### Goals of the evaluation

- Are we evaluating in order to improve the design? **Formative evaluation!**
- Are we evaluating in order to determine if the prototype/system fulfills the customer's requirements and can be delivered? **Summative evaluation**

#### Evaluate against the goals/requirements
- Effect goals, result goals, product goals
  - ex: increase sales, reduce dropouts ...
- Requirements
  - data, functionality, qualities

### How to measure goal fulfilment

- When in the design process?
- Available user?

#### Two types
- User testing
- inspection methods

### When to use what?

#### Applicable testing method vs system representation

- Sketches, scenarios, storyboards
- Paper prototypes
- Computer prototypes
- Working system

- Inspection methods
- User testing
- Inspection methods
Inspection methods (usually formative, but summative possible)

- Cognitive walkthrough
- Action Analysis (back-of-the-envelope version)
- Heuristic Analysis (evaluation)

Nielsen’s usability heuristics

- Visibility of system status
- Match between system and real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Esthetic and minimalistic design
- Help users recognize, diagnose and recover from errors

Heuristic evaluation

- Select set of heuristics to use
- 1-5 usability experts receive the representation of the system (prototype/specification/scenario...)
- Expert explores system to become familiar
- Expert checks system against heuristics and notes discrepancies and problems
- Collect and analyze data
User testing (formative or summative)

- Setting a user in front of a system or prototype.
  - Determine task to use to test requirement
  - Determine what to look for (measurements, observations etc)
  - Identify, contact and brief subjects
  - Run pilot test, and adjust preparations if necessary
  - Run test and collect data
  - Analyse data

Roles in user testing

- Test leader - welcomes the subject, introduces the session and purpose, presents test tasks (one by one). After tasks asks interview questions or administers questionnaire. May try to take notes, but lower prio.
- Observer - sits quietly observing and taking notes on whatever the user is doing and saying, signs of confusion (facial expressions, hesitations etc).
- Subject - the user.

User testing a paper prototype

- Four roles:
  - Test leader
  - Observer
  - Subject
  - COMPUTER

Think-aloud testing

- The subject is asked to think aloud during execution of the test tasks
- In order to allow the observer to track the reasoning.
- BUT: when tasks get too complex, it is natural to fall silent. Test leader may prompt: What are you thinking about? Please speak?
- Test leader MAY NOT help!
Exercise: user test your mobile

- Function to test: Telephone books for storing telephone numbers
- Should require no learning, “walk-up-and-use”
- Requirement/user goal: Easy to use
  - Summative: A new user should be able to make a new entry into the phone book with less than 2 seconds of hesitation at each possible choice.
  - Formative: usability problems?

Task: user test your mobile

- Form groups of 3
- Select one mobile phone
- Choose roles (test leader, observer and subject):
  - If possible, the subject should not be familiar with the type of phone chosen
  - The owner should be the observer

Exercise: user test your mobile

- User task 1:
  - Starting from "start screen" (top screen), add your own name and telephone number to the telephone book.

Exercise: user test your mobile

- User task 2:
  - Again starting from the "start screen", call yourself
    - note: don’t answer ;-)
Test leader:

- Make sure the phone is showing top screen.
- Deliver task 1
- Remind the subject to think aloud. If you like you may try to take notes during the test.
- When the subject considers him/herself done: Check that the number is properly added and return the phone to top screen.
- Deliver task 2 (note; the user should not answer, that would cost the owner money!)

Test leader: Interview questions

- Was it easy?
- Did any step confuse you?
  - If so, why, what were you expecting? Why would you choose or expect that?
- Could it be made simpler?
  - What would you gain by doing that?
- How similar is this phone to your own (if you have one)?

Observer:

- Make observation protocol on a sheet of paper
  - write a short description (one to two words per step) on what steps need to be taken for each task, leaving lots of space around the text. This is the “ideal path”.
  - Think about possible mistakes and make notes at appropriate places in the ideal path.
- When observing:
  - note when user does not follow the ideal path
  - look for hesitations (note that a hesitation of more than 2 second in this context means the requirement is not fulfilled). Also shorter hesitations may be significant.

