

A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations

Sivaji, Ashok; Clemmensen, Torkil; Nielsen, Søren Feodor

Document Version
Accepted author manuscript

Published in:
International Journal of Human-Computer Interaction

DOI:
[10.1080/10447318.2016.1243928](https://doi.org/10.1080/10447318.2016.1243928)

Publication date:
2017

License
Unspecified

Citation for published version (APA):
Sivaji, A., Clemmensen, T., & Nielsen, S. F. (2017). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human-Computer Interaction*, 33(5), 357-370.
<https://doi.org/10.1080/10447318.2016.1243928>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 03. Jul. 2025



A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations

Ashok Sivaji, Torkil Clemmensen, and Søren Feodor Nielsen

Journal article (Post print version)

CITE: A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. / Sivaji, Ashok; Clemmensen, Torkil; Nielsen, Søren Feodor. In: International Journal of Human-Computer Interaction, Vol. 33, No. 5, 2017, p. 357-370.

The final publication is available at Springer via <http://dx.doi.org/10.1080/10447318.2016.1243928>

Uploaded to [Research@CBS](#): June 2017

A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations

Ashok Sivaji^{1*}, Torkil Clemmensen², Søren Feodor Nielsen³

¹ UX/Usability Lab, MIMOS Technology Solutions, MIMOS Berhad Kuala Lumpur Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia

* Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi Selangor, MALAYSIA

² Department of IT Management, Copenhagen Business School, Solbjerg Plads 3, DK-2000 Frederiksberg, Denmark

³ Department of Finance, Copenhagen Business School, Solbjerg Plads 3, DK-2000 Frederiksberg, Denmark

¹ Phone: +6012-2119154

*Corresponding Author: ashok.sivaji@mimos.my

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations

Abstract: The usability movement has historically always sought to empower end-users of computers, so that they understand what is happening and can control the outcome. In this paper, we develop and evaluate a ‘Textual feedback’ tool for usability and UX evaluation that can be used to empower well-educated, but low-status, users in UX evaluations in countries and contexts with high power distances. The proposed tool contributes to the HCI community’s pool of localized UX evaluation tools. We evaluate the tool with 40 users from two socio-economic groups in real-life UX usability evaluations setting in Malaysia. The results indicate that the Textual Feedback tool may help participants to give their thoughts in UX evaluation in high power distance contexts. In particular, the Textual Feedback tool helps high status females and low status males express more UX problems than they can do with traditional CTA alone. We found that classic concurrent think aloud UX evaluation works fine in high power contexts, but only with the addition of Textual feedback to mitigate the effects of socio-economic status in certain user groups. We suggest that future research on UX evaluation look more into how to empower certain user groups, such as low status female users, in UX evaluations done in high power distance contexts.

Keywords: Empirical Studies of User Behaviour; Human-Computer Interaction in Organizations and Society; Human-Computer Interaction Theory; Evaluation/Methodology; User Interface Management Systems (UIMS); Global and Cultural Considerations

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

1 Introduction

The usability movement has historically always sought to empower end-users of computers, so that they understand what is happening and can control the outcome (Nielsen, 2005). Researchers have argued that we should pay attention to how uneven power relations are enacted in design practice by studying power dynamics (Irani, Vertesi, Dourish, Philip, & Grinter, 2010; Pineda, 2014), power distance (Y. Kim & Mutlu, 2014; Xie, Rau, Tseng, Su, & Zhao, 2009) and power dimension (Clemmensen & Plocher, 2007; Jano et al., 2015; Marcus & Alexander, 2007). In this paper, we develop and evaluate a ‘Textual feedback’ tool for usability/UX evaluation (hereafter “UX evaluation” to cover both terms). UX professionals can use the tool to empower well-educated, but low-status, users in UX evaluations in countries with high power distances. The proposed tool contributes to the HCI community’s pool of localized UX evaluation tools, e.g., the “Bollywood method” (Chavan, 2005), the “Harambee method” (Oyugi, Abdelnour-Nocera, & Clemmensen, 2014), and the “Community approach” (Winschiers-Theophilus, Chivuno-Kuria, Kapuire, Bidwell, & Blake, 2010).

Empowering the user in usability tests is not simple. The classic advice to the evaluator is to tell the user “it is not you that are being tested, but the system”, and then assume that the outcome is valid and interesting. This may not work as intended. For example, if the evaluator is downplaying power distances in an effort to create a more equitable, collaborative environment, this may unintentionally increase distrust in the users with high power distance expectations (Oyugi, et al., 2014). In particular, when doing usability evaluation with people of low socio-economic status, the evaluator-user power gap is obvious. When the evaluator has a higher socioeconomic status than the participant and, in addition perhaps is foreign to the local community, this leads to strong feelings of intimidation and performance anxiety at the user’s side (Sherwani, Ali, Rosé, & Rosenfield, 2009). To recognize the existence of a problematic power gap in UX evaluation in some contexts is a start to tackle it.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

The background for this paper is that research in usability and user experience (UX) evaluation methodologies continues to be an important and active area of research (James R Lewis, 2015). Classical concurrent think aloud (CTA) usability testing in a lab remains one of the most used UX evaluation methods among UX professionals worldwide. CTA emerged in interface design for urban and highly educated adults in Europe and USA (Chavan, 2005; C. Lewis, 1982), but today the method is applied in most regions of the world and for diverse groups of people (McDonald, Edwards, & Zhao, 2012; Rosenberg & Kumar, 2011). Usability and UX is becoming institutionalized around the world (Smith et al., 2007). This movement expands and challenges our understanding of what is good quality UX evaluation. Many emerging economies are trying to become powerhouses for IT innovation, and need to not only follow western knowledge and practices, but also build on their own strengths. In this process, it is of particular importance to go beyond simply using classic UX evaluation methods, and instead aim to adapt, develop and assure the quality of UX evaluation methods for use in national test labs and national product certification bodies to do inter-laboratory comparison, global IT development and local evaluation.

Applying UX evaluation in new cultural contexts and with a diversity of users may not be valid and reliable without modifying the method, or it may even be necessary to modify the concepts of usability and UX to fit the new context (Chavan, 2005; Winschiers & Fendler, 2007). These and similar insights have led to the proposal for a social-psychological theory – cultural model usability (CMU) - for explaining cultural usability and UX phenomena, and as a framework for deploying UX evaluations in multi-cultural settings (Clemmensen, 2009; Pineda, 2014). The CMU theory says among other things that usability and UX findings in a UX evaluation depends on what the user finds socially and culturally appropriate to say in the test situation. The social appropriateness of using accessible and available cultural knowledge comes into question in particular when the evaluator and test participants come from different sociocultural backgrounds. Sharing knowledge and coordinating descriptions of usability and UX defects depend on the

mutual perception of belonging to a group, and test users might ask themselves implicit questions

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

regarding the social appropriateness of sharing the available knowledge with the evaluator in the UX evaluation situation. On a societal level, the issue of power in UX evaluation has been addressed by using Hofstede’s national cultural dimensions (Hofstede & Minkov, 2013) by Andy Smith (2011), who suggested a model for fitting variants of UX evaluation methods to countries, based on national values. Smith identified in particular one of Hofstede’s cultural dimensions as important for UX evaluation, the Power Distance (PD) dimension, which can be defined as the degree of generally accepted dependence between boss and subordinate. In this paper, we address the users’ socio-economic status in the test situation as our independent variable, and take into account national cultural values by focusing on a high power distance (PD) country. The paper is an attempt to answer the question: *Can a ‘textual feedback’ tool that change the user’s think aloud from oral to written text, mitigate the influence from high power distance on the quality of the user feedback in UX evaluation?*

In the following, we discuss the related work on theory behind usability and UX testing, the critique of the missing social and cultural context of UX evaluation, and the issue of power distance in UX evaluation. We present our case of Malaysia, and the proposed tool that supports UX professionals to get design insights by empowering end-users in CTA in high power distance contexts. The evaluation section evaluates the tool by comparing the outputs of think aloud verbally to the textual feedback in a lab experiment. We discuss the results, and conclude that the textual feedback tool can reduce the effect of high power distance between test participants and UX professionals in UX evaluation situations.

2 Related work

The classical CTA is based on Ericsson & Simon’s information processing theoretical model and substantial empirical evidence for verbal protocol analysis as method for collecting verbal reports of thought sequences. They explain how valid data on thinking can be collected by instructing participants to verbalize their thoughts in a way that does not change the sequence of the thoughts that they have when completing

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

a task (Ericsson & Simon, 1984). The verbal data can later be analysed and used to model task-directed human thinking. When applying CTA in UX evaluation, however, the aim is not to provide a model of what goes on in the head of the human, but to find a list of usability problems related to the interaction with the new design being tested (C. Lewis, 1982).

According to Ericsson and Simon, CTA gives a valid expression of users' thoughts and leaves task performance unaffected as long as users verbalize information in their current focus of attention (verbalization at levels 1 and 2) and refrain from providing explanations and retrieving additional information from memory (verbalization at level 3). The current evidence suggest that CTA does not degrade the user's task performance, i.e., does not change the task the user does (reactivity), except perhaps for Asian ethnic groups (Fox, Ericsson, & Best, 2011; H. S. Kim, 2002). There is little doubt that the classic CTA is scientifically valid method under the theoretical conditions specified by Ericsson and Simon (1984), see (Fox, et al., 2011).

The CTA was introduced to HCI and usability testing by Clayton H. Lewis (1982) at IBM in 1982, as one of many inventions from US and European cognitive psychology imported into the then emerging field of HCI. The assumption is that the use situation – individual task directed performance – in the CTA usability test is a universally important use situation for future users of the new design, hence the method should be useful all over the world. In a usability test, an individual user thinks aloud, while using a single application on a single device, with a specific task or goal in mind. The UX professional interpret the user's verbal and non-verbal behaviour, and later communicate the scientifically valid insights from CTA to the developer/designer as UX defects (i.e., issues, problems, design opportunities) to be corrected in the next design iteration. However, two strands of research have in particular challenged the spread of the use of classic CTA for UX evaluation: The research on the social situation and research on the cultural context.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

2.1 The socio-cultural context of CTA UX evaluation

A CTA UX evaluation session is a social situation. In most UX evaluation situations, the user is not alone with the computer. Thinking aloud is often not restricted to verbalization at levels 1 and 2, during which the person thinking out loud suspends any awareness of listeners; instead thinking aloud includes elements of conversation and interview, and often with the evaluator supporting the user's thinking aloud with active listening techniques (Boren & Ramey, 2000). Zhao and McDonald (2010) compared classic and interactive CTA usability testing and found that both produced explanations, opinions and recommendations (level 3 verbalizations), and conclude that the social context of usability testing may override the instruction to think aloud only. Krahmer & Ummelen (2004) found that the relaxed protocols produced more task completion and a "more natural interaction, where participant and test administrator are engaged in a dialogue". However, both Zhao & McDonald (2010) and Krahmer & Ummelen (2004) pointed out that the relaxed protocol of evaluator's active listening a la Boren and Ramey creates validity problems for CTA. The choice of protocol appears to be a balance between validity (e.g., are the insights really about the information that the user needs when interacting with the current design?) and pragmatic usefulness (e.g., can the information that the user gives be used to identify usability and UX problems? And to give design suggestions?). While the evidence on which protocol to use in a CTA usability evaluation is not conclusive (James R. Lewis, 2014), the evidence appears to be in favour of choosing the protocol that gives the most useful information for the UX professional's purpose, i.e., to use a relaxed protocol for CTA when the purpose is UX evaluation.

A CTA UX evaluation session is a cultural practice. The traditional assumption is that the UX evaluation use situation is universally important for future users of the new design. In contrast, it is now increasingly accepted that cultural matters are of importance to the practice and outcome of such evaluations (James R. Lewis, 2014). It appears that all the major elements of a UX evaluation may be susceptible to cultural contextual influences: 1) instructions and tasks, 2) verbalization, 3) reading the user, and 4) the overall

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

relation between the evaluator and test user (Orngreen, Clemmensen, & Pejtersen, 2009). Staying within the cognitive psychology approach used by classic CTA, Clemmensen et al (2009) used the literature on cultural cognition, e.g., (Nisbett, Peng, Choi, & Norenzayan, 2001), to argue that the CTA is based on a view of humans as analytic thinkers, while many people evidently tend to think more holistically and take contextual factors into account. They concluded that when applying the classic CTA usability method with Asian users (Nisbett et al. argues that Asians tend to be holistic thinkers), the test leader (evaluator) should be sensitive towards several aspects of the CTA method. Similar issues have been raised by others, e.g., (Chavan, 2005; Evers, 2002; Goh et al., 2013; Jano, et al., 2015; Kamppuri, 2011; Pineda, 2014; Shi, 2008; Sivaji & Ahmad, 2014; Smith, 2011; Winschiers-Theophilus, et al., 2010; Xie, et al., 2009). There is a growing number of suggested modifications to UX evaluation based on particular regional user groups or contexts.

2.2 The power issue in UX evaluations

The power distance or power gap between the evaluator and the user appears to be a significant socio-cultural issue in UX evaluation in several countries, e.g., India (Chavan, 2005), China (Sun & Shi, 2007), Namibia (Winschiers-Theophilus, et al., 2010), Kenya (Oyugi, et al., 2014), and Malaysia (Yeo, 1998). Sun and Shi (2007) studied power and language issues in usability tests using participants from China, Sweden, and Denmark. Results indicated that power distance matters; when talking to evaluators with higher rank, users tended to use more gestures to express themselves, indicating problems with face-to-face verbalization. Furthermore, there was a language effect (Chinese vs. English). Speaking Chinese made the evaluator give more help in detail, and encourage users more frequently; speaking English made the evaluator and user look at each other more often to make themselves understood, and evaluators paid more attention to checking the task list. In summary, there was an effect both from the user-evaluator overall relation and one related to verbalization. The overall relation between the evaluator and the user, and the language/verbalization used in the test, appears to be two elements of UX evaluation susceptible to power distance influences.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

2.2.1 Overall user-evaluator relations

The overall relation between the evaluator and the user is perhaps the most obvious element of a UX evaluation to be influenced by a high power distance context. Oyugi, Dunckley, & Smith (2008) studied the use of CTA with UK, Kenyan and Indian users, and examined the effectiveness of methods based on think-aloud protocols to elicit users' views. Depressingly, they found that the classic CTA was less effective with users from Non-Western countries. The Kenyan users appeared to perceive the evaluation session as a test of them, they perceived every task that was incomplete as personal failure, whereas for the UK users the opposite was true and they placed the blame on the design. Oyugi, et al. (2008) suggested that the large differences between the UK and the other countries/cultures (East Africa, India) were related to Power Distance issues. Oyugi et al. explain this by suggesting that evaluator-user relation is like traditional medical doctor-patient relation and non-Western users appeared to be more sensitive towards this in terms of giving less critique of the design. Another example is the study by Petrie, Walsh, Odutola, & Ang (2015) who studied Nigerian, UK and Chinese users, and found that levels of agreement cannot be considered comparable across different cultural groups. Petrie et al. argue that in particular the findings in Nigeria – users did not respond negatively to the new design, as expected by the researchers - can be explained with this country being a high Power Distance culture, and hence the users did not want to offend the high status researchers.

Shi (2008) studied CTA in China and found that the overall evaluator test user relation is important, and that the relation should be trustful in order for the participant (the user) to think aloud accurately and to a reasonable extent. However, Shi (2008) also found that “All users took the evaluator as an interviewer, facilitator, or guide”, though they did not report pressure or being uncomfortable with the evaluator. Furthermore, she found that “the users often kept silent and did not speak out actively, so probing questions and reminding users to speak out tended to be very necessary in Chinese thinking aloud usability/UX evaluations.” These users were all highly educated, urban, Chinese users, accustomed to being

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

participants in a diversity of marketing and product tests etc. Accordingly, even under the best of conditions, Chinese users may see the evaluator as a teacher or guide and themselves as students. Asian students are often described as talking less and being more reserved and quiet than Western students even in Western class room contexts, which in part can be ascribed to (self-)stereotyping Asians as less talkative in class rooms (Devos & Yokoyama, 2014). Shi (2008) found that her Chinese users preferred, instead of giving the classic concurrent think aloud report, to wait until after the event, and then give a complete and thoughtful report to the evaluator. In contrast to Shi's findings, the classic CTA UX evaluation assumes that the evaluator-test user relation is a learner-teacher relation, with the evaluator as the learner and the user as the teacher.

2.2.2 Verbalization

Verbalization appears to be a problem in UX evaluation in high power distance countries. Xie et al. (2009) found that people accustomed to high power distance situations could not communicate as effectively as people accustomed to low power distance situations in an interface design communication session with verbal interaction. There was no significant difference between people accustomed to high power and low power situations in dealing with communication without verbal interaction, e.g. situations relying on non-verbal communication and tasks without interaction. Apala Chavan (2005), in India, developed the "Bollywood Method" based on nine rasas (emotional ticket) derived from the Nāṭyasāstra, an ancient work of dramatic theory, written during the period between 200 BC and 200 AD. The informality and 'fun' value of the tickets made users less defensive and thus give more frank answers. H.S. Kim (2002) showed that requiring Asian user to think aloud directly hinders task performance, while this is not the case for western users. Evers (2002) found that verbalization is easier for North Americans than for Japanese users who feel uncomfortable verbalizing their thoughts. In sum, there are indications that Asian users may benefit from avoiding the classic CTA socio-cultural situation with face-to-face verbalization, when giving feedback to the evaluator.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

2.3 Media choice: Talk or write the feedback

The purpose with the user's verbalization in CTA is to allow the evaluator to 'read the user'. Because the verbal data collected from the user in CTA may anyway be analysed as text (Lacity & Janson, 1994), we suggest to consider text as a supplementary media for data collection in UX evaluation. Textual feedback written on a computer is a form of computer-mediated communication (CMC), since participants focus on communicating to the evaluator through the computer. We provide three arguments, taken from the rich literature of media choice, see (George, Carlson, & Valacich, 2013), for adding textual feed back to the data collection methods.

First, textual feedback may reduce situational constraints and social expectations. We know from studies of CMC in work place situations that text-based CMC reduces situational constraints and social expectations, and that this may help increase participation rate and confidence in participants who otherwise tend to suppress self-expression (Bazarova & Yuan, 2013). This is also known as the "equalization phenomenon" (Dubrovsky, Kiesler, & Sethna, 1991), i.e., that text-based CMC reduces the differences in rate of participation between group members of different social status. Participants are less concerned with social differences and attributes because they do not expect to be socially evaluated in the absence of contextual and nonverbal cues. Hence, more of the participants will act without considering the power distance to the evaluator in a situation with lean, text-based media.

Second, Amichai-Hamburger, McKenna, & Tal (2008) argue that there is an empowering effect of CMC, such that being in CMC situations lower participants' social expectations and related risks of social sanctions, and thereby stimulate participation. Participants in UX evaluations in high power countries might feel more confident when, instead of having to think aloud face-to-face, they will be allowed to state their thoughts in textual feedback media.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

Third, CMC can conceal nonverbal clues that present contradicting or disturbing information to the receiver of the communication, who in this case is the evaluator. Thus CMC may take off mental load from the participants, which might be particularly beneficial to people with a holistic way of thinking and hence difficulties expressing thoughts in concurrent think aloud (H. S. Kim, 2002).

UX evaluation is, ideally, a simple communicative situation – give your thoughts while you work on the tasks; hence a lean media such as writing a short text on the computer should theoretically be the participant's ideal media choice.

3 The case of Malaysia

Malaysia has for long been known for its so-called multimedia corridor that presents impressive infrastructure and low costs, and thus makes it attractive to global software companies that seek to test and develop their products for small and emerging markets, or may want to support the digitization of government, commercial, health, and educational sectors. Malaysia is also known for the multilingual skills of Malaysian workers in English, Hindi, Malay-Indonesian, Chinese, and Thai, which allow global companies working from Malaysia to launch their products in other markets (Ong, 2005). In many ways, it is of interest to study everyday practice in the Malaysian ICT industry and sector.

Within HCI, several studies indicate that the Malaysian cultural context and user groups are important for UX of websites and other IT products used, produced, or localised in Malaysia. Marcus and Gould (2000) compared websites from Malaysia, a high PD country, and from the Netherlands, a low PD country, and found that PD impacts website design significantly. Gould, Zalcara & Yusof (2000) found that Malaysian websites show characteristics of high PD and collectivism by incorporating the organization chart, logo of the institution and chancellor conferring degree to a graduate. Nasrul, Nor, Masrom, & Syarief (2012) found that Malaysian websites have high PD characteristics such as axial symmetry, official symbol, monumental buildings and picture of people of higher authority. Ahmed, Mouratidis, & Preston (2009), while comparing

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

Malaysian and British websites, found that Malaysian websites displayed characteristics of a high PD country, as there was a high tendency to address the leaders' with official title such as 'Dato', 'Datuk', 'Tunku', 'Tan Sri', 'Tun' or 'Dr', and showcase awards and recognition to display the achievements of the organisation. In general, relatively many research papers on website usability are about Malaysia (Nawaz & Clemmensen, 2013). Together, these studies indicate the importance of considering power issues while designing and developing IT for a high PD country such as Malaysia.

The application of classic UX evaluation methods in Malaysia has been studied by Yeo (1998, 2000, 2002). He found that the imported usability assessment tools like CTA does not work in Malaysia, but that a possible solution would be to "apply imported UATs [usability assessment tools] in a realistic setting, that is, apply the UATs to evaluate the usability of a localised software application" (Yeo, 2002, p57). The tools employed in usability evaluation would only be effective if the imported UX evaluation methods were localised for the Malaysian context (Yeo, 2002). Yeo and others, e.g., Zaman & Winschiers-Theophilus (2015), have continued this line of research, but with a focus on rural users in Malaysia, and, because classic CTA UX evaluation did not work with these user groups, have developed a community participatory design approach. However, Yeo's original line of work that focused on localising UX evaluation methods to the classic UX user group: the urban, big city user, continues to be important, as these users make up the major part of the population of IT users in Malaysia. Recent studies support that lab based CTA UX evaluation are needed in Malaysia, and that the adaptation of CTA UX evaluation requires the development of a variety of UX evaluations methods in combination (Sivaji & Ahmad, 2014; Yammiyavar & Clemmensen, 2008). We believe that our case of Malaysia is of interest not only to UX professionals researchers in Malaysia, who may benefit from using our proposed tool, but our Malaysian case study should be of interest to UX researchers and practitioners worldwide, as power distance, e.g., high and low power status,

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

is a global issue. Our proposed method might be applied around the world, e.g., in Europe, for UX evaluation with low status groups.

4 The textual tool

Mi-UXLab 1.0 is a tool developed by the MIMOS Usability/UX Lab at Technology Park Malaysia, Kuala Lumpur¹. The lab is a national test lab for ensuring the quality of products for the Malaysian market, and it has been accredited by the Department of Standards Malaysia for ISO17025:2005. The purpose of the tool is to support UX research, design, development, verification and validation and reporting². The tool enables both remote and lab based UX testing. The features and functionalities include being able to gather the user demographic, creating tasks to be performed by the users, and measuring usability and UX, Figure 1. Most importantly, the tool can textually gather feedback during UX research and/or test of websites.

To demonstrate how the tool works, Figure 2 details the user's journey when conducting UX testing. Let us imagine the case when a user has been asked by the evaluator to do a task on a website. The test involves the moderator who will facilitate the test. For a step by step description, see Figure 2.

¹ Mi-UXLab 1.0 is a tool developed by the MIMOS Usability/UX Lab at Technology Park Malaysia, Kuala Lumpur, as part of the Test in the Cloud strategy, under the Eleventh Malaysia plan (Abolfazli et al., 2015), and of the MIMOS Berhad initiative to establish an environment of an 'Information Society acceptable to all citizen' to allow richer communication and collaboration by anyone, anytime anywhere (Stephanidis et al., 1999; Suffian, Fahrurazi, & Ibrahim, 2014).

² Researchers, who are interested, can write to the first author, to request for a trial use. At the time of publishing this article, the tool was hosted on <https://usability.mimos.my/mi-uxlab/home.php>

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

Insert Figure 1 about here

Insert Figure 2 about here

The use of the tool Mi-UXLab 1.0 may provide the UX researcher and practitioner with various insights into the website and software design for the Malaysian context. In this paper, we focus on tools' textual feedback capture capability, hence the name 'Textual feedback' tool.

5 Method

*As indicated by discussions in the research literature, e.g., Fischer (2011), linking culture and individual behaviour is not straightforward. The relation between society trends such as high power distances and individual behaviour may be quite complex. We cannot expect that individuals should be representative of society trends, but at the same time role expectations, such as expectations as to how to behave as a test user and as a low status person, may follow societal expectations. So when we talk about usability test, we talk about individual behaviour that should follow universal role expectations, but as we point out in this study, there seem to be some important – and largely invisible and unstudied – relations between standard usability tests and the societal role expectations/power relations. We attempt to demonstrate that classic CTA with its specific – and largely invisible - power relations does not produce the expected results in societies and contexts with other expectations.

We have chosen to use Malaysia as a case of a cultural context where expressions of power, e.g., Hofstede's power distance, may be quite different from the contexts where the usability testing method originally was developed. We acknowledge that power relations and power contexts are always there in usability tests, but we want to point out that these are often invisible power relations - taken for granted -

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

in established usability test. The expectation is that you as a test user are outright critical and put forward your immediate thoughts, but there are indications that established usability tests do not straightforward work as expected in high power contexts. When it comes to our participants, we have used formal job titles and job experience as an expectation of social status in the usability tests.

5.1 Research design

This research applies a repeated within subjects design, where every participant provided two types of feedback – CTA and Textual Feedback - to each of a number of tasks as they participated in industry UX evaluations that were conducted at the MIMOS lab. *The usability process practices performed by the lab have been accredited by the Department of Standards Malaysia for ISO17025:2005, specifically under ISO/IEC 25010:2011 Systems and software Quality Requirements and Evaluation. The lab has a good track record of performing usability testing for both Malaysian government and private sector. We gathered the participants' qualitative feedbacks verbally for CTA and textually for textual feedback using the Mi-UXLab tool.

5.2 Participants

The targeted users were Malaysians from different segments such as students, system administrators, business owners, test engineers, executives, and supervisors. *We selected the participants to fit one of two groups according to their relative status vis-a-vis to the perceived high status of the evaluators. The first group (low status) consisted of undergraduates and junior executives. The second group (high status) consisted of senior executives, management and senior management. Based on these criteria, potential participants were screened in personal interviews, and participants who were motivated and had relevant skills in using the particular IT system under test were selected. For example, for the test of the job portal, undergraduate users who were expected to enter the labour market soon were selected through a university internship unit. For the Human Resource Information System under test, MIMOS Corporate

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

Human Resources members of staff were selected. All selected participants could read and write and had substantial experience in the use of computers in general.

5.3 Material and setup

The websites on which the participants did their assigned tasks were real-life websites. All of the websites chosen were designed and developed in Malaysia. All of the websites were in English, and in addition some had text in Malay. The various websites evaluated in this study were intended to be used by consumers. As suggested by Nawaz and Clemmensen (2013), the study included a broad selection of websites of government and private sector websites. The websites included Employment portal for students, Cloud Computing Website for System Administrators, Cloud Computing Website Application for Business Owners, Defect Management Website Application for Test Engineers, Human Resource Information System for Executives, Human Resource Information System for Supervisors, Document Management System Desktop Application for Executives, Wireless Portal for System Administrators, Financial Planning Website for consumers, and Integrity Measurement Software for consumers. Data were collected for each website over a period of 25 days.

The lab set up was a formal setting, the UX lab in MIMOS, consisting of a testing room, an observer room and waiting room. This UX evaluation situation was a classic usability test configuration similar to what participants would know as a school exam situation, with an evaluator present in the room to instruct, facilitate, and record the test session including the CTA and textual data collection, and with the participant following instructions.

5.4 Instructions to participants

At the start of the UX evaluation, the evaluator would greet the user and explain the purpose of the test as a standard usability test of several websites for Malaysian companies and institutions. Neither the evaluator nor the user was made aware of the fact that the CTA and Textual Feedback method would be compared. The website was automatically opened on a PC and the evaluator would describe the first task

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

to the user. While performing the task, the user was asked to think aloud and perform the task. Upon task completion, the evaluator would ask the user to provide feedback in written form, Figure 2. After this, the evaluator briefly interviewed the participant to clarify issues mentioned during the CTA and textual.

5.5 Data collection

To compare the verbalization from concurrent think aloud and textual feedback, CTA verbalizations were transcribed. The written input in Textual Feedback was collected. The material for analysis was CTA transcriptions and Textual feedback writings from each participant.

