

# Usability testing

From Wikipedia, the free encyclopedia

**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  
(<http://jerz.setonhill.edu/design/usability/intro.htm>)

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?)

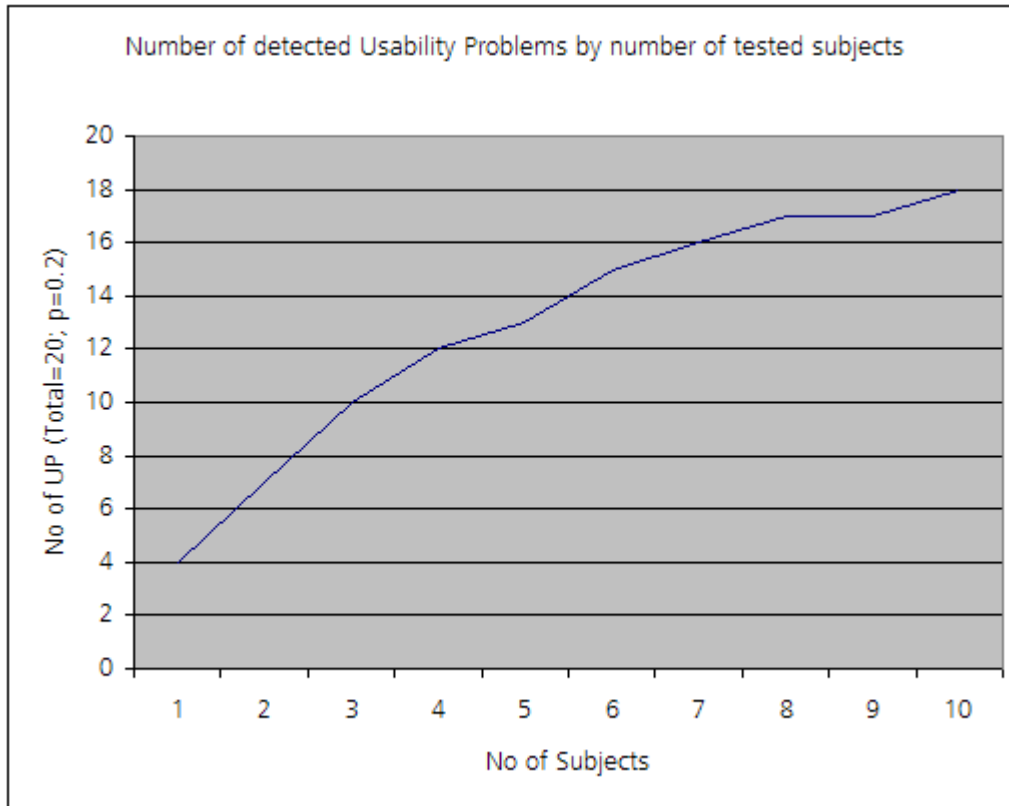
In the early 1990s, Jakob Nielsen, at that time a researcher at Sun Microsystems, popularized the concept of using numerous small usability tests -- typically with only five test subjects each -- at various stages of the development process. His argument is that, once found that two or three people are totally confused by the 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  
(<http://www.useit.com/alertbox/20000319.html>). Nielsen subsequently published his research and coined the term

heuristic evaluation.

The claim of "Five users is enough" was later described by a mathematical model (Virzi, R.A., Refining the Test Phase of Usability Evaluation: How Many Subjects is Enough? Human Factors, 1992. 34(4): p. 457-468.) which states for the proportion of uncovered problems U

$$U = 1 - (1 - p)^n$$

where p is the probability of one subject identifying a specific problem and n the number of subjects (or test sessions). This model shows up as an asymptotic graph towards the number of real existing problems (see figure below).



In later research Nielsen's claim has eagerly been questioned with both empirical evidence 3 (<http://citeseer.ist.psu.edu/spool01testing.html>) and more advanced mathematical models (Caulton, D.A., Relaxing the homogeneity assumption in usability testing. Behaviour & Information Technology, 2001. 20(1): p. 1-7.). Two of the key challenges to this assertion are: (1) since usability is related to the specific set of users, such a small sample size is unlikely to be representative of the total population so the data from such a small sample is more likely to reflect the sample group than the population they may represent and (2) many usability problems encountered in testing are likely to prevent exposure of other usability problems making it impossible to predict the percentage of problems that can be uncovered without knowing the relationship between existing problems. Most researchers today agree that, though 5 users can generate a significant amount of data at any given point in the development cycle, in many applications a sample size quite larger than five is required to detect a satisfying amount of usability problems.

Bruce Tognazzini advocates *close-coupled testing*: "Run a test subject through the product, figure out what's wrong, change it, and repeat until everything works. Using this technique, I've gone through seven design iterations in three-and-a-half days, testing in the morning, changing the prototype at noon, testing in the afternoon, and making more elaborate changes at night." 4 (<http://www.asktog.com/columns/001closecoupleddtesting.html>) This testing can be useful in research situations.

**See also**

- Software testing
- Educational technology
- Universal usability
- Commercial eye tracking

## External links

- Usability.gov (<http://www.usability.gov/index.html>)

Retrieved from "[http://en.wikipedia.org/wiki/Usability\\_testing](http://en.wikipedia.org/wiki/Usability_testing)"

Categories: Usability | Software testing | Educational technology

- 
- This page was last modified 23:24, 5 September 2006.
  - All text is available under the terms of the GNU Free Documentation License. (See **Copyrights** for details.)  
Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc.