Usability Evaluation – Lab classes

https://www.interaction-design.org

Beatriz Sousa Santos, 2016/2017
• Usability is, according to ISO 9241-11:

“the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”

• How to measure it??
(Cokton, 2013):

“Put simply, **usability evaluation** assesses the extent to which an interactive system is easy and pleasant to use.

Things aren’t this simple at all though, but ...:
- Usability is an inherent measurable property of all interactive digital technologies
- ... developed evaluation methods that determine whether or not an interactive system or device is usable.
- ...usability evaluation methods also determine the extent of its usability, through the use of robust, objective and reliable metrics
- Evaluation methods and metrics are thoroughly documented ...”

http://www.interaction-design.org/encyclopedia/usability_evaluation.html
Evaluation Methods

- Analytical (without users)
  - Heuristic Evaluation
  - Cognitive Walkthrough
  - Model based methods
  - Review methods
  - ... (we are going to use)

- Empirical (involving users)
  - Observation
  - usability tests
  - Query
  - Controlled Experiments
  - ...
Heuristic Evaluation (Nielsen and Molich 1990)

• A “discount usability engineering method” for quick, cheap, and easy evaluation of a UI design

• Most popular usability inspection method; yet is subjective

• It is a systematic inspection of a design for usability

• Meant to find the usability problems in the design so that they can be attended to as part of an iterative design process

• Involves a small set of analysts judging the UI against a list of usability principles ("heuristics")
How to perform HE

• Should be performed by several evaluators (one person will never be able to find all the problems)

• Evaluators should work independently:
  – First get a general idea of the UI
  – Then perform a detailed inspection using a set of heuristics
  – List usability problems (heuristics not followed and severity degree)

• Findings of all evaluators should be integrated in the same report

• The report should help the development team to prioritize problem fixing
• Nielsen proposed **10 general usability heuristics**, yet there are other sets (e.g., for web, mobile, visualization applications, for seniors or children...)

• More details on how to conduct a heuristics evaluation at:

• And how to rate the severity of the usability problems found:

• The list of problems and severity rates should help the development team to prioritize problem fixing
List of recognized usability principles ("the heuristics")

https://www.nngroup.com/articles/ten-usability-heuristics/

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation
Number of problems found by several evaluators
Example: “the magic number five”

- Heuristic evaluation of a banking system:
  - 19 evaluators
  - 16 usability problems

black square - problem found
white square – not found

http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation

- Conclusion: in general 3 to 5 evaluators seems reasonable, but be careful!

- More evaluators find more problems but cost more
First assignment – Heuristic evaluation

- Select a user interface (UI) of any interactive system/device/application:
  - S/W (IDEs, Operating systems, other professional S/W)
  - web applications, mobile apps, etc.
  - consumer electronics (TV, home audio, vehicle electronics, appliances, phones…)
  - wearable devices (smart watches, fitness bands, etc.)
  - office (copier, printer, scanner, fax, etc.), scientific or medical equipment…)

- The selected UI should not be too simple; if too complex evaluate only part of it

- Propose your choice and ask for validation at the next Lab class
• Perform a heuristic evaluation (the four students must work independently in a first phase)

• Using the 10 heuristics by Nielsen or any other set you consider adequate (after asking for permission to use it)

• Find usability problems and assign a severity degree to each problem (use the scale proposed by Nielsen)

• If you cannot find many usability problems, point out positive aspects (complying with the heuristics).
Assignment submission and presentation
14,15,16 March

• Prepare a **15 minute presentation** (~15 slides) briefly describing:
  - the UI and intended usage (target users, main tasks...)
  - the methods used in the evaluation (heuristics sets)
  - the main results obtained
  - a table with a summary of problems found by each and all evaluators
  - your overall appreciation of the UI usability and UX

• The presentation file should have a **name**: “PX_name of system evaluated“
  (e.g. P1_SmartWatch)

• Submit the presentation **through Moodle**: March 14 (Tuesday classes)
  March 15 (Wednesday classes)
  March 16 (Thursday classes)
Cognitive Walkthrough  (Wharton, et al., 1992)

• Usability inspection method (thus not involving users)

• Based on the fact that users usually prefer to learn a system by using it (e.g., instead of studying a manual)

• Focused on assessing learnability (i.e., how easy it is for new users to accomplish tasks with the system)

• May produce results quickly at a low cost

• Applicable at early phases, before any coding
How to perform a cognitive walkthrough

1- Task analysis: sequence of steps or actions required by a user to accomplish a task, and the system responses

2- Designers and developers walkthrough as a group, asking themselves a set of questions at each step

3- Data gathering during the walkthrough: answering the questions for each subtask usability problems are detected

4- Report of potential issues

5- UI redesign to address the issues identified
CW Four questions:

• Will the user try to achieve the effect that the subtask has? (Does the user understand this subtask is needed to reach the goal?)

