

Putting awareness into practice: practical steps for conducting usability tests

**Jenny Craven, Research Associate, CERLIM
Helen Booth, formerly Research Associate, CERLIM**

Abstract

Methods/Approach: This paper describes the methods and findings of two research projects undertaken to explore user behaviour and usability issues relating to the use of Web-based resources by people with disabilities. Whilst the aims of the research differed, the methods adopted were similar – including observation, talk aloud protocols, questionnaires and focus groups.

Findings: The findings from both studies provided evidence of the problems faced by disabled users when using Web-based resources. It provided evidence of the types of features users liked and disliked, how they overcame navigational problems and what types of features enhanced their experience, all of which can be fed back into recommendations for the design of electronic resources.

Practical implications: The richness of the data collected for both studies confirmed the importance of involving users in accessibility and usability assessments. Analysis of both the findings and the methods for both studies also enabled a checklist to be developed, outlining issues that should be taken into consideration when planning a usability test.

Value of the paper: Document analysis of past and current usability studies revealed that suggested methods are often directed more towards expert usability testing, rather than managing the user testing in-house. Although it may be preferable to engage an expert, in reality practical guidance is also needed. The checklist aims to address gaps in usability advice identified, by providing practical guidance to enable providers and developers of web-based resources to conduct their own effective usability testing.

Introduction

Advances in technology and use of the web have provided more choices in the delivery of and access to information and resources. In the educational arena this has been used effectively to deliver a range of resources such as lecture notes, reading lists, course programmes, and full-text journals and, in more recent years, the development of virtual learning environments (VLEs). Not only does this widen access for users from a variety of locations, but for disabled people this often provides them with access to resources they were previously denied or that they had to wait for in order to be transcribed into an alternative format appropriate to their requirements.

Widening access can only be achieved if resources are designed in a way that is accessible to as many people as possible. Designers of web-based content also need to be aware of the wider issues surrounding accessibility, otherwise they will

not necessarily consider it important or perhaps relevant to their target audience. Access is not just an ethical issue, since disability legislation such as the Disability Discrimination Act (Great Britain, 1995) and the Special Educational Needs and Disability Act (Great Britain, 2001) in the UK now requires public and education service providers to ensure reasonable steps have been taken to ensure equal access for all - including access to their web-based resources.

Putting awareness into practice is a growing concern for some resource providers and designers who may be aware of accessibility recommendations but are not necessarily sure how to implement them. Accessible design guidelines such as those developed by the World Wide Web Consortium Web Accessibility Initiative (<http://www.w3.org/WAI/>) are a step in the right direction, but to aid a better understanding of accessibility and help foster a culture of social inclusion and universal design, the users themselves also need to be considered.

Usability studies and user testing are emerging as an important feature of service design and development and a number of experts are able to offer comprehensive advice on usability testing (see for example: Coyne and Nielsen, 2001; King et al, 2004). Methods could include cognitive walkthroughs, heuristic evaluation, expert usability evaluations and usability audits.

Practical guidance, however, is also needed to enable developers of web-based resources to conduct their own effective usability testing.

The Centre for Research in Library and Information Management (CERLIM) has been involved in a number of studies relating to usability and user testing: the Non-Visual Access to the Digital Library (NoVA) project (Craven and Brophy, 2003) explored user behaviour in digital environments through the observation of a group of sighted users and a group of visually impaired users; and DIAMONDS (Kendall and Booth, 2003) involved user testing of two WebCT tutorials, with students who had specific learning disabilities. Both studies used a combination of methods, including observation, talk aloud protocols, questionnaires and focus groups.

The aim of this paper is to provide practical methods for conducting usability testing, drawing on the methods adopted by the studies undertaken by CERLIM. The experiences and findings of the user-testing phase of both studies will be described, and a usability-testing checklist will be provided that can be used by anyone wishing to develop or undertake usability testing of their own.

The DIAMONDS project

A user testing study involving students with specific learning difficulties was carried out in the Department of Information and Communications at Manchester Metropolitan University 16th-25th June 2003, as part of the Diamonds (Developing Interactive and Accessible Material for Online Delivery to Students) project. The project aimed to extend the use of a virtual learning environment, WebCT, through the development of generic online tutorials (Kendall and Booth 2003).

