

Technical Report No: 2002/05

***Low Cost Remote Evaluation with Synchronous
Groupware***

Lynne Dunckley

Lucia Rapanotti

Jon G. Hall

2002

***Department of Computing
Faculty of Mathematics and Computing
The Open University
Walton Hall,
Milton Keynes
MK7 6AA
United Kingdom***

<http://computing.open.ac.uk>



Low Cost Remote Evaluation with Synchronous Groupware

Lynne Dunckley
Thames Valley University,
UK

L.Dunkley@tvu.ac.uk

Lucia Rapanotti
The Open University, UK

L.Rapanotti@open.ac.uk

Jon G. Hall

The Open University, UK

J.G.Hall@open.ac.uk

Abstract

The paper reports a refinement of the remote evaluation method of [2], that includes synchronous groupware with audio over the Internet. The method applies ideas from contextual inquiry and participatory design for the formative evaluation of a prototype linked to a subsequent redesign process. The investigation involves the Open University – the major provider of distance education in the UK and Europe.

The paper describes the refined method, and discusses the findings of an investigation into its characteristics.

Keywords

Remote evaluation, contextual inquiry, synchronous groupware.

1. Introduction

Information systems are being developed to support increasingly complex tasks. Many of these systems are distributed, involving geographically remote users. The growth of network and, particularly, Web technology means that users are communicating with central systems using a wide range of machines and operating systems.

Evaluating and testing such systems is difficult: the network itself and the remote work setting are important parts of the system and produce usage patterns difficult to reproduce in a laboratory. At the same time, bringing developers and users together in remote locations can be impractical, and often prohibitively expensive. These problems are particularly acute where the software is developed for an international market, typically involving usability evaluations in different countries.

In order to overcome such difficulties, remote usability evaluation methods have been developed. These include subjective techniques, such as reports from users, user-identified critical incidents, questionnaires, interviews and ethnographic techniques, as well as objective techniques that involve automatically collecting data about the application and its users [1,3,4,5].

Recently, we have developed a low cost subjective remote evaluation method for the formative evaluation of high fidelity prototypes linked to the subsequent redesign

process. The method, initially called the ‘write-along’ method [2], was strongly influenced by ideas of contextual inquiry and participatory design, and aimed at simulating a conversation between developer and remote user. Rather than the verbalisation of their experience with a (semi-) working prototype, a user located remotely would record their experiences through a commentary, prompted by a questionnaire, and written concurrently. An investigation [2] found the method highly effective in identifying usability problems. However, it was noted that it suffered from a

‘temptation for users to explore and interact before they have completed a written answer’

with a concomitant loss of the conversational nature of the dialogue the method was attempting to simulate. Not unrelated was the fact that critical incidents caused unrecoverable breakdowns of the evaluation, whenever the user could not overcome usability problems with the interface. These facts prompted refinements of the method to use synchronous Internet-based communication technology to re-establish a conversation between user and evaluator. The paper reports on some of the findings of further investigations into the new method.

The paper is structured as follows. Section 2 describes the prototype evaluated in the study. Section 3 summarizes the method. Section 4 and 5 describe the trials of the method mediated by two Internet conferencing tools, Lyceum and NetMeeting™. Section 6 reports on the findings. Finally, Section 7 concludes the paper.

2. The prototype

The investigation we describe involves the Open University (OU), the major provider of distance-learning education at university level in the UK and Europe. Increasingly, the OU has moved to the electronic provision of its teaching services via the Internet. A component of the course materials, the continuous assessment system, provides the major vehicle for driving the students’ distance learning experience: students complete assessments as they study a course; the scripts are marked by associate lecturers (aka tutors) located remotely both from the student and the university campus; marked scripts are returned to students for feedback on their progression, via the OU.

The move to electronic provision for continuous assessment has been facilitated by the development of software for script marking, namely the eTMA (electronic Tutor Marked Assessment) Marking Tool. The original design of the eTMA Marking Tool was subject to conflicting requirements. In operation, the software should require minimum support and technical backup. The usage pattern would be one of long gaps followed by intensive use for short periods. Visibility and affordance of design were therefore crucial issues.

