Usability Testing

Refs: Lecture note from D. Petkovic

Objective

Ä In order to help you with designing and executing usability testing, I am summarizing main points from class 3 and adding a few more practical guidelines.

Ä Recommend book:

Â

Technical Communication Library, 1994

Usability Testing

Ä Process that employs representative target population participants to evaluate product usability using specific usability criteria

Ä Usability testing is not a guarantee for product success, but it should identify at least the key problems

Basic components

1. Development of specific problems statements and test plans and objectives
2. Use of representative sample of end users
3. Representation of the actual work environment
4. Observation of end users during product use or review
5. Collection of quantitative and qualitative measurements
6. Analysis of results and recommendations

Types of usability tests

Exploratory:

Ä Early in the process

Ä Can be based on any form of the GUI (sketch, wire diagrams etc.)

Ä Evaluate preliminary, basic design concept

Ä

Ä Informal test methodology, a lot of interaction

Ä Discuss high level concepts

Types of usability tests

Assessment

Ä Done After fundamental concepts are done

Ä Evaluates usability of lower level operations

Ä The users actually perform set of well defined tasks

Ä Less interaction with test monitor

Ä Quantitative measurements are collected

Types of usability tests

Validation

Ä Done late in development cycle, close to release

Ä

Ä Often the first time when the whole product si tested (including help and docs)

Ä Evaluate wrt. some predetermined usability standard or benchmark

Ä Standards come from previous testing, competitive information, marketing etc.

Ä Very specific quantitative tests

Ä Can establish standards for future products

Ä Can be also done by beta customers

Types of usability tests

Comparison

Ä Can be done at any point in the development cycle

Ä Compare alternatives using objective measures

Ä Can be informal of formal, depending when it is done

Ä Often, best of alternative designs is combined

Test environments

- Ä Simple single room setups
  - Â Observer/monitor close to evacuator
  - Â Observer removed from evaluator
- Ä Electronic observation room
- Ä Classic elaborate usability lab
- Ä Mobile lab

Typical test plan format

Ä Purpose: what is the main purpose of the test

Ä Problem statement: specific questions you want resolved

Ä Test plan and objectives: tasks the user will do

Ä User profile: who will be the users

Ä Method and test design: how will you observe it, how will you collect the data etc.

Ä Test environment and equipment

Ä Test monitor role

Ä Evaluation measures and data to be collected: how will you collect the feedback and how will you evaluate it

Ä Report: what will final report contain

Task selection for evaluation

- Ä Tasks to be evaluated are functions users want to do with the product. Focus is on user view of the tasks and NOT at the components and details that you used to implement it. Examples:
  - Â Create and file document
  - Â Import several images
  - Â Find the right document
- Ä Objective is to indirectly expose usability flaws by asking the user to perform typical tasks and NOT telling them exactly how to do it.
- Ä Choose key and most frequently done tasks
- Ä The task has to be specific and measurable (quantitatively or qualitatively)

Task components

TASK	DESCRIPTION
Task	Load paper into copier
Machine state	Paper tray empty
Successful completion criteria	Paper properly loaded
Benchmark	Completed in 1 minute

Selection of evaluators and test groups

Ä Evaluators should be representative of the targeted users

Ä Independent groups or within-subject design (but be careful to avoid exposing users to same tests since this will bias the results)

Ä Adequate numbers of testers

Ä Offer motivation and rewards

Measurements and Questionnaires

Ä Performance data: measures of user behavior such as error rates, number of accesses to help, time to perform the task etc.

- Â Usually can and should be objectively and automatically measured

Ä Preference data: measures of user opinion, thought process such as rankings, answers to questions, comments etc.

- Â Use questionnaires.

Some performance measures (measure what can be measured)

Ä Time to complete each task

Ä Number and percentage of tasks completed successfully/unsuccessfully

Ä Time required to access information

Ä Count of incorrect selections

Ä Count errors

Ä Time for system to respond

Ä

Data should be collected automatically or manually in an objective way.

Questionnaires (for preference data)

Likert scale

Ä I found GUI easy to use (check one)

__ Strongly disagree __ Disagree

__ Neither agree or disagree

__ Agree __ Strongly agree

(can also assign numbers from 2 to 2)

Semantic differentials

Ä I found File Open menu (circle one)

Simple 3 2 1 0 1 2 3 Complex

Questionnaires

Fill in questions

Ä I found the following aspects of GUI particularly easy to use (list 0-4 aspects)

Questionnaires

Check-box

Ä Please check the statement that best describes your usage of spell check

--- I always use spell check

--- I use spell check only when I have to

--- I never use spell check

Questionnaires

Branching questions

Ä Would you rather use advanced search

--- NO (skip to question 19)

--- YES (continue)

Ä What kind of advanced search would you like? (check one)

--- Boolean

--- Relevance

Summarizing performance results

- Ä Performance data
  - Â Mean time to complete
  - Â Median time to complete
  - Â Range (high and low)
  - Â Standard deviation of completion times
  - Â System response time statistics
- Ä Task accuracy
  - Â % of users completing the task within specified time
  - Â % of users completing the task regardless of time
  - Â Same as above, with assistance
  - Â Average error rate

Summarizing preference results

- Ä For limited choice questions
  - Â Count how many participants selected each choice (number and %)
  - Â For Likert scale or semantic differentials provide average scores if there are enough evaluators
- Ä For free form questions
  - Â List questions and group answers into categories, also into positive and negative answers
- Ä For free comments
  - Â List and group them at the end of the report

Analyzing Data

- Identify and focus on tasks that did not pass the test or showed significant problems
- Identify user errors and difficulties
- Identify sources of errors
- Prioritize problems by criticality = severity AND probability of occurrence
- Analyze differences between groups (if applicable)

- Provide recommendations at the end

Problems statements and performance data to collect

Problem Statement	**Performance Data Collected**
How effective is the tutorial	Compare error rates of users who used and not used this
How easy is it to perform task X	Error rate OR Number of steps needed

Note: this is Performance data measurement only. You also need to asses user Preference data (see next slide)

Problems statements and preference data to collect

Problem Statement	**Preference Data Collected**
How effective is the tutorial	Ask user to rate it from very ineffective to very effective (Lickert scale or semantic differentials) + free comments
How easy is it to perform task X	Ask user to rate it from very easy to very difficult (Lickert scale or semantic differentials) + free comments

Relate problem statements with tasks

Problem Statement	**Task**
How effective is the tutorial	GroupA: Import image w/o using tutorial GroupB: Same but use tutorial first
How easy is it to Create Virtual Machine	Create Virtual machine with "this" properties using New VM Wizard

Task components

TASK	**DESCRIPTION**
Task	Create VM using New VM Wizard
Machine state	VMware WS SW just loaded
Successful completion criteria	Working VM created
Benchmark	Completed in 30 sec.

Example questionnaire

Question 15:

It was easy to create virtual machine

●	○	○	○	○
Strongly disagree	Disagree	Neutral	Agree	Strongly agree

Comments:

<< Previous

Next >>

END

Some Suggested GUI issues to cover in preference data collection

Use GUI principles as general measures of quality to evaluate

Ä Screen layout matches tasks

Ä Amount of information is adequate

Ä

Ä Good use of colors

Ä Proper grouping of related info

Ä Navigational problems

Ä Users get lost in the system

Ä Organized by user tasks

Ä Icons are self-explanatory

Ä Consistency

Note: ask these questions in the context of concrete user tasks not in vacuum