(Almost) Everything You’ll Need to know about Usability Testing

Section 1 – What is Usability and Usability Testing?

**Usability testing** is a means for measuring how well people can use some human-made object (such as a web page, a computer interface, a document, or a device) for its intended purpose, i.e. usability testing measures the usability of the object. Usability testing focuses on a particular object or a small set of objects, whereas general human-computer interaction studies attempt to formulate universal principles.

If usability testing uncovers difficulties, such as people having difficulty understanding instructions, manipulating parts, or interpreting feedback, then developers should improve the design and test it again. During usability testing, the aim is to observe people using the product in as realistic a situation as possible, to discover errors and areas of improvement. Designers commonly focus excessively on creating designs that look "cool", compromising usability and functionality. This is often caused by pressure from the people in charge, forcing designers to develop systems based on management expectations instead of people's needs. A designers' primary function should be more than appearance, including making things work with people.

Simply gathering opinions on an object or document is market research, rather than usability testing. Usability testing usually involves a controlled experiment to determine how well people can use the product. 1

Rather than showing users a rough draft and asking, "Do you understand this?" usability testing involves watching people trying to use something for its intended purpose. For example, when testing instructions for assembling a toy, the test subjects should be given the instructions and a box of parts. Instruction phrasing, illustration quality, and the toy's design all affect the assembly process.

Setting up a usability test involves carefully creating a scenario, or realistic situation, wherein the person performs a list of tasks using the product being tested while observers watch and take notes. Several other test instruments such as scripted instructions, paper prototypes, and pre- and post-test questionnaires are also used to gather feedback on the product being tested. For example, to test the attachment function of an e-mail program, a scenario would describe a situation where a person needs to send an e-mail attachment, and ask him or her to undertake this task. The aim is to observe how people function in a realistic manner, so that developers can see problem areas, and what people like. Techniques popularly used to gather data during a usability test include think aloud protocol and eye tracking.

**What to measure**

Usability testing generally involves measuring how well test subjects respond in four areas: time, accuracy, recall, and emotional response. The results of the first test can be treated as a baseline or control measurement; all subsequent tests can then be compared to the baseline to indicate improvement.

- **Time on Task** -- How long does it take people to complete basic tasks? (For example, find something to buy, create a new account, and order the item.)
- **Accuracy** -- How many mistakes did people make? (And were they fatal or recoverable with the right information?)
- **Recall** -- How much does the person remember afterwards or after periods of non-use?
- **Emotional Response** -- How does the person feel about the tasks completed? (Confident? Stressed? Would the user recommend this system to a friend?)

In the early 1990s, Jakob Nielsenn, at that time a researcher at Sun Microsystems, popularized the concept of using numerous small usability tests -- typically with only five test subjects each -- at various stages of the development process. His argument is that, once found that two or three people are totally confused by the
(Almost) Everything You’ll Need to know about Usability Testing

home page, little is gained by watching more people suffer through the same flawed design. "Elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford." 2. Nielsen subsequently published his research and coined the term heuristic evaluation.

http://en.wikipedia.org/wiki/Usability_testing

Section 2 – What’s the Best Way to Test Website Usability

There are many variations on Web site usability testing, but a simple way to picture it is to imagine a real user, almost (but not quite) picked off the street, sitting in front of a PC and accomplishing a short list of tasks on a Web site. As they work through the tasks they talk about them, how difficult or easy they are, and a psychologist listens and watches.

The process is repeated with a handful of different users. When the sequence is complete, the psychologist who's watched the tests should have a good idea of the site's weaknesses and strengths. The weaknesses can then be rectified and the site becomes more successful.

That's a very simple picture. In practice there are many variations on testing, beginning with the basic division between testing live sites that already exists on the Web, and testing new sites still at an early stage of development. The two are approached in different ways.

Testing existing sites is the more common of the two, perhaps because there are relatively few new sites in the process of creation compared to the number that already exist, and also because sites need to be repeatedly tested through their lifetime as they change and incorporate new content.

