REMOTE WEB SITE USABILITY TESTING
- BENEFITS OVER TRADITIONAL METHODS

JESSICA GARDNER
Statistical Division
United Nations Economic Commission for Europe (UNECE)[1]
Geneva, Switzerland
jessica.gardner@unece.org

Abstract

Web sites have become a key communication medium, and usability is an essential factor in good web site design. Usability testing is an inexpensive way to gather valuable feedback from representative users, which can help web designers and content creators make their site more usable and relevant to its audiences. This paper examines the benefits of remote online testing over more traditional face-to-face methods. Remote online testing provides access to a larger pool of potential testers, cuts out travel time, and can significantly lower the cost of usability testing. Although the benefit of face-to-face contact is lost, research shows this method is just as effective in identifying usability issues as traditional testing. The UNECE Statistical Division recently conducted tests of its current web site (www.unece.org/stats) as a basis for redesigning the site's information architecture and establishing a benchmark for future usability studies. Tests were conducted remotely using online conferencing software, allowing testers to be truly representative of our geographically dispersed users and significantly reducing costs.

Keywords: remote usability testing, usability testing, UNECE

1. Testing Web Site Usability – Why and How?

Good web site design is generally achieved by maximising two factors: usability and visual appeal [Beaird, 2007]. Usability is important for anything with a user interface, whether it be a potato peeler or a computer application. Factors affecting web site usability include information architecture, navigation, jargon and terminology used, design of user-input forms, and page-layout. In essence, users must be able to easily conduct typical tasks for that web site, whatever they may be.

> "On the Web, usability is a necessary condition for survival. If a website is difficult to use, people leave."
>
> Dr. Jakob Nielsen (2003)

Usability tests provide insight into how users interact with a web site. Without this testing, it is not easy to tell if a web site is really usable or not. Observing how people actually use your web site can reveal issues that designers and programmers are simply unable to recognize themselves. Watching a user independently conducting a task is an excellent way to see whether the site really works as intended.

---

[1] The views expressed are those of the author and do not necessarily reflect those of the UNECE secretariat.

"...I've spent a lot of time watching people use the Web, and the thing that has struck me most is the difference between how we think people use Web sites and how they actually use them."

Steve Krug [2000], author of *Don't Make Me Think!*

A strong advantage of usability tests is they can be relatively inexpensive and easy to conduct. At one extreme, usability tests can be held in a specially designed laboratory with one-way glass and systems enabling facilitators to interact with and observe the tester, recording their on-screen actions, facial expressions, and verbal feedback. This approach may be more precise, but there are costs associated with hiring or owning a laboratory, recruiting testers and sending staff to facilitate and observe the testing. A simpler approach is to conduct the test at a user's desk with a facilitator simply observing and taking notes with a pen and paper.

It is recommended to test three to six users in order to identify the majority of usability issues with a web site [Krug, 2006; HHS & GSA, 2006; Nielsen, 2003]. Although research by Nielsen [2000] shows that 15 users are needed to discover all of the usability problems, he states that testing five users will reveal 85% of the problems and is the most cost-effective approach. However, a single test can reveal around one third of usability problems, so testing one user is better than conducting no tests at all [Krug, 2006]. If a web site is aimed at several types of users who conduct different tasks, then tests of each user group are recommended.

The pervasiveness of online communication technology today has made remote usability testing an attractive option. Remote testing uses online software to observe user behaviour as they work with the web site from another physical location, such as their workplace or home. The tester and facilitator communicate over the telephone throughout the test, as verbal feedback is important to understanding usability problems.

2. Remote Testing – Lower Your Costs without Compromising Results

Each method of usability testing has advantages and disadvantages. A well equipped laboratory provides a controlled environment that enables detailed recording and observation of the tester, making it easier to analyse results. However, laboratory facilities are expensive and create an unnatural environment for a user. Furthermore, travel to the test facility may be inconvenient and time-consuming for both testers and facilitators.

Informal 'pen and paper' usability testing, conducted at the desk of the facilitator or tester, can reduce travel time for one party, and provide an environment closer to the tester's natural operating mode. As with laboratory testing, this method has the advantage of tester and facilitator meeting face-to-face, making it easier to build rapport, prevent miscommunication, and observe non-verbal feedback.

