

Evaluating Electronic Textbooks: A Methodology

R. Wilson¹ and M. Landoni²

¹CDLR: Centre for Digital Library Research, University of Strathclyde

²Department of Information Science, University of Strathclyde

email: ruth.m.wilson@strath.ac.uk,
monica@dis.strath.ac.uk

Abstract. EBONI (Electronic Books ON-screen Interface) [1] builds on the premise to emerge from the Visual Book [2] and WEB Book projects [3], that appearance is important in the design of electronic textbooks, and offers an evaluation model, or general methodology, from which ebook usability experiments in a range of areas can be extracted and remain comparable at a basic level. The methodology sets out options for selecting material, participants, tasks and techniques, which vary in cost and level of sophistication. Results from each study will feed into a set of best practice guidelines for producing electronic textbooks on the Web, reflecting the requirements of students and academics throughout the UK.

1 Introduction

The Visual Book and the WEB Book experiments have demonstrated that appearance is a critical factor in the design of effective electronic textbooks [4]. The Visual Book experiment studied the application of the paper book metaphor to the design and production of electronic books, particularly focusing on the role of visual components such as size, quality and design, as well as the way in which people interact with them. It was found that users felt familiar with the representation of the book on screen and were able to rely on their experience with paper books to interact with the electronic book. The WEB Book experiment, on the other hand, focused on the impact of appearance on the usability of textbooks on the Web. Two electronic versions of the same text, one in a very plain scrolling format, the other made more "scannable" according to Morke's and Nielsen's Web design guidelines [5], were selected as the material for evaluation, and the scannable text proved to be 92% more usable.

At a time when the place of the book in the digital world is changing [6], it is important that the role of appearance in the design of ebooks is explored thoroughly in order that commercial publishing developments are fully informed from a design, as well as a content and technology perspective, and are delivered to the end-user in a form which maximises their usability.

EBONI (Electronic Books ON-screen Interface), funded under the JISC DNER Programme for Learning and Teaching [7], builds on the experience of the Visual Book, the WEB Book and related projects [8], and will conduct a large-scale evaluation of electronic textbooks, assessing the usability requirements of students and academics throughout the UK. This will be achieved through one core experiment in one academic discipline, and a number of "satellite" studies in other disciplines. In order to provide cohesiveness to the project, a general ebook evaluation model has been developed, from which each experiment will be extracted. This comprises a wide variety of techniques including "low cognitive skill" tasks measuring participants' ability to retrieve and recall information, "high cognitive skill" tasks set by lecturers to measure students' understanding of concepts in the

texts, and questionnaires designed to measure user satisfaction. In addition, some users will be filmed during the experiment and others will be asked to participate in “think-aloud” procedures. As such, it will aim to measure “usability” comprehensively and at a variety of levels, incorporating traditional IR concepts as well as users’ satisfaction and lecturers’ pedagogical objectives.

This evaluation model is outlined below. The quantity and combination of elements employed will depend on:

- Appropriateness to the specific objectives of the study. For example, an investigation into pure HCI issues when reading texts on the Web may wish to measure the speed and accuracy of retrieval of information, and users’ subjective satisfaction, but be less concerned with the completion of high cognitive skill tasks set by lecturers. Studies into the application of material for educational purposes, however, will adopt measures which reflect the requirements of learners and teachers in Higher Education; and
- Availability of resources. Every study varies in the resources (space, time, people and tools) it has at its disposal. Therefore, for some experiments adopting evaluation techniques such as interviews and think aloud sessions, which are costly in terms of space and time, may not be possible.

2 Selection of Material

Electronic books offer a diverse array of material for evaluation. The term “electronic book” is used throughout professional literature and popular culture to refer variously to hardware, software and content. As Tony Cawkell notes [9], the traditional concept of the book includes novels, dictionaries, telephone books, textbooks, anthologies, instruction manuals, proceedings of meetings, and directories. The phrase “electronic books”, however, has been applied to some types of CD-ROM systems, palm-top CD players, on-demand text, electronic documents systems of various kinds, and nearly any kind of computer-based text system that needs “hyping up” for marketing purposes.