Heuristic Evaluations

The simplest and cheapest usability studies don't include testing. They're a straightforward audit of the site by a usability expert - also referred to as a heuristic evaluation. The definition of heuristic used here belongs more to psychologists than to programmers. It means expertise gained through experience, but without the programming implications of dogged improvement through trial and error, usually involving wasted time.

A heuristic evaluation of a Web site is an inexpensive way to identify its most obvious usability flaws. It may be the only option for companies that don't have the budget for proper testing, which can cost from ten thousand to a hundred thousand dollars.

Testing

Not everybody trusts experts - in the field of usability as elsewhere. As the joke goes, there was Adam and there was Eve, and then there was the snake, the first consultant. But most companies who are wary of experts will be convinced about opinions if they're coming from their customers, and that's the aim of usability testing. It aims to interpret the experiences of target users and turn this into information that's actionable.

****It's not the same as market research, which might be interested in whether a site is attractive to the user or has a memorable logo. This kind of sentiment may come out of a usability test as a useful by-product, but the primary focus is on pure performance - how easy the site is to use.****
(Almost) Everything You’ll Need to know about Usability Testing

Choosing Users

The aim is to recruit real people from the site's target demographic to carry out tasks typical of those they would do in the real world, for example logging in and purchasing items.

In practice, it's been found that only a small number of people are needed for the tests to be effective - between three and ten, depending whose expert opinion you choose to believe. The Usability Company opts for eight. Such small numbers might rankle with theoretical statisticians looking for statistical significance and expecting hundreds of testers, but experience shows that it works. The first eight testers will invariably identify a site's worst flaws, and after this number the law of diminishing returns rapidly kicks in. Additional testers rarely discover further issues of crucial importance.

Some clients prefer to test with their own registered users, but most prefer the testing company to find a selection for them. The Usability Company subcontracts this selection to agencies, which is more expensive than keeping a database of users itself, but avoids the problem of database junkies who turn up as a housewife one month and a stockbroker the next. The non-database "free-found" option works out slightly more expensive, but the cost of hiring this kind of participant remains a relatively small proportion of the total testing cost - not much above 5%.

Each of the eight participants is, in turn, asked to accomplish the same sequence of tasks, and since most users begin to tire after an hour, the sequence is aimed for a one hour duration. Clearly, on a big site with thousands of pages, one hour doesn't get you very far, so it's crucial to choose the right tasks. A big transaction including lots of online forms could take up half an hour on its own.

Evaluations tend to divide into two types - broad-brush tests looking at multiple aspects of a site, but not in great detail, and more focused evaluations covering just one element of the site but with great thoroughness.

Choosing the right tasks is part of the skilled work of the testing company. The Usability Company often begins question formulation with online questionnaires added at an early stage to the live site. These help to identify areas where visitors may be experiencing problems, and are likely to be combined with path data from analyzed Web logs.

Typical tasks include registration, finding specific information, entering a competition or finding and buying a specific product. The number of tasks is usually low - a dozen at most and sometimes just two or three. Practical examples of tasks for a mortgage site under test included - You currently have the following mortgage (details given), see how an online mortgage from this site would compare. And - Use this site to apply online for the mortgage described.

The exact phrasing of the tasks is also crucial, and it can take time to explain to clients why one particular phrase has the correct psychological implications while another suggested phrase would only produce low quality data. It's important that the phrasing doesn't give specific instructions on *how* to accomplish the tasks, only what end result is required.

Pre-Test Questionnaire

Before testing begins, users are given a pre-test questionnaire. This usually requests demographic information such as age, gender, and level of Internet experience, to confirm the user comes from the right demographic.
(Almost) Everything You’ll Need to know about Usability Testing

Test Laboratory

The user sits at a regular computer screen, with a video camera to pick up their facial expressions and a microphone to record speech. Their screen is captured by a separate, synchronized online video, including cursor movements. In practice most users ignore the camera and microphone after just a few minutes. It is possible to track eye movements too, but rarely necessary.