5.6 Content analysis

*The quality of feedback obtained from both CTA and Textual Feedback methods were coded and analysed systematically. We assessed the quality of the usability feedback by coding the content of the participants' feedback, instead of simply counting raw number of e.g., words, or using severity assessments for which the interpretation depends on the design phase (Hertzum 2006). Following Clemmensen, Hertzum, Yang, & Chen (2013), we used Han, Yun, Kwahk, & Hong's (2001) and Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, & Sinnelä's (2011) defect classifications. Han et al.'s classification scheme classifies the feedback according to whether the feedback is related to objective performance (perception/cognition, memorization/learning and control/action) or subjective image or impression (basic sense, image description and user's evaluative feeling). In addition to Han's subjective-objective scheme, we used Kujala et al.'s system-oriented classification into general UX issues, attractiveness, ease of use, utility, and degree of usage over time. *Table 1 shows the 'Han' and 'Kujala' classifications that were used to categorize the usability feedback.

Insert Table 1 about here

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

The first author applied the two classifications on the CTA feedback and the textual feedback. An example feedback received from one of the user, Case Id 1: “Ada information pasal syarikat dan peluang pekerjaan. Live support terlalu kecil dan tak Nampak. Kalendar hanya lambangkan calendar sahaja”; or translated into English: “Information is shown about company. Live support is too small and has low visual clarity. Calendar feature only shows calendar instead of events.” This feedback was classified as Performance: “Information is shown about company” (memorization/learning - informativeness), and “Calendar feature only shows calendar instead of events” (memorization/learning - informativeness), while “Live support is too small and has low visual clarity” (description of image - image/icon clarity) was classified as Image /Impression Dimension.

5.7 Statistical analysis

*The observations are in the form of counts: How many comments of a specific type are given by a specific participant when completing a specific task? Each observation is a count (e.g., how many Han-Performance comments did XX make in task YY); the smallest counts are 0, the largest count is 16. As the counts are small, the statistical analysis is based on Poisson regression models (Gardner & Shaw, 1995). Moreover, the counts are dependent as participants provide both CTA feedback and textual feedback on each task they are presented with (paired observations) as well as feedback from different tasks (repeated measurements).

To account for this dependence, we include Gaussian random effects for participants and tasks and thus use mixed effects log-linear Poisson regressions (Tuerlinckx, Rijmen, Verbeke, & Boeck, 2006) to model the data. The random effects also handle participant heterogeneity (some participants tend to provide more feedback than others do), task heterogeneity (some tasks tend to provoke more feedback than others do) as well as any overdispersion in the counts. To eliminate potential carry-over effects - participants may recall their CTA feedback when providing textual feedback - we use a ratio of ratios approach similar to the

"difference in differences" methods used in econometrics (Ashenfelter, 1978). Though difference in

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

differences estimation is usually applied to non-experimental data, the bias it is designed to eliminate is mathematically similar to the carry over effect in an experimental design such as ours. We use ratios to measure the effects rather than differences as Poisson models are multiplicative rather than additive.

All models are fitted using the lme4-package (Bates, Maechler, Bolker, & Walker, 2013) in R version 3.2.1. (R Core Team, 2014).

6 Results

6.1 Demographics

Out of the 171 usability tests tasks done in total, 91 (~53%) were from low power status participants and the remaining 80 recordings were from high power status participants (~47%). The recruited participants comprised of 73% Malay ethnic, 18% Indian, 5 % Chinese and another 5 % other ethnics. Most participants used mainly Malay as their first choice of language and English as the second choice of language. Table 2 shows the distribution of gender on the participant groups and tasks.

Insert Table 2 about here

6.2 Feedback from CTA and Textual

Table 3 shows the amount of feedback per participants per task corrected for differences between tasks and participants. The amounts of feedback are given for each category of feedback as well as for the total (i.e. summed) feedback ("Overall"), for males and females separately and for each feedback tool.

Insert Table 3 about here

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

The results in the table indicate that participants in the low status group generally produce less feedback than high status participants do, regardless of the category of feedback, the gender of the participant, and the type of feedback. Hence, the data exhibits the expected power distance effect. We also see that the textual method is not able to eliminate the power distance effect.

The table also shows feedback ratios, i.e. the amount of textual feedback divided by the amount of CTA feedback. With just two exceptions (Kujala "General", for low status females and high status males), the feedback ratios are all above 1, ranging from just above 1 to almost 5. Thus, the textual feedback method produces more useful feedback than CTA does. As the textual feedback is provided after the CTA feedback, we would expect to see a carry-over effect, i.e. that participants would tend to remember the feedback they produced in the first part of the experiment and repeat this in the second part. However, the fact that more feedback is produced is an indication that textual feedback may be useful either alone or in combination with CTA.

Insert Table 4 about here

Table 4 shows ratios of feedback ratio estimates, i.e. the feedback ratio of high status participants divided by the feedback ratio of low status participants, for each combination of gender and feedback category. The potential carry over effect, that is expected to inflate the feedback ratios, cancels out, when looking at ratios of ratios. A ratio of ratios of 0.75, as we get for the overall Han score for male participants, shows that the feedback ratio of high status males is only three quarters of that of low status males. Thus, the textual feedback tool has a larger effect on low status males than high status males, when we look at the overall Han score. We see that for males the ratios of ratios are consistently smaller than 1, indicating that

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

the feedback ratio of low status participants is consistently larger than for high status individuals. Thus male low status participants benefit more from the textual feedback method than high status individuals do. This indicates that the textual feedback tool is able to lessen the effect of the power distance for males. The effect is, however, only statistically significant for Kujala "General" ($p < 0.0001$). For females, we observe the opposite relationship: High status females benefit much more than low status females do (with the exception of Kujala "Degree"). For females the difference is significant for both overall feedback amounts ($p < 0.0001$) as well as for Han "Performance" ($p < 0.0001$) and Kujala "General" ($p < 0.0001$) and "Utility" ($p = 0.0005$). The significant differences remain significant after applying a Bonferroni correction for multiple testing.

Returning to the ratios in Table 3, we note that low status females have small feedback ratios (except for Kujala "Degree") regardless of the feedback tool. Thus the effect we see in the ratios of ratios in Table 4 is due to high status females producing much more feedback when using the Textual method, while low status females produces little feedback when using either method. In other words, while the textual feedback tool is very useful for high status females, it makes little difference for low status females.

Figure 3 shows how the overall amount of Han (left) and Kujala (right) feedback depend on gender, power status and type of feedback tool after adjusting for differences between tasks and individuals. The two classification schemes lead to similar conclusions. The figure show that the Textual tool yields more feedback than CTA does, and that the difference is particularly striking for high status females. The figure also illustrates that difference between the amount of CTA feedback and textual feedback is more pronounced for low status males than for high status males.

7 Discussion

In this paper, we raise the question: Can a “textual feedback” tool that changes the user’s think aloud from oral to written text, mitigate the influence from high power distance on the quality of the user feedback in

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

UX evaluation? We have proposed a tool for collecting usability evaluation data in high power contexts, the Textual Feedback tool, and evaluated the tool in a UX usability evaluation setting in Malaysia. The results indicate that the Textual Feedback tool may help participants to give their thoughts in UX evaluation in high power distance contexts. In particular, the Textual Feedback tool helps high status females and low status males to express more UX problems than they can do with traditional CTA alone.

Insert Figure 3 about here

7.1 Status matters for UX evaluation in high power contexts

In our study, no matter which data collection method (CTA and Textual feedback) or analysis method (the Han and the Kujala UX defect classification) was used, the results indicated that low status users tend to produce less of the useful type of feedback in usability evaluation than high status users do. This is consistent with the literature about power distance in UX evaluation that says that when the evaluator has a higher socioeconomic status than the participant this leads to strong feelings of intimidation and performance anxiety at the user's side (Sherwani, et al., 2009), that when talking to evaluators with higher rank, users tend to use more gestures rather than verbalization to express themselves (Sun & Shi, 2007), that low status users perceive every task that are incomplete as a personal failure (Oyugi, et al., 2008), and that low status users do not want to offend the evaluator (Petrie, et al., 2015). The results support the general idea that the social context of usability testing may override the instructions given to the user in how to think aloud (Zhao & McDonald, 2010), and the CMU theory's idea that what the user in a given socio-cultural situation finds socially appropriate to say is an important aspect of how we identify UX issues (Clemmensen, 2009). However, our study shows that the Textual Feedback tool may change the socio-cultural test situation and hence mitigate the power distance in UX evaluations, which we discuss in the sections below.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

7.2 The proposed tool generates more useful feedback than CTA verbalization

The results indicate that the Textual Feedback tool may help participants to give their thoughts in high power distance UX evaluation situations. The textual method generally results in more useful feedback being produced with useful feedback ratios ranging from just above 1 to almost 5 times as much useful feedback from the textual method than from CTA, with the most significant improvements for low status males and high status females. This can be interpreted in different ways.

Textual Feedback may be just another example of a relaxed protocol a la Boren and Ramey's (Boren & Ramey, 2000) active listening protocol, which produces more useful feedback than classic CTA in UX evaluations. Zhao & McDonald (2010) found that compared to classic CTA their interactive think-aloud led to more useful feedback, such as examples of problem formulation, causal explanation and recommendation. We found that for both Han's UX and Kujala's UX classifications, Textual feedback produced more useful feedback such as information related to performance or impression than CTA did. Krahmer & Ummelen (2004) found that their relaxed protocols led to more task completion and a more natural social interaction. In our study, post-test interviews with our participants revealed that they found giving the textual feedback natural, as they felt it was like filling out a customer feedback form digitally after a service has been received. For some, giving textual feedback was a relief, as they felt that their views were being documented systematically for the upgrade of the system, so that what they said was listened to and accepted as important. Our results, on the significant improvement in feedback ratios with Textual Feedback for low status males and high status females, indicate that it is more socially appropriate, or 'relaxed', for these user groups to express their thoughts on UX defects as textual feedback.

