

# User Testing

November 28, 2011

# Announcements

- A4 is posted
  - Due Dec 7<sup>th</sup> by 11:30am

# Experiments, user testing & usability testing

- *Experiments* test hypotheses to discover new knowledge by investigating the relationship between two or more things
  - Conditions
  - Variables
- *User testing* is applied experimentation
  - Developers check that the system being developed is usable by the intended user population for their tasks
- *Usability testing* uses a combination of techniques, including user testing & user satisfaction questionnaires

# User testing vs. research testing

## **User testing**

- Aim: improve products
- Few participants
- Results inform design
- Not perfectly replicable
- Controlled conditions
- Procedure planned
- Results reported to developers, stakeholders

## **Research experiments**

- Aim: discover knowledge
- Many participants
- Results validated statistically
- Replicable
- Strongly controlled conditions
- Experimental design
- Scientific paper reports results to community

# User testing

- Goals & questions focus on how well users perform tasks with the product
- Comparison of products or prototypes
- Should be a major part of usability testing
- Focus is on time to complete tasks & number and types of errors
- Informed by video & interaction logging
- User satisfaction questionnaires provide data about users' opinions

# Testing environment

- Usability lab or other controlled space
- Major emphasis is on
  - selecting representative users
  - developing representative tasks
- 5-10 users typically selected
- Tasks usually last no more than 30 minutes
- The test conditions should be the same for every participant
- Informed consent form explains ethical issues

# Case Study: Testing MEDLINEplus

- Goals and questions
  - Identify range of usability problems
  - Particularly testing whether topic categorization on homepage works
  - Also testing whether navigation within site is effective



# Case Study: Testing MEDLINEplus

- Selection of participants
  - 9 participants from health care practices in DC area
  - Asked to fill out factual questionnaire
    - Screened based on frequency of web usage (> two times per month)
  - 7 Females, 2 Males

# Case Study: Testing MEDLINEplus

- Five tasks were developed
  - Wanted to check categorization and navigation support
  - Task 1: Find information about whether a dark bump on your shoulder might be skin cancer
  - Task 2: Find information about whether its safe to use Prozac during pregnancy
  - Task 3: Find information about whether there is a vaccine for hepatitis C
  - Task 4: Find recommendations about the treatment of breast cancer
  - Task 5: Find information about the dangers associated with drinking alcohol during pregnancy

# Case Study: Testing MEDLINEplus

- Test procedure
  - 5 scripts prepared in advance
  - Done in lab setting
  - Goals of the study and test procedure were explained

# Case Study: Testing MEDLINEplus

Thank you very much for participating in this study.

The goal of this project is to evaluate the interface of MEDLINEplus. The results of our evaluation will be summarized and reported to the National Library of Medicine, the federal agency that has developed MEDLINEplus. Have you ever used MEDLINEplus before?

You will be asked to use MEDLINEplus to resolve a series of specific, health-related information needs. You will be asked to "think aloud" as you search for information with MEDLINEplus.

We will be videotaping only what appears on the computer screen. What you say as you search for information will also be recorded. Your face will not be videotaped, and your identity will remain confidential.

I'll need you to review and sign this statement of informed consent. Please let me know if you have any questions about it. *(He hands an informed consent form similar to the one in Box 11.3 to the participant.)*

Figure 14.2 The script used to greet participants in the MEDLINEplus study.

# Case Study: Testing MEDLINEplus

We'll start with a general overview of MEDLINEplus. It's a web-based product developed by the National Library of Medicine. Its purpose is to link users with sources of authoritative health information on the web.

The purpose of our work today is to explore the MEDLINEplus interface to identify features that could be improved. We're also interested in finding out about features that are particularly helpful.

In a few minutes I'll give you five tasks. For each task you'll use MEDLINEplus to find health-related information.

As you use MEDLINEplus to find the information for each task, please keep in mind that it is MEDLINEplus that is the subject of this evaluation—not you.

You should feel free to work on each task at a pace that is normal and comfortable for you. We *will* be keeping track of how long it takes you to complete each task, but you should not feel rushed. Please work on each task at a pace that is normal and comfortable for you. If any task takes you longer than *twenty* minutes, we will ask you to move on to the next task. The Home button on the browser menu has been set to the MEDLINEplus homepage. We'll ask you to return to this page before starting a new task.

As you work on each task, I'd like you to imagine that it's something you or someone close to you needs to know.

All answers can be found on MEDLINEplus or on one of the sites it points to. But if you feel you are unable to complete a task and would like to stop, please say so and we'll move on to the next task.

Before we proceed, do you have any questions at this point?

Fall 2011 **Figure 14.3** The script used to explain the procedure.

