Week 13

Evaluation (Part 2)

Goals

- Explain how to do usability testing

- Outline the basics of experimental design

- Describe how to do field studies

Usability testing

- Involves recording performance of typical users doing typical tasks
- Controlled settings
- Users are observed and timed
- Data is recorded on video, and key presses are logged
- The data is used to calculate performance times and to identify and explain errors
- User satisfaction is evaluated using questionnaires and interviews
- Field observations may be used to provide contextual understanding

Quantitative performance measures

- Number of users successfully completing the task

- Time to complete task

- Time to complete task after time away from task

- 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 type of error

Source: Wixon and Wilson, 1997

Usability lab with observers watching a user and assistant

Tobii Glasses Mobile Eye-Tracking System

Source: Dalton et al., 2015, p.3891. Reproduced with permission of ACM Publications.

Portable equipment for use in the field

Setup used in the Chicago usability testing sessions

Source: iPad App and Website Usability Study. Used courtesy of the [Neilsen Norman Group](#).

Testing the iPad usability

- First study was conducted quickly in two cities: Fremont, CA and Chicago, IL

- Tests had to be done quickly, as information was needed by third-party app developers

- Also needed to be done secretly so that the competition was not aware of the study before the iPad was launched

- Seven participants with over three months experience with iPhones

Testing the iPad usability *(continued)*

- Signed an informed consent form explaining:
    - What the participant would be asked to do
    - The length of time needed for the study
    - The compensation that would be offered for participating
    - Participants' right to withdraw from the study at any time
    - A promise that the person's identity would not be disclosed
    - An agreement that the data collected would be confidential and available to only the evaluators

- Participants were asked to explore the iPad

- Next, they were asked to perform randomly-assigned specified tasks

Examples of the tasks used in the iPad evaluation

App or website	Task
iBook	Download a free copy of *Alice's Adventures in Wonderland* and read through the first few pages.
Craigslist	Find some free mulch for your garden.
eBay	You want to buy a new iPad on eBay. Find one that you could buy from a reputable seller.
Time Magazine	Browse through the magazine and find the best pictures of the week.
Epicurious	You want to make an apple pie for tonight. Find a recipe and see what you need to buy in order to prepare it.
Kayak	You are planning a trip to Death Valley in May this year. Find a hotel located in the park or close to the park.

Adapted from Budiu and Nielsen, 2010
Source: iPad App and Website Usability Study. Used courtesy of the Neilsen Norman Group.

Problems and actions

- Examples of problems detected:
  - Accessing the Web was difficult
  - Lack of affordance and feedback
  - Getting lost in an application
  - Knowing where to tap

- Actions by evaluators:
  - Reported to developers
  - Made available to public on [Neilsen Norman Group](.).

Problems and actions *(continued)*

- Accessibility for all users is important

- Study did not address how iPad would be used in people's everyday lives

- Another study was done a year later to examine this and other issues that there was insufficient time to address in the first study

Usability testing conditions

- Usability lab or other controlled space

- Emphasis on:
  - Selecting representative users
  - Developing representative tasks

- 5-10 users typically selected

- Tasks usually around 30 minutes

- Test conditions are the same for every participant

- Informed consent form explains procedures and deals with ethical issues

How many participants is enough for user testing?

- The number is a practical issue

- Depends on:
  - Schedule for testing
  - Availability of participants
  - Cost of running tests

- Typically 5-10 participants

- Some experts argue that testing should continue until no new insights are gained

Usability testing and Experiments

- Usability testing is applied experimentation

- Developers check that the system is usable by the intended user population by collecting data about participants' performance on prescribed tasks

- Experiments test hypotheses to discover new knowledge by investigating the relationship between two or more variables

Usability testing and research

Usability Testing

- Improve products

- Few participants

- Results inform design

- Usually not completely replicable

- Conditions controlled as much as possible

- Procedure planned

- Results reported to developers

Experiments for Research

- Discover knowledge

- Many participants

- Results validated statistically

- Must be replicable

- Strongly controlled conditions

- Experimental design

- Scientific report to scientific community

Experiments

- Test hypothesis

- Predict the relationship between two or more variables

- Independent variable is manipulated by the researcher

- Dependent variable influenced by the independent variable

- Typical experimental designs have one or two independent variables

- Validated statistically and replicable

Experimental designs

Different participants (between subjects):

Single group of participants is allocated randomly to the experimental conditions

Same participants (within subjects):

All participants appear in both conditions

Matched participants (pairwise):

Participants are matched in pairs, for example, based on expertise, gender, and so on

Different, same, matched participant design

Design	Advantages	Disadvantages
Different	No order effects	Many subjects and individual differences a problem
Same	Few individuals, no individual differences	Counter-balancing needed because of ordering effects
Matched	Same as different participants, but individual differences reduced	Cannot be sure of perfect matching on all differences

Field studies

- Field studies are done in natural settings
- "In the wild" is a term for prototypes being used freely in natural settings
- Seek to understand what users do naturally and how technology impacts them
- Field studies are used in product design to:
  - Identify opportunities for new technology
  - Determine design requirements
  - Decide how best to introduce new technology
  - Evaluate technology in use

A field study of a pain-monitoring device

- Monitoring patients' pain is a known challenge for physicians
- Goal of the study was to evaluate the use of a pain-monitoring device for use after ambulatory surgery
- Painpad is a keypad device
- It was usability tested extensively in the lab before brought into two hospitals
- Goal was to understand how Painpad was used in the natural environment and as part of routines in two UK hospitals.
- How pain-monitoring differed with Painpad

Painpad

A tangible device for inpatient self-logging of pain

Source: Price et al., 2018. Reproduced with permission of ACM Publications.

Data collection and participants

- Two studies in two hospitals involving 54 people

- 13 males, 41 females

- Privacy was a important concern

- Hospital stay ranged from 1-7 days, mean and median age 64.6, 64.5

- Patients given Painpad after surgery and prompted to report pain levels every two hours

- Nurses also collected scores

- All data entered into charts

- Patients in one hospital were given a user-satisfaction survey when they left

- Also rated Painpad on a 1-5 Likert scale

Data analysis and presentation

- Three types of data were collected:
    - Satisfaction with Painpad was based on questionnaire responses
    - Patients' compliance with the two-hour routine
    - How data collected from Painpad compared with data collected by nurses
- Data showed:
    - Satisfaction with Painpad 4.63 on Likert scale
    - Patience compliance was mixed: some liked it while others disliked or didn't notice the prompts
    - Patients recorded more scores with Painpad than through the nurses

Summary

- Usability testing takes place in controlled usability labs or temporary labs

- Usability testing focuses on performance measures, for example, how long and how many errors are made when completing a set of predefined tasks

- Indirect observation (video and keystroke logging), user satisfaction questionnaires, and interviews are also collected

- Affordable, remote testing systems are more portable than usability labs

- Many also contain mobile eye-tracking and other devices

Summary *(continued)*

- Experiments test a hypothesis by manipulating certain variables while keeping others constant

- The experimenter controls independent variable(s) in order to measure dependent variable(s)

- Field studies are evaluation studies that are carried out in natural settings to discover how people interact with technology in the real world

- Field studies that involve the deployment of prototypes or technologies in natural settings may also be referred to as 'in-the-wild' studies

- Sometimes the findings of a field study are unexpected, especially for in-the-wild studies that explore how novel technologies are used by participants in their own homes, places of work, or outside