

An Overview of
Usability Evaluation and
the World Wide Web

Jeremy Bradbury
(bradbury@cs.queensu.ca)
March 7, 2001

Overview

- What Exactly is Usability? Usability Evaluation?
- Why is Usability Evaluation Needed?
- Four Usability Evaluation Techniques.
 - Heuristic Evaluation
 - Published Guidelines
 - Cognitive Walkthrough
 - Usability Testing
- Two Techniques that don't Work

What is Usability?

- Usability can be considered a part of quality.
- Usability can be defined as “...*the ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component*” [IEEE 90].
- “*To some extent, usability is a narrow concern compared to the larger issue of system acceptability, which basically is the question of whether the system is good enough to satisfy all the needs and requirements of the users and other potential stakeholders ...*” ((Nielsen, 1993)).

What is Usability Evaluation?

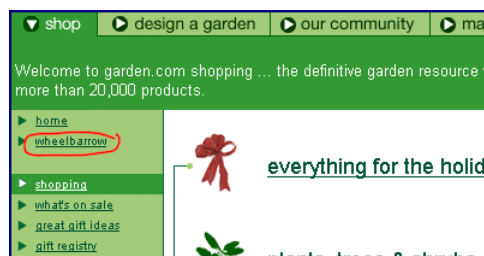
- Sun’s Usability Labs approach to usability evaluation:
 - define usability in terms of its 6 general attributes:
 - Utility
 - Learnability
 - Efficiency
 - Retainability
 - Errors
 - Satisfaction
 - discuss usability evaluation as part of a usability engineering approach to development in which iterative design and evaluation occur.

Why is Usability Evaluation needed at all?

- Lack of user involvement into the design process is one of the top two reasons that software development projects fail
- Approximately 30% of software development projects fail to be completed because of inadequate user design input
- \$80 billion is lost each year because of these failed projects
- 46 % of all new product development costs go to failures
- examples:*
 - RCA Videodisk lost \$500 million total
 - IBM PCjr lost \$40 million in marketing alone

Why is Usability Evaluation needed on the Internet?

- One example of why usability evaluation is needed on the world wide web is the <http://www.garden.com> site.
- Metaphors are a popular technique in e-commerce. The idea of a shopping cart is widely used to represent where a shopper can buy products on a website.
- Is a wheelbarrow really a good metaphor? Users of <http://www.garden.com> didn't think so since the site in this form no longer exists.



Why is Usability Evaluation needed on the Internet?

- The new version of gardens.com uses the traditional shopping cart metaphor



Usability Test Coverage

- Site-level Usability*
 - ease of use of search utility, site navigation, page layout and templates, overall writing and graphical style, site-wide design standards.
- Page-level Usability*
 - ease of understanding of links, forms, error messages, headlines, etc.
 - inclusion of useless information and the exclusion of useful information
 - relevance of graphics and icons used

Four Techniques for Usability Evaluation

- *Heuristic Evaluation*

Usability experts examine the website and look for properties that they recognize from their own experience as leading to usability problems.

- *Using Published Guidelines*

Allows non-experts to conduct a “heuristic evaluation” in which the guidelines are a substitute for experience.

- *Cognitive Walkthrough*

Involves walking through the user interface in the context of a user trying to perform a specified task.

- *Usability Testing*

Allows for a website to be tested by observing users.

Usability Test Coverage

- *Site-level Usability*

Tested through usability testing (i.e.. User work through scenarios/tasks and are observed by testers)

Cognitive walkthrough could also be used.

- *Page-level Usability*

Heuristic evaluation (conducted by usability experts)

Guidelines evaluation (usually conducted by testers/evaluators who are novices)

Heuristic Evaluations...

...on the World Wide Web

Heuristic Evaluations & Severity Ratings

- Evaluators individually evaluate a portion of the web site with a heuristic evaluation.
- After evaluators conduct a heuristic evaluation then the entire set of findings is put into one big list.
- Independently each evaluator then rates the entire list including his/her subset of the list.
- Each usability problem is given a final rating based on the mean of the set of ratings by the evaluators.
- Usually a set of ratings from 3 evaluators is sufficient for performing the severity ratings.

Usability Problems & Severity Ratings

- How severe a usability problem is depends on:
 - *Frequency*: How often the problem occurs?
 - *Impact*: How hard is it for users to overcome this problem?
 - *Persistence*: Is this a one-time problem or a recurring problem?