The purpose of the usability study was to gain feedback from students with specific learning difficulties on two of the WebCT tutorials created as part of the DIAMONDS project. The first was a WebCT version of the Internet Detective tutorial, which builds upon the original developed by Netskills. This tutorial aims to give students the skills to assess information they find on the Internet. The second was the Citing Proficiency Test tutorial, which is a fun and interactive approach to learning how to cite references correctly in academic work.

Both the tutorials were developed using accessibility guidelines and were used by undergraduate and postgraduate students in the Department for the first time in the academic year 2002-2003. These students gave feedback and some changes were made to the tutorials as a result. It is expected that the tutorials will be used in the future by other departments in the University, either in their entirety, or with sections adapted to suit each department.

A group of participants were selected that included students with dyslexia, as this is the most common disability at Manchester Metropolitan University. In fact, it is estimated that up to ten percent of the population may be affected by dyslexia (Wright and Stephenson, 2003). Conducting the testing during the summer vacation enabled any changes to the tutorials to improve their usability, both for this group and for all students in general, to be made for the next academic year. One of the major issues to arise was that the dyslexic students would like to customise the appearance of the tutorials on the screen by using their preferred fonts, type sizes and colours, but many of the participants in the study were not aware of how to do this. A set of instructions were provided to accompany the tutorials for the next academic year on how to alter these settings in the browser.

Recruitment

An attempt at undertaking a user testing study with disabled participants had been made in the autumn of the previous year. Announcements asking for volunteers for the project were sent out via email by the Learning Support Unit to students registered as disabled. This proved unsuccessful at attracting participants, possibly because it was not possible to offer any incentives to take part and the autumn term is a busy time for students.

Funding was later secured specifically to conduct user testing. The Learning Support Unit was contacted again and kindly agreed to send out postal advertisements to students registered with them as having specific learning difficulties. The advertisement stipulated that students must not be from the Department of Information and Communications, as these students had already completed the tutorials. It also stated that they would be asked to undertake ten hours of paid user testing and that, apart from two meetings where students needed to come into the University, this work could be done from home. Students were asked to respond within one week.

The participants

The number of students that could take part was limited by the funding. Ten students were accepted for the study on a first-come-first-served basis. Six of the students were female and four male. Three were aged 18-21, six aged 22-40 and one over 40. Five students were in their first year, three in their second year and two in their third year and were from a range of faculties within Manchester Metropolitan University. Nine of the participants were dyslexic and one was dyspraxic.

Methods

A number of methods were considered for the study, including observation and talk-aloud protocol analysis. These methods were rejected because students were required to spend some time working independently through both the tutorials. It was decided to use a combination of short preliminary interviews, feedback and log sheets and a focus group. The preliminary interview questions and the formats and phrasing of the questions for the sheets were adapted after studying the literature on

dyslexia and dyspraxia. Some further questions about student characteristics, Internet and WebCT use were added after examining student questionnaires from the Sole (Students' Online learning Experiences) Project (<http://sole.ilrt.bris.ac.uk/>).

The feedback sheets and log sheets were developed using the same questions that had been used for the undergraduate and postgraduate students who had completed the tutorials so it was not necessary to pilot test these. Any new questions were piloted with staff at the university. This helped to highlight any potential problems such as with the language used, or the structure of the questions.

It was decided to give the students a ten day period to spend eight hours working through the tutorials so they would be able to give feedback on the entire experience. This had an additional benefit for the participants, as many of them commented at the focus group that having completed the tutorials would be very useful for their future studies. The use of log sheets and feedback sheets for the students to complete as they worked through the sections of the tutorial enabled them to work at their own pace.

The use of these methods meant that the participants could work at home if they chose, on their own PC, with any assistive technology or other tools they may need to help them read text on-screen already set up.

The introductory workshop

All the students attended an introductory workshop, where they were given an introduction to the study and shown how to log on to WebCT. It was stressed that they could work at their own pace and that there was no pressure on them to complete both tutorials, but that it would be possible to track their use of WebCT and see how much of the content they had accessed. In case any students needed extra support, they were told that they could arrange to come into University to work on the tutorials at times when members of staff would be available.

Student packs

Students were each given a pack that contained a handout of the introductory talk, which included relevant staff contact details, the URL of WebCT and details of the final group meeting. The pack also contained a user consent form, log sheets and tutorial feedback sheets. These were in two formats: printed and on disk. The contents of the pack were explained to the students and they were told that they could handwrite their feedback and log sheets, or use a PC and submit them on disk.