Broadly, the term can be applied to the following categories:

- **Hardware devices.** Electronic books can be read on electronic handheld devices, which replicate the size and portability of paper books. Some, such as Gemstar’s REBs [10], the eBookman [11] and the goReader [12], are dedicated primarily to reading books. Increasingly, reader software is also available on Pocket PCs or PDAs, which typically perform a host of other functions;
- **Ebook reader software.** A variety of proprietary formats, such as Microsoft Reader [13], Adobe Acrobat Ebook Reader [14] and the TK3 Reader [15], have been developed, designed to make electronic texts easier to read by preserving the logical structure of the paper book and some of its visual features such as typefaces, colour images and page layout. This software can run on any laptop or desktop PC and is often intended for use on handheld devices, such as the PDAs and Pocket PCs described above;
- **Web books.** Electronic books are accessible via the Web in a number of forms, for free, to borrow, or at a price. Some are simply scrolling pages of text; some incorporate paper book features such as tables of contents, indexes, and page numbers; others exploit Web technology and features of HTML through hyperlinks and frames, and by incorporating search facilities.

EBONI is interested in each of these categories, and the methodological model outlined here can be adapted to them all. However, the core experiment to be extracted from the model is concerned with Web books and, in particular, educational texts produced on the Internet. This is free material published by lecturers and academics to aid the learning and

teaching of students at all stages and in all disciplines. Because it is produced by individuals, groups or departments, this material is characterised by its diversity: rather than adhering to a proprietary format, resources employ a variety of styles and techniques, such as frames, hypertext, tables of contents, indexes, internal search mechanisms, navigation icons, interactive elements, and so on. Inevitably, material which displays so much diversity in its presentation will also vary in the success with which it provides help to students. EBONI's aim is to identify those styles and techniques which maximise information intake by users, in order to provide the creators of online learning and teaching content with guidance on how to produce usable, effective material.

The process of selecting particular textbooks for evaluation will be directed by the objectives of the investigation and it is important that the chosen material will enable those objectives to be met in full. In an e-book evaluation, texts may be selected for comparison according to three parameters:

- Format/appearance;
- Content;
- Medium.

For example, the Visual Book experiment compared the same text in electronic and paper media, while the WEB Book experiment compared two electronic versions of the same text, each of which exhibited different styles of presentation (or formats). EBONI's core study, on the other hand, has selected texts which vary according to two parameters: appearance and content. Several texts (different in content) representing different styles have been selected for evaluation, all of them aimed at the same level of students in the same discipline.

In all, the possible combinations of objects selected for comparative evaluation in an e-book usability study include:

- The same text in electronic and paper formats;
- Different electronic versions of the same text;
- Different electronic texts exhibiting different styles/techniques, in the same discipline;
- Different electronic texts exhibiting similar (or the same) styles/techniques, in different disciplines;
- The same electronic texts on different monitors or hardware devices (such as portable e-books).

3 Selection of Actors

Ebook evaluations will vary in terms of the effort and skills required to set up an experiment. In general, though, four main roles, or possible actors, can be distinguished:

3.1 The Participant

Participants interact with the selected texts during a structured evaluation session and their feedback, derived through a variety of techniques, forms the results of the experiment.

Different experiments will have different requirements in terms of the characteristics of the groups of individuals they wish to select. For example, in the WEB Book study participants were asked to read and perform tasks using a textbook on Information Retrieval; therefore, it was important that participants had some knowledge of the subject matter in order that they could engage effectively with the content of the textbook. Other ebook evaluations might require participants of a certain age, or with a certain level of

experience with the Internet, or to be undergraduates or postgraduates in a particular discipline.

Preliminary questionnaires can be used to glean more specific information about participants. They can help to distinguish between different “groups” of participants, identify significant differences in population samples, and record any other user needs, characteristics or abilities which may affect the results of the experiment. The information gathered as a result of a pre-questionnaire will usually be factual and, in the case of electronic book evaluations, is likely to fall into two broad categories, both of which may be relevant when interpreting results: background information about the user, and details of the equipment he or she is using to conduct the experiment.