Severity Ratings

- The following is a 0 to 4 rating scale

Rating	Problem	Description
0	There is no problem	The evaluator doesn't agree that there is a problem
1	Cosmetic problem	Fixing this problem is not necessary unless extra time is available
2	Minor Usability problem	Low priority fix
3	Major Usability problem	High priority fix
4	Usability Catastrophe	Has to be fixed prior to release! Highest priority

Using Guidelines...

...on the World Wide Web

Common Heuristics/Guidelines for the Web

- Recognition rather than recall
 - Make actions, and options visible
 - Don't rely on the user remembering information
- Aesthetic and minimalist design
 - Don't include irrelevant information
- User control and freedom
 - Support undo and redo features
- Help and Documentation
 - Make system documentation easily available

Cognitive Walkthroughs...

...on the World Wide Web

Cognitive Walkthroughs

• *“...uses a more explicitly detailed procedure to simulate a user’s problem-solving process at each step through the dialogue, checking if the simulated user’s goals and memory content can be assumed to lead to the next correct action.”*

-Jakob Nielsen

• Not much mention of the use of cognitive walkthroughs on the Internet.

• Some reasons for this could be a result of the fact that:

-Cognitive walkthroughs are tedious

-The use of usability experts instead of users is needed

Usability Testing...

...on the World Wide Web

How long does User Testing of a Website take?

- A beginner can conduct a qualitative usability study in less than a week
- An experience usability expert can do it in just 2 days

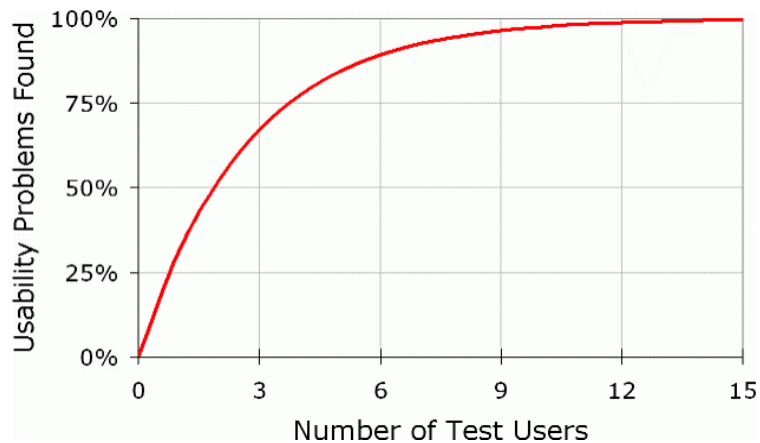
Time	Task
1-2 hours	write test tasks
0 hours (\$1000)	focus group company to recruit 5 test users
< 24 hours	actual tests
2-3 hours	analysis of results
2 WORK DAYS	ENTIRE USABILITY STUDY

How Many Users are Needed for Usability Testing?

- Source of data is research conducted by Jakob Nielsen and Thomas Landauer
- Let n be the number of user
- Let N be the total number of usability problems
- Let L be the proportion of usability problems discovered in testing by a single user (usually around 31%)
- Let $N(1-(1-L)^n)$ be the number of usability problems found in a usability test with n users.

How Many Users are Needed for Usability Testing?

- Plotting the function $N(1-(1-L)^n)$ with $L = 31\%$ gives the following graph.



How Many Users are Needed for Usability Testing?

- From this graph we can see that
 - 0 users find 0% of usability problems
 - 2 users find 50% of usability problems
 - 5 users find 80% of usability problems
- The conclusion by Nielsen and Landauer is that although testing with 15 users would uncover almost all problems, testing with **5 users** in an iterative approach with **3 rounds** of design and testing makes more sense.
- Later rounds of testing will be able to find more problems with structure rather than surface-level usability problems that are mainly encountered in the first round of testing.
- *NOTE:* for quantitative usability testing that involves metrics, 20 users are suggested to get a tight confidence interval on the results.