Remote usability testing conveniently allows both tester and facilitator to work from their office or home. For a web site that has decentralized users, remote testing significantly increases the pool of potential testers and avoids the costs usually associated with testing across dispersed geographical locations. This is particularly valuable for organizations with a global audience. A further advantage of remote testing is the opportunity to observe the web site on a variety of computer configurations, including different operating systems, screen resolutions, and internet

connections speeds - a bonus for organizations that have limited resources for web site testing.

The varied configurations of remote computers can also be a source of problems. A loss of control over the remote technical environment could lead to unanticipated difficulties that inhibit testing, such as the tester being unable to download required plug-ins, or there being restricted internet access on the remote machine [Brush et al, 2004]. Another downside of remote testing is that it can be more challenging to build rapport and communicate with the user when you are not in the same room, and unless the user has a webcam, monitoring facial expressions and other non-verbal feedback is not possible. Also, if the tester lacks a speaker phone or headset, holding the telephone receiver can make it more difficult for them to complete some online tasks.

Despite the disadvantages, several studies of traditional versus remote usability testing have found "no reliable differences between lab-based and remote testing in terms of the number and type of usability issues identified" [HHS & GSA 2006:196]. In fact, some have reported more positive results from remote testing [Thompson et al, 2004]. A comparative study conducted by Bolt Peters in 2002 [Houck-Whitaker, 2005] noted significant differences in cost and in time for recruitment and testing, with laboratory testing being 50% more expensive and taking seven times longer to complete than remote testing.

A study by Brush, Ames and Davis [2004] found qualitative differences between remote and local usability testing, but no significant variation in the usability issues identified. Their study involved eight testers who participated in both remote and face-to-face testing of a software interface. Following the tests they were asked a series of questions on their perspective of the different testing methods. Despite the authors' assumption that testers would be more comfortable talking to the facilitator wen face-to-face, results showed that 75% were just as comfortable in either environment. When asked their preference between the two methods, 50% preferred remote testing and the remaining half considered them "about equal", with no respondent indicating a preference for local (face-to-face) testing.

3. The UNECE Experience

3.1. Background

The UNECE Statistical Division website (www.unece.org/stats) was established in 1995, primarily as a mechanism for disseminating documents to meeting participants. It has grown considerably since its inception and now includes hundreds of pages covering general information about division activities, statistical standards, publications, databases, and methodological and meeting papers.

All websites need regular review to ensure they align with the goals of an organisation, provide information and services expected by their clients, and take advantage of changes in technology. The UNECE website needed a systematic review to ensure it is meeting users' needs. A plan for redesigning the website was developed involving a range of user-centric design activities, including website usability testing, which was conducted remotely in November/December 2006.

UNECE chose remote usability testing over traditional methods for obvious reasons. As an international organization with users of its web site located in more than 56 different countries, the ability to involve users across the globe was extremely valuable. It enabled the recruitment of truly representative users from national statistical offices, highlighted different interpretations of terminology, and ensured

International Journal of Public Information Systems, vol 2007:2
www.ijpis.net
Page 65

testing of the English interface by both native and non-native speakers. Limited experience and budget for testing usability made traditional methods prohibitive. Remote testing provided an option for gathering extremely valuable feedback in a short time frame and without additional costs.

3.2. Methodology

WebEx[2] web conferencing software was already being used in the UNECE, and this platform was found suitable for facilitating remote testing. Amongst other features, WebEx allows applications being used on a remote computer to be viewed and/or controlled over the internet, and it easy to use and set up. Thompson, Rozanski and Haake [2004] and Bolt [2006] provide a brief evaluation of other software options for remote usability testing.