3. The method

The low cost evaluation method attempts to capture the user's response to prototype interfaces in a contextual manner and provides a framework to simulate a remote conversation between the developer and the user.

The method was strongly influenced by the ideas of contextual inquiry [6] and participatory design [10]. Its aim is to conduct the evaluation in the context in which the software is used, and to promote developer-user conversations that the developer does not dominate. A key part of the method is to establish a conversation between user and developer supported by a series of questions structured within Norman's seven stages of action [7]. This enables users to articulate the way in which they would use a prototype interface to complete their normal tasks.

Operationally, we use a mix of synchronous and asynchronous remote evaluation sessions. The synchronous remote sessions are mediated by Internet-based groupware. These require users and evaluators to be on-line at agreed times. The asynchronous sessions are supported by a package that provides an electronic, user-completed journal to capture users' responses during interaction with the prototype. This was based on the evaluation method used in [2]. The asynchronous sessions involve users only, working on their own and in their own time.

In the investigation we used two groupware systems: Lyceum, an OU proprietary system; and NetMeeting™, a commercial system by Microsoft Inc. The key functionality provided by Lyceum is many-to-many voice conferencing and shared collaborative workspaces; that of NetMeeting™ is to allow the evaluator to view the user's desktop at the same time as peer-to-peer voice conferencing. The different capabilities of the two systems allowed us to adopt different strategies in conducting the on-line sessions. These are detailed in Sections 4 and 5.

In general, reported remote evaluation experiments have taken place in organisational settings, where the work environment could be controlled, and audio and video equipment used [6]. In contrast, our experiment involved users at widely distributed locations; their social context was the home, with background family noise.

Video recording and video conferencing were not feasible options, as most users would not have the equipment; also, communication bandwidth was limited to 56kbs at most.

We used the same prototype of the eTMA Marking Tool used in [2], as its usability problems had already been identified by that study, and through subsequent conventional usability evaluations. This allowed us to measure the effectiveness of the method against the previous studies.

The set of users was drawn from the same population of OU tutors as those of [2]. This is a large population, so we were able to select users with no previous experience of the electronic assessment tool and check that their user profile was comparable to that of the previous experiment. By and large, the users were new to Lyceum and NetMeeting™.

The same question and answer framework as in [2] was used, but the questions were adapted to a briefer style needed for electronic and audio communication over the Internet. We also used the same task scenario, which consisted of four tasks:

1. selecting courses and scripts to mark;
2. setting part marks for the standard mark scheme;
3. marking sample scripts;
4. storing completed marked scripts.

Users were asked to complete tasks 1. to 4. on-line, then explore further completion of the tasks off-line, using the write-along method, and complete a summative questionnaire.

4. Lyceum trial

Lyceum is a client-server voice groupware system developed at the OU, designed to support real-time collaborative eLearning [8]. Lyceum users can participate in real-time voice conferencing and share collaborative spaces and tools over the Internet. The Lyceum client is designed to run on mid-range Windows PCs with standard multi-media support.

Among Lyceum's features, the following were particularly relevant to our investigation:

1. Participants meet in 'virtual rooms', where they talk to each other and share collaborative tools. During the experiment, one room was used for plenary discussions, and separate rooms were used for individual evaluation activities.
2. Users have access to a collaborative whiteboard and text editor. The whiteboard was used during plenary discussions, while the text editor contained the tasks description and the user's commentary during the individual evaluations.
3. The Lyceum voice tools allow participants to talk and facilitate the moderation of plenary sessions.

4.1 The sessions

The Lyceum sessions were designed to allow a number of users and evaluators to work together. We divided the users into groups, each group made up of 4 or 5 people. Each user group engaged in one training and one evaluation session for a total of just over two hours on-line. Training and evaluation sessions were structured as follows. Each started with a plenary session for all participants, followed by individual evaluation activities in separate virtual rooms, followed by a plenary discussion to close. Each session was run by two evaluators.

