CS 5150 Software Engineering

Evaluation and User Testing

William Y. Arms
Usability:
The Analyze/Design/Build/Evaluate Loop
Evaluation

If your system has users, the schedule should include time for user testing and time to make changes after the user testing is completed.

**When to do evaluation**

- Iterative improvements *during development*.
- Making sure that a system is usable *before launching it*.
- Iterative improvements *after launch*.

**Methods of evaluation**

- Empirical evaluation with users (*user testing*)
- Measurements on operational systems
- Analytical evaluation: without users (*not in CS 5150*)
Evaluation

How do you measure usability?

Usability comprises the following aspects:

Effectiveness
The accuracy and completeness with which users achieve certain goals
Measures: quality of solution, error rates

Efficiency
The relation between the effectiveness and the resources expended in achieving them
Measures: task completion time, learning time, number of clicks

Satisfaction
The users' comfort with and positive attitudes towards the use of the system
Measures: attitude rating scales

From ISO 9241-11
Evaluation based on Measurement

**Basic concept:** log events in the users’ interactions with a system

**Examples from a Web system**
- Clicks (when, where on screen, etc.)
- Navigation (from page to page)
- Keystrokes (e.g., input typed on keyboard)
- Use of help system
- Errors

May be used for statistical analysis or for detailed tracking of individual user.
Evaluation based on Measurements

Analysis of system logs

• Which user interface options were used?
• When was the help system used?
• What errors occurred and how often?
• Which hyperlinks were followed (click through data)?

Human feedback

• Complaints and praise
• Bug reports
• Requests made to customer service
## The Search Explorer Application: Reconstruct a User Session

### Search Sessions

<table>
<thead>
<tr>
<th>Session Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mon Apr 28 23:26:26 EDT 2003</td>
</tr>
<tr>
<td>2. Tue Apr 29 00:37:27 EDT 2003</td>
</tr>
<tr>
<td>3. Tue Apr 29 10:56:34 EDT 2003</td>
</tr>
<tr>
<td>4. Tue Apr 29 10:58:50 EDT 2003</td>
</tr>
<tr>
<td>6. Thu May 01 15:09:26 EDT 2003</td>
</tr>
<tr>
<td>7. Thu May 01 20:42:31 EDT 2003</td>
</tr>
<tr>
<td>8. Mon May 05 17:25:02 EDT 2003</td>
</tr>
</tbody>
</table>

### Queries for session 2

<table>
<thead>
<tr>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. haley's comet</td>
</tr>
<tr>
<td>2. haleys comet</td>
</tr>
</tbody>
</table>

### Results for query 2 [20 results]

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Interacting electrons in magnetic fields: Tracking p</td>
</tr>
<tr>
<td>4. Comets</td>
</tr>
<tr>
<td>5. Comets and meteor showers</td>
</tr>
<tr>
<td>Click #2</td>
</tr>
<tr>
<td>6. Comet Definitions</td>
</tr>
<tr>
<td>Click #3</td>
</tr>
<tr>
<td>7. Comet Observation Home Page</td>
</tr>
<tr>
<td>Click #5, 7</td>
</tr>
<tr>
<td>8. Comets</td>
</tr>
</tbody>
</table>

### Comet Definitions

The following are generally accepted definitions for terms related to cometary astronomy.

#### absolute magnitude (Ho)

The brightness of a comet when it is at 1 AU from both the Earth and Sun. As this virtually never happens, this quantity is calculated from the comet’s light curve. Unfortunately, this quantity is far from absolute. It can be different pre- and post-perihelion. It can also change from apparition to apparition (for periodic comets).

http://encke.jpl.nasa.gov/define.html
Evaluation with Users

Stages of evaluation with users:

1. Prepare
2. Conduct sessions
3. Analyze results

User testing is time-consuming, expensive, and essential.
Evaluation with Users: Preparation

Determine **goals** of the usability testing

“Can a user find the required information in no more than 2 minutes?”

Write the **user tasks**

“Given a new customer application form, add a new customer to the customer database.”

