Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Usability
Evaluation Methods

Beatriz Sousa Santos

- 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 a measurable property of all interactive digital technologies

- Evaluation methods determine if an interactive system or device is usable

- And the extent of its usability, through 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
- **Empirical** (involving users)
  - Observation
  - Query
  - Controlled Experiments

Observation, Query > usability tests ✓

Controlled Experiments ↙

(✓ - have used in Lab classes)

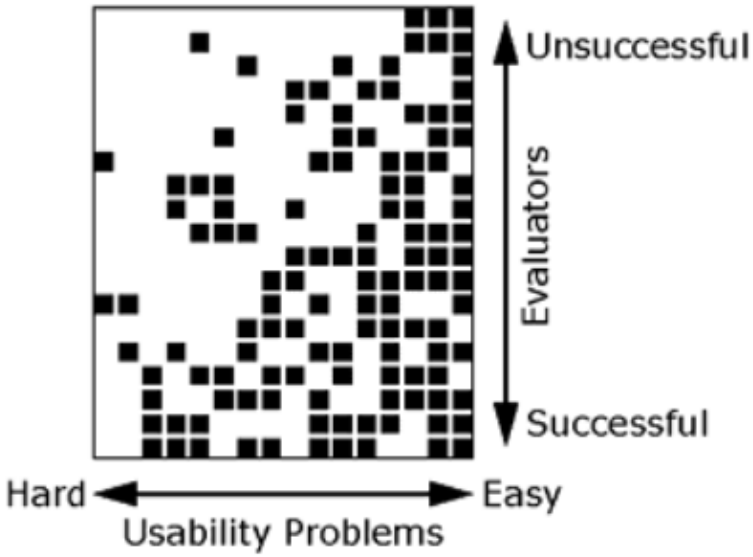
Heuristic Evaluation (Nielsen and Molich 1990)

- A “discount usability engineering method” for quick, cheap, and easy evaluation of a UI design
- The most popular of the usability inspection methods
- 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").

- Is difficult for a single individual to do; one person will never be able to find all the problems
- Involving multiple evaluators improves the effectiveness of the method significantly
- Nielsen generally recommends to use three to five evaluators
- not much gain by using larger numbers

Example:

- 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/#sthash.OmTrV7Og.6ZrkgzXB.dpuf

This suggests that in general 3 to 5 evaluators may be reasonable

How to select the number of evaluators for a specific case?

- Consider the following criteria:
  - Complexity of the user interface
  - Experience of the evaluators
  - Expected costs /benefits
  - How critical is the system (cost of user errors)
  - …

How to perform Heuristic Evaluation

- Each evaluator:
- First make a general analysis to get to know the UI
- Then, make a systematic analysis having in mind the heuristics
- Take note of each potential problem, the heuristic and the severity grade
- Finally, compile all the potential problems

http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation
sthash.OmTrV7Og.dpuf

Ten Nielsen's 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

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

- **Main advantages of heuristic evaluation**:
  - May produce useful results with modest investment
  - Simple to apply even by not very experienced evaluators
  - May be used along the development process from early phases

- **Main limitations**:
  - Subjective (partially overcome with more and more experienced evaluators)
  - Tends to find many small problems which may not be very important
  - Can't find all usability problems

-> **evaluation involving users is needed!**

Cognitive Walkthrough (Wharton, et al., 1992)

- Usability inspection method not involving users (analytical)
- 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)
- 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 may not know the optimal way 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?

  comprises the 3 first questions of CW

  - 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

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

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 hiden slides

  user: anyone familiar with a previous version

Look for a person's phone number and email address at the University of Aveiro Web site

User: any student from the University

Task analysis:

- find the icon  (search);
- input part of the person's name and search in "Pessoas"
- 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!

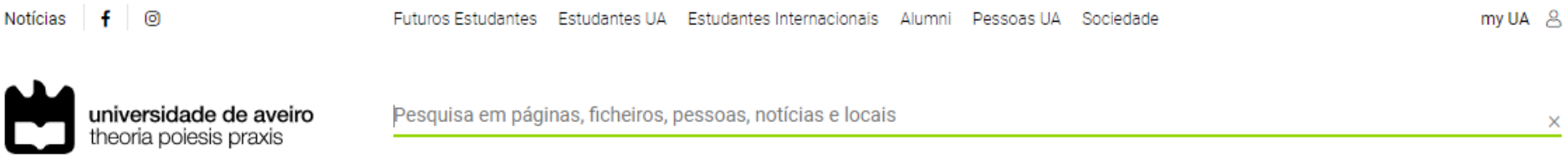
Previous action for foreign students: Select the English version seems easy (it is a "standard" way to do it in sites)