3.2 The Evaluator

The evaluator coordinates all aspects of the experiment, from the selection of material and participants, to the design of the methodology and the selection of evaluation techniques and tasks. More than one evaluator may be needed to supervise studies involving a large number of participants, or if it is felt appropriate for more than one to be present at an interview or think-aloud session.

In a less sophisticated experiment, the evaluator may also take on the role of task developer and assessor. For example, in the WEB Book experiment, the same person devised tasks, arranged the experiment and assessed the feedback. This is possible in cases where:

- The tasks are simple Scavenger Hunt or memory tasks which don't require a subject expert to set them;
- The evaluator has sufficient understanding of the experimental material and with the requirements of lecturers and students to set high cognitive skill tasks himself.

3.3 The Task Developer

The task developer devises and sets tasks for the participant to carry out, using the selected texts. The role of task developer is particularly important in developing high cognitive skill tasks, intended to reflect the more complex uses of material by HE students (see below). The task developer should be a lecturer or tutor in the appropriate discipline who:

- Is aware of the goals of the experiment;
- Understands the level of knowledge and ability of the participants;
- Is willing to devise tasks which will require participants to use the test material critically, in a way which meets the learning requirements for the level of participants selected.

For developing Scavenger Hunt and memory tasks, it may be possible for this role to be performed by the evaluator.

3.4 The Task Assessor

The assessor accumulates and interprets the evaluative data provided by the participant. Like the task developer, the assessor of high cognitive skill tasks will be a teacher in the appropriate discipline who understands the objectives of the study and the expected level of ability of the participants. He will adopt a marking scheme which is consistent both within the realms of the experiment and which enables results to be compared across other ebook studies derived from this methodology.

4 Selection of Tasks

A number of methods for measuring the usability of user interfaces have been inherited from the field of Human Computer Interaction [16]. Certain of these tasks and techniques are particularly appropriate to extracting evaluative information from electronic books, and these are described below.

Tasks are a way of bringing together participants and the test material in a structured manner, enabling quantitative data about particular aspects of interacting with the electronic book to be gathered. Three types of task are outlined here, intended to measure usability at different levels:

4.1 “Scavenger Hunt”

Scavenger Hunts involve participants in hunting through the material selected for evaluation in search of specific facts without using the browser’s “Find” command. This method is suggested by Spool et al [17] as a means of learning how easy it is for participants to find information in a Web site, and is of particular relevance to electronic textbooks, which are often used for the retrieval of facts and pieces of information. Scavenger Hunts are appropriate in all cases where it is important to the goals of the experiment to determine the accuracy and/or speed with which information can be retrieved from the test material, and the level of difficulty of tasks set will change according to the level of knowledge and expertise of the participants. Coupled with participant observation or the think-aloud technique, they can provide valuable information about how people search an electronic text and what factors help or hinder them.

The results of the Scavenger Hunt will feed directly into two measures of usability:

- **Task time:** the number of seconds it takes users to find answers to the search tasks and one judgment task;
- **Task success:** a percentage score based on the number of correct answers users give in the search tasks.

Although the cost of implementing Scavenger Hunts varies according to the type and quantity of material selected for evaluation, overall they are the least expensive of the types of tasks outlined in this methodology. They can easily be implemented in a group laboratory evaluation session, and require little time and expertise to assess. The questions participants are asked can be derived by the task developer directly from the test material, and so should be quick and easy to devise. Further, responses can be quickly checked against a list of correct answers by the task assessor.

4.2 Memory Tasks

Memory tasks involve the participant reading a chapter or a chunk of text for a short period of time, learning as much as possible in preparation for a short exam. The exam will comprise a set of multiple-choice questions (measuring recognition) and a question asking them to recall a list of items. These tasks are suggested by Morkes and Nielsen as a method of measuring a participant’s ability to recognise and recall information from an electronic text, after spending a specified amount of time reading it [18]. Data gathered from such tasks can be used to infer how the appearance of information on screen affects users’ ability to remember that information.