The plenary discussions allowed the users to debate prototype usability issues, and were moderated by an evaluator, who also recorded the users' comments. During the individual evaluation activities, the two evaluators would visit the users' separate rooms in a round-robin fashion. As they worked through their tasks, the users completed write-along documents, which were visible to the evaluators.

During all the sessions, the users were connected to the Lyceum server via an ISP, while the evaluators worked through a LAN.

5. NetMeeting™ trial

NetMeeting™ is an audio, video and data conferencing system developed and distributed by Microsoft Inc. NetMeeting™ supports real-time Internet conferencing and provides collaborative groupware. Users install the NetMeeting™ application on their desktop and communicate with each other in peer-to-peer mode¹ through the Internet. As with Lyceum, NetMeeting™ runs on mid-range Windows PCs with standard multimedia support.

In our investigation, the most relevant features of NetMeeting™ were peer-to-peer voice communication and the ability to share the user's desktop across the Internet.

5.1. The sessions

In each NetMeeting™ session one user and one evaluator would work together to perform a remote usability evaluation synchronously. In the experiment, we adapted a face-to-face method described in [11] for use with NetMeeting™, which we used to guide the user/evaluator interaction.

In each session, user and evaluator were connected peer-to-peer, which allowed them to converse and share collaborative tools. The user's desktop was shared during

the session allowing the evaluator to observe events occurring on the user's PC.

Difficulties with users' machine base and ISPs meant that it was not possible to obtain audio of an acceptable quality. (Similar problems were reported in previous experiments with low cost conferencing software, including NetMeeting™ [9].) Therefore, the NetMeeting™ sessions were conducted on a campus LAN with evaluator and user located in separate rooms, and with two high specification laptops, appropriately configured and tested for NetMeeting™ use. However, even in this configuration, the audio quality suffered excessive distortion.

6. Findings

Evaluation data were collected during the on-line sessions by the evaluators, as well as extracted from the user-compiled evaluation commentary and summative questionnaire.

In terms of identifying usability problems with the prototype interface, an analysis of the data showed no significant differences in the number of usability problems identified by the combined synchronous/asynchronous methods compared to the write-along method of [2], once the difference in the number of users is taken into account (see Table 1). For the NetMeeting™-based method, sessions with the 4 users located on campus seemed to identify fewer usability problems compared to the write-along method. Note, however, that because of the tendency for the audio to break up, fewer tasks were completed on-line.

	Write-Along	Lyceum	NetMeeting™
<i>No. of users</i>	13	9	4
<i>Total of usability problems</i>	34	23	8
<i>Mean of usability problems</i>	6.23	7.55	5.75
<i>Std. deviation</i>	3.76	2.78	1.89

Table 1 - Comparison between the write-along and the on-line methods.

Although the differences in the number of usability problems do not appear to be significant, there was some difference in the nature of the usability problems identified. As can be seen from Table 2, the severity ratings in the Lyceum trial (the R index in the table) are the same or slightly higher than those in the write-along trial, which implies that the more severe problems were identified in the Lyceum sessions. Also, the Lyceum plenary sessions identified four usability problems not identified by the previous evaluation.

¹ NetMeeting™ can also be used in client/server mode; we did not consider this mode in our experiment.

eTMA Marking Tool Windows	Write- Along (13 users)	Lyceum (9 users)		
		On-line	Off-line	To tal
Marking Scheme	10 (R= 3.5)	2 (R= 3.5)	3 (R= 3.6)	5
Main Window	7 (R= 2.9)	2 (R= 3.0)	3 (R= 3.0)	5
Score Allocation	17 (R=3.4)	9* (R= 3.8)	4 (R= 3.0)	13
<i>Total</i>	34 (R= 3.2)	13 (R= 3.5)	10 (R= 3.5)	23

Table 2 – the severity of usability problems discovered with the write-along and Lyceum-based on-line methods. Rating R ranges from 0 (generally usable) to 4 (catastrophic usability problem). * indicates new usability problems.