• Will the user notice that the correct action is available? (E.g. is the button visible?)

• Will the user understand that the wanted subtask can be achieved by the action? (E.g. the button is visible but the user doesn’t understand the text and will not click on it)

• Does the user get feedback? Will the user know that they have done the right thing?
Common issues

- The evaluator doesn't know how to perform the task; the method involves the optimal sequence of actions.

- Involves an extensive analysis and documentation and often too many potential issues are detected, resulting very time consuming.

Thus:

Lighter variants of Cognitive Walkthrough were proposed to make it more applicable in S/W development companies.
Streamlined Cognitive Walkthrough (Spencer, 2000)

- Only two questions:
  - Will the user know what to do at this step?
  - If the user does the right thing, will they know that they did the right thing, and are making progress towards their goal?

- And a set of rules to streamlining the walkthrough and trade-off granularity for coverage

comprises the 3 first questions of CW
According to Spencer the method can be applied successfully if the usability specialist:

- takes care to prepare the team for the walkthrough,
- avoids design discussions during the walkthrough,
- explicitly neutralizes defensiveness among team members,
- streamlines the procedure by collapsing the first three questions into one question,
- and captures data selectively
**Example:** Evaluation of a desktop photocopier UI

- **Machine UI:**
  - numeric keypad,
  - "Copy" button,
  - push button on the back to turn on the power

  The machine automatically turns itself off after 5 min inactivity

- **Task:** copy a single page

- **User:** any office worker

- **Actions needed:** turn on the power,
  put the original on the machine,
  press the "Copy" button

[http://hcibib.org/tcuid/chap-4.html#4-1](http://hcibib.org/tcuid/chap-4.html#4-1)
• Story for action number one:
   “the user wants to make a copy and knows that the machine has to be turned on. So she pushes the power button. Then she goes on to the next action”

Not convincing!

• why shouldn't the user assume that the machine is already on?
  That is often the case

• Will the user figure out that the machine is off, and find the power switch?
  etc. etc.
Practice the Streamlined Cognitive Walkthrough:

- Look for a phone number at the University of Aveiro Web site
  user: any student from the University

- Create a pdf of a PowerPoint file using the Print option but not printing the hidden slides
  user: anyone familiar with a previous version
Look for a person’s phone number at the University of Aveiro Web site
user: any student from the University

Task analysis:
- look for the icon (directório);
- input part of the person’s name and search
- get the phone number
But the defined user profile (any student from the University) includes foreign students, thus a previous action is needed:
- select the English version
For each action we need to ask the two questions and put ourselves in the shoes of the user!
First action: find the icon

Q1 - Will the user know what to do at this step?

Even reading the tooltip (directório) possibly the correct icon is not recognizable!

Q2 - If the user does the right thing (selects the icon), will they know that they did the right thing, and are making progress towards their goal?

Probably yes; this looks a familiar search bar and it is adequately labeled (lista telefónica; pesquisar)
Second action: input part of the person’s name

Q1 - Will the user know what to do at this step?

Probably yes; the tooltip lets the user know he/she should input the person’s name and select “pesquisar”

Q2 - If the user does the right thing (selects the icon), will they know that they did the right thing, and are making progress towards their goal?

Probably yes; however, some users might not recognize 24117 as a phone number (it only has 5 digits, as it is internal, and not 9 as possibly expected)
Another example: Smart TV

How to access the Internet?

(before reading the manual?)

(we see the symbol at the screen only after pressing it on the control!)
Practice the Streamlined Cognitive Walkthrough:

Analysing interactive systems/applications that should be very intuitive:

• Turn on and off the video projector in your Lab using the remote control or directly on the projector
  user: any student from the University

• Create a pdf of a PowerPoint file but printing the hidden slides
  user: anyone familiar with previous or current versions of Office

• Change the Channel using the box of your TV service (not the remote control)
  user: anyone having a TV box
Limitations of Analytical Methods

– Are subjective
– Involve several usability experts
– Cannot find all usability problems

Thus, empirical methods (involving users) are needed

observation
query
controlled experiments (scientific approach)

Usability test (engineering approach)
Ethics in applying empirical methods

Involving users implies specific cautions:

- Asking for explicit consent
- Confidentiality
- Security (avoid any risk)
- Freedom (users may give up at any time)
- Limit stress

It’s the system that is under evaluation not the user!
Observation

Has many variants from very simple to very complex and expensive:

• **Direct**: observer takes notes

• **Undirect**: through audio/ video – more complex and time consuming

• **Think Aloud**: users are asked to explain what they are doing

• **Logging**: users activity is logged by the system

• **Combinations of the previous, etc**
Query

• Two main variants:
  – Questionnaire  (reach more people; less flexible)
  – Interview

• Should be carefully prepared and tested

• Collected data should be carefully analyzed
Controlled experiments

• The work horse of science ...

