COMMONWEALTH OF AUSTRALIA

Copyright Regulations 1969

WARNING

This material has been copied and communicated to you by or on behalf of the University of New South Wales pursuant to Part VB of the Copyright Act 1968 (the Act).

The material in this communication may be subject to copyright under the Act. Any further copying or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice.

Usability Evaluation Techniques

COMP3511/9511 Human Computer Interaction

Alex Mednis

Adapted from slides by Dr Daniel Woo and Dr Nadine Marcus

Readings/References

- Interaction Design , Ch12,14 &15
- Rubin and Chinsell (2008) Handbook of Usability Testing How to Plan, Design and Conduct Effective Tests (2nd Ed)
- Nielsen (1993), Usability Engineering
- Wolfe, Hiser

why, what, and when

why test

- Ensure application can be used
- Ensure application works as expected
- Ensure application meets particular criteria
- Measure productivity gains between using this application and another

what to test

- Part of an application
- All of an application
- Competing designs
- Icons/graphics
- Online help

when to test

- Prototypes (starting early)
- Alpha/beta code
- Deployed application

ethical issues when testing

- Participants must be assured of the following:
  - that this is a 'test' of the software, not of them
  - they may stop at any time
  - their performance and opinions will be kept anonymous (if desired)

ID 13.2.5

usability testing phases

- Planning the sessions
- Running the sessions
- Analysing, documenting, and presenting the results

planning the test

planning activities

- Set goals and measures
- Determine what and how to measure
- Define scenarios and tasks
- Decide how to collect data
- Select participants and define test roles

planning activities - continued

- Prepare materials
- Invite and brief users and observers
- Select environment and set up equipment
- Select and obtain software and hardware
- Conduct dry run

Copyright Nadine Marcus @CSE UNSW

setting goals

- Defining goals
  - learnability
  - efficiency
  - memorability
  - minimal errors
  - satisfaction
  - ...

the most appropriate goals

- The nature of the application determines goals and criteria
  - an air traffic control system must be error-proof
  - a word processing application must be easy to learn or quick to use

criteria for achieving goals

- Usability goals should be measurable, objective, and concrete
  - number of errors per task
  - number of errors per hour
  - length of time to complete task (relative, not absolute, measure)

acceptability of results

Usability exceeds expectations

Usability meets expectations

Minimal usability

Usability is below expectations

what to measure

- Initial reactions
- Users' exploration
- Successful completion of tasks
- Timely completion of tasks
- How well tasks are supported
- Users' satisfaction

possible metrics

- Time required to learn
- Time required to complete tasks
- Number of errors
- Severity of errors
- Amount of assistance required
- Ratings and comments

Copyright Nadine Marcus @CSE UNSW

determining what to test

- Those 20% of the functions used 80% of the time
- Problematic or critical functions (even if used infrequently)
- Functions difficult to design
- Functions difficult to document
- Functions difficult to teach
- Consider the scientific method

putting functions into context

- Task Scenarios
  - users and their goals, usage context
  - actions and artifacts used
  - whether action is independent or part of another goal
  - special processing conditions

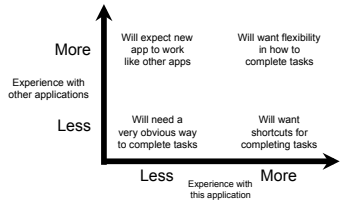
putting functions into context

- Tasks
  - set activities presented in a logical order
  - derived from scenarios

selecting who to test

- Representative users from each category
- Wide range of experience
- Wide range of usage
- Cross-section of user population
- Try to select randomly
  - Consider the scientific method

implications of experience



how many users

- Be realistic about time and budget
- Four to five per category
- When trends emerge, stop
- If no trends emerge, continue

- We will talk about the difference with scientific method later…

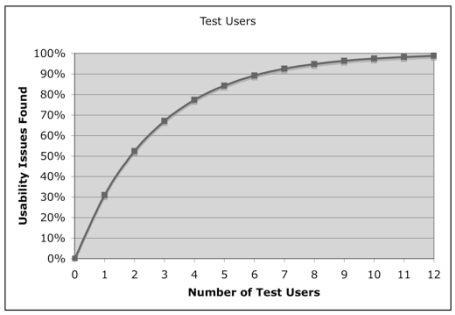
Issue Table

	P1	P2	P3	P4
Selected wrong menu item in view menu	X			
Did not understand zoom icons	X		X	
Choices pop up menu needed Other... option to add		X		
Confused about modifier keys	X	X	X	X
Incorrectly used date field		X		X
...				