Textual feedback may provide a better fit to many users' perception of the overall user-evaluator relation, compared to the assumption in classic CTA about an active user-passive evaluator relation. In textual feedback, the user is writing in silence under the evaluator's surveillance, which may fit many users' perception of the evaluator as a high status person. In particular, in text-centred and teacher-centred

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

cultures of learning, teachers should be open to innovation and new materials to support the students express themselves (Jin & Cortazzi, 2006). Shi (2008) found that her Chinese users preferred, instead of giving the classic concurrent think aloud report, to wait until after the event, and then give a complete and thoughtful report to the evaluator. Shi (2008) found that the users talked about the evaluator as an interviewer, facilitator, or guide, though they did not report pressure or being uncomfortable with the evaluator. Oyugi, et al. (2008) suggested that the evaluator-user relation for their Kenyan user groups was like traditional medical doctor-patient relation and thus these user groups appeared to be sensitive towards this in terms of giving less critique of the design. Petrie, et al. (2015) speculated that their low status users in Nigeria did not want to offend the evaluator, whom they may have seen as a figure of authority and expertise.

The Textual feedback tool may replace or complement the classic CTA verbalization as a way for users to express their thoughts. The Textual feedback tool is a technology that allows the user to communicate through the computer to the evaluator instead of in face-to-face dialogue. In this sense, Textual feedback is different from Boren and Ramey's and others' relaxed protocol that implies face-to-face speech communication. In the speech communication situation, the user is supported by the evaluator's speech and non-verbal gestures. Zhao & McDonald's (2010) use of acknowledgment tokens helped to make participants feel that their opinions were being listened to and valued, thereby giving them more confidence to make comments. In contrast, the textual feedback tool reduces social expectations and removes social cues, and thereby stimulates a different (less socially dependent) kind of participation. Textual feedback may be understood as a CMC tool that empower the user in UX evaluation personally and on the interpersonal level. Amichai-Hamburger, et al. (2008) argue that there are empowering effects of CMC. In CMC, the user is more in control because they can edit out what they do not want to share and they can disclose emotions more freely. The risk of users doing self-stereotyping, e.g., see themselves as someone who should be less active in the given social context (Devos & Yokoyama, 2014), or as predicted

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

by CMU theory (Clemmensen, 2009) find it socially inappropriate to give their thoughts, may be reduced in CMC situations. For example, in our study, it was remarkable how much more high status females participated in textual feedback, which related to less self-stereotyping and a CMC social situation where it is more appropriate for high status females to express themselves. Xie et al. (2009) found that people accustomed to high power distance situations could not communicate as effectively as people accustomed to low power distance situations in a communication session with verbal interaction. In contrast, there was no significant difference between people accustomed to high power and low power situations in dealing with communication without verbal interaction, i.e., situations relying on non-verbal communication and tasks without interaction. Textual feedback is a situation with little focus on verbal interaction, and we found significant improvement in feedback ratios for people accustomed to high power situations in dealing with textual feedback.

7.3 There is a gender effect in who the Textual feedback tool helps

The textual feedback method is more effective for high status females and for low status males. Thus as intended, the textual feedback method helps to counter the status effect for males. For females it seems that even if both low and high status individuals provide more feedback, it is particularly the high status females that benefit from being asked to give the feedback in writing.

From reading the literature, we did not expect this finding. Current literature about gender in HCI is scarce. Barkhuus & Rode (2007) found that many CHI studies did not use a gender-balanced sample, or simply did not mention the gender distribution altogether. An exception is Kotamraju (2011), who reported on her experiences with UX evaluation as a female evaluator, and tells how playing stupid helped her open up a male participant and get him to give feedback. While playing stupid to elicit information is a time-honoured tactic in Human Factors used by both men and women, Kotamraju's story illustrates that people involved in UX evaluation do in fact pay attention to gender. Gender relations could be a possible factor to explain the

low amount of feedback given by the female users in our CTA. Textual feedback may work to increase the

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

amount of feedback through the “equalization phenomenon” (Dubrovsky, et al., 1991), i.e., that text-based CMC reduces the differences in rate of participation between group members of different social status.

There are suggestions that HCI if combined with knowledge of gender theory might question some existing hypotheses in the field (Rode, 2011). Gender positionality (Alcoff, 1988) is a framework that suggests that women are conceptualized relative to a constantly changing society, i.e., female gender roles are constantly changing (Rode, 2011). This presents a view of being a woman as a perspective from which values, including women’s relations to computers, are interpreted and constructed. Malaysia is one of the countries in the world where women earn the majority of science degrees among all graduates (Charles, 2011), and this development has happened rapidly in recent years. We did not study how female gender roles in UX evaluation are performed, but clearly our very different results for females and males suggest that something is going on. In our study, what it means to perform as a Malaysian high status woman in a UX evaluation in Malaysia appears to be less related to being able to think aloud orally, and more related to being able to express one’s immediate thoughts in writing.

7.4 CTA may hinder feedback for user groups in high power situations

Textual feedback may simply be better than CTA for the user populations (Malaysian users) that we studied. In contrast to Zhao & McDonald (2010) and Krahmer & Ummelen (2004), our results do indicate that classic CTA, in order to be valid, in some situations should be adapted to a more relaxed protocol that includes Textual Feedback in UX evaluations. From the classic CTA theory, the otherwise solid current evidence that CTA does not lead to reactivity (CTA does not degrade the user’s task performance) is not conclusive for Asian ethnic groups (Fox, et al., 2011; H. S. Kim, 2002). Krahmer & Ummelen (2004) found that their relaxed protocols led to more task completion. We do not have data on task completion, only on amount of useful feedback, but given the idea that more task completion leads to more useful feedback, we may speculate that Textual feedback may enhance task completion over a series of tasks. The fact that more useful feedback is produced by Textual feedback than by CTA is an indication that textual feedback, **To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.**

or textual feedback and CTA in combination, may be a more valid UX evaluation method than CTA alone in high power contexts.

7.5 Mixed power and gender issues

According to Smith's (2011) model for fitting UX evaluation to countries, CTA UX evaluation (or 'task-based usability evaluation' as he calls it) should not be doable in a country like Malaysia with values that go against openly voicing your critique and against doing tasks individually and in a sequence. In contrast, we found that task based CTA UX evaluation works fine in Malaysia, but only with the addition of Textual feedback to mitigate the effects of socio-economic status in certain user groups.

For the Han classification, for the Impression scores, all groups provide more textual feedback than CTA. In both scores (performance and impression), for men the difference between high status and low status is less in Textual feedback than in CTA feedback. For women it is the other way round. Again, the high status females provide a lot more textual feedback than CTA. Oyugi et al. (2014) studied the effect of culture on the quality of usability evaluation results among Kenyan young professionals and Kenyan farmers. The expectation was that a high power distance would influence the results. People from high power distance (such as those from Kenya) believe that it is natural and beneficial for some members of a group or society to exert considerable control over their subordinates. Therefore, respect and formal differences for high status people such as bosses, parents and elders are valued. Oyugi et al. used field studies to develop a method called 'Harambee', depicting working together during the evaluation exercise, and evaluated the 'Harambee' and Retrospective Protocol methods. The findings indicated that despite being from the same culture-group, the effect of culture on quality of usability results differed among high-end and low-end users. Consequently, when adapting UEMs to local country conditions, there is a need to go beyond the national cultural level comparisons and directly focus on the user-type (sub-groups' socio-economic status) as an independent variable.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

Yeo (1998, 2000, 2002) found that a possible solution to the problem of adapting UX evaluation methods would be to apply imported UX evaluation methods in realistic settings. However, while Yeo looked at PD as amount of negative/positive comments (more negative comments equals more useful comments) from higher status participants, we applied two established classifications of UX to classify a richer set of feedback. Yeo also did not look into the gender effects, which turned out to be surprisingly important in our study.

7.6 Limitations

We have focused on establishing if Textual Feedback is better than CTA in high power contexts, and we have not ventured into the finer aspects of the procedure about how to do Textual Feedback data collection. However, elsewhere, the second author has tentatively proposed a theoretically based procedure that combines CTA and Textual feedback based on Eastern philosophy (Law et al., 2015).

In this study we did not focus on factors related to ICT proficiency between low status and high status users having an effect on the textual to CTA differences. However, we did consider participants' varying levels of ICT proficiency, from e.g., seamless use demonstrating self-efficacy to only general knowledge and functional skills, in the selection of participants. As mentioned in Section 5.1., all selected participants could read and write and had substantial experience in the use of computers in general.

8 Conclusion

In this paper we raise the question: can a “textual feedback” tool that change the user’s think aloud from oral to written text, mitigate the influence from high power distance on the quality of the user feedback in UX evaluation?

We see a clear power distance effect in the sense that low status participants produce less feedback than high status participants do, when being in a high power context. We also see that the textual feedback

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

method yields more feedback than CTA does, but acknowledge that this may be due to the fact that the textual feedback always follows CTA in our experiment. This, unfortunately, is unavoidable: There is no way of obtaining textual feedback before CTA as CTA is ‘concurrent’.

We conclude that low status males benefit more from the Textual method than high status males do. In this sense it seems that the textual method may be able to eliminate a part of the power distance effect. We also see that low status females do not benefit from the textual method, neither when compared to high status females nor in absolute terms. We conclude that the textual method is not able to help low status females overcome the power distance effect. Interpreting gender as another aspect of power distance, we conjecture that the textual method is able to help high status females overcome the power distance effect.

Future research will have to show how we can help low power females in such UX situations. Furthermore, gender positionality research examines the practice of gender as a relational concept in which gender roles are mutually constituted. While participating in a UX evaluation may be a relational performance, in the sense the high status female users interacting with female evaluators would perform different when giving feedback in UX evaluations, compared to when interacting with male evaluators, is something that should be studied by future research.