The UNECE approach to usability testing consisted of five phases: (a) determine users and their needs; (b) prepare test scenarios for each user group, representing typical tasks conducted on the website; (c) recruit testers that represent each user group; (d) conduct tests; and (e) analyze the results. Each are described in more detail below.
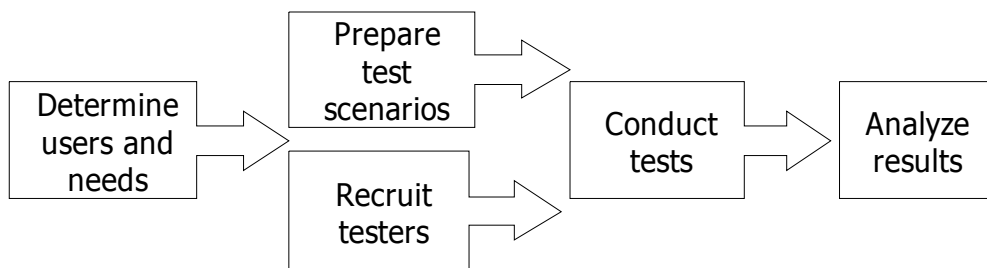
Figure 1. Phases of UNECE usability testing

3.2.1. Determine users and their needs

A fundamental requirement for usability testing is to know who the web site users are and what are their needs. This information should be the basis for web site design decisions. Once this is determined, consideration can be given to how each type of user would use the website – what are the kinds of tasks they conduct and what information are they likely to be looking for? What does it mean for them to be satisfied by their experience with our web site?

The process used by the UNECE Statistical Division to clarify the profile of our users was to develop a proposed list of user groups based on staff experience and analysis of past user enquiries. A table of users and their needs was drafted and circulated to division staff to solicit their feedback. Users were divided into five groups: (1) staff in statistical offices, (2) meeting delegates, (3) policy makers, (4) researchers, educators and students, and (5) the media.

---

[2] WebEx is an online meeting application owned by Cisco Systems Inc. (see www.webex.com for more information).

Due to time limitations it was decided that the first round of usability testing would focus on the highest priority user groups: staff in statistical offices, meeting delegates, and policy makers.

3.2.2.   Prepare test tasks

Test tasks (or scenarios) need to be representative of typical tasks conducted on the website. According to Kaufman [2006], the tasks should be:

o **short** – the user should spend most of their time completing the task rather than reading the scenario
o **specific** – clearly worded with a specific end goal
o **realistic** – typical of the activities of an average user on the site
o **understandable/clear** - in the user's language and related to the user's context.

With these principles in mind, 25 tasks that are typical for the UNECE Statsitical Division web site were prepared. There was some overlap in needs between user groups, so some tasks were considered relevant to more than one user group. Five to eight tasks were devised for each user group. In addition, there were five general tasks asked of every test participant.

Your manager has asked you to find out what you can about Statistical Metadata. What resources exist on the UNECE website?
Download the report of the most recent meeting on Consumer Price Indices.
Can the UNECE website help you find out what the population of Bulgaria was in 1995?
You are going to be going to Geneva to attend a meeting next March, but you don't know anything about the city. Can you find any information on the UNECE website?
Can you find a page that explains what Gender Statistics is, and why we need them?
Find a list of UNECE member countries.

Table 1. Example of some of the 25 tasks used to test usability of the UNECE web site

The tasks were tested by colleagues within the UNECE Statistical Division to ensure they were achievable, could be completed within the expected time limit (45 – 60 minutes), and would be likely to reveal issues with usability.

3.2.3.   Recruit testers

The UNECE Statistical Division conducted three different tests, each using two or three testers who were representative of the high-priority user groups (a total of eight testers).

Testers were recruited from national statistical offices and international organizations through existing networks. It was preferred if testers had minimal experience using the UNECE Statistical Division website, but were experienced in using the internet, in accordance with our typical web site user profile.

Twenty potential testers were suggested by UNECE Statistical Division staff and these people were sent an email inviting them to participate. The majority responded positively, but availability was an issue for some. It was practical to invite more testers than necessary so the target of eight testers was easily reached. Those people that were not able to participate in this round of testing will be invited to be involved in future tests.

This method of recruitment was successful because the user groups had been clearly identified in an earlier phase and the testers fit these profiles. In other circumstances it could be important to ask potential testers to complete a brief survey or telephone interview, verifying that they have the required internet skills and knowledge of statistics, if applicable.