No prioritisation/analysis , just capturing the observations

Approximation

- Nielsen and Landaur (1993, in Nielsen, Usability Engineering) proposed that an approximation to assessing the number of usability problems found is given by:
- Problems_found(i) = $N(1 - (1 - L)^i)$
  - i = number of test users
  - N = total number of usability problems in the interface
  - L = probability for finding any single problem with any single users



Based on Nielsen p173, N=41, L=31%

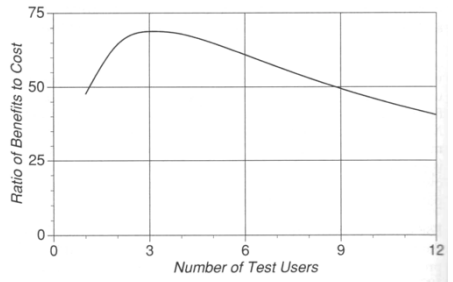


Figure 18 *The pay-off ratio (how much larger the benefits are than the costs) for user tests with various numbers of test users under the assumptions for a "typical" medium-sized project described in the text.*

p174, Nielsen, Usability Engineering, 1993

test roles

- Test co-ordinator
- Test facilitator
- Observer(s)
- Camera/equipment operator

data collection techniques

- Verbal protocols
- Active intervention
- Co-discovery

verbal protocols

- Useful for establishing learnability and effectiveness
  - users participate individually
  - users work through scenarios or tasks
  - users think aloud
  - facilitator may assist if user is frustrated

active intervention

- Useful for establishing effectiveness
  - users participate individually
  - facilitator probes and asks specific questions
  - requires skill in knowing when to intervene

co-discovery

- Useful for assessing general design of early prototypes
  - two or more participants
  - participants encouraged to talk to each other as they explore

co-discovery

- Tips
  - find people who know each other
  - if one is dominant, let the other drive
  - allow only one to 'drive' at a time

eliciting verbal feedback

- 'What are you thinking?'
- 'Describe the screen for me.'
- 'What do you think will happen?'
- 'Did you expect that to happen?'

error handling

- Give users a hint after x minutes
- Give users a hint after x errors
- Give answers after x minutes or errors and go on to next task
- If it is difficult, it is the interface (not the user) at fault!

Copyright Nadine Marcus @CSE UNSW

questionnaires

- Capture demographic information
- Capture feedback on usefulness and usability of the application
- Ensure appropriateness of terminology
- Capture opinions not expressed during the test

format of questions

- Rating scales for usefulness and usability
- Rating scales for terminology
- Series of statements about the application, rating how much they agree with each statement
- Open ended questions (e.g., what are the three things you like best/worst?)

tips on creating questionnaires

- Keep questionnaires short
- Use plain language

tips on creating questionnaires

- When using rating scales:
  - use five-point scales
  - provide n/a option (not applicable) or no opinion
  - provide anchors on scales (1= very difficult, 5 = very easy)
- Provide space for additional comments

briefing participants

- Application background
- Importance of involvement
- High-level overview of procedure
- Incentive for participation
- Results will remain anonymous
- It's a test of the interface, not the user

logging sheets

- Two basic kinds of (manual) logging sheets:
  - task-specific log sheets (with correct steps listed)
  - general log sheets (with space for errors, accuracy, and time)
- Be sure to allow plenty of room to record comments

Copyright Nadine Marcus @CSE UNSW

consent forms

- Necessary for any data collection
  - what will be observed
  - what will be recorded
  - how the data will be used

appropriate environment

- Test on appropriate environment
  - minimum hardware/software configuration
  - typical hardware/software configuration
- Identify (and test with) applications expected to co-exist with the application

'dry run'

- Ensures that scenarios/tasks can be performed with current state of software and hardware
- Ensures tasks make sense
- Validates usability of questionnaire and logging sheets
- Provides an estimate of time required for each participant

running the test

steps to running the test

- Ensure equipment is functional
- Ensure materials are available
- Ensure observers have assembled
- Introduce user to test procedure
- Emphasise that participant not being tested

running the test - continued

- Participant signs consent form
- Complete demographic questionnaire
- Facilitator instructs participant from a script (so each participant given same information)
- Participant completes tasks
- Complete feedback questionnaire
- Thank participant