Experiences

- Test leader
  - keeping quiet? Not helping?
  - Prompting?
- Observer
  - keeping up with what happens?
- Subject
  - embarassment?
Try again:
- Test leaders switch groups
- Test leaders become subjects in the new group
- Previous observers become test leaders
- Previous subjects become observers (continue using the same observation protocol that previous observer used)

Analyse observations
- Observer: Look at the notes, think back to what happened and add any points you missed.
- What were your observations?
  - problems, hesitations?
- Look at the interview answers.
- Often differences.

Remember:
- When contacting presumptive subjects:
  - contact a few more than you need. There will be drop-outs.
  - Tell them about the course
  - Tell them the purpose of their participation
  - Tell them it is the system/prototype that is being tested, not themselves. Remind them if necessary.
  - Informed consent: You must ask if you may record and also if you may use (de-identified) quotes in your report.
- When writing task instructions:
  - Text should say what to accomplish; not how
  - One task per page (no distraction)

When user-testing a prototype
- Should properly not be done by the developers...
- NEVER defend your design in front of the user (you will ruin your chances for critique)
- If you test a paper prototype remember to practice being the "computer" so that the response to user actions is smooth and rapid.
How to measure goal fulfilment when doing user testing

- User tasks focus on important functions
- Prototypes illustrate types of data provided, not exact content - data requirements only partially possible.
- Desired qualities must be tested for.
  - General usability qualities (ease of use, ease of learning ...)
  - Specific qualities indicated by personas or stakeholders
    - inspiring
    - relaxing
    - engaging
    - ...
- Prototypes illustrate types of data provided, not exact content - data requirements only partially possible.

Desired qualities must be tested for.

- General usability qualities (ease of use, ease of learning ...)
- Specific qualities indicated by personas or stakeholders
  - inspiring
  - relaxing
  - engaging
  - ...

Ex: H&M and the Shopping mom

- Effect goals: increase online shopping by:
  - reducing drop-out rate
  - attract more customers to the online shop
  - sell more to each online shopper.
- User Goals:
  - Shop for herself and her family online (problem: product presentation, filtering not good)
  - Be inspired and get tips on garments that fit together
  - Be able to see other’s ratings of children’s clothes
  - Price sensitivity: relatively high
  - Quality importance: relatively high
  - Trendiness importance: middling
- General usability goals

Measurements:

- Is it possible to shop (for self/family)?
- Does the subject feel inspired?
- Are combination items noted by the subject?
- Are ratings by other customers found?
- Does the subject feel that bargains can be found?
- Is the rate of failed buying-tasks lower than the previous drop-out rate?
- (Is the SUS measure greater than 90?)

Test tasks: ex. shopping

- Task focus: finding clothing for self/family
- Test task: Your son is outgrowing several old sweaters, you need to order at least one new. He is 140 cm tall, has several brown pants that should go with the new sweater (thus it should be brown or beige). He likes Spiderman and Cars.
Measurements for formative eval

- Measure task completion
- Observe what is happening to find problems
  - observe hesitations, suboptimal choices, muttered questions, confusion...
  - Try to understand why the user is confused...
- Identify problems
  - was the intended item found, why not? Did it take a lot of effort? Was there information missing? Functions not found?
  - (why)

Measurements cont’d

- Questions afterwards for qualities:
  - Intention to use: How probable is it that you would buy clothes from this site? Why?
  - How inspiring was the presentation of the clothes? Why?
  - To what extent do you feel you’ve made a good bargain today? Why?

Measurements for summative evaluation

- Task completion
  - at least 90% of the test users should complete task 1 (find the spiderman sweater).
  - At least 90% of the users should complete at least 5 out of 6 buying-tasks. (compare to current drop-out)
- We will not measure time to completion, longer browsing might induce more sales.
- Experience: Interview or questionnaire
  - How inspired did you feel by the things you found? not at all - a little - rather - very
  - at least 75% of the users should feel "very inspired"

Summative cont’d

- SUS - system usability scale. Research shows that a system that gets at least 90 will be recommended to other users (thus recruiting more users).
Calculating the SUS value

- SUS gives a summary value. The values of individual questions give no information.
- Algorithm: Each question contributes 0-4 points. Question 1, 3, 5, 7, 9 are calculated as mark-1. Question 2, 4, 6, 8, 10 calculated as 5-mark
- Multiply w 2.5 for total value: 0-100

Testing your persona hypothesis

- Remember: they're just guesswork!
- When you do user testing, you have a user within reach...
- Compare to reality: ask relevant questions of your subjects before testing.