Ethical issues

The Ethical Framework of the University (<http://www.rdu.mmu.ac.uk/ethics/mmuframwork.htm>) was consulted and a consent form was developed to take into account the issues involved with carrying out a study using participants with learning difficulties. The consent form contained the following sections:

- I have been given an explanation of the nature and purpose of the study.
- I understand what I will need to do to take part and my questions have been answered satisfactorily.
- I have the contact details for the staff involved.
- I understand that I may withdraw myself and my data at any time, without consequences.
- I am satisfied with the arrangements to ensure that it will not be possible for me to be identified when the results are made available.

Staff involved in the testing went through this information with each person in turn and then asked them to sign the form.

The preliminary interviews

A short interview was conducted with each person to collect from them the following information:

- Basic demographics
- Learning difficulty
- Technology, tools or other methods used to help read text on-screen
- Previous Internet and WebCT use
- Where the student would be likely to work on the tutorial
- Any extra support or help needed
- Whether the student would like the pack printed on different coloured paper.

Log sheets

Participants were asked to complete one log sheet for each tutorial, filling in the time spent, the section they were working on and any comments they had as they were working through.

Feedback sheets

Participants were asked to complete one feedback sheet for each tutorial. This involved responding to statements in three sections, 'starting off', 'the tutorial' and 'the quizzes' or 'the self tests', depending on which tutorial they were using. The questions focused on:

- Logging on
- Finding their way to the start of the tutorial
- Understanding instructions for using WebCT
- Finding their way around the tutorial
- The organisation of the content
- Understanding the content
- Fonts
- Graphics
- Access, instructions and understanding of questions for the quizzes or self tests.

Participants selected one of five responses by placing a cross in the appropriate box. The responses were 'agree strongly', 'agree', 'neither agree nor disagree', 'disagree' and 'disagree strongly'. The text for these responses was accompanied by an image showing a suitable facial expression for each response (i.e. a smiley face for 'agree strongly'). Space was given on the feedback sheets next to each response for participants to add their comments. Underneath the statements participants were asked: 'Can you help us to improve the tutorial? Please put your comments or suggestions here.'

The final group meeting

At the end of the ten day period the participants attended a final group meeting. The proceedings of the meeting were taped and notes were made on a flipchart. The meeting lasted about an hour and questions were asked on the following areas:

- Advantages of learning online
- Disadvantages of learning online
- When in the academic year should the tutorials be introduced to students?

- Would you like to have the marks from the quizzes count as part of your assessment?
- If the marks were to count was there any extra support you would have wanted to have?
- Would you have liked to have someone sit with you and show you how the quizzes worked before you had to do them?
- What could we do to make it easier for people to log on?
- Would you have liked to have a handout to explain how to log on and to use WebCT, or do you prefer to have the instructions online?
- Any further comments or questions.

Data transcription and analysis

The time available for data transcription needed to be taken into account when selecting the methods to be used. The recording of the group meeting took one working day to transcribe. As this was a small study, data analysis software was not used. The quantitative data from the log sheets and feedback sheets was entered into tables and the comments from each question were entered into a Word document.

Potential problems identified

As it had been decided not to use observations and to allow students to work remotely, staff had to rely on them to remember to complete the log sheets and to write their comments and give feedback about any problems they may have had.

It was felt that there may have been a problem with asking participants with specific learning difficulties to work on the tutorial and fill out the log sheet at the same time but all the participants were able to do this.

Some participants initially had trouble logging in to the tutorials from home and needed to contact the relevant staff for help. One of the participants who had said in the preliminary interview that they were not very confident with using the Internet and had never used WebCT before dropped out of the study in the early stages.

A few of the participants were reluctant to speak in front of the group at the group meeting. It tended to be the same three or four people who did the majority of the talking. This illustrates the need to use a combination of methods for the study rather than relying solely on focus groups.

The NoVA project

The Non-Visual Access to the digital library (NoVA) project was undertaken to explore the usability of web-based resources in digital libraries, with a particular focus on blind and visually impaired users. The overall objective of the project was to develop an understanding of the searching or browsing behaviour of visually impaired users who could not read or interact with a screen without the aid of assistive technologies. To achieve this, a series of search and retrieval experiments were undertaken in order to map their approaches and to highlight usability issues.