As with Scavenger Hunts, memory tasks are quick to devise and assess, and require little expertise on the part of the task developers and assessors. They are derived directly from

the test material, and answers can be checked quickly and effortlessly. Memory tasks are more demanding in terms of the environment required to implement them, however, because exam conditions must be enforced. It is still possible for many participants to complete the tasks at once, but quiet surroundings will be required with sufficient personal space for each participant, and more input will be required from the evaluator, who will be required to supervise the exam and enforce exam conditions.

These tasks will comprise two more measures: recognition and recall.

- **Recognition memory:** a percentage score based on the number of correct answers minus the number of incorrect answers to the multiple-choice questions;
- **Recall memory:** a percentage score based on the number of items correctly recalled minus the number incorrectly recalled.

4.3 High Cognitive Skill Tasks

Scavenger Hunts and memory tasks involve participants in interacting with the text in a relatively straightforward manner, searching for facts, or reading and trying to remember information. Textbooks, however, are often put to more complex uses by students and lecturers, and high cognitive skill tasks are intended to measure participants' ability to engage with the selected material in a manner which requires a higher degree of cognitive skill. In devising such tasks, the roles of task developer and task assessor become key.

Lecturers in the appropriate discipline are asked to read the test material and to assign tasks, the results of which will indicate students' understanding of the concepts in that material. These "high cognitive skill" tasks should reflect the more complex uses to which HE material is often put and measure the effectiveness of the electronic textbook in enabling participants successfully to engage in processes involving skills appropriate to their Higher Education course. A lecturer will also be asked to assess the results of these tasks. Adoption of the same marking scheme across all evaluations which implement high cognitive skill tasks will enable results from different experiments to be compared easily and effectively.

High cognitive skill tasks are the most costly of the types of task outlined here, primarily due to the time and expertise required to develop and assess them. Because they are intended to reflect the learning requirements of students and teachers, at least one lecturer in the relevant discipline will be heavily involved in the development of tasks which will elicit responses from participants that indicate their ability to use the text material critically. A lecturer will also be required to assess participants' responses. This is a time-consuming process that requires heavy communication between the evaluator and the lecturer(s) developing tasks in order to:

- Discuss the goals of the experiment;
- Discuss the learning requirements of the lecturer; and
- Incorporate the lecturer's requirements into the high cognitive skill tasks.

5 Selection of Evaluation Techniques

The following procedures are suggested by EBONI for obtaining qualitative feedback about the selected material:

5.1 Subjective Satisfaction Questionnaire

Questionnaires are a relatively inexpensive evaluation technique. They can be completed by many participants at once and, as demonstrated by the WEB Book experiment, it is possible to distribute them over the Internet, effectively reducing space costs to zero.

Satisfaction is measured after participants have used the test material and carried out any tasks which form part of the experiment, so that their responses are informed and based on experience. Morkes and Nielsen suggest using indices such as quality, ease of use, likability and user affect to measure satisfaction [19]. Studies especially concerned with learning and teaching aspects of the test material (such as those employing high cognitive skill tasks) may find it appropriate to engage the help of a lecturer in the relevant discipline in devising the questionnaire; he or she may be able to advise, for example, on items to include in an index measuring participants' satisfaction with the educational elements of the test material.

5.2 Behaviour observation

Observation is appropriate to investigations which are particularly concerned with HCI issues, and can be used in e-book evaluations to examine closely the interaction between users and the test material. While interviewing and think-aloud procedures discover information about participants' thoughts, views and opinions, covert observation enables participants' physical behaviour to be studied and draws attention to specific problems. Using video as an observation tool will enable the evaluator to investigate interaction issues that are not easily studied by any other method, and will therefore provide additional data to that derived from other evaluation techniques. Observation is costly in terms of the equipment (a video camera) required to record participants' behaviour, and the time involved in scrutinising the footage after the evaluation session.