A dedicated laboratory is useful because of the amount of recording involved. It also helps to cut down on distractions. Apart from the user, the only other person in the room should be the usability specialist, who sits alongside and watches. If clients are present, they sit behind a one-way mirror in an adjoining soundproof room. Clients are encouraged to send staff along to watch the tests, because it helps them get more out of the results. They also get the opportunity to ask the specialist questions between tests.

The Specialist

The specialist/analyst who sits in on the test will almost certainly be a behavioral psychologist, with cognitive psychology skills (the process of learning and understanding) and knowledge of HCI (Human Computer Interaction). They will also be a usability expert, but it's likely that their background will be in psychology rather than site design. When testing, psychology is far more important than the rational mechanics of good information architecture, though it's clearly desirable to understand both.

This same specialist helps to formulate the tasks, runs the tests, analyses the video, writes the test report and makes the final presentation to the client.

The specialist generally watches and listens to the test user, but may decide to interfere if they feel that time is being wasted, even to the extent of suggesting how to complete a task. This is one of the skills of the specialist - judging when interference will increase the amount or quality of data, rather than reduce it.

Users are encouraged to think aloud, to articulate thoughts about the site. They may also be prompted for thoughts by the specialist - for example if they get stuck at a particular stage.

Allowing for questionnaires and breaks, a test schedule usually fits in four tests on one day, followed by four the next to complete the set of eight. The site under test shouldn't be changed over the two days, though there are amusing tales of clients behind the glass on their mobile phones getting live changes made to the site while a user is still part way through a task, or leaving the laboratory to go back to the office and completely redesign the live site between one test day and the next.

Post Test

The final element of each test session is for the user to fill in a post-test questionnaire. These are short and take less than 15 minutes to fill out. They include questions like - How easy was the site to use, would you be likely to use it in the future, which areas did you like, and so on. Many of the questions don't relate to the specific tasks that were performed, because that information should already have been picked up by monitoring the user at the screen.

Results

The client company gets the results as a video, a report, and a presentation. They don't generally want to watch the entire eight hours of testing, but will probably be interested in seeing repeated reactions to problems areas,
which can be sewn together as a clip with a split screen showing the computer screen, including cursor, and the participant’s face, along with their comments on the soundtrack.

The report will offer recommendations for actions in each problem area, so the client can take immediate action. It's surprising how much information can be generated by a handful of tasks taking one hour in total. A sample report had over 70 recommendations, and the mortgage application mentioned earlier generated even more.

Return on Investment-

Andrew Starling
April 15, 2002

The Return on Investment (ROI) for usability testing is very high. The Usability Company quotes an average figure of 800% over a twelve month period, meaning that companies can expect an increase in revenue equal to eight times the cost of the usability testing.

Clearly, if a site has no decent revenue streams to begin with, usability testing isn't going to make the situation much better, but financial service and ecommerce sites usually find that usability testing is a very profitable exercise.

A practical example was Smile-on, the UK's leading online provider for dental practices. The company noted some customers had difficulty using its online E-store. They commissioned a single usability evaluation with eight users, concentrating on tasks associated with purchasing through the E-store, starting with registration and working all the way through search and the shopping basket system to completion of order. The evaluation identified 46 issues and made 60 recommendations.

Smile-on subsequently relaunched itself and its site. Monthly orders improved by over 100%. The number of new clients increased by over 240%. Some of this improvement was down to factors other than improved usability, but the vastly improved conversion rate showed how well the usability study had worked. The return on investment (within one year) for the usability project was estimated at better than 10:1.

Sites outside the finance and pure e-commerce sectors have also benefited - mtv.co.uk gained a 30% increase in core traffic after usability testing.


Section 3 – Three Common Missteps and How to Avoid Them
Mistake #1: Do You Know Why You're Testing?