If your contact with user groups is minimal and it is difficult to develop a pool of potential representative testers, there are companies that will recruit testers on your behalf. Another approach is the one used by the U.S. Energy Information Administration, which recruits their own testers by inviting volunteers through their website (http://www.eia.doe.gov/neic/aboutEIA/recruit.html). Although using representative testers is likely to give more accurate results, Krug [2006:135] argues that it is "more important to test early and often" than spend too much time looking for testers that fit a precise profile.

3.2.4. Conduct tests

The tests were scheduled and testers were sent an email containing a link to the web conferencing website and some basic instructions about joining the test. The questions were not sent beforehand to prevent users from trying the scenarios before the test and biasing the results. However, not having the questions in advance proved to cause some delays and technical difficulties, as a file containing the questions had to be downloaded and printed before the test could commence.

The tests took 45-60 minutes each and were facilitated by three staff in the User Services Section of the UNECE Statistical Division, one of which had prior experience with usability testing. The test started when the facilitators telephoned the tester at the agreed time, and then proceeded through the following stages:

- o Introduction – introduce facilitators and explain how the test will work
- o Technicalities – connect to web conferencing software (involves downloading a browser plug-in) and download and print document containing tasks to be completed
- o Pre-test questions – a series of questions about the tester's operating environment (browser, monitor, operating system, etc) and their general experience with the UNECE Statistical Division website
- o Work through the tasks – testers were asked to work through the tasks at their own pace, but to speak about what they were thinking as they used the site and made decisions to click on certain links
- o Post-test questions – general questions about the tester's impression and experience using the site

A vital part of usability testing, whether conducted remotely or face-to-face, is to ask the tester to "think-aloud". This provides the most useful feedback as testers reveal why they are thinking of clicking on a particular link, or what they are expecting to find if they click there. The facilitators can ask probing questions to ascertain the

testers understanding of particular terms and their impressions of web site appearance and usability.

3.2.5.　　　Analyse results

The tests were observed by two to three staff who are directly responsible for design and maintenance of the UNECE Statistical Division web site. Notes were taken during testing to record whether the tester was successful in completing the task, and what usability issues were detected. There was no screen or voice recording used during the test, and it was quite challenging to take meaningful notes while observing user behaviour. After each test the facilitators discussed the outcomes and agreed on the main findings.

After all three rounds of testing was complete, the facilitators' notes were written up into a 15 page report, using a template from www.usability.gov[3]. This report comprises the following sections:

1. Overview
2. Key findings and recommendations
3. Methodology
4. Detailed findings and recommendations for each task

Preparing the report was a time consuming exercise, taking several days to complete. While the report serves as a good record, the most valuable part is the key findings and recommendations. It groups and prioritises all the usability issues that were found, providing clear support for recommendations to improve website design.

In *Don't Make Me Think!* [2006:159], web site usability expert Steve Krug recommends rather than writing detailed analytical reports, it is more effective to simply discuss the findings immediately after conducting a round of testing to agree on what should be fixed.

This was the first in a series of tests planned throughout the UNECE Statistical Division website redesign project. It has provided valuable feedback about issues with the current site design and will provide a benchmark against which future tests, to be conducted after site redesign, can be compared.

3.3.　Findings

The tests revealed important usability issues with the UNECE Statistical Division website. The key findings were:

o Confusion between UNECE and the Statistical Division pages
o UNECE banner navigation menu was often first preference
o Statistical Division navigation menu (upper left hand side) was rarely used
o All users commented on there being too much information on the home page
o Search results were sometimes useful, but not consistently. One issue was the inability to sort results by file type and date
o Meaning of some terminology was not clear

