COMPUTER SOCIETY OF INDIA

Playing away from home – usability testing in a global world

John Murphy, Steve Howard**, Jesper Kjeldskov*

1. Introduction

Usability testing is one of a suite of techniques used during the design and development of a product to uncover usability problems. It has been termed the 'gold standard' of testing [4] for determining relevant problems during design. It may be distinguished from other testing and evaluation approaches based on expert reviews, such as design walkthroughs and heuristic evaluations, by the involvement of participants who are in some sense representative of the target audience [8].

Usability tests are created by usability test designers, conducted by usability testers (often the same person) and conducted on test subjects sampled from the target audience of a current or envisioned product. In the case of the design of international products, however, there are a number of additional complications in the design of both the product and the usability testing approach. Not only must the design accommodate different languages and a multitude of elements including customary beliefs, social norms and material traits for particular groups known as 'cultures' [1], but also usability testing may well be required internationally to ensure representative participants or test subjects. In this instance, cultural and language aspects of communication as well as the physical distance between the foreign usability test designer and the local test subject must be taken into consideration. Whilst there is some literature to inform the usability practitioner of the many options and factors involved in designing and conducting international usability tests [7], there appears to be little to assist in the evaluation and trade–off of these options when faced with the challenge of international testing. Without the careful planning and design of an international test, the usability tester may encounter major communication problems, spend an inappropriate amount of project money and be unable to identify significant usability problems.

Section 2 of the paper develops a framework for the usability test designer to create effective international usability tests by enabling comparison of different test situations against a set of criteria. These criteria were informed by the international literature, and developed on the basis of real world practice. In section 3 we turn our attention to remote usability testing. We conclude with some lessons learned.

2. Case 1 - International usability Testing

The framework proposed below articulates experiences gained during usability testing of a multi-lingual informational web site of some 10,000 pages. The design team, based in Australia, comprised a multi-disciplinary group including technical, business, human computer interaction (HCI) and creative/graphic design skills.

Testing in New Zealand was conducted in Auckland with test data recorded in English. Testing in Singapore was conducted in Singapore with test data recorded in English. Testing in the United States was conducted in San Francisco with test data recorded in English. Testing in Germany was conducted in Frankfurt with test data recorded in

**Novell Pty Ltd, 574 St. Kilda Road, Vic 3004, Australia, john.murphy@novell.com*

***Department of Information Systems, University of Melbourne, Vic 3010, Australia, showard@unimelb.edu.au*

****Department of Computer Science, Aalborg University, DK-9220 Aalborg East, Denmark, jesper@cs.aau.dk*

German. Finally, testing in Japan was conducted in Tokyo with test data recorded in both Japanese and English through live audio translation.

Usability Testing Models

Three different usability testing models, 'local', 'foreign' and 'mixed' were defined based on the persons conducting the test. Thus the 'local' test model utilised 'local' testers very familiar with the native language and culture of the test country and most likely to have lived in the country for a number of years, the 'foreign' model utilised 'foreign' test persons less familiar with the culture and language of a country and the mixed model employed a combination of the two.

Local usability test model (Testing in native language). For this model, written test instructions, questions and background information were translated from English into the native language. The local tester is trained and instructed remotely (by the foreign test designer using telephone and e-mail), conducts the usability test in the subjects' native language, transcribes native language results from audio/visual data, translates results into English language, and then the foreign test designer analyses the data.

Foreign usability test model (Testing in English). For this model, written test instructions, questions and background information remain in the English language. No training is required as the foreign test designer travels to the local site and conducts the test in English, transcribes the results from audio/visual data, and analyses the data.

Mixed usability testing model (Testing in native language). For this model, written test instructions, questions and background information were translated from English into the native language. The local tester is trained and instructed by a foreign test designer either remotely or on the local site, then some combination of the local tester and foreign test designer conduct the test in the native language. The results are transcribed and translated from the audio/visual data by some combination of local and foreign persons and then the foreign test designer analyses the data.

Framework to support Usability Test Models

We are interested in minimising the communication gap between the usability test designer, tester and the test subjects. Our previous industry experience had shown that the quality of communication between personnel is a significant factor in the overall effectiveness of the usability test. Thus, the notion of a 'communication distance' proved to be a useful metaphor.