• Important issues to consider:
  
  – Hypothesis
  – Variables (input or independent; output or dependent)
  – Secondary variables
  – Experimental design (within groups; between groups)
  – Participants (number, profile)
  – Statistics
Usability tests

• Involve observation and query

• Main aspects:
  – Participants
  – Tasks
  – Test facilities and systems
  – Experimental design
  – Usability measures
  – Data analysis

• Have a complex logistics

• Standard: Common Industry Format (CIF) for usability test reports
Participants

• The total number of participants to be tested
  (a valid statistical analysis implies a sufficient number of subjects)

• Segmentation of user groups tested, if more than one

• Key characteristics and capabilities of user group
  (user profile: age, gender, computing experience, product experience, etc.)

• How to select participants

• Differences between the participant sample and the user population
  (e.g. actual users might have training whereas test subjects were untrained)
Tasks

• The task scenarios for testing

• Why these tasks were selected
  (e.g. the most frequent tasks, the most troublesome tasks)

• The source of these tasks
  (e.g. observation of users using similar products, product specifications)

• Any task data given to the participants

• Completion or performance criteria established for each task
  (e.g. n. of clicks < N, time limit)
Experimental design

• Procedure: the logical design of the test

• Participant general instructions and task instructions

• The independent variables and control variables

• The usability measures to be used:
  – a) for effectiveness (completeness rate, errors, assists)
  – b) for efficiency (times)
  – c) for satisfaction
Test Facilities and equipment

• The setting and type of space in which the evaluation will be done
  (e.g. usability lab, cubicle office, meeting room, home office, home family room, manufacturing floor, etc.)

• Any relevant features or circumstances that can affect the results
  (e.g. video and audio recording equipment, one-way mirrors, or automatic data collection equipment)

• Participant’s Computing Environment
  (e.g. computer configuration, including model, OS version, required libraries or settings, browser name and version; relevant plug-in, etc.)

• Display and input devices characteristics

• Any questionnaires to be used
Common Industry Format (CIF) for usability test reports
ISO/IEC 25062:2006

• Specifies the format for reporting the results of a **summative** evaluation

• The most common type of usability evaluation is **formative**, (i.e. designed to identify problems that can be fixed)

• A summative evaluation produces usability metrics that describe how usable a product is when used in a particular context of use

• The CIF report format and metrics are consistent with the ISO 9241-11
  
  [http://www.usabilitynet.org/prue/cif.htm](http://www.usabilitynet.org/prue/cif.htm)
  [https://www.nist.gov/itl/iad/industry- usability-reporting](https://www.nist.gov/itl/iad/industry- usability-reporting)

• Top ten things to know about the CIF (a list of do and don’t)
  
Well-known usability questionnaires

- System Usability Scale (SUS)

- Questionnaire for User Interface Satisfaction (QUIS)

• SUS provides a “quick and dirty”, reliable tool for measuring the usability
• It includes 10 questions with five response options

• QUIS is a measurement tool designed to assess a computer user's subjective satisfaction with the UI
• It is designed to be configured according to the needs of each UI analysis by including only the sections that are of interest to the user
• It includes questions with ten response options

• Both questionnaires should be completed following use of the UI in question
System Usability Scale (SUS)

- Provides a “quick and dirty”, reliable tool for measuring the usability
- It includes 10 questions with five response options
- It allows to evaluate a wide variety of products and services (H/W, S/W, mobile devices, websites and applications)
- Has become an industry standard, with references in over 1300 publications

Benefits of using a SUS
- Is a very easy scale to administer to participants
- Can be used on small sample sizes with reliable results
- Is valid – it can differentiate between usable and unusable systems

SUS Questions

- I think that I would like to use this system frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

Scoring SUS

Let $R(n)$ be the answer to Question $n$:

$$
SUS = \left( \sum_{n=1}^{5} R(n) - 1 + 5 - R(n*2) \right) * 2.5
$$

0... 100; SUS > 68 would be considered above average
QUIS - Questionnaire for User Interface Satisfaction