5.3 Think-aloud

The "think-aloud" technique involves at least one participant in thinking aloud to explain what he or she is doing at each stage of performing the tasks, and why, to at least one evaluator. This provides qualitative information about the participant's cognitive processes, explanations of how he or she is navigating the test material, and reasons for difficulties. The evaluator can observe the participant's behaviour at the time of the task, which adds another source of data.

Implementing this technique is more expensive than questionnaires and covert observation in terms of space. Sessions require quiet surroundings, separate from other participants in the evaluation, in order that:

- The "thinking-aloud" procedure does not distract the other participants;
- The think-aloud participant can be heard clearly by the evaluator; and
- The think-aloud participant is not influenced or distracted by the other participants.

Think-aloud sessions will be conducted while the participant completes the tasks although, as explained above, they cannot easily be integrated into a laboratory session with other participants present; therefore, they must occupy a separate "time-slot" and are relatively costly. Moreover, this technique requires the presence of at least one evaluator for every think-aloud participant, in order to record everything that occurs during the session. However, this method can be used effectively with little training on the part of the evaluator

[20, 21]. The more participants selected to take part in this procedure, the more expensive it will be in terms of people, time and space.

5.4 Interviews

Interviews will be conducted on a one-to-one basis after any tasks have been performed and the subjective satisfaction questionnaire has been completed. They use a "script" or set of clear instructions and cover a list of questions in a predetermined order, but the interviewer and respondent are free to follow leads [22]. Patrick Dilley suggests structuring the flow of questions to lead the conversation pointedly yet comprehensively toward the larger research questions of the study [23]. Even if the interviewer deviates from the script later, a printed list of questions serves as a guide to return to should he lose his way in the course of listening to answers. They can, therefore, be used to elicit full feedback on selected aspects of the experiment, and to follow leads on additional themes raised by the participant.

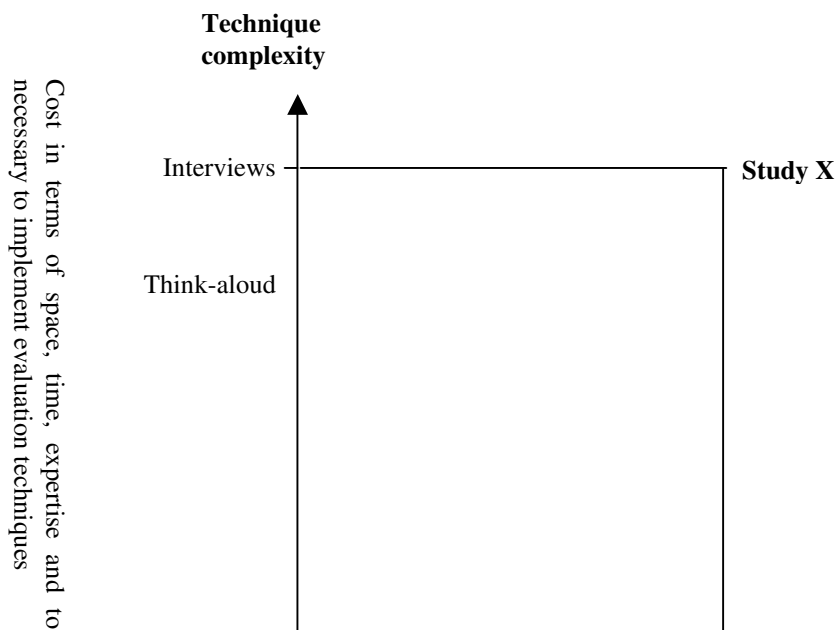
Of all the evaluation techniques discussed in this methodology, interviews are the most expensive. As with think-aloud sessions, interviews are costly in terms of space, requiring quiet surroundings, separate from the other participants. They will be conducted after the completion of tasks and questionnaires, and in-depth discussions will be time-consuming for the evaluator. Transcribing the interview later will be an additional cost. Interviews require the presence of at least one evaluator for every participant, skilled in interview techniques. Like think-aloud sessions, the more participants interviewed, the more expensive this technique will be in terms of people, time and space.

