Testing & modeling users

The aims

- Describe how to do user testing.
- Discuss the differences between user testing, usability testing and research experiments.
- Discuss the role of user testing in usability testing.
- Discuss how to design simple experiments.
- Describe GOMS, the keystroke level model, Fitts' law and discuss when these techniques are useful.
- Describe how to do a keystroke level analysis.

Experiments, user testing & usability testing

- Experiments test hypotheses to discover new knowledge by investigating the relationship between two or more things – i.e., variables.
- User testing is applied experimentation in which 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 is not research

User testing

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

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 common
- Major part of usability testing
- Focus is on time to complete task & number & type of errors
- Informed by video & interaction logging
- User satisfaction questionnaires provide data about users' opinions

Testing conditions

- Usability lab or other controlled space
- Major emphasis 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

Type of data (Wilson & Wixon, '97)

- 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

Usability engineering orientation

- Current level of performance
- Minimum acceptable level of performance
- Target level of performance

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

Experiments

- Predict the relationship between two or more variables
- Independent variable is manipulated by the researcher
- Dependent variable depends on the independent variable
- Typical experimental designs have one or two independent variable

Experimental designs

- Different participants - single group of participants is allocated randomly to the experimental conditions
- Same participants - all participants appear in both conditions
- Matched participants - participants are matched in pairs, e.g., based on expertise, gender

Advantages & disadvantages

Design	Advantages	Disadvantages
Different	No order effects	Many subjects & 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

Predictive models

- Provide a way of evaluating products or designs without directly involving users
- Psychological models of users are used to test designs
- Less expensive than user testing
- Usefulness limited to systems with predictable tasks - e.g., telephone answering systems, mobiles, etc.
- Based on expert behavior

GOMS (Card et al., 1983)

- Goals - the state the user wants to achieve e.g., find a website
- Operators - the cognitive processes & physical actions performed to attain those goals, e.g., decide which search engine to use
- Methods - the procedures for accomplishing the goals, e.g., drag mouse over field, type in keywords, press the go button
- Selection rules - determine which method to select when there is more than one available

Keystroke level model

GOMS has also been developed further into a quantitative model - the keystroke level model.

This model allows predictions to be made about how long it takes an expert user to perform a task.

Response times for keystroke level operators (excerpt from

Operator	Description	Time (sec)
K	Pressing a single key or button Average skilled typist (55 wpm) Average non-skilled typist (40 wpm) Pressing shift or control key Typist unfamiliar with the keyboard	 0.22 0.28 0.08 1.20
P P1	Pointing with a mouse or other device on a display to select an object. This value is derived from Fitts' Law which is discussed below. Clicking the mouse or similar device	0.40 0.20
H	Bring 'home' hands on the keyboard or other device	0.40
M	Mentally prepare/respond	1.35
R(t)	The response time is counted only if it causes the user to wait.	t

Fitts' Law (Paul Fitts 1954)

- The law predicts that the time to point at an object using a device is a function of the distance from the target object & the object's size.
- The further away & the smaller the object, the longer the time to locate it and point.
- Useful for evaluating systems for which the time to locate an object is important such as handheld devices like mobile phones

Key points

- User testing is a central part of usability testing
- Testing is done in controlled conditions
- User testing is an adapted form of experimentation
- Experiments aim to test hypotheses by manipulating certain variables while keeping others constant
- The experimenter controls the independent variable(s) but not the dependent variable(s)
- There are three types of experimental design: different-participants, same- participants, & matched participants
- GOMS, Keystroke level model, & Fitts' Law predict expert, error-free performance
- Predictive models are used to evaluate systems with predictable tasks such as telephones