Usability Metrics

- A method of quantifying a qualitative evaluation process
- Allows for tracking of progress between releases
- Allows for assessment of competitive position
- Allows for an educated decision to be made regarding release dates

Usability Metrics

- USABILITY is measured relative to a users' PERFORMANCE on a scenario or task set.
- Some measurements that could be used in usability metrics are:
 - time required to complete a task
 - Error rate
 - Subjective level of satisfaction from the user
 - Percentage of time that a user follows an optimal navigation path
 - Number of times a user has to backtrack

Usability Metrics Example

- Macromedia did a usability study of a Flash site. It compared the original design to a redesign that was based on guidelines. The following are the study's results:

	Original Design	Redesign
Task 1	12 seconds	6 seconds
Task 2	75 seconds	15 seconds
Task 3	9 seconds	8 seconds
Task 4	140 seconds	40 seconds
Satisfaction Score*	44.75	74.50

* Scale of 12 (unsatisfactory) to 84 (excellent)

- Is the redesign a lot better than the original design?

Usability Metrics Example

- It depends....
- If satisfaction wasn't as important to the users (ex: stock site) as performance and if Task 3 was actually used 99.9% of the time and the other 3 tasks were used 0.1% of the time then the improvement isn't as impressive over the original as if each task was used evenly and satisfaction was the most important design issue.
- Use of weighted values for each element of metric or use of mean differences of tasks is necessary only if usage of tasks differs
- Performance vs. Satisfaction must be taken into account. A stock website would rank performance as the most important while an entertainment site might rank satisfaction as the most important attribute of the website for users.

Manual Testing vs. Automatic Testing

- A computer program that follows links on a website and keeps track of the number of clicks is not an adequate substitute for users working through a set of tasks.
- The subjective component of usability can't be automatically measured.
- Automated methods can be used in testing websites, however their use should be restricted to aspects of usability such as:
 - Calculating response times
 - Validating HTML code
 - Locating "Linkrot"

Opinion vs. Use

- Marketing research methods are not a substitute for usability evaluation.
- Techniques such as focus groups, surveys, having a panel of users fill out a questionnaire don't work because:
 - Users often tell what they think you want to hear or what is socially acceptable, not what they want to say (especially in groups)
 - Users tell what they remember they did and memory is not reliable
 - Users have a tendency to unconsciously rationalize their behavior after the fact

Conclusions

- Usability testing is useful on the world wide web.
- The early usability problems are found the better. Usability problems that are found late usually cost \$\$\$.
- Heuristic evaluations and guidelines are best used for uncovering page-level usability issues.
- Usability testing and cognitive walkthroughs are best used for uncovering site-level usability issues.
- Marketing research methods are not a substitute for usability evaluation.

References (1)

- R. Jeffries, J. Miller, C. Wharton, K. Uyeda. *User Interface Evaluation in the Real World: A Comparison of Four Techniques*, ACM (1991) 119-124.
- Constance Petersen, *Usability Testing*, Borland Community Developers News (Nov. 19, 1999) <http://community.borland.com/devnews/>.
- *Sun Usability Labs' Web Site*, Sun Microsystems, <http://www.sun.com/usability/>
- M. Levi, F. Conrad, *Usability Testing of World Wide Web Sites*, Bureau of Labor Statistics (May 1998) <http://stats.bls.gov/>

References (2)

- E. Boling, *Usability Testing for Web Sites*, Indiana University (Hypermedia '95 Conference) <http://www.indiana.edu/~iirg/ARTICLES/usability/usability.main.html>
- Jakob Nielsen, *Voodoo Usability*, Alertbox (Dec. 1999) <http://www.useit.com/alertbox/991212.html>
- Jakob Nielsen, *Cost of User Testing a Website*, Alertbox (May 1998) <http://www.useit.com/alertbox/980503.html>
- Jakob Nielsen, *Severity Ratings for Usability Problems*, <http://www.useit.com/papers/heuristic/severityrating.html>

References (3)

- Jakob Nielsen, *Usability Metrics*, Alertbox (Jan. 2001)
<http://www.useit.com/alertbox/20010121.html>
- Jakob Nielsen, *Why You Only Need to Test With 5 Users*, Alertbox (Mar. 2000)
<http://www.useit.com/alertbox/20000319.html>
- Jakob Nielsen, *Summary of Usability Inspection Methods*,
http://www.useit.com/papers/heuristic/inspection_summary.html
- Jakob Nielsen, *Ten Usability Heuristics*,
http://www.useit.com/papers/heuristic/heuristic_list.html

References (4)

- Jakob Nielsen, *How to Conduct a Heuristic Evaluation*,
http://www.useit.com/papers/heuristic/heuristic_evaluation.html

Presented by Jeremy Bradbury (bradbury@cs.queensu.ca) on March 7, 2001