9 References

- Abolfazli, S., Sanaei, Z., Tabassi, A., Rosen, S., Gani, A., & Khan, S. U. (2015). Cloud Adoption in Malaysia: Trends, Opportunities, and Challenges. *Cloud Computing, IEEE*, 2(1), 60–68.
- Ahmed, T., Mouratidis, H., & Preston, D. (2009). Website Design Guidelines: High Power Distance and High Context Culture. *International Journal of Cyber Society and Education*, 2(1), 47-60.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

- Alcoff, L. (1988). Cultural feminism versus post-structuralism: The identity crisis in feminist theory. *Signs*, 405-436.
- Amichai-Hamburger, Y., McKenna, K. Y. A., & Tal, S.-A. (2008). E-empowerment: Empowerment by the Internet. *Computers in Human Behavior*, 24(5), 1776-1789.
- *Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *Review of Economics and Statistics*, 60(1), 47-57.
- Barkhuus, L., & Rode, J. A. (2007). *From Mice to Men-24 years of Evaluation in CHI*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1 – 16.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, 1(4).
- Bazarova, N. N., & Yuan, Y. C. (2013). Expertise recognition and influence in intercultural groups: Differences between face-to-face and computer-mediated communication. *Journal of Computer-Mediated Communication*, 18(4), 437-453.
- Boren, M. T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3), 261-278.
- Charles, M. (2011). What gender is science? *Contexts*, 10(2), 22-28.
- Chavan, A. (2005). Another culture, another method. *Proceedings of the 11 th International Conference on Human-Computer Interaction (Las Vegas, Nevada, United States, July 22-27, 2005)*. CD-ROM.
- Clemmensen, T. (2009). Towards a Theory of Cultural Usability: A Comparison of ADA and CM-U Theory. In M. Kurosu (Ed.), *Human Centered Design* (Vol. 5619, pp. 416-425): Springer Berlin Heidelberg.
- Clemmensen, T., Hertzum, M., Yang, J., & Chen, Y. (2013). Do Usability Professionals Think about User Experience in the Same Way as Users and Developers Do? *Human-Computer Interaction—INTERACT 2013* (pp. 461-478): Springer.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

- Clemmensen, T., & Plocher, T. (2007). The cultural usability (CULTUSAB) project: Studies of cultural models in psychological usability evaluation methods. In N. Aykin (Ed.), *Usability and Internationalization, Pt 1, Proceedings* (Vol. 4559, pp. 274-280). Berlin: Springer-Verlag Berlin.
- Devos, T., & Yokoyama, Y. (2014). Silent or Talking in the Classroom: Implicit Self-Stereotyping Among Asian and White Students. *Basic and Applied Social Psychology*, 36(5), 386-396.
- Dubrovsky, V. J., Kiesler, S., & Sethna, B. N. (1991). The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. *Human-computer interaction*, 6(2), 119-146.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis. Verbal reports as data*: Cambridge Massachusets.
- Evers, V. (2002). *Cross-cultural applicability of user evaluation methods: A case study amongst Japanese, North-American, English and Dutch users*. Paper presented at the CHI'02 Extended Abstracts on Human Factors in Computing Systems, pp. 740 – 741.
- *Fischer, R. (2011). About chicken and eggs: four methods for investigating culture–behaviour links. In F. J. R. van de Vijver, A. Chasiotis & S. M. Breugelmans (Eds.), *Fundamental Questions in Cross-Cultural Psychology* (pp. 190-213): Cambridge University Press.
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological bulletin*, 137(2), 316.
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression Analysis of Counts and Rates: Poisson, Overdispersed Poisson, and Negative Binomial Models. *Quantitative Methods in Psychology*, 118(3), 392-404.
- George, J. F., Carlson, J. R., & Valacich, J. S. (2013). Media selection as a strategic component of communication. *MIS Q*, 37(4), 1233-1251.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

- Goh, K. N., Chen, Y. Y., Lai, F. W., Daud, S. C., Sivaji, A., & Soo, S. T. (2013, April). A comparison of usability testing methods for an e-commerce website: A case study on a Malaysia online gift shop. In *Information Technology: New Generations (ITNG), 2013 Tenth International Conference* (pp. 143-150). IEEE.
- Gould, E. W., Zalcara, N., & Yusof, S. M. M. (2000). *Applying culture to Web site design: A comparison of Malaysian and US Web sites*. Paper presented at the Professional Communication Conference, 2000. Proceedings of 2000 Joint IEEE International and 18th Annual Conference on Computer Documentation (IPCC/SIGDOC 2000), pp. 161 – 171.
- Han, S. H., Yun, M. H., Kwahk, J., & Hong, S. W. (2001). Usability of consumer electronic products. *International Journal of Industrial Ergonomics*, 28(3), 143-151.
- *Hertzum, M. (2006). Problem prioritization in usability evaluation: From severity assessments toward impact on design. *International Journal of Human-Computer Interaction*, 21(2), 125-146.
- Hofstede, G., & Minkov, M. (2013). VSM 2013. *Values survey module*. URL: <http://www.geerthofstede.nl/vsm2013> (2013).
- Irani, L., Vertesi, J., Dourish, P., Philip, K., & Grinter, R. E. (2010). *Postcolonial computing: a lens on design and development*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA. CHI '10, pp. 1311-1320.
- Jano, Z., Noor, S. M., Ahmad, R., Md Saad, M. S., Saadan, R., Bokhari, M., & Abdullah, A. N. (2015). Website usability and cultural dimensions in Malaysian and Australian universities. *Asian Social Science*, 11(9), 1-10.
- Kamppuri, M. (2011). *Theoretical and methodological challenges of cross-cultural interaction design (phD Dissertatation)*. University of Eastern Finland.
- Kim, H. S. (2002). We talk, therefore we think? A cultural analysis of the effect of talking on thinking. *Journal of personality and social psychology*, 83(4), 828.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

- Kim, Y., & Mutlu, B. (2014). How social distance shapes human-robot interaction. *International Journal of Human Computer Studies*, 72(12), 783-795.
- Kotamraju, N. P. (2011). Playing stupid, caring for users, and putting on a good show: Feminist acts in usability study work. *Interacting with Computers*, 23(5), 439-446.
- Krahmer, E., & Ummelen, N. (2004). Thinking about thinking aloud: A comparison of two verbal protocols for usability testing. *Professional Communication, IEEE Transactions on*, 47(2), 105-117.
- Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers*, 23(5), 473-483.
- Lacity, M. C., & Janson, M. A. (1994). Understanding qualitative data: A framework of text analysis methods. *Journal of Management Information Systems*, 137-155.
- Law, E. L.-C., Silpasuwanchai, C., Ren, X., Bardzell, J., Clemmensen, T., & Liu, Y. (2015). *Leveraging and Integrating Eastern and Western Insights for Human Engagement Studies in HCI*. Paper presented at the Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, Seoul, Republic of Korea. CHI EA '15, pp. 2433-2436.
- Lewis, C. (1982). Using the "Thinking-aloud" Method in Cognitive Interface Design (pp. 6). NY: IBM Thomas J Watson Research Center.
- Lewis, J. R. (2014). Usability: Lessons Learned ... and Yet to Be Learned. *International Journal of Human-Computer Interaction*, 30(9), 663-684.
- Lewis, J. R. (2015). Introduction to the Special Issue on Usability and User Experience: Methodological Evolution. *International Journal of Human-Computer Interaction*, 31(9), 555-556.
- Marcus, A., & Alexander, C. (2007) User validation of cultural dimensions of a website design. *Vol. 4560 LNCS. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 160-167).

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

- Marcus, A., & Gould, E. W. (2000). Crosscurrents: cultural dimensions and global Web user-interface design. *interactions*, 7(4), 32-46.
- McDonald, S., Edwards, H. M., & Zhao, T. (2012). Exploring think-alouds in usability testing: An international survey. *Professional Communication, IEEE Transactions on*, 55(1), 2-19.
- Nasrul, M. A., Nor, K. M., Masrom, M., & Syarief, A. (2012). Website Fit: an overview. *Procedia-Social and behavioral Sciences*, 40, 315-325.
- Nawaz, A., & Clemmensen, T. (2013). Website Usability in Asia “From Within”: An Overview of a Decade of Literature. *International Journal of Human-Computer Interaction*, 29(4), 256-273.
- Nielsen, J. (2005). Usability: Empiricism or Ideology? Retrieved from Nielsen Norman Group website: <http://www.nngroup.com/articles/usability-empiricism-or-ideology/>
- Nisbett, R. E., Peng, K. P., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108(2), 291-310.
- Ong, A. (2005). Ecologies of expertise: assembling flows, managing citizenship. *Global assemblages: Technology, politics, and ethics as anthropological problems*. (pp. 337-353): Blackwell Publishing Ltd.
- Orngreen, R., Clemmensen, T., & Pejtersen, A. M. (2009). Virtual Teams and Human Work Interaction Design - Learning to Work in and Designing for Virtual Teams. In T. Gross, J. Gulliksen, P. Kotze, L. Oestreicher, P. Palanque, R. O. Prates & M. Winckler (Eds.), *Human-Computer Interaction - Interact 2009, Pt II, Proceedings* (Vol. 5727, pp. 983-984). Berlin: Springer-Verlag Berlin.
- Oyugi, C., Abdelnour-Nocera, J., & Clemmensen, T. (2014). *Harambee: a novel usability evaluation method for low-end users in Kenya*. Paper presented at the Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. NordiCHI '14, pp. 179-188.