The usability tests undertaken for the NoVA project provided an insight to the type of problems faced by users. Interestingly, although the focus of the project was on the information seeking behaviour of blind and visually impaired people, the control group of sighted users also highlighted usability problems of their own as well as

problems experienced by the experimental group, thus reinforcing the importance of involving all types of user in any design and development project.

The study showed that although awareness of web accessibility is increasing, all types of user can be faced with navigational problems. Some problems experienced are due to accessibility and usability conflicts such as inappropriate or unhelpful use of alternative text, or poor use of language. Other problems are due to a lack of understanding of the different ways users interact and navigate around web-based resources. Careful consideration must therefore be given not only to ensure conflicts between accessibility and usability are addressed, but in the layout and navigation of a site and to the ways different assistive technologies interact with them. The study revealed that different assistive technologies presented different problems for the users and that success in navigation might not only depend on experience in searching, but also on experience in the use of assistive technology - which raises training issues both for users and trainers.

The final report of the NoVA project (Craven and Brophy, 2003) describes the methodology, findings and conclusions and outlines a number of recommendations for digital library system design and is available to download from the NoVA project website in Word, PDF and HTML. The website address is:
<http://www.cerlim.ac.uk/projects/nova.html>

Participants

The NoVA study sought to obtain in-depth data on the way people search electronic resources and was not aiming for a truly representative sample. The number of participants was therefore set at forty as this number would provide comparable data of the information seeking behaviour of twenty 'sighted' users (the control group) and twenty 'visually impaired' users' (the experimental group) and also be a manageable figure in terms of data collection, transcription and analysis, taking into account the timescale of the project and the number of staff involved in conducting the study.

Participants were selected according to the following criteria:

- they could understand the basic concepts of the internet and searching the web
- they could interact with a screen unaided using a keyboard, mouse, keystrokes and/or assistive technology
- they had used at least one search engine to look for information.

The control group was defined as 'sighted users'. This group was made up of participants who met the above criteria but could interact with a standard computer screen without the aid of assistive technology. The experimental group was defined as 'visually impaired' users. This group was made up of participants who needed to use some form of assistive technology to interact with a standard size computer screen.

The NoVA study managed to recruit participants from each of the following areas:

- a mix of novice and expert
- screen reader users
- magnification users
- users who need to be make screen adjustments
- one user with a cognitive disability.

Recruitment

The NoVA study did not have funding to pay participants and therefore had to rely on people being willing to give up their time to participate. Participants for the usability testing were gathered partly on a voluntary basis and partly on a selection basis. An email was sent out to specific groups of users such as a distribution list of university students, library staff, staff and students from a university department and by making contact with users via organisations such as the National Library for the Blind and Action for Blind People who contacted potential participants on behalf of the project team. This method relied on participants responding to the request and it took around two months to finally secure the number needed for the study. A potential problem could have been the amount of time it took to gather volunteers for the study. Fortunately this had been built in to the NoVA project time frame, but it is worth bearing in mind that a substantial amount of time and effort may be needed for this.

NoVA methods

To meet the objectives of the NoVA project, methods selected included:

- Semi-structured tasks
- Observation and verbal dialog
- Transaction logging
- Pre-and post-task questions.

The NoVA study conducted a pilot test using five people, including both 'sighted' and 'visually impaired' users. It was a vital part of the study because it highlighted a problem in the number of tasks proposed (there were too many!), it enabled the facilitator to give participants in the actual test an idea of the length of time the session might take, and it also ensured that the instructions given to participants were unambiguous.

Semi-structured tasks

When devising usability tasks it is important to select resources and tasks that are appropriate for the study. In many cases this may be a specific resource or web site, or in the case of the NoVA study, to observe how people interact with web-based resources. Four information seeking tasks were devised using four different web-based resources:

- Search engine.
- Library OPAC.
- Online shopping site.
- Directory of Internet resources.

The resources were chosen because they displayed elements in their design that were appropriate to one of the objectives of the NoVA project: to observe navigation using resources that were designed in a non-linear way (e.g. using tables and frames).

Each task was consistently set so that comparisons could be made between the sighted and visually impaired participants. It was recognised that success in performing searches could be influenced by previous knowledge or experience, either of the technology, the site visited, the subject matter of the task, or by individual interpretation and approach to a task. In an attempt to obtain a balanced picture the tasks set covered a fairly broad subject base such as weather forecasts, shopping for clothes and travel information.