# Case Study: Testing MEDLINEplus

Before we begin the tasks, I'd like you to explore MEDLINEplus independently for as long as ten minutes.

As you explore, please "think aloud." That is, please tell us your thoughts as you encounter the different features of MEDLINEplus.

Feel free to explore any topics that are of interest to you.

If you complete your independent exploration before the ten minutes are up, please let me know and we'll proceed with the tasks. Again, please remember to tell us what you're thinking as you explore MEDLINEplus.

Figure 14.4 The script used to introduce and describe the initial exploration task.

# Case Study: Testing MEDLINEplus

Please read aloud this task before beginning your use of MEDLINEplus to find the information.

After completing each task, please return to the MEDLINEplus home page by clicking on the “home” button.

Prompts: “What are you thinking?”

“Are you stuck?”

“Please tell me what you’re thinking.”

[*If time exceeds 20 minutes:* “I need to ask you to stop working on this task and proceed to the next one.”]

Figure 14.5 The script used to direct participants’ behavior.

# Case Study: Testing MEDLINEplus

How did you feel about your performance on the tasks overall?  
Tell me about what happened when [cite problem/error/excessive time].  
What would you say was the best thing about the MEDLINEplus interface?  
What would you say was the worst thing about the MEDLINEplus interface?

Figure 14.6 The debriefing script used in the MEDLINEplus study.

# Case Study: Testing MEDLINEplus

- Data collection
  - Completion times automatically recorded and calculated from the video and log data
  - Search moves of participants were collected
  - Data collection contained:
    - Start and completion times of tasks
    - Page count (i.e. pages accessed during the search task)
    - Medical publications accessed during the search task
    - Users' search path
    - Negative comments, mannerisms
    - User satisfaction questionnaire data

# Case Study: Testing MEDLINEplus

- Data analysis categorized according to:
  - Website organization such as arrangement of topics, menu depth, organization of links, etc.
  - Browsing efficiency such as navigation menu location, text density
  - Search features such as search interface consistency, feedback, etc.
- Conclusions and reporting the findings
  - Reaching external sites was often difficult
  - Analysis of search moves revealed that participants experienced difficulty finding health topics devoted to different types of cancer
  - Post-test questionnaire showed opinion was fairly neutral
  - Rated well for ease of learning but poorly for ease of use

# Case Study: Testing MEDLINEplus

Table 14.2 Performance data for task 1: Find information about whether a dark bump on your shoulder might be skin cancer. Mean (M) and standard deviation (SD) for all subjects are also shown.

Participant	Time to nearest minute	Reason for task termination	MEDLINEplus Pages	External sites accessed	MEDLINEplus searches	MEDLINE publication searches
A	12	Successful completion	5	2	0	2
B	12	Participant requested termination	3	2	3	0
C	14	Successful completion	2	1	0	0
D	13	Participant requested termination	5	2	1	0
E	10	Successful completion	5	3	1	0
F	9	Participant requested termination	3	1	0	0
G	5	Successful completion	2	1	0	0
H	12	Successful completion	3	1	0	6
I	6	Successful completion	3	1	0	0
M	10		3	2	1	1
SD	3		1	1	1	2

# Creating tasks

- A task is designed to probe a problem
- Tasks should be straightforward and require the user to find certain items, or do certain operations
- They can be more complex such as solving particular problems
- Sample tasks for a weather network web site:
  - What is the forecasted weather for Winnipeg?
  - What is air quality in Los Angeles today?
  - What is the level of humidity in Winnipeg?
  - What is the forecast for Ottawa for the upcoming weekend?

# How many participants is enough for user testing?

- The number is largely a practical issue
- Depends on:
  - schedule for testing
  - availability of participants
  - cost of running tests
- Typical 5-10 participants
- Some experts argue that testing should continue until no new insights are gained

# Activity

- You are developing a user test for the new CS web page. Identify 6 tasks for the test:
  - Task 1: Identify the instructor for Comp 3020
  - Task 2: Find the e-mail address of the Comp 3020 prof
  - Task 3: Find the admission requirements for the M.Sc. Program
  - Task 4: Find out the first day of classes next term
  - Task 5: Locate the requirements for being a Co-op student
  - Task 6: Identify whether the graduate Graphics course is a “fundamentals” course

# Elements of user testing

- Adopt DECIDE framework
- Determine the goals and Explore the questions
  - Not suitable for testing low-fidelity prototypes
  - Goals can be broad but need more specific questions
    - Can users complete a certain task within a certain time?
    - Can users find an item within a certain time frame?
- Choose the approach and techniques
  - “user testing” approach
  - Record data using a combination of video and interaction logging, user satisfaction questionnaires, and interviews