Initial analysis of the relationship between the foreign test designer and the local test subject, denoted $\mathbf{R}_{D\text{-}S}$, led to the definition of two dimensions or distances of communication that had a direct bearing on the quality of the usability test results. These distances were verbal language (Lang-v) and cultural understanding (Culture). Distances were minimised when all participants in the test shared both language and culture, and maximised when language and culture diverged. Thus four combinations of communication distances based on these two dimensions emerge:

$R1_{D\text{-}S}$-poor verbal match (✗ Lang-v) & little cultural match (✗ Culture)

$R2_{D\text{-}S}$ - poor verbal match (✗ Lang-v) & close cultural match (✓ Culture)

$R3_{D\text{-}S}$ - good verbal match (✓ Lang-v) & little cultural match (✗ Culture)

$R4_{D\text{-}S}$ - good verbal match (✓ Lang-v) & close cultural match (✓ Culture)

For relationships $R1_{D\text{-}S}$ and $R2_{D\text{-}S}$, we found it necessary to introduce a local tester to mitigate the verbal communication distance between the foreign test designer and the local subjects. For testing relationship $R3_{D\text{-}S}$, we found a local tester or foreign test designer may conduct the test depending on the magnitude of the cultural distance and the test designers' ability to mitigate the communication distance through the test design. In the case of testing relationship $R4_{D\text{-}S}$, we found the foreign test designer should conduct the test, as a local tester would simply add unnecessary communication layers between the test designer and the test subject.

Where these test designer-subject communication distances were found to warrant the use of a local tester to conduct the test in lieu of the foreign test designer an additional set of relationships, denoted $\mathbf{R}_{D\text{-}T}$, and thus a new set of communication distances were introduced. The communication distances between the foreign test designer and the local tester, again based on industry experience, were described as verbal language (Lang-v), written language (Lang-w) and usability testing method knowledge (HCI-k).

The set of possible combinations of communication distances are:

R1D-T - poor verbal match (✗ Lang-v), poor written skills (✗ Lang-w), poor knowledge of testing (✗ HCI-K)

R2D-T - poor verbal match (✗ Lang-v), poor written skills (✗ Lang-w), good knowledge of testing (✓ HCI-K)

R3D-T - poor verbal match (✗ Lang-v), good written skills (✓ Lang-w), good knowledge of testing (✓ HCI-K)

R4D-T - poor verbal match (✗ Lang-v), good written skills (✓ Lang-w), poor knowledge of testing (✗ HCI-K)

R5D-T - good verbal match (✗ Lang-v), good written skills (✓ Lang-w), good knowledge of testing (✓ HCI-K)

R6D-T - good verbal match (✗ Lang-v), good written skills (✓ Lang-w), poor knowledge of testing (✓ HCI-K)

R7D-T - good verbal match (✗ Lang-v), poor written skills (✓ Lang-w), poor knowledge of testing (✗ HCI-K)

R8D-T - good verbal match (✗ Lang-v), poor written skills (✓ Lang-w), good knowledge of testing (✓ HCI-K)

Clearly the ideal foreign test designer/local tester relationship is $R5_{D-T}$. In this case the test designer was able to easily instruct the local tester in the purpose of the testing and was confident of retaining good control over the usability testing process and receiving good quality transcriptions and test result translations. The other foreign test designer/local tester relationships were found to give rise to various issues around communicating the purpose of the test, maintaining control over the process of the test and communicating data from the test.

Lessons Learned

Table 1 maps the foreign test designer/local subject and foreign test designer/local tester relationships as two axes. The table cells are populated with proposed test models drawn from the set of local, foreign and mixed model types, as defined above. Rather that discuss all cells of the table, we will highlight some of the more interesting and leave the reader to explore the rest.

The 'Foreign' tester model was found to be suitable for all test relationships in column $R4_{D-S}$ where there was good verbal communication and a good cultural understanding between the foreign test designer and the local test subject. In the case study, this model was adopted for the New Zealand testing where language communication and cultural differences between the test designer and the test subject were not found to compromise the quality of the testing. Testing for Singapore was conducted by a foreign test designer/tester given that the user goals, scenarios and storyboards had been adapted for the local culture, the travelling distances from Australia were relatively small and there was no local tester available at the time of testing. Conversely, in the United States a local tester was used, based on availability and the significant travelling distances from Australia.