To try and create a relaxed environment every attempt was made to dispel feelings among the participants that they were being tested in any way (although inevitably this still occurred to some extent). This included offering tea, coffee and biscuits, chatting informally prior to the tasks and explaining that the purpose of usability testing was to highlight web usability issues rather than to test information seeking skills. Naturally, participants were assured that all their responses would be kept strictly anonymous. At this stage they were encouraged to ask any questions, or if they needed the facilitator to clarify anything for them.

To ensure everyone started from the same place, participants were required to commence each task using the stated resource, but were then allowed to choose whether they used the search facility or browsed through the site for relevant links. For example, to look for a national weather forecast for the UK, participants were required to use the search engine, but could choose whether to use the search facility or browse the site for a relevant weather link.

Whether or not to time the users was a decision that needed to be made at the planning stage. There are pros and cons to timing. Timing can evoke un-natural user behaviour as they may feel that they are being tested and try to complete each task as quickly as possible. However, it may not be desirable to allow participants to continue with one task for a long period of time as this may be counter-productive and difficult to transcribe apart from being stressful for the participant. Also, if the participant has been told that the session will last a specific amount of time (e.g. an hour) they may be unable to complete all of the tasks required of them and thus reduce the number of results to analyse.

To try and encourage real-life information seeking, participants in the NoVA study were not given a time limit to complete each task. At the beginning of the session they were told that they could stop the task at any time and were given examples such as "if you are satisfied that you have found the information", "if you are not satisfied, but think you have found all the information there is", or "if you are fed up with the task". This was done to help ensure completion was determined by the participant rather than the facilitator. Only on a very few occasions did the facilitator prompt a tactful termination of the task because after a considerable amount of time the participant was clearly not going to find the information. On these few occasions the facilitator generally said something like "do you want to stop, or are you happy to carry on?", or "are you ok, or shall we move on to the next task?" to reassure the participant that it was still up to them whether to continue or not.

Observation and verbal dialog

Observation is a useful method for obtaining an insight into the usability of a resource. Ideally the session should include a facilitator and an observer. The facilitator sits next to the participant and directs the testing while the observer watches the process remotely, either through a one-way mirror, or by using a video to record the session which the observer watches in another room. This method enables the observer to concentrate fully on what the participant is doing and saying, rather than on the actual running of the experiment.

The NoVA project did not have sufficient resources to remotely observe and/or video the sessions. Instead, the usability tests were conducted on a one-to-one basis with the facilitator acting also as an observer.

Participants were asked to undertake a series of tasks as well as providing a verbal dialog. The aim of this was to provide an insight into not only what they were doing, but also why they were doing it and how they felt about it.

During the tasks, the facilitator tried to avoid interaction with the participant. However, some participants found it difficult to provide a verbal description of what they were doing, or forgot to continue talking once they had started a task. In these cases it was necessary for the facilitator to probe a little. For example, if someone was staring silently at the screen the facilitator would not direct them in any way, but would simply ask something like 'what are you thinking about now?' or 'where did you think that link would take you?'. Some users also liked to ask the facilitator questions and often tried to engage in conversation during the task. In these cases, the facilitator acknowledged the question but did not answer it directly. For example, if the subject asked what they should do next, the facilitator would follow this with a question such as 'what do you think you should do?' or 'what would you do if you were looking for this on your own'.

During the NoVA testing, some usability problems kept recurring. In these cases it was not felt to be productive to let everyone struggle with the same problem (and possibly give up). The usability problem had clearly been identified, so if this happened to about two or three participants the facilitator provided a hint to allow users to progress. This was logged in the transcription notes.

Transaction logging

Transaction logging refers to in-depth recording of on-screen interaction (see Griffiths et al 2002). This could include keystrokes, mouse clicks, general user interactions and even a verbal dialog. Transaction logging can be performed manually or automatically. A number of products are available on the market which will capture on-screen activity, these include ScreenCam (<http://www.lotus.com/products/product2.nsf/wdocs/screencam>), Snag It (<http://www.techsmith.com/products/snagit/default.asp>) and Camtasia (www.camtasia.com/).