Compared to the write-along method, the on-line sessions stopped the users becoming ‘lost’ during the evaluation tasks, for instance, by trying to perform a task in the wrong window. In both the Lyceum and NetMeeting™ sessions, the evaluators were able to correct problems as they arose. On the other hand, when such situations occurred in the write-along investigation of [2], they often resulted in a breakdown of the evaluation. This points to one strength of on-line methods over the write-along.

Users’ impressions of the method and tools used in the on-line evaluations were captured during the plenary discussion and in the post-evaluation questionnaire. Mostly, Lyceum’s users found the experience interesting and worthwhile. The voice communication tools were quickly used effectively in the plenary discussion. Users spoke naturally in terms of their own virtual evaluation room and the plenary room. Some users would have liked the chance to prepare more before the on-line evaluation.

Users of NetMeeting™ did not express the same level of satisfaction with their sessions, although most of the problems were due to technical failures. Users reported having problems hearing the evaluator: this made the tasks very demanding. They also expressed some anxiety related to the feeling of being constantly under observation:

‘...it’s difficult not to feel that it’s the user who’s on test and not the system...’ (User A)

However, the ability to observe the user’s desktop was considered as a real asset by the evaluators, and proved a great advantage when dealing with critical incidents.

7. Conclusion

This paper has reported on an investigation into the development of an effective remote evaluation method applicable to users in their natural working environment. The method uses Internet communication and collabora-

tion technology to facilitate the conversation between remote user and evaluator.

Two on-line approaches were discussed in the paper, based on adaptations of the write-along method of [2]. The investigation found that the write-along method was usefully extended by the addition of voice conferencing, which allowed for better support of the user/evaluator conversation and proved effective when critical incidents occurred.

Among the tools’ features, the following were considered of particular value:

- many-to-many voice and moderation tools, to support focus group based evaluation;
- virtual rooms, to support task based evaluation; and
- desktop sharing, to allow the evaluator to deal with critical incidents.

Acknowledgement

We thank our colleagues at the Open University, in particular the Lyceum development team, for their support, and the LTIC for funding this project.

References

- [1] Castillo, J.C., Hartson, H.R., & Hix, D. (1998), “Remote Usability Evaluation: Can users report their own critical incidents?”, *Proceedings of CHI’98*, ACM Press, 253-254.
- [2] Dunckley, L., Taylor, D., Story, M., & Smith, A. (2000), “Low Cost Remote Evaluation for Interface Prototyping”, *People & Computers*, XIV, 389-404.
- [3] Hammontree, M., Weiler, P., & Nayak, N. (1994), “Remote Usability Testing”, *Interactions*, July, 1994, 21-25.
- [4] Hartson, H.R., Castillo, J.C., Kelso, J., & Neale, W.C. (1996), “Remote evaluation: the network as an extension of the usability laboratory”, *CHI’96, Proceedings of ACM*, 228-235.
- [5] Hilbert, D.M., & Redmiles, D.F. (1998), “An approach to the Large-Scale Collection of Application Usage data over the Internet. Separating the Wheat from the Chaff”, *Internet-mediated user feedback, Proceedings of the Workshop on Internet-based Groupware for user participation in product development*, 1998.
- [6] Holtzblatt, K., & Jones, S. (1993), “Conducting and analysing a contextual interview”, Schuler, D., & Namioka, A. (eds.), *Participatory design: principles and practices*, Lawrence Erlbaum Associates, 177-210.
- [7] Norman, D.A. (1988), *The design of everyday things*, MIT Press.
- [8] Rapanotti, L., & Hall, J.G. (2000), “Lyceum: the system and its architecture”, *Proceedings of ED-ICT2000*, Vienna, 43-52.

- [9] Shah, D., Candy, L., & Edwards, E. (1998), "An investigation into supporting collaboration over the Internet", *Computer Communication*, **20**, 16, 1458-1466.
- [10] Shuler, D., Namioka, A. (1993), *Participatory design: principles and practices*, Erlbaum, Hillsdale, N.J.
- [11] Wright, P.C., & Monk, A. F. (1991), "A cost-effective evaluation method for use by designers", *International Journal of Man-Machine Studies*, **35**, 6, 891-912.