Table 1: Summary of the relations between test designers, testers and subjects

		R1 $_{D-S}$ ✗ Lang-v $_{D-S}$ ✗ Culture $_{D-S}$	**R2** $_{D-S}$ ✗ Lang-v $_{D-S}$ ✓ Culture $_{D-S}$	**R3** $_{D-S}$ ✓ Lang-v $_{D-S}$ ✗ Culture $_{D-S}$	**R4** $_{D-S}$ ✓ Lang-v $_{D-S}$ ✓ Culture $_{D-S}$
R1 $_{D-T}$	✗ Lang-v $_{D-T}$ ✗Lang-w $_{D-T}$ ✗ HCI-k $_{D-T}$	Avoid this situation	Avoid this situation	Foreign	Foreign
R2 $_{D-T}$	✗ Lang-v $_{D-T}$ ✗ Lang-w $_{D-T}$ ✓ HCI-k $_{D-T}$	Mixed (purpose & data issues)	Mixed (purpose & data issues)	Foreign	Foreign
R3 $_{D-T}$	✗ Lang-v $_{D-T}$ ✓ Lang-w $_{D-T}$ ✓ HCI-k $_{D-T}$	Mixed (purpose issues)	Mixed (purpose issues)	Foreign	Foreign
R4 $_{D-T}$	✗ Lang-v $_{D-T}$ ✓ Lang-w $_{D-T}$ ✗ HCI-k $_{D-T}$	Mixed (purpose & control issues)	Mixed (purpose & control issues)	Foreign	Foreign
R5 $_{D-T}$	✓ Lang-v $_{D-T}$ ✓ Lang-w $_{D-T}$ ✓ HCI-k $_{D-T}$	Local	Local	Foreign or Local	Foreign
R6 $_{D-T}$	P Lang-v $_{D-T}$ P Lang-w $_{D-T}$ ✗ HCI-k $_{D-T}$	Mixed (control issues)	Mixed (control issues)	Foreign or Mixed (control issues)	Foreign
R7 $_{D-T}$	✓ Lang-v $_{D-T}$ ✗ Lang-w $_{D-T}$ ✗ HCI-k $_{D-T}$	Mixed (control & data issues)	Mixed (control & data issues)	Foreign or Mixed (control & data issues)	Foreign
R8 $_{D-T}$	✓ Lang-v $_{D-T}$ ✗ Lang-w $_{D-T}$ ✓ HCI-k $_{D-T}$	Mixed (data issues)	Mixed (data issues)	Foreign or Mixed (data issues)	Foreign

'Foreign or Mixed' models were proposed for ($R6_{D-T}$ $R3_{D-S}$), ($R7_{D-T}$ $R3_{D-S}$), and ($R8_{D-T}$ $R3_{D-S}$) as issues of control of the testing process and quality of the data were variously present in the communication distances between the foreign test designer and the local tester.

A 'Local' test model was found to be suitable for situations ($R5_{D-T}$ $R1_{D-S}$) and ($R5_{D-T}$ $R2_{D-S}$) because, given poor communication between the foreign test designer and the local subject, the relationship between foreign test designer and local tester was ideal with good verbal, written language and HCI communication. Relationship situation ($R5_{D-T}$ $R1_{D-S}$) was mapped to the German testing in the case study and a local tester was used.

'Mixed' test models were used for the majority of the relationship situations in columns $R1_{D-S}$ and $R2_{D-S}$ where the foreign test designer needed to mitigate the problems caused by confusion over purpose, control and data quality, and to maintain the effectiveness of the test. For case study 1, ($R7_{D-T}$ $R1_{D-S}$) was mapped to the Japanese testing. In this case, whilst the local tester's level of method knowledge was relatively high, it was orientated more towards traditional marketing methods such as focus groups, rather than usability testing. This meant that whilst the Japanese tester was experienced at probing into responses, this was usually in the context of a focus group rather than a one on one interview and so control of the testing was seen as an issue.

3. Case 2 - Remote Usability Testing

Case 1 considers situations where usability testing is conducted in the country of residency of the test subjects. This is not always practicable however and occasions arise where testing is 'done for', rather than 'with' a design team working in some other part of the world. The costs savings gained from the fast turn around times made possible when time multiplexing is deployed makes such distributed usability practice highly attractive. Such 'outsourcing' (or 'insourcing', essentially outsourcing to other units within the same company) is of great commercial and research interest at the moment.