• The QUIS contains:
  
  – a demographic questionnaire,
  – a measure of overall system satisfaction,
  – a measure of specific UI factors (e.g. screen visibility, terminology and system information, learning factors, and system capabilities)

• QUIS has pen and paper and PC software versions for administration
• Uses a 10-point scale to rate 21 items relating to the system's usability
• These ratings produce data for the overall reaction to a system's usability on 6 factors.
• It is easy to use and analyse.

https://ext.eurocontrol.int/ehp/?q=node/1611
Example questions of QUIS

OVERALL REACTIONS TO THE SOFTWARE

terrible 0 1 2 3 4 5 6 7 8 9 wonderful

difficult 0 1 2 3 4 5 6 7 8 9 easy

frustrating 0 1 2 3 4 5 6 7 8 9 satisfying

inadequate power 0 1 2 3 4 5 6 7 8 9 adequate power

dull 0 1 2 3 4 5 6 7 8 9 stimulating

rigid 0 1 2 3 4 5 6 7 8 9 flexible

SCREEN

Characters on the computer screen

hard to read 0 1 2 3 4 5 6 7 8 9 easy to read

Highlighting on the screen simplifies task

not at all 0 1 2 3 4 5 6 7 8 9 very much

Organization of information on screen

confusing 0 1 2 3 4 5 6 7 8 9 very clear

USABILITY AND USER INTERFACE

Use of colors and sounds

poor 0 1 2 3 4 5 6 7 8 9 good

System feedback

poor 0 1 2 3 4 5 6 7 8 9 good

System response to errors

awkward 0 1 2 3 4 5 6 7 8 9 gracious

System messages and reports

poor 0 1 2 3 4 5 6 7 8 9 good

System clutter and UI “noise”

poor 0 1 2 3 4 5 6 7 8 9 good
Example questions of QUIS (cont.)

**TERMINOLOGY AND SYSTEM INFORMATION**

Use of terms throughout system
inconsistent 0 1 2 3 4 5 6 7 8 9 consistent
Computer terminology is related to the task you are doing
never 0 1 2 3 4 5 6 7 8 9 always
Position of messages on screen
inconsistent 0 1 2 3 4 5 6 7 8 9 consistent
Messages on screen which prompt user for input
confusing 0 1 2 3 4 5 6 7 8 9 clear
Computer keeps you informed about what it is doing
never 0 1 2 3 4 5 6 7 8 9 always
Error messages
unhelpful 0 1 2 3 4 5 6 7 8 9 helpful

**LEARNING**

Learning to operate the system
difficult 0 1 2 3 4 5 6 7 8 9 easy
Exploring new features by trial and error
difficult 0 1 2 3 4 5 6 7 8 9 easy
Remembering names and use of commands
difficult 0 1 2 3 4 5 6 7 8 9 easy
Tasks can be performed in a straight-forward manner
never 0 1 2 3 4 5 6 7 8 9 always
Help messages on the screen
unhelpful 0 1 2 3 4 5 6 7 8 9 helpful
Supplemental reference materials
confusing 0 1 2 3 4 5 6 7 8 9 clear
Practice preparing a summative usability test

• Consider the application/site/device you have evaluated using analytical methods

• Define the potential users and context of use

• Define a list of tasks (e.g. task that seemed more difficult or more important)

• Select a set of usability measures (times, errors, complete rates...)

• Prepare all the materials:
  - declaration of consent
  - list of tasks and perceived difficulty
  - final questionnaire
  - list of tasks for the observer to take notes

• Run a pilot test
Materials to use during the test (examples in Moodle)

1- Informed consent form

2- List of tasks to the user

2- Post-task questionnaire

2- Observer’s script
Bibliography – Books and links

• Alan Dix, Janet Finlay, Gregory Abowd, Russell Beale, *Human-Computer Interaction*, 3rd edition, Prentice Hall, 2004

• Jackob Nielsen, *Usability Engineering*, Morgan Kaufmann, 1993


• Standard ISO 9241-11 - *Ergonomic requirements for office work with visual display terminals Part 11 : Guidance on usability*


• What is CIF? - [http://www.usabilitynet.org/prue/cif.htm](http://www.usabilitynet.org/prue/cif.htm)

• UXPA Code of Professional Conduct [http://uxpa.org/resources/uxpa-code-professional-conduct](http://uxpa.org/resources/uxpa-code-professional-conduct)

Bibliography - Papers


  http://www.interaction-design.org/encyclopedia/usability_evaluation.html