On-screen interaction for the NoVA study was logged using a combination of:

- ScreenCam Software which records on-screen activity plus verbal dialog
- Sound recording and manual note taking (for technical and practical reasons it was decided not to use ScreenCam for the visually impaired sample).

Both methods used for logging data were pilot tested to ensure valid and reliable data could be transcribed.

Each step of the search process (i.e. keystrokes or mouse clicks) was logged using a combination of on-screen data capture (Lotus ScreenCam), sound recording of verbal dialog and note taking.

Pre- and post-task interviews

Interviews were conducted to provide further information about the participants themselves, for providing qualitative data relating to the participant's experience and feelings toward the resources tested, and to clarify any actions or comments made during the test.

Pre and post task interviews were conducted for the NoVA study to gather data of a more qualitative nature. Interviews began with a set of **user profile questions** to

help draw up a profile of user characteristics. Questions included details such as age range, use of the internet, location of access and a brief description of their visual impairment if this applied. The nature of a participant's visual impairment was gathered to provide a clearer picture of the range of participants involved in the study rather than as an indication of the problems faced by different impairments. Likewise, age and experience were used as an indication of demographic spread rather than to analyse behaviour according to a person's age or ability.

Semi-structured interviews were conducted to provide data on emotion, feelings and experience. These comprised **general questions**, such as how to tell a page is loading, initial comments about the interface and the type of information provided; and **usability questions**, such as their experience in finding resources required, correcting errors, knowing where to input information, online help facility. Interviews were conducted before (general) and after (usability) each task to help ensure the electronic resource and the task performed were still fresh in the participant's mind before moving on to the next resource and task.

Transcription

Pilot testing for the NoVA study revealed that, on average, it took at least twice as long to transcribe the data collected from a session as the session itself took. For this reason it was necessary to reduce the number of tasks and resources to four in order to allow enough time to transcribe and analyse the data within the project's time-frame.

The sample size and quantitative data gathered for the NoVA project was relatively small. Therefore analysis was performed manually.

Qualitative data was transcribed in two ways:

- Each step of the search process was logged and coded according to the action performed by the user. For example, the action of clicking on a link was coded as CO, tabbing down a page was coded as TD and the TI code was assigned to the action of typing in terms.
- Pre- and Post-task questions were transcribed word for word.

Analysis

Data from the searches and questions were entered into the Atlas-ti data analysis software tool and coded accordingly. This method of analysis was chosen because it enabled large amounts of data to be analysed qualitatively and to some extent quantitatively. Although it was initially time consuming to assign categories and codes to each transaction, once established it offered quick and easy access to the data.

Qualitative data analysis for the NoVA study concentrated on the text retrieval and code and retrieve functions available in Atlas-ti. The "text retrieval" function was used to analyse the pre- and post-task interviews. Data from the interviews was coded according to instances of words and phrases.

Potential problems identified:

An important part the NoVA study was the experience of subjects who needed to use assistive technology to access electronic information sources. Planning ahead was therefore vital. Having established what assistive technologies were required well in

advance, it might have been possible to provide appropriate software on the PCs that were to be used for the usability testing. However, because some participants used different versions of the same assistive technology it was necessary to think of some alternatives. For the NoVA testing, the solution was to visit subjects *in situ* – i.e. at their place of work, their public library, at home – wherever they had access to the assistive technology (and version) of choice.

Another alternative could have been to include a sample of 'remote' participants. This would comprise sending participants the URL of the site for testing, a task list and a set of pre and post task questions. They would then run through each task and answer the questions provided. Although the observation element of the testing would be lost, it could still be a useful way of gathering usability feedback from a wider number of participants.

It was very difficult to create a natural environment for the participants to perform the tasks, and although every attempt was made to make them feel comfortable and to dispel any feelings that their ability was being tested, inevitably at times this did occur. However, this problem was probably unavoidable for the capture of qualitative data.

Summary

The usability tests undertaken for the DIAMONDS and NoVA projects provided a rich picture of the types of problems faced by users when using Web-based resources. It provided evidence of the types of features users liked and disliked, how they overcame navigational problems and what types of features enhanced their experience, all of which can be fed back into recommendations for the design of electronic resources.

Usability testing should be used alongside accessibility checking to provide a rich picture of the accessibility and usability of a website, which will help designers and developers to ensure their sites embrace universal access and access for all.