Copyright Nadine Marcus @CSE UNSW

facilitator

observers

debriefing

- Ideally done immediately after each participant
- Allows rapid turnaround of results
- Facilitates collation of observations
- Provides opportunity to summarise
- Helps categorise and prioritise
- Helps obtain consensus on issues

analysing, documenting, and reporting the results

after the test

- Collate results
- Interpret results
- Document results
- Deliver results

data analysis

- Log sheets
- Results of debriefing
- Questionnaire results
- Video (if used)

Copyright Nadine Marcus @CSE UNSW

diagramming the results

prioritising the results

- Assign severity ratings
  - frequency of problem
  - impact if problem occurs
  - persistence of problem

Sample Severity Ratiings

- 4 = showstopper (cannot release until fixed)
- 3 = major problem (must fix)
- 2 = minor problem (should fix)
- 1 = cosmetic problem (fix if there is time)
- 0 = not a usability problem (e.g., bug)

reporting the results

- Summarise and present findings immediately
- Document for future reference
  - executive summary
  - introduction
  - method
  - results and recommendations
  - appendices

problem reporting

- Use problem reporting structure of development environment
- Lends more credibility to usability issues
- Makes someone accountable for addressing problems

a bit of advice

- When presenting results of a usability test, always try to find something positive to say!

other testing issues

testing desktop vs. web sites

- A lot of similarities
  - ensure proper users
  - ensure proper tasks
  - understand the business and user drivers

testing desktop vs. web sites

- Many differences
  - desktop applications focus primarily on functionality while visual design also becomes important on the web
  - users' expectations of the web is different
  - clicking and interactivity

to video or not to video

- Reviewing user interactions
- Creating a highlights tape for development teams
- Multi-media reports
- Teaching purposes

comparison of techniques

	effectiveness	impact	effort
Live Observation	high	high	none (automatic)
Video Highlights	medium	high	lots (weeks)
Written Report	low	medium	days

usability testing labs

- Observation & test areas
- Audio/video recording and editing equipment
- Monitors for observing
- Logging software

lab example

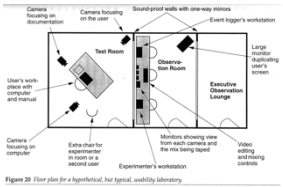


Figure 20 *Floor plan for a hypothetical, but typical, usability laboratory.*

p156, Nielsen, Usability Engineering

Sun Microsystems (site)

usability testing labs

- Pros
  - Facilities in place
  - Many ways to capture data
  - Higher quality results
  - Minimises interruptions
  - Shows commitment to usability

Cons

- Expensive
- May be overkill
- Requires users to travel
- Artificial environment
- Can be intimidating

users' work area

Pros

- Captures true context of use
- Most convenient to users
- Inexpensive

Cons

- Moving equipment
- Environment may be inappropriate

other sites

other sites

other sites

Pros

- Flexibility in location of users
- Cost effective
- Minimises interruptions

Cons

- Difficult to organise from off-site
- Moving equipment
- Contextual information is minimal

do-it-yourself testing

- Users install software
- Users step through tasks and record responses on logging sheets
- Users complete questionnaires

do-it-yourself materials

- Software
- Documentation
- Tasks
- Logging sheets
- Questionnaire

indirect observation

- Diaries
  - ask users to provide a record of what they did
- Interaction logging
  - tracking users' actions (e.g., key presses, mouse or other device movements)
  - Other counters (e.g., page hits)
  - Consider the ethical issues

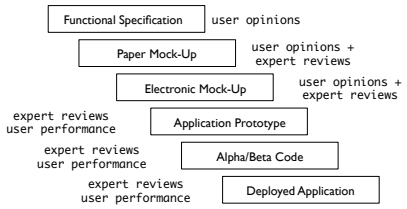
remote testing ideas

- Via the web
- Via teleconferencing facilities

the value of evaluations

- Usability evaluations are essential for ensuring an application meets its users' needs
- Usability evaluations should be conducted throughout the development of an application

evaluations throughout development



evaluations throughout development