---

[3] http://www.usability.gov/templates/index.html#usareports

The most unexpected usability issue was the consistent tendency for testers to click on the UNECE corporate banner navigation bar, rather than the upper left menu specific to the pages of the Statistical Division. This caused confusion as they were redirected to other sections of the UNECE website, without necessarily realising they were leaving the Statistical Division pages. This clarified how much of an impact navigation can have on usability and addressing this particular problem will be a key priority in the site's redesign.
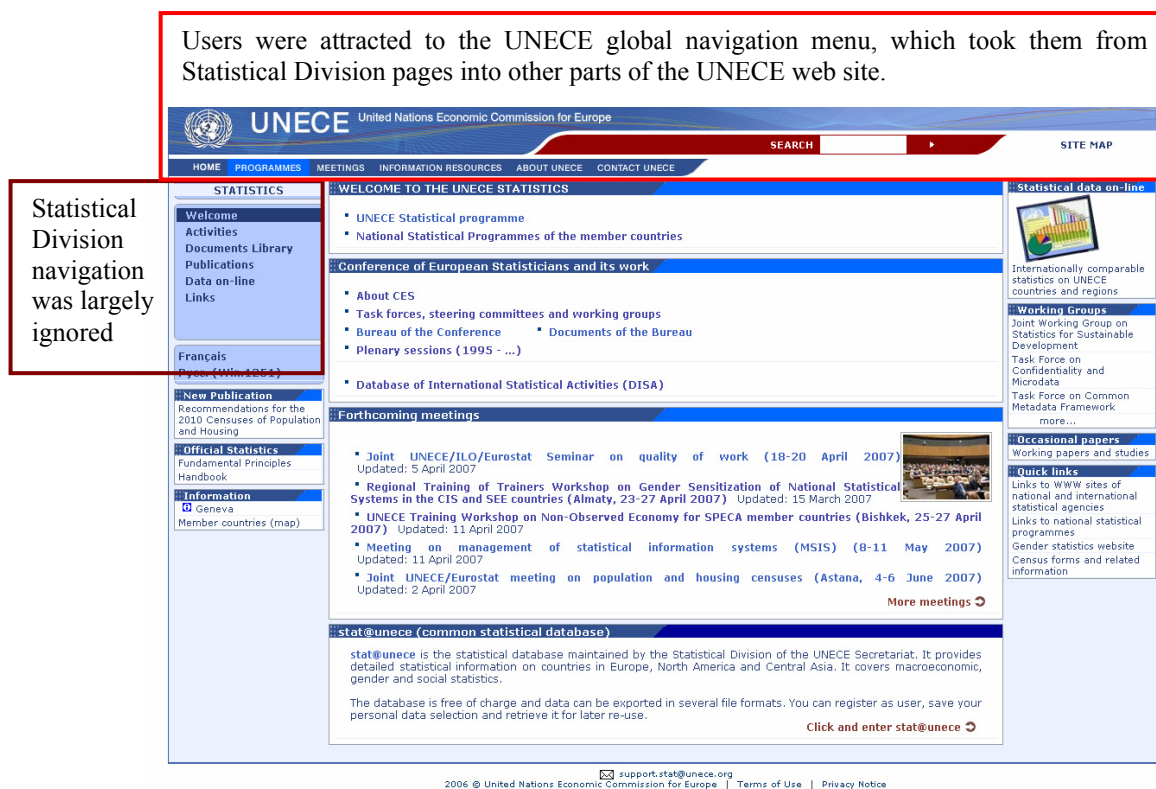

Figure 2. UNECE Statistical Division home page at time of usability testing. Source: [www.unece.org/stats]

It was particularly interesting to observe the different paths and techniques testers used to find the same piece of information. Most testers browsed by choosing a link in the navigation bar or towards the top centre of the home page (for example, UNECE Statistical Programme was a popular choice). When the resulting page was different to what they expected, testers often clicked on the browser's back button to return home and choose another alternative. Few testers (25%) had a tendency to use the search functionality, with one tester citing their reluctance to use this feature being because they were "so often disappointed with the results given by in-site search engines". To help navigate through pages which contained a lot of text, two testers often used "Ctrl + F" to find a particular term within the page.

After considering all the findings, the recommendations for improving the site include:

o There should be a clear distinction between the UNECE corporate banner and the Statistical Division web pages. It should be obvious to users that links in the banner will take them away from Statistical Division content.
o Design the site so users can browse by topic from the home page.