The following checklist outlines some issues that should be taken into consideration when planning a usability test:

Usability testing checklist

Decide on the objectives of the usability test

Before setting up a usability test, decide on the objectives of the study as this will influence the size and type of participants, the methods chosen for the study and the presentation of findings. Consider also the amount of staff time available and allocation of funds for the study.

Methods chosen should be appropriate to the focus of the testing. Usability testing may be undertaken for a number of reasons, for example:

- To gain an overall impression of a website or resource.
- To test specific areas, such as:
 - Searching: simple and advanced
 - Using the thesaural pages
 - Help facilities
 - Results
 - Navigation (where you are, how to move back and forth etc)

- Spell checking
 - Interface: changing the look and feel
 - Interaction: using assistive technologies
 - Language used throughout the site
- To test the utility of a resource.
 - To test users perceptions of a resource or service.

Select number and type of participants

The number of participants to include in a usability test will depend on the type of testing to be undertaken and the objectives of the test. For example, if the objective is to provide a comprehensive insight to the usability of a website with statistically significant data, then it will be necessary to select a truly random sample that is representative of the user population.

If, however, the objective is to pin-point possible usability errors of a particular feature or resource, then it might be more appropriate to select a smaller number of participants. Nielsen (2000) recommends using at least 15 participants to discover all the usability design problems in one site; however he also suggests that excellent results can be gained from using as few as five users.

Recruiting participants

Recruitment of participants can be time consuming – particularly if it is not possible to offer some kind of incentive such as paying participants for their time.

Ideally a range of users should be included, for example:

- a mix of novice and expert users
- users from different age groups
- users from different ethnic minorities
- users with a range of disabilities
- users who use a variety of assistive technologies.
- users who need to make screen adjustments.

Pilot testing

Having agreed the objectives, selected the methods and drafted a framework for conducting a usability test (e.g. the tasks, interview questions etc) it is essential to run a pilot test. This will give the facilitator a better idea of the length of time taken to undertake the process and will highlight any problems that may not be apparent in the planning stages, such as ambiguous instructions or technical problems. The number of participants to include in a pilot test will depend on the size and complexity of the proposed testing and also on the amount of time that can be reasonably set aside.

Ethical issues

Organisations may have their own ethical framework, so it is important to check this first. In the case of MMU, participants were asked to sign a consent form, giving the facilitator permission to tape and record subjects, and to use the transcript data for the purposes of the project only. It also affirmed that all subjects would remain anonymous and could withdraw at any time.

Undertake the usability testing

Prior to the usability testing session, participants should be given a full explanation of what is expected of them. This should include details about the aims of the usability test, the resources or sites to be tested and how the session will proceed. At this stage subjects should be given the opportunity to ask any questions or have anything clarified for them.

Transcription of data

Transcription and analysis of data collected from usability tests can be a time consuming process, therefore adequate time must be set aside for this. If time is a problem then it may be necessary to reduce the sample size, bearing in mind a smaller sample will produce statistically less significant results (but useful results none the less).

Data analysis

Having transcribed the data it will be necessary to perform some analysis in order to draw conclusions and recommendations from the user testing. For quantitative data analysis, this may be a case of counting responses or actions, recording and comparing data. Analysis can be performed manually or by using an automated tool such as SPSS (<http://www.spss.com/>). If it is a relatively small sample it would be perfectly feasible to use manually generated counts, percentages, averages etc. For a larger sample it may be more manageable to use an automated tool which, although time consuming to set up, is relatively easy to input and extract data.

Qualitative data analysis requires a different view of the data gathered, looking beyond counts or measures to provide a richer picture of responses and actions. This could include grouping of keywords or phrases under specific headings or the grouping of different types of responses (e.g. positive and negative). Analysis could be undertaken in a word file, using the Find facility to extract keywords and phrases for grouping. For larger amounts of data it may be more manageable to use an automated qualitative analysis tool such as Atlas-ti (<http://www.atlasti.de/>) or NUD*IST (<http://www.qsr.com.au/>).

Presentation of findings

Naturally, findings from usability testing will depend on the type of testing that has been undertaken and on the remit of the study. For example it could simply be a checklist of recommendations for further work; a summary of the main usability issues; or an in-depth report outlining each stage of the testing, together with conclusions and recommendations.