The first mistake we see constantly is teams don't understand when usability testing can help and when it can't. Usability testing is a tool to produce information. However, it can't effectively produce *all* types of information.

These teams often make the mistake of using usability tests to see how the users "feel" about the design. They want to know if the tested participants will favor the design, want to use it again, and share it with their friends.

While these are all important things to find out, a standard usability test is not the way to do it. We've seen instances where users were extremely frustrated with the design -- couldn't complete a single task -- yet told us they loved it. We've also come out of tests where the users completed every task quickly and effectively, but hated the design, even though they also told us they'd use it again. It's very hard to know what to change when you get results like these.

Because a usability test allows you to observe the user's actual behavior, its real forte is in telling you where the interface causes frustration. The observation of how users flow through the design provides far more actionable information than asking them if they like it or not.

You can avoid this first mistake by being clear what you want to get out of the test. Posing a behavioral question or two, such as "Can our users apply for a mortgage without confusion?" or "Will the content reduce calls to our support center?" will dramatically improve the test results you get. The more detailed the question, the better your results will be. You'll know when the design is working and what to do if it's not.

Mistake #4: Not Designing the Right Tasks

Years ago, we helped with a study of Ikea.com, looking at how people found products on the site. When we got there, they'd already started the testing process and were using tasks like "Find a bookcase." Interestingly, every participant did exactly the same thing: they went to the search box and typed "bookcase".

Upon our suggestion, the team made a subtle change to the instructions they were giving their participants: "You have 200+ books in your fiction collection, currently in boxes strewn around your living room. Find a way to organize them."

We instantly saw a change in how the participants behaved with the design. Most clicked through the various categories, looking for some sort of storage solution. Few used Search, typing in phrases like "Shelves" and "Storage Systems". And, nobody searched on "bookcase".

The way you design tasks could have a dramatic outcome on the results, without your even realizing it. In a testing situation, the participants really want to please you by following your directions. If the tasks direct participants to take a certain path, that's the way they'll go. If it's not what real users do in the true context of the design's use, then you may get distorted results.
(Almost) Everything You’ll Need to know about Usability Testing

You can get around this mistake by constantly exploring the "context of use." When designing tasks, ask yourself, "What events or conditions in the world would motivate someone to use this design?" Use the answers as the primary formation of the tasks you create.

Mistake #7: Not Iterating to Test Potential Solutions

Usability Testing is great for identifying problems. Yet, it's horrible at identifying solutions. Fortunately, we've never run into a design team that couldn't generate half dozen possible solutions to any problem, within moments of its discovery.

The problem comes with choosing which solution is best to implement. You can't tell from the initial test, which pinpointed a problem, what solution will work. You need to test again, this time with a working implementation of the solution. (Often a paper mockup will suffice.)

We see teams skipping this step. Either they don't schedule a second round of testing to work out solutions or they cut corners in their process due to overconfidence. The results can be disastrous – the solution may actually be worse than the original implementation!

Planning a round of testing, to validate any yet-to-be-discovered potential solutions, is the antidote to this problem. You need to do this before you even know what the problems will be. Of course, if you don't have any problems, then you can always cancel the testing. (Yeah, like *that* will ever happen!)

http://www.uie.com/articles/usability_testing_mistakes/

Update as of 11/3/06

Effective Short Reports
SWW: Chapter 15 — Key Sections
1. Guidelines for Writing Short Reports: pp. 631-634
   a. See especially comparison between poorly written, formatted, and organized report vs. revised version pp. 635-637 written, org
2. Progress Reports: pp. 638-644
   a. See also BWC: 88-91, Progress and Activity Reports
3. Test Reports: pp. 652-656
   a. Part of this will be helpful when you describe the actual Usability Test development, design, implementation, and results
4. Document design: BWC: 149-156

Usability Tests
1. See Handout from Friday, November 3, 2006, [(Almost) Everything You’ll Need to know about Usability Testing]  
2. Class website: Links page, Reading on the Web & Usability Tests -