- The banner logo should link to the homepage of either the UNECE or the UNECE Statistical Division, not www.un.org as it currently does.
- Long lists of links should be better formatted to enable quick scanning.
- Search results should be sorted by date (most recent to least) and type – eg HTML pages first, then PDF documents, with the ability to filter the results for the different types.
- The terms 'Conference of European Statisticians' and 'Bureau of the CES' were confusing or meaningless to testers. Provide a brief description.
- Cease to maintain our own page about Geneva and integrate our content with the UNECE page.
- Instead of using the title 'Links' consider something like 'Other statistical offices' or 'Related websites'.
- Make the Statistical Division contact information more prominent.

The redesigned home page and new global navigation is expected to be launched in October 2007, with further improvements to the web site being introduced systematically over the next twelve months.

4. Conclusion and Lessons Learned

Remote usability testing is a valuable tool, particularly when users are dispersed across a wide geographical area. Planning and conducting tests of this nature is not resource intensive and the feedback is highly valuable. It is difficult, if not impossible, to obtain this quality of information about an interface without some form of usability testing. Lessons learned through this round of usability testing include:

- have contingency plans in case of technical problems
- provide questions to the participants by email well before the test is scheduled to start. It saves time and ensures the tester knows what to expect
- recording the screen movements and verbal feedback would provide more material for detailed analysis of the tests. It is challenging to take good notes and observe user behaviour at the same time
- make sure testers have access to a speaker phone or headset for the test.

Our first experience with usability testing was a valuable learning experience both in terms of identifying web site usability issues, and gaining experience in usability testing. Now that test scenarios have been developed, they can be reused for future rounds of testing, making the process even more efficient.

With or without access to a testing laboratory, remote testing provides a practical solution for testing usability that can be implemented quickly and easily. Organizations with limited resources can now reconsider the possibility of introducing usability testing into their web site and software development processes.

References

Beaird, J. (2007). *The Principles of Beautiful Web Design*, SitePoint Pty Ltd

Bolt, N. (2006). *Guide to Remote Usability Testing*, OK/Cancel, accessed on 1 August 2007 at http://www.ok-cancel.com/archives/article/2006/07/guide-to-remote-usability-testing.html

Brush, A.J., Ames, M. and Davis, J. (2004). *A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface* from CHI 2004 Proceedings, 1179-1182 accessed on 26 July 2007 at http://research.microsoft.com/~ajbrush/papers/brushremoteusability.pdf

Gough, D. and Phillips, H. (2003). "Remote Online Usability Testing: Why, How and When to Use It", boxes and arrows, accessed on 26 July 2007 at

http://www.boxesandarrows.com/view/remote_online_usability_testing_why_how_and_when_to_use_it

Houck-Whitaker, J. (2005). "Remote Testing versus Lab Testing", accessed on 23 March 2007 at http://www.boltpeters.com/articles/versus.html

Kaufman, J. (2006). "Practical Usability Testing", in *Digital Web Magazine* accessed on 29 March at http://www.digital-web.com/articles/practical_usability_testing/

Krug, S. (2006). *Don't Make Me Think! A Common Sense Approach to Web Usability*, Second edition, New Riders

Neilsen, J (2000). "Why You Only Need to Test With 5 Users", Jakob Nielsen's Alertbox of 19 March 2000, accessed on 26 July 2007 at http://www.useit.com/alertbox/20000319.html

Neilsen, J (2003). "*Usability 101: Introduction to Usability*", Jakob Nielsen's Alertbox of 25 August 2003, accessed on 29 March 2007 at http://www.useit.com/alertbox/20030825.html

The Wiki for Remote Usability (2007). accessed on 15 August 2007 at http://remoteusability.com/

Thompson, K.E., Rozanski, E.P. and Haake, A.R. (2004). "Here, There, Anywhere: Remote Usability Testing That Works", in *Conference on Information Technology Education Proceedings*, pp. 132-137, accessed on 26 July 2007 at

http://portal.acm.org/citation.cfm?id=1029567&dl=&coll=&CFID=15151515&CFTOKEN=6184618

United States Department of Health and Human Services (HHS) & United States General Services Administration (GSA) (2006), *Research-Based Web Design & Usability Guidelines*, Version 2, U.S. Government Printing Office.