Recruit **participants**

Use the descriptions of users from the requirements phase to determine categories of potential users and user tasks
Usability Laboratory

Concept: monitor users while they use system

Evaluators

User

one-way
mirror
Evaluation with Users: Sessions

Conduct the session
  Usability Lab
  Simulated working environment

Observe the user
  Human observer(s)
  Video camera
  Audio recording

Inquire satisfaction data
Evaluation with Users: Results Analysis

**Test the system, not the users**

Respect the data and users' responses. Do not make excuses for designs that failed.

If possible, use statistical summaries.

Pay close attention to areas where users:

- were frustrated
- took a long time
- could not complete tasks

Note aspects of the design that worked and make sure they are incorporated in the final product.
Eye Tracking at Google

Dan Russell
2007
Eye Tracking at Google

1. Unicycle UK.com - F.A.Q. - What size?
   12" wheeled unicycle: this is a small children's unicycle size. It's good for children who are too small to ride a 16" unicycle, but it needs smooth ground ...
   www.unicycle.uk.com/FAQ.asp?Category=53 - 23k - Cached - Similar pages

2. Select a unicycle. Unicycle.com NZ: buy a unicycle or learn ...
   16" wheel unicycle: this is a children's unicycle, the small wheel makes it only suitable for smooth areas. Best used indoors or on smooth ground; ...

3. 100 Miles for Kids - The Goal
   "The Afghan Mobile Mini Circus for Children is an established ... attempt to break the GUINNESS WORLD RECORD for the ONE HOUR UNICYCLE DISTANCE RECORD ..."
   www.unicycle4kids.org/ - 9k - Cached - Similar pages

4. Unicycles page at Jugglingworld
   This is a children's unicycle, the small wheel makes it only suitable for very smooth areas. Best used indoors or on smooth ground; not so good outdoors ...
   www.jugglingworld.biz/shop/products_unicycles.html - 100k - Cached - Similar pages

5. Buy a Unicycle. Unicycle.com AU: buy a unicycle or learn unicycling
   Check out a Unicycle Leaners Pack for an easy and economical way to take your first steps into the One Wheeled World ... Suitable as a Children's Unicycle ...
   www.unicycle.au.com/View.php?action=Page&Name=Unicycles - 10k - Cached - Similar pages

6. Article - News - A unicycle ride for children
   Adam Brody, 21, of San Juan Capistrano, led a charity event Saturday that benefits the Orangewood Children's Foundation. The Unicycle Club of Southern ...

Dan Russell
2007
Evaluation Example: Eye Tracking
How we’re user testing:

- One-on-one, 30-45 min user tests with staff levels
- Specific tasks to complete
- No prior demonstration or training
- Pre-planned questions designed to stimulate feedback
- Emphasis on testing system, not the stakeholder!
- Standardized tasks / questions among all testers
How we’re user testing:

Types of questions we asked:

- Which labels, keywords were confusing?
- What was the hardest task?
- What did you like, that should not be changed?
- If you were us, what would you change?
- How does this system compare to your paper based system?
- How useful do you find the new report layout? (admin)
- Do you have any other comments or questions about the system? (open ended)
What we’ve found: Issue #1, Search Form Confusion!
A CS 5150 Project: Results

What we’ve found: Issue #2, Inconspicuous Edit/Confirmations!
A CS 5150 Project: Results

What we’ve found: Issue #3, Confirmation Terms
A CS 5150 Project: Results

What we’ve found: Issue #4, Entry Semantics

Category: Reference
Label: 1 to 5 minutes
Medium: In Person

Notes:

☑ Referred to a Librarian/Unit for this question?

Record this question how many times? 1
A CS 5150 Project: Results

What we’ve found: #5, Search Results Disambiguation & Semantics
Do not allow evaluators to become designers

Designers are poor evaluators of their own work, but know the requirements, constraints, and context of the design:

- Some user problems can be addressed with small changes
- Some user problems require major changes
- Some user requests (e.g., lots of options) are incompatible with other requests (e.g., simplicity)

Designers and evaluators need to work as a team