- Oyugi, C., Dunckley, L., & Smith, A. (2008). *Evaluation methods and cultural differences: studies across three continents*. Paper presented at the Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges. NodiCHI '08, pp. 318-325.
- Petrie, H., Walsh, T., Odutola, O., & Ang, L. (2015). Cross-Cultural Issues in Working with Users in the Design of Interactive Systems *ICoRD'15—Research into Design Across Boundaries Volume 1* (pp. 515-525): Springer.
- Pineda, R. G. (2014). *Technology in culture: a theoretical discourse on convergence in human technology interaction (PhD dissertation)*. Jyväskylä Studies in Computing [187], University of Jyväskylä.
- R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012: ISBN 3-900051-07-0.
- Rode, J. A. (2011). A theoretical agenda for feminist HCI. *Interacting with Computers*, 23(5), 393-400.
- Rosenberg, D., & Kumar, J. (2011). Leading global UX teams. *interactions*, 18(6), 36-39.
- Sherwani, J., Ali, N., Rosé, C. P., & Rosenfield, R. (2009). Orality-grounded HCID: Understanding the oral user. *Information Technologies and International Development*, 5(4), 37 - 49.
- Shi, Q. (2008). *A field study of the relationship and communication between Chinese evaluators and users in thinking aloud usability tests*. Paper presented at NordiCHI '08 (Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges), pp. 344-352
- Sivaji, A., & Ahmad, W. F. W. (2014). Benefits of Complementing Eye-Tracking Analysis with Think-Aloud Protocol in a Multilingual Country with High Power Distance. In Horsley, M., Knight, M., & Reilly, R. (Eds). *Current Trends in Eye Tracking Research*. Springer, pp. 267-278.
- Smith, A. (2011). Issues in Adapting Usability Testing for Global Usability. In I. Douglas & Z. Liu (Eds.), *Global Usability*. Springer London, pp. 23-38.
- Smith, A., Joshi, A., Liu, Z., Bannon, L., Gulliksen, J., & Li, C. (2007). Institutionalizing HCI in Asia. *Lecture notes in computer science*, 4663, pp. 85-99.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. International Journal of Human–Computer Interaction, 1-14.

- Stephanidis, C., Salvendy, G., Akoumianakis, D., Arnold, A., Bevan, N., Dardailler, D., Emiliani, P. L., Iakovidis, I., Jenkins, P., Karshmer, A., Korn, P., Marcus, A., Murphy, H., Oppermann, C., Stary, C., Tamura, H., Tscheligi, M., Ueda, H., Weber, G., Ziegler, J. *International Journal of Human-Computer Interaction* 1999, 11, 1-28.
- Suffian, M. D. M., Fahrurazi, F. R., & Ibrahim, S. (2014). The Design and Execution of Performance Testing Strategy for Cloud-based System. *International Journal of Software Engineering and Technology*, 1(2), 19-25.
- Sun, X., & Shi, Q. (2007) Language issues in cross cultural usability testing: A pilot study in China. Vol. 4560 *LNCS. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 274-284.
- Tuerlinckx, F., Rijmen, F., Verbeke, G., & Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology*, 59(2), 225-255.
- Winschiers-Theophilus, H., Chivuno-Kuria, S., Kapuire, G. K., Bidwell, N. J., & Blake, E. (2010). *Being participated: a community approach*, in PDC '10 (Proceedings of the 11th Biennial Participatory Design Conference), pp. 1-10.
- Winschiers, H., & Fendler, J. (2007). Assumptions Considered Harmful. In N. Aykin (Ed.), *Usability and Internationalization. HCI and Culture* (Vol. 4559, pp. 452-461): Springer Berlin / Heidelberg.
- Xie, A., Rau, P.-L. P., Tseng, Y., Su, H., & Zhao, C. (2009). Cross-cultural influence on communication effectiveness and user interface design. *International Journal of Intercultural Relations*, 33 (1), 11-20.
- Yammiyavar, P., & Clemmensen, T. (2008). Extracting Users' Data: Towards development of a cultural and semantically sensitive combinatorial methodology. In T. Clemmensen (Ed.), *Cultural Usability and Human Work Interaction Design – techniques that connects: Proceedings from NordiCHI 2008*

Workshop Sunday October 19. 2008 Working Papers from Department of Informatics, Frederiksberg, Denmark: Copenhagen Business School, pp. 47-49.

Yeo, A. W. (1998). *Cultural effects in usability assessment*. Paper presented at the CHI '98 Conference Summary on Human Factors in Computing Systems, Los Angeles, California, USA, pp. 74-75.

Yeo, A. W. (2000). Are usability assessment techniques reliable in non-western cultures? *The Electronic Journal of Information Systems in Developing Countries*, 3, 1, 1-21.

Yeo, A. W. (2002). *Exploring Usability Evaluation of Localised Software in Malaysia (PhD dissertation)*. The University of Waikato, Hamilton, New Zealand.

Zaman, T., & Winschiers-Theophilus, H. (2015). Penan's Oroo'Short Message Signs (PO-SMS): Co-design of a Digital Jungle Sign Language Application *Human-Computer Interaction–INTERACT 2015* (pp. 489-504): Springer.

Zhao, T., & McDonald, S. (2010). *Keep talking: an analysis of participant utterances gathered using two concurrent think-aloud methods*. Paper presented at the Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI2006), pp. 581 – 590.

Table and Figure Captions

List of Figures

| Figure # | Description – Caption |
|----------|---|
| Figure 1 | Snapshots of adding task and measuring ISO9241-11 metrics using Mi-UXLab 1.0 (Soo Shi Tzuaan, Chuan Ngip Khean, Norfarhana Abdollah, Francis Goon, Catherine Ng, 2015) |
| Figure 2 | The CTA and the Textual Feedback (red square) UX evaluation process. |
| Figure 3 | Amount of feedback per individual per task according to the Han classification (left) and the Kujala classification (right) adjusted for differences between tasks and individuals. |

List of Tables

| Table # | Description – Caption |
|---------|---|
| Table 1 | ‘Han’ and ‘Kujala’ classifications (Adapted from Clemmensen, Hertzum, Yang, & Chen, 2013). |
| Table 2 | Status and Gender. |
| Table 3 | Amount of feedback per task and per individual corrected for differences between individuals and tasks, as well as feedback ratios (textual feedback to CTA feedback) |
| Table 4 | Ratio of feedback ratios: The ratio of the feedback ratio (textual to CTA) of high status individuals to the feedback ratio of low status individuals. |

Biographies

Ashok Sivaji is pursuing PhD at National University of Malaysia. He currently works as a Senior Staff Engineer at MIMOS Technology Solutions (MIMOS Berhad) as the Usability/UX Lab Lead. He is interested in cultural usability/UX testing, other areas evaluation methodologies, cloud testing and automation.

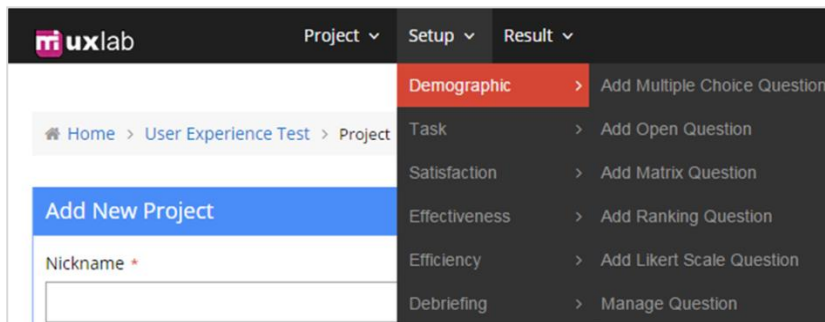
Torkil Clemmensen is a professor with special duties in Human-Computer Interaction at the Department of IT Management, Copenhagen Business School, Denmark. His interest is in psychology as a science of design, including human work interaction design.

Søren Feodor Nielsen is a professor of statistics and the director of the Center for Statistics at Copenhagen Business School, Denmark. His main research interest is in developing models, methods and theory suitable for incomplete data problems but he has also contributed to nonparametric statistics and econometrics.

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

Figure 1

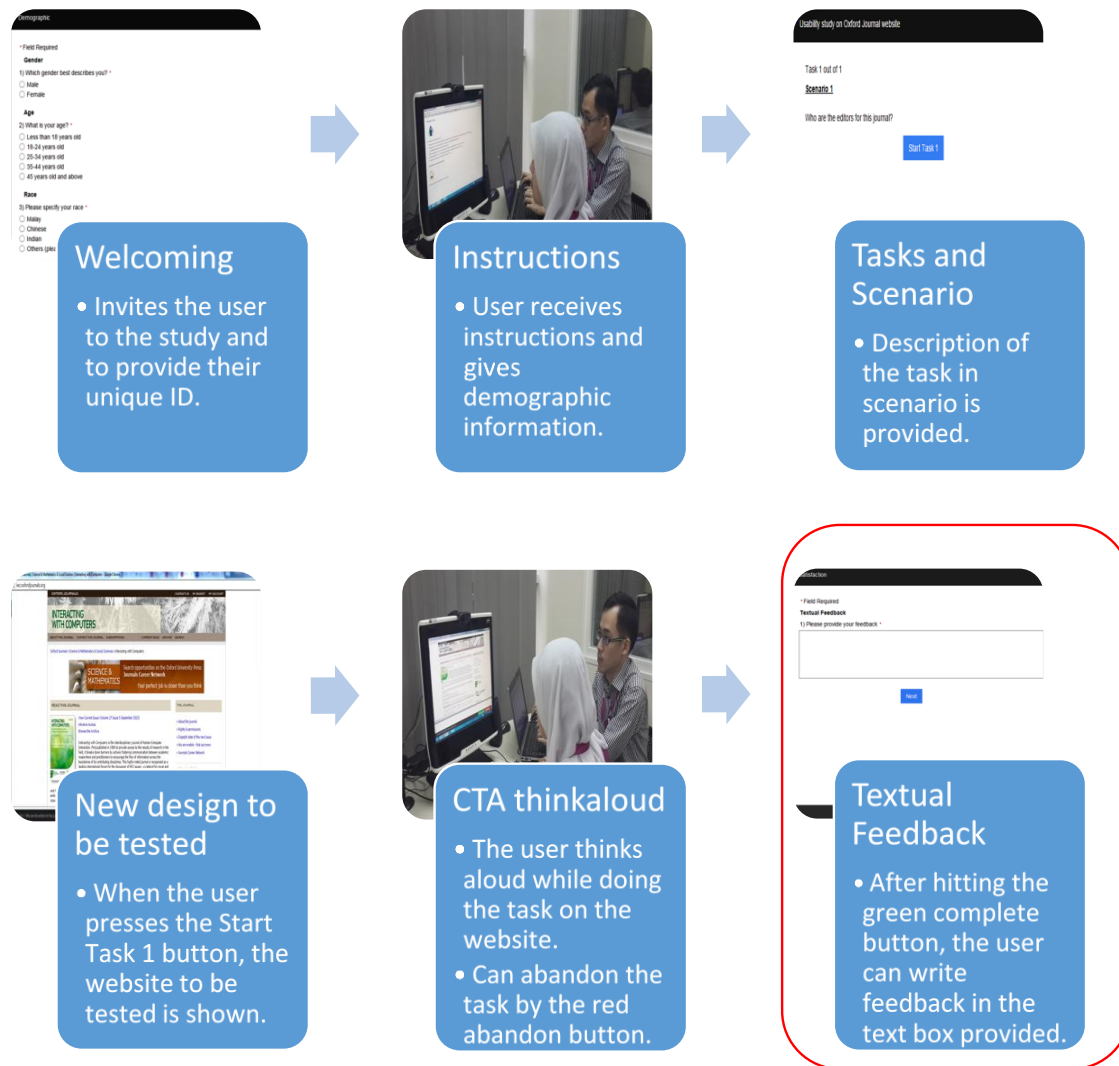
Snapshots of adding task and measuring ISO9241-11 metrics using Mi-UXLab 1.0 (Soo Shi Tzuaan, Chuan Ngip Khean, Norfarhana Abdollah, Francis Goon, Catherine Ng, 2015)



