Further, in the established HCI research field of computer supported cooperative work (CSCW), the emerging HCI field of human-information interaction (HII) [22], the technology enhanced learning paradigm of computer supported collaborative learning (CSCL), and the emerging field of computer supported intercultural collaboration (CSIC), users are interacting with each other as well as with the computer. Currently, the Internet is undergoing a profound shift towards a participatory mode of interaction. With the advent of fundamentally social software such as social networking sites (Orkut, Facebook, MySpace, Mixi etc), cultural usability needs to be expanded to include technological intersubjectivity [41]. An initial attempt to address theoretical and methodological

issues on intra- and inter- cultural computer supported collaboration can be found in [41]. Specifically, [41] attempted to develop a theory-based empirical study of socio-technical affordances and their appropriation and an empirically informed theory of the structures and functions of technological intersubjectivity. In my opinion, there is a need for a real-time and real-space interactional account of cultural cognition for the field of cultural usability. In sum, this is a call for the field of cultural usability to widen its research horizons.

5. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under CAREER award 0093505 to Dan Suthers. The author was partially supported in the writing of this article by NSF award 0535036 to Scott Robertson.

6. REFERENCES

1. Beu, A., Honold, P. and Yuan, X. How to Build Up an Infrastructure for Intercultural Usability Engineering. *The International Journal of Human-Computer Interaction*, 12 (3&4). 347-358.
2. Bhawuk, D.P.S. and Brislin, R.W. The measurement of intercultural sensitivity using the concepts of individualism and collectivism. *International Journal of Intercultural Relations*, 16 (4). 413-436.
3. Bourges-Waldegg, P. and Scrivener, S.A.R. Applying and testing an approach to design for culturally diverse user groups. *Interacting with Computers*, 13 (2). 111.
4. Chin, J.P., Diehl, V.A. and Norman, K. Development of an instrument measuring user satisfaction of the human-computer interface. in *SIGCHI '88*, ACM., New York, 1988, 213-218.
5. Clarke, A.C. *Profiles of the future: an inquiry into the limits of the possible*. Harper & Row, 1962.
6. Clemmensen, T. and Goyal, S., Cross cultural usability testing. in *The First All India Human Computer Interaction Conference*, (Bangalore, India 2005).
7. Clemmensen, T., Hertzum, M., Hornbæk, K., Shi, Q. and Yammiyavar, P. Cultural Cognition In The Thinking-Aloud Method For Usability Evaluation *International Conference on Information Systems - ICIS2008*, Paris, Accepted for publication.
8. Clemmensen, T. and Plocher, T. The Cultural Usability (CULTUSAB) Project: Studies of Cultural Models in Psychological Usability Evaluation Methods. *Lecture Notes in Computer Science*, 4559. 274.
9. Clemmensen, T. and Qingxin, S., What is part of a usability test? in *CHI '08 extended abstracts on Human factors in computing systems.*, (2008).
10. 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. *Proceedings of Conference of Human Computer Interface, International. Beijing, PR China*.
11. Day, D. and Evers, V., Questionnaire Development for Multicultural Data Collection. in *Proceedings of the International Workshop on Internationalisation of Products and Systems*, (1999).
12. Day, D. and Evers, V. The Role of Culture in Interface Acceptance. *Human Computer Interaction, INTERACT' 97*.
13. Evers, V., Cross-cultural Understanding of Metaphors in Interface Design. in *Attitudes toward Technology and Communication*, (London, 1998).