Background

A major international commercial company was developing a new product, which was intended to support collaborative work amongst non-technical commercial workers. For this product to succeed, non-technical users must be able to use the tool easily. A significant component of the ease of use of the product was the users' ability to create a clear and coherent mental model of the system. In order to evaluate the design, we conducted a typical usability evaluation. A secondary goal of the evaluation was to determine whether the interface, largely screen design and dialogue flow, supported the individual tasks of file creation and sharing, and the mechanisms for achieving that, e.g. issuing and accepting an invitation to share.

The company has offices in Australia that, aside from day-to-day business are involved in HCI based research in collaboration with the Universities of Melbourne (Australia) and Aalborg (Denmark). This program has been running for over four years and encompasses research collaboration on developing usability techniques, industry projects, teaching and sponsorship of a state of the art usability laboratory in The University of Melbourne, Department of Information Systems.

Challenges to the evaluation

In this section we focus on three challenges that we found particularly problematic: location, location and location!

Location – Geography

Conducting a remote usability evaluation places a particular burden on communication and the maintenance of situation awareness [5,2]. Multiplexed time zones can aid in rapid turn around of results but only if synchronous interaction is not required at times of unavailability, or indeed uncivilised hours, and only if the disparate teams are 'talking the same language'. Prior to commencing the evaluation, and drawing on a mix of local knowledge, documentation, email and teleconferencing skills, we harvested as much understanding of the remote situation as we were able. Conductors of such remote tests face numerous hurdles, including:

Elevated expectations on rapid turn around time and streamlined reporting requirements

Preferences for and bias toward different data collection methods and data types than are present at the remote site

Interest in the process (how the evaluation was conducted) as opposed to merely the product and the findings from the evaluation. Remote customer sites, unable to experience the evaluation directly, will often request a process debrief, thus ensuring that they understand the origin of the findings.

Location - Sector

Combining multiple sectors (in this case industry practitioners and university researchers and research students) is a real strength of our approach. The established and ongoing relationship between the company and the Universities of Melbourne and Aalborg allows us to respond rapidly to emerging opportunities under the rubric of a tested agreement. However, as a cross sectoral collaboration it is not without its frustrations (but see Lambert, 2003 for some solutions).

Location – Development phase

Usability evaluators, be they located in industry or universities, are unfortunately rather experienced at being introduced too late into the lifecycle to have a major impact on the product. It was therefore rewarding to be invited to comment at a relatively early stage in a product's development (see [8] for a discussion of the importance of life cycle positioning). However, an opportunity to comment early should not be confused with an occasion for unbridled creativity! We took great care in:

Gauging the degrees of freedom available to the

development team in responding to the identified usability flaws.

Streamlining the return of findings and feedback. It is hard to overstate the importance of the representational form of any feedback provided to the design team.

Balancing a critical perspective on the present design with a constructive account of the next.

Evaluation Description

The product usability evaluation was conducted over two days at a state-of-the-art usability laboratory at The University of Melbourne, Australia. The evaluation was done in a collaborative working environment with real life scenarios and tasks requiring the use of other software such as an e-mail client and folder and file manipulation tools. Two independent usability evaluations were conducted; a user-based evaluation and a heuristic walkthrough. The user-based evaluation was based on think-aloud protocol, involving three triads of test subject working collaboratively through the product. The test subjects were physically separated from each other and could only collaborate using the product and e-mail. The user-based evaluation sessions were recorded on digital video capturing overviews of all three test subjects and their respective computer monitors.

Secondly, three Doctoral students specializing in Human-Computer Interaction conducted a Heuristic Walkthrough of the product software using the scenarios described above. The Heuristic Walkthrough session lasted approximately ninety minutes and was facilitated by the first author who recorded usability problems by the expert reviewers for later analysis and comparison with the user based data.

Reporting the results

The evaluation had several audiences - project stakeholders in the form of product managers and senior product development staff, company HCI professionals based in the United States and most importantly, product engineers actually working on the product. Each of the different audiences required different information; the project stakeholders were mostly concerned with the feasibility of outsourced usability evaluation in terms of costs, resources and overall effectiveness; the HCI professionals were concerned to validate the evaluation process and results to both ensure the quality of the results for the product work ahead, but more importantly to investigate how and whether this process and resource might be able to support on-going company HCI work; and the product engineers wanted "design ready" findings. From a product engineering perspective, it was understood that the reporting of problems would not be useful without some accompanying proposal of a solution, particularly in the case of significant or complex problems.