6 Conclusions

The tasks and evaluation techniques described above differ in terms of cost, level of sophistication, and the criteria they intend to measure or evaluate. Therefore, depending on the objectives of each ebook evaluation and the resources at its disposal, the implementation of different combinations of tasks and/or evaluation techniques will be appropriate, and the total expense of each experiment will vary across two dimensions:

- Task complexity, ranging from simple retrieval (Scavenger Hunt) tasks to more complex high cognitive skill tasks; and
- Technique complexity, from inexpensive questionnaires to interviews requiring time and expertise.

Although the cost of each element is not quantifiable, Figure 1 provides a rough indication of the relationship between these two dimensions, and is a crude indication of the relative total costs of implementing different combinations of tasks and techniques in an experiment derived from this methodology.



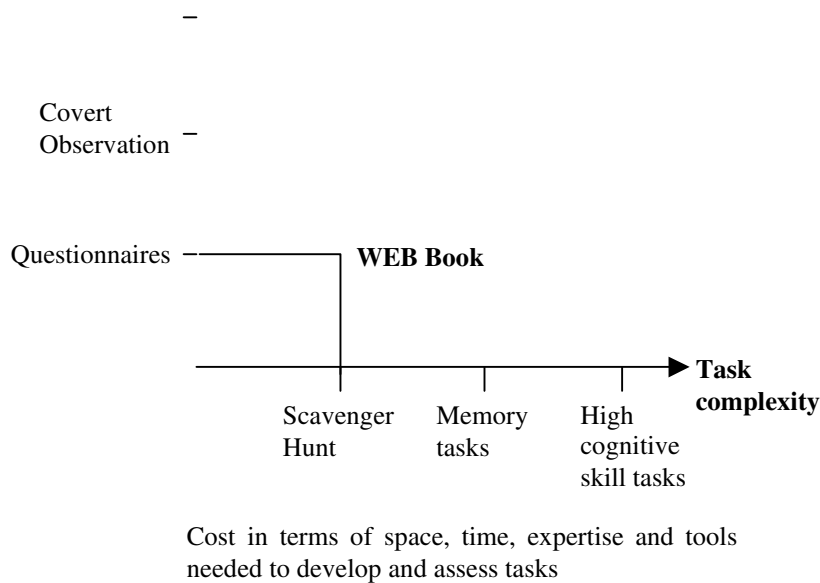


Fig. 1. Relative total costs of the WEB Book experiment and Study X

An experiment implementing a Scavenger Hunt and a questionnaire (such as the WEB Book) will have low scores on both dimensions, indicating a relatively inexpensive experiment. On the other hand, a study which employs high cognitive skill tasks and conducts interviews with participants (Study X) will score highly in terms of task complexity and technique complexity, indicating a relatively costly experiment.

Note that the incremental costs of implementing more than one element from a dimension (for example, questionnaires and think-aloud sessions) are not indicated in this Figure. EBONI's core experiment, therefore, which employs all the tasks and all the evaluation techniques, scores very highly on both axes, yielding a very large total area.

EBONI's hypothesis is that reductionist principles can be applied to the evaluation model so that very sophisticated experiments with unlimited resources which comprise many of the tasks and techniques described above, can be broken down to their constituent elements and "mapped" to simple, unsophisticated experiments employing only one task or technique. Following this logic, the results of EBONI's core study, The WEB Book, and the various ebook satellite studies which will employ other combinations of elements from the two dimensions, will all be comparable at some level.

This will enable the results of all user evaluations to feed directly into a set of best practice guidelines for producing educational material on the Web. Available in January 2002, these will be sent to relevant organisations, targeting publishers of electronic material, similar or related programmes, libraries and museums involved in digitizing collections and interested parties in the HE community in general. In addition, they will be available from the project Web site, together with an example of a text on which they have been implemented.