To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

Figure 2

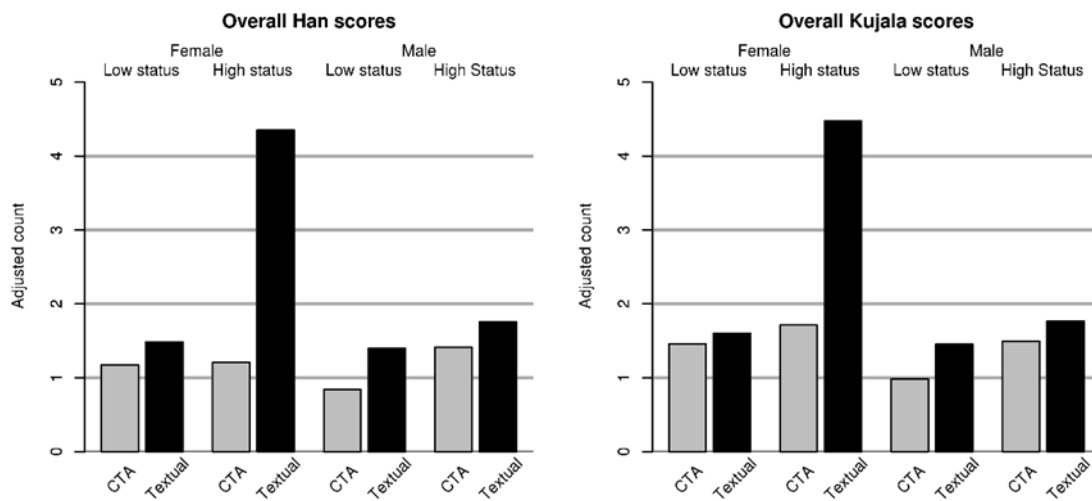
The CTA and the Textual Feedback (red square) UX evaluation process.



To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

Figure 3

Amount of feedback per individual per task according to the Han classification (left) and the Kujala classification (right) adjusted for differences between tasks and individuals.



To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

Table 1

TABLE 1.
‘Han’ and ‘Kujala’ classifications (Adapted from Clemmensen, Hertzum, Yang, & Chen, 2013).

| <u>‘Han’ classification^a</u> | <u>Category definitions</u> |
|--|--|
| Performance | How well users perceive and interpret the interface of a system, how fast the users get used to the product and how well they remember it, and the users control activity and its results; |
| Image/Impression | Basic sense (the primitive and direct image/impression stemming from the design characteristics), description of image of a system, and evaluative feeling |
| <u>‘Kujala’ classification^b</u> | <u>Category definitions</u> |
| General relationship and user experience with the system | Any experience with systems that users find meaningful and important |
| Attractiveness of the system | General attractiveness (appeal) of the product in the users’ own eyes and in those of their friends; more than users’ rational or practical experiences |
| Ease of use of the system | Product is easy and effortless to use |
| Utility of the system | Product serves an important function for the user |
| Degree of usage of the system | Degree of usage which affects user experience over time, related to quality of experience |

Note. ^a ‘Han’ classification (Han, Yun, Kwahk, & Hong, 2001)

^b ‘Kujala’ classification (Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, & Sinnelä, 2011)

Table 2

| TABLE 2 Status and Gender. | | | | | |
|-------------------------------|--------|--------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Gender | Status | No of participants | Average no of tasks per participant | Minimum no of tasks per participant | Maximum no of tasks per participant |
| Female | Low | 14 | 10.00 | 6 | 12 |
| Female | High | 6 | 7.33 | 6 | 8 |
| Male | Low | 7 | 12.00 | 6 | 22 |
| Male | High | 13 | 11.54 | 4 | 22 |

Note. N=171 tasks

Table 3

TABLE 3.

Amount of feedback per task and per individual corrected for differences between individuals and tasks, as well as feedback ratios (textual feedback to CTA feedback)

| Score | Gender | Status | CTA | Textual | Ratio |
|-------------------------------|-------------|-------------|-------|---------|-------|
| Han ^a | Overall | low | 1.175 | 1.485 | 1.264 |
| | | Female high | 1.209 | 4.351 | 3.600 |
| | | Male low | 0.844 | 1.398 | 1.657 |
| | | high | 1.414 | 1.758 | 1.243 |
| | Performance | Female low | 0.845 | 0.870 | 1.029 |
| | | high | 0.643 | 3.047 | 4.737 |
| | | Male low | 0.623 | 0.815 | 1.308 |
| | | high | 0.953 | 1.008 | 1.058 |
| | Impression | Female low | 0.304 | 0.594 | 1.957 |
| | | high | 0.609 | 1.371 | 2.250 |
| | | Male low | 0.196 | 0.523 | 2.667 |
| | | high | 0.378 | 0.629 | 1.667 |
| Kujala EaseOfUse ^b | Overall | low | 1.457 | 1.604 | 1.101 |
| | | Female high | 1.713 | 4.476 | 2.612 |
| | | Male low | 0.985 | 1.453 | 1.475 |
| | | high | 1.495 | 1.768 | 1.183 |
| | General | Female low | 0.406 | 0.390 | 0.960 |
| | | high | 0.625 | 1.245 | 1.992 |
| | | Male low | 0.344 | 0.393 | 1.140 |
| | | high | 0.436 | 0.410 | 0.941 |
| | Attractive | Female low | 0.126 | 0.136 | 1.077 |
| | | high | 0.136 | 0.272 | 2.000 |
| | | Male low | 0.098 | 0.228 | 2.333 |
| | | high | 0.114 | 0.202 | 1.769 |
| | Utility | low | 0.471 | 0.511 | 1.086 |
| | | Female high | 0.382 | 0.892 | 2.333 |
| | | Male low | 0.215 | 0.334 | 1.556 |
| | | high | 0.257 | 0.257 | 1.000 |
| | Degree | low | 0.341 | 0.365 | 1.069 |
| | | Female high | 0.426 | 1.803 | 4.231 |
| | | Male low | 0.184 | 0.231 | 1.250 |
| | | high | 0.407 | 0.502 | 1.233 |
| | Degree | Female low | 0.019 | 0.074 | 4.000 |
| | | high | 0.094 | 0.125 | 1.333 |
| | | Male low | 0.056 | 0.093 | 1.667 |
| | | high | 0.080 | 0.121 | 1.500 |

Note. Score categories ^{a b} have been previously used by used by Clemmensen, Hertzum, Yang, & Chen (2013)

To cite this article: Sivaji, A., Nielsen, S. F., & Clemmensen, T. (2016). A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations. *International Journal of Human–Computer Interaction*, 1-14.

Table 4

TABLE 4.

Ratio of feedback ratios: The ratio of the feedback ratio (textual to CTA) of high status individuals to the feedback ratio of low status individuals.

| Q | Scores | estimate | Female | estimate | Male |
|---------------------|-------------|----------|-----------------|----------|------------------|
| | | | (lower ; upper) | | (lower ; upper) |
| Han ^a | Overall | 2.849*** | (1.800 ; 4.510) | 0.750 | (0.460 ; 1.222) |
| | Performance | 4.602*** | (2.558 ; 8.277) | 0.809 | (0.446 ; 1.466) |
| | Impression | 1.150 | (0.537 ; 2.464) | 0.625 | (0.260 ; 1.500) |
| Kujala ^b | Overall | 2.373*** | (1.567 ; 3.592) | 0.802 | (0.500 ; 1.286) |
| | General | 2.074*** | (2.061 ; 2.087) | 0.826*** | (0.821 ; 0.831) |
| | Attractive | 1.857 | (0.500 ; 6.899) | 0.758 | (0.234 ; 2.451) |
| | EaseOfUse | 2.149 | (0.875 ; 5.279) | 0.643 | (0.229 ; 1.809) |
| | Utility | 3.958*** | (1.835 ; 8.538) | 0.987 | (0.358 ; 2.718) |
| | Degree | 0.333 | (0.039 ; 2.874) | 0.900 | (0.166 ; 4.866) |

Note. Score categories ^{a b} have been previously used by used by Clemmensen, Hertzum, Yang, & Chen (2013) *** denotes significant at the 0.001-level.