# Types of data (Wilson & Wixon, '97)

- Identify practical issues – what data to collect?
  - Time to complete a task
  - Time to complete a task after a specified time away from the product
  - Number and type of errors per task
  - Number of errors per unit of time
  - Number of navigations to online help or manuals
  - Number of users making a particular error
  - Number of users completing task successfully

# Elements of user testing

- Identify practical issues – select typical users
  - make sure you have appropriate representation
    - i.e. e-recipe primarily for families but 90% sample are single people
- Identify practical issues – prepare testing conditions
  - Lab preferably
- Identify practical issues – plan to run tests
  - Have scripts in place
  - Test equipment
  - Have recording material prepared
- Deal with ethical issues
  - Consent form

# Elements of user testing

- Evaluate, analyze, and present data
  - Report on times to complete task, number of errors
  - Provide simple statistical measures: mean, median, std dev.

# Questionnaires

- Earlier in the term we discussed questionnaire design for gathering requirements
- Most user satisfaction questionnaires consist primarily of *closed* questions
  - Participants encouraged to leave their comments in space provided on the page, or in the margins
- More on designing closed questions...



# Question and response format – Likert scales

- Steps for designing Likert scales:
  - Gather a pool of short statements about the features of the product that are to be evaluated
  - Divide the items into groups containing the same amount of positive and negative statements
  - Create logical/conceptual groups
  - Decide on the scale (5-point/3-point/9-point)
  - Select items for the final questionnaire and reword as necessary

# Likert scales – response options

- Odd/Even
  - if possible to have 'neutral' response, then use odd number of options (central = neutral place)
  - if judging something is good/bad, male/female then look at two response options
  - even numbers 'force' respondents in one way or another
    - end up with random responses between middle items
- How wide (1 to 3, 1 to 5, or even 1 to 12?)
  - how will the majority distinguish between flavours of meaning
  - if majority fairly uninformed about the topic, then use small number
  - if dealing with experts, then you can use a much larger set

# Anchors

- Anchors are the verbal comments above the numbers ('strongly agree', etc.)
- How many to include?
- In factual statements
  - considered good to use anchors above all options → will give you accurate results
    - News:            Daily    Weekly    Monthly    Never
- In opinion or attitude work
  - asking respondent to express their position on a scale of feeling from strong agreement to strong disagreement
  - helpful to indicate the central (neutral) point if meaningful, having numerous anchors may not be so important
    - Navigation on links is clear:
      - 1 (strongly agree)            3 (neutral)            5 (strongly disagree)

# Guidelines for questionnaire design

- See notes on from earlier in the term (recall)
  - Conciseness: questions should be clear and specific
    - e.g. should the system include a users manual? (YES/NO)
  - Closed questions: when possible ask closed questions and offer a range of answers
    - e.g. How often do you print checks? (1: very often – 5: never)
  - Alternate option: Consider including a “no-opinion” option for questions that seek opinions
    - e.g. the payroll module is essential (...N/A)
  - Order: think about the ordering of questions. General questions should precede specific ones
    - e.g. a question about a specific feature say in a payroll module should come after asking whether the payroll module is essential

# Guidelines for questionnaire design

- Break up multiple questions: Avoid complex multiple questions
  - e.g. is the payroll system and attendance manager efficient?
- Proper scales: when scales are used make sure the range is appropriate and do not overlap
  - e.g. 10...30, 31...40, ....
- Language: avoid jargon
  - e.g. should the display be based on bezier curves?
- Instructions: provide clear instructions on how to complete the questionnaire
  - e.g. please rate the performance of the following items
- Compactness: a balance must be struck between white space and the need to keep the questionnaire as compact as possible



# Analyzing questionnaire data

- Helps to think of analysis of questionnaire even before its design
  - Present results clearly - tables can be used for proper structure
  - Simple statistics can say a lot, e.g., mean, median, mode, standard deviation
  - Percentages are useful but give population size
  - Bar graphs show categorical data well
  - More advanced statistics can be used if needed

# User Testing: Recap

- User testing is a form of applied experimentation
- User testing supplemented with satisfaction questionnaires is known as usability testing
- Key components:
  - Representative users
  - Representative tasks
- Most common data collected
  - Task completion times
  - Errors
- Tasks should be focused and specific
- Plan procedure carefully and apply the same procedure same for all participants

# User Testing: Recap

- User satisfaction questionnaires:
  - Consist primarily of closed questions
  - Likert scales often used to collect opinions, attitudes, beliefs
- Data analysis: transforming raw data into something more meaningful and digestible
  - Use aggregation, visualization and statistical analysis
    - E.g., Means, standard deviations, graphs, tables, percentages
- Drawing conclusions
  - What do the results mean for your system?
    - i.e., What does your analysis say about what worked and what didn't