Overall Conclusions

This paper has described the methods and findings of two research projects undertaken at the Centre for Research in Library and Information Management (CERLIM): One study to explore user behaviour in digital environments through the observation of a group of sighted users and a group of visually impaired users; and a study involving the usability testing of two online tutorials with students who had specific learning disabilities. Whilst the aims of the research projects differed, the methods adopted were similar – including observation, talk aloud protocols, questionnaires and focus groups.

The findings from both studies provided evidence of the problems faced by disabled users when using Web-based resources. It provided evidence of the types of features users liked and disliked, how they overcame navigational problems and what types of features enhanced their experience, all of which can be fed back into recommendations for the design of electronic resources. The studies also revealed limitations of this type of testing, such as the difficulty of creating a natural environment for the participants to perform the tasks, and of encouraging full participation in focus group sessions, as well as specific problems relating to assistive technologies used by the participants. However, the benefits of the rich data collected for both studies far outweighed the limitations identified and confirmed the importance of involving a variety of users in any accessibility and usability assessments.

Document analysis of past and current usability studies revealed that advice and guidance provided, such as cognitive walkthroughs, heuristic evaluation, expert usability evaluations and usability audits, are often directed more towards expert usability testing rather than managing the user testing in-house. Whilst it may be preferable to engage an expert to undertake the work, in reality practical guidance is also needed. Analysis of both the findings and methods for both studies described in the paper enabled the usability testing checklist to be developed, outlining issues that should be taken into consideration when planning for this type of testing. The aim is to address gaps in usability advice identified, with practical guidance to enable providers and developers of web-based resources to conduct their own effective usability testing, and therefore help widen access for all.

Further Reading

This section refers back to references in the text as well as providing suggestions for further reading:

COYNE, K. and Nielsen, J., 2001. *Beyond ALT text: making the web easy to use for users with disabilities*. Fremont, CA:, Nielsen Norman Group.

CRAVEN, J. and BROPHY, P., 2003. *Non-visual access to the digital library: the use of digital library interfaces by blind and visually impaired people*. *Library and Information Commission Research Report 145*, Manchester: Centre for Research in Library and Information Management.

GREAT BRITAIN, 1995. *Disability Discrimination Act*. London: HMSO.

GREAT BRITIAN, 2001. *Special_Educational Needs and Disability Act*. London: The Stationery Office Limited.

GRIFFITHS, J.R. et al., 2002. "An improved method of studying user-system interaction by combining transaction log analysis and protocol analysis." *Information Research*.

KENDALL, M.A. and BOOTH, H. (2003) Developing generic online tutorials as a strategy for extending the use of WebCT. O'Reilly, U. (ed.) *4th annual conference of the LTSN Centre for Information and Computer Sciences*, 26th –28th August 2003, National University of Ireland Galway, Ireland. [Online] URL <http://www.ics.ltsn.ac.uk/pub/conf2003/index.htm>

Library review, 2006, vol. 55, no. 3. pp.179-194

KING, N., et al. (2004). An incremental usability and accessibility evaluation framework for digital libraries. In Brophy P. et al eds. *Libraries without walls 5: the distributed delivery of library and information services: proceedings of the 5th Libraries Without Walls conference. Lesvos, Greece 19-23 September 2003*. London: Facet Publishing, pp 123-131.

KRUG. S., 2000. *Don't make me think!: a common sense approach to web usability*. Indiana: New Riders.

Manchester Metropolitan University's Academic Ethical Framework [online] <http://www.rdu.mmu.ac.uk/ethics/mmuframework.htm> [Accessed 26/5/04]

NIELSEN, J., 2000. *Designing web usability*. Indianapolis: New Riders Publishing.

RAVDEN, S. and JOHNSON, G., 1989. *Evaluating usability of human-computer interfaces: a practical method*. Chichester: Ellis Horwood Ltd.

Sole (Students' online learning experiences) Project [online] <http://sole.ilt.bris.ac.uk/> [Accessed 26/5/04]

TAMLER, H. M., 1998. How (much) to intervene in a usability testing session. *Common Ground*. 8(3): pp11-15.

WRIGHT, L and STEPHENSON, J, 2003. *Issues around the use of WebCT in on-line learning for dyslexic students at the University of East London*. [online] [accessed 7th July 2003] http://www.uel.ac.uk/lds/internal_resources/docs/SoTL_Paper01.PDF