Given these different audiences and reporting requirements, a number of different reporting mechanisms were employed. A telephone conference was used to report high level findings, costing and an overall project feasibility to stakeholders and HCI staff. A short highlights video of the usability laboratory, equipment and 'snippets' of the actual evaluation was prepared to present the evaluation process to the company HCI staff and stakeholders. A written evaluation report was prepared explaining the results in detail for product engineers and company HCI staff. It was structured with a usability problem summary table, a discussion of each of the usability issues, user interface design solution ideas and a description of the test.

Lessons Learned

The product software was still under development and prone to errors at time of testing. These factors led to a significant increase in the standard level of support and intervention required for usability testing. For instance, participants required support where the ability of a user was significantly different to the other team members and needed to maintain timely collaboration with colleagues. In cases where participants acted as team leaders, sharing files and occasionally becoming entangled in Microsoft file-sharing, they were assisted back to the product environment to maintain the flow of the task.

In relation to the process of evaluating the product, significant contextual knowledge is required to ensure the testing is effective. Budgets, timelines for product development and intended audience are all used to support the design of the evaluation. Other subtler issues concern target market share, future plans to integrate with other products, competing products and number and skill of

engineers available to rework the product post evaluation.

The video highlights were found to be extremely valuable as a fast effective mechanism of providing a significant amount of information to the project stakeholders and company HCI staff. The video highlights viewed in conjunction with the teleconference meant that the presentation and ensuing discussion quickly became informed and focused.

4. Concluding Comments

We have described a framework (Case 1) to assist usability test designers who are tasked with designing and conducting international usability tests, where the testing takes place in the subjects' country of residence. Secondly, we reflected on a further mode of working (Case 2), where the tests are conducted away from the participants' country of residence. Amongst the many options available for international usability testing, it is hoped that usability practitioners may be able to use this framework and the reflections to understand the impact of communication distance, and thus balance the use of foreign and local test persons, and remote working opportunities, to create more effective usability tests.

Acknowledgements

We thank the test subjects who participated in the product evaluation and the Doctoral students at the University of Melbourne who conducted the heuristic walkthrough: Jeni Paay, Sonja Pedell and Jan Skjetne. Thanks also to Steve Goschnick for assistance with the set up for Case 2. This work is contributing to Novell's 'Customers of the Future' strategy, which acknowledges support from the Australian Research Council through Linkage Grant LP0347459. Parts of this text have appeared previously in Murphy [5] and Murphy et al [6].

References

1] Del Galdo, E.M. & Nielsen, J. (Eds.): *International User Interfaces*. United States: John Wiley & Sons, 77-78.

2] Hartson, H., Castillo, J., Kelso, J., Kamler, J., & Neale, W. (1996) Remote Evaluation: The Network as an Extension of the Usability Laboratory. *Proceedings of CHI'96*. http://www.acm.org/sigchi/chi96/proceedings/papers/Hartson/hrh_txt.htm (accessed 17/8/2005 3:46 PM)

3] Lambert (2003) Lambert Review of Business-University Collaboration http://www.hm-treasury.gov.uk/media/EA556/lambert_review_final_450.pdf (accessed 17/8/2005 3:46 PM)

4] Landauer, T.K. (1995). *The Trouble with Computers-Usefulness, usability, and productivity*. MIT Press, Cambridge, Mass.

5] Murphy, J. (2001) Modelling 'Designer – Tester – Subject' Relationships in International Usability Testing. *IWIPS 2001 Designing for Global Markets* 3 July 12-14 Milton Keynes, United Kingdom Editors, Donald L. Day and Lynne Dunckley

6] Murphy J., Howard S., Kjeldskov J. and Goschnick S. (2004) Location, Location, Location: Challenges of Outsourced Usability Evaluation. In *Proceedings of the Workshop on Improving the Interplay between Usability Evaluation and Interface Design, NordiCHI 2004*. Tampere, Finland

7] Nielsen, J. (1999). *Designing Web Usability - The Practice of Simplicity*. New Riders Publishing, Indianapolis, USA, 333-344.

8] Rubin, J. (1994) *Handbook of usability testing: how to plan, design, and conduct effective tests*. New York: Wiley.