EBONI is based at the Centre for Digital Library Research [24], University of Strathclyde. The project welcomes feedback at all stages, and interested parties are invited to join the project mailing list [25].

References

1. EBONI (Electronic Books ON-screen Interface), <http://eboni.cdldr.strath.ac.uk/>
2. Landoni, M.: The Visual Book system: A study of the use of visual rhetoric in the design of electronic books. Glasgow: Department of Information Science, University of Strathclyde (PhD Thesis) (1997).
3. Wilson, R.: The importance of appearance in the design of WEB books. Glasgow: Department of Information Science, University of Strathclyde (MSc Diss.) (1999).
4. Landoni, M., Wilson, R. and Gibb, F.: From the Visual Book to the WEB Book: the importance of good design. Proceedings of the Fourth European Conference on Research and Advanced technology for Digital Libraries, Lisbon, Portugal, 18-20 September 2000.
5. Morkes, J. and Nielsen, J.: Concise, SCANNABLE, and objective: how to write for the Web, <http://www.useit.com/papers/webwriting/writing.html> (1997)
6. Lynch, C.: The battle to define the future of the book in the digital world. First Monday. 6 (6) (2001)
7. JISC (Joint Information Systems Committee) DNER (Distributed National Electronic Resource) Programme for Learning and Teaching. Available at: <http://www.jisc.ac.uk/dner/> (Last visited 14/06/01)
8. Landauer, T., Egan, D., Remde, J., Lesk, M., Lochbaum, C. and Ketchum, D.: Enhancing the usability of text through computer delivery and formative evaluation: the SuperBook project. In: McKnight, C., Dillon, A. and Richardson, J. (eds.): Hypertext: A Psychological Perspective. Available: <http://telecaster.lboro.ac.uk/HaPP/happ.html> (Last visited 14/06/01)
9. Cawkell, T.: Electronic books. Aslib Proceedings. 51 (2), 54058. (1999)
10. Gemstar REBs. Available: <http://www.ebook-gemstar.com/> (Last visited 14/06/01)
11. Franklin's eBookMan. Available: <http://www.franklin.com/ebookman/> (Last visited 14/06/01)
12. goReader. Available: <http://www.goreader.com/> (Last visited 14/06/01)
13. Microsoft Reader. Available: <http://www.microsoft.com/READER/> (Last visited 14/06/01)
14. Adobe Acrobat eBook Reader. Available: <http://www.adobe.com/products/ebookreader/main.html> (Last visited 14/06/01)
15. Night Kitchen's TK3 Reader. Available: <http://www.nightkitchen.com/> (Last visited 14/06/01)
16. Hilbert, D and Redmiles, D.: Extracting usability information from user interface events. ACM Computer Surveys. 32 (4). Baltimore (2000)
17. Spool, Scanlon, Schroeder, Snyder and DeAngelo.: Web site usability: a designer's guide. California: Morgan Kaufmann (1999)
18. Morkes, J. and Nielsen, J. (1997). *op.cit.*
19. *ibid.*
20. Jorgensen, A.K.: Thinking-aloud in user interface design. Human-Computer Interaction – INTERACT '90. D. Diaper et al. (eds.), Elsevier Science Publishers B.V. (North-Holland) (1990) 351-356
21. Wright, P.C., and Monk, A.F.: The use of think-aloud evaluation methods in design. SIGCHI Bulletin. 23 (1) (1991) 55-57
22. Russell, Bernard, H.: Social research methods: qualitative and quantitative approaches. London: Sage Publications (2000) 189-225
23. Dilley, P.: Conducting successful interviews: tips for intrepid research. Theory into Practice. 39 (3) (2000) 131-7
24. Centre for Digital Library Research (CDLR). Available: <http://cdlr.strath.ac.uk/> (Last checked 14/06/01)
25. For details of how to join the EBONI mailing list, see the JISCmail Web site: <http://www.jiscmail.ac.uk/lists/open-eboni.html> (Last checked 14/06/01)