14. Evers, V., Kukulska-Hulme, A. and Jones, A., Cross-Cultural Understanding of Interface Design: A Cross-Cultural Analysis of Icon Recognition. in *Proceedings of the International Workshop on Internationalization of Products and Systems*, (Rochester, 1999).
15. Fernandes, T. *Global Interface Design*. Academic Press, 1995.
16. Garfinkel, H. *Studies in Ethnomethodology*. Prentice-Hall, Englewood Cliffs, NJ, 1967.
17. Hall, E. *Beyond Culture*. Anchor Books, New York, 1977.
18. Harper, B., Slaughter, L. and Norman, K. Questionnaire administration via the WWW: A validation & reliability study for a user satisfaction questionnaire. www.lap.umd.edu/webnet/paper.html.
19. Herman, L. Towards Effective Usability Evaluation in Asia: Cross-Cultural Differences. *Proceedings of the 6th Australian Conference on Computer-Human Interaction (OZCHI'96)*. 135.
20. Honold, P. 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). 327 - 345.
21. House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W. and Gupta, V. *Culture, Leadership and Organizations: The GLOBE study of 62 societies*. Sage Publications, Newbury Park, CA, 2004.
22. Jones, W., Pirolli, P., Card, S.K., Fidel, R., Gershon, N., Morville, P., Nardi, B. and Russell, D.M. "It's about the information stupid!": why we need a separate field of human-information interaction *CHI '06 extended abstracts on Human factors in computing systems*, ACM, Montreal, Quebec, Canada, 2006, 65-68.
23. Khaslavsky, J. Integrating Culture into Interface Design. *CHI 98 conference summary on Human factors in computing systems*. 365-366.
24. Marcus, A. and Gould, E.M. Cultural Dimensions and Global Web User-Interface Design. *Interactions*, 7 (4). 32 - 46.
25. Molich, R., Ede, M.R., Kaasgaard, K. and Karyukin, B. Comparative usability evaluation. *Behaviour & Information Technology*, 23 (1). 65 - 74.
26. Nisbett, R.E. and Norenzayan, A. Culture and Cognition. in Medin, D.L. ed. *Stevens' Handbook of Experimental Psychology*, 2002, 561-597.
27. Orne, M.T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Prevention & Treatment*, 5 (35).
28. Reeves, B. and Nass, C. *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press New York, NY, USA, 1996.
29. Ross, N. *Culture and Cognition: Implications for Theory and Method*. Sage Publications, 2004.
30. Russo, P. and Boor, S. How Fluent is your Interface? Designing for International Users. *Proceedings of the conference on Human factors in computing systems*. 342-347.
31. Schwartz, S.H., Melech, G., Lehmann, A., Burgess, S., Harris, M. and Owens, V. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology*, 32 (5). 519-542.
32. Shi, Q. and Clemmensen, T. Relationship Model in Cultural Usability Testing. 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, 422-431.
33. Stoffregen, T.A. Affordances and Events. *Ecological Psychology*, 12 (1). 1-27.
34. Stoffregen, T.A. Affordances as Properties of the Animal-Environment System. *Ecological Psychology*, 15 (2). 115-134.
35. Suinn, R.M., Ahuna, C. and Khoo, G. The Suinn-Lew Asian Self-Identity Acculturation Scale: Concurrent and Factorial Validation. *Educational & Psychological Measurement*, 52 (4). 1041-1046.
36. Suthers, D., Vatrappu, R., Medina, R., Joseph, S. and Dwyer, N. Beyond Threaded Discussion: Representational Guidance in Asynchronous Collaborative Learning Environments. *Computers and Education*, 50 (4). 1103-1127.
37. Tannen, D. *Gender and Discourse*. Oxford University Press, 1996.
38. Triandis, H.C., Kashima, Y., Shimada, E. and Villareal, M. Acculturation Indices as a Means of Confirming Cultural Differences. *International Journal of Psychology*, 21. 43-70.
39. Turvey, M.T. Affordances and Prospective Control: An Outline of the Ontology. *Ecological Psychology*, 4 (3). 173-187.
40. Vatrappu, R. Cultural Considerations in Computer Supported Collaborative Learning. *Research and Practice in Technology Enhanced Learning*, 3 (2). 159-201.
41. Vatrappu, R. Technological Intersubjectivity and Appropriation of Affordances in Computer Supported Collaboration *Communication and Information Sciences*, University of Hawaii at Manoa, Honolulu: Available at <http://iilt.ics.hawaii.edu/~vatrappu/docs/Vatrappu-Dissertation.pdf>, 2007, 538.
42. Vatrappu, R. and Pérez-Quinones, M. Culture and Usability Evaluation: The Effects of Culture in Structured Interviews. *Journal of Usability Studies*, 1 (4). 156-170.
43. Vatrappu, R. and Suthers, D. Culture and Computers: A Review of the Concept of Culture and Implications for Intercultural Collaborative Online Learning. in Ishida, T., Fussell, S.R. and Vossen, P.T.J.M. eds. *Intercultural Collaboration I : Lecture Notes in Computer Science*, Springer-Verlag 2007, 260-275.
44. Vatrappu, R. and Suthers, D. Technological Intersubjectivity in Computer Supported Intercultural Collaboration. in *ACM 2nd International Workshop on Intercultural Collaboration*, ACM Digital Library, in press.
45. Vatrappu, R., Suthers, D. and Medina, R., Usability, Sociability, and Learnability: A CSCL Design Evaluation Framework. in *16th International Conference on Computers in Education (ICCE2008)*, (Taipei, Taiwan, in press), (CD-ROM).
46. Wertsch, J. *Vygotsky and the social formation of mind*. Harvard University Press, Cambridge, MA, USA, 1985.
47. Yammiyavar, P. and Goel, K.M., Emphasis on non-verbal cues for interpreting cognitive processes in protocol analysis. in *Indo-Danish HCI Research Symposium* (Guwahati, India, 2006), The Indian Institute of Technology Guwahati, Guwahati 781039. Assam India.
48. Yeo, W.A. Cultural Effects in Usability Assessment. *Proceedings of CHI 98 conference summary on Human factors in computing systems*. 74-75.
49. Yeo, W.A. Global-Software Development Life Cycle: An Exploratory Study. *Proceedings of the SIG-CHI Conference on Human Factors in Computing Systems*. 04 -11.