First action in the Portuguese version: find the icon 

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

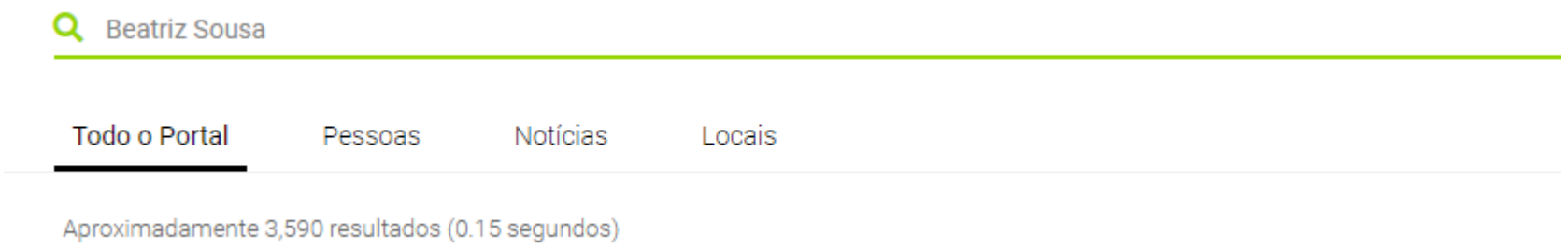
Even without tooltip the correct icon seems recognizable (it is "standard 

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?



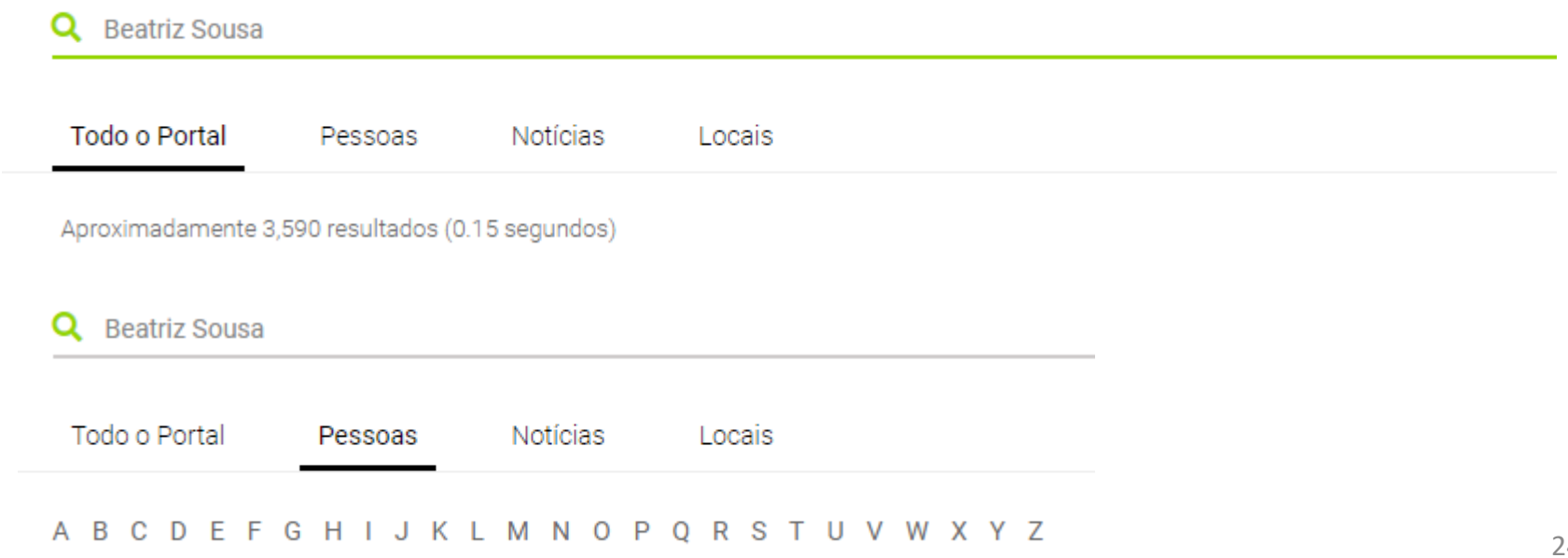
Probably yes; while it may not look a search bar, it is adequately labeled (Pesquisa em páginas, ...)

Second action: input part of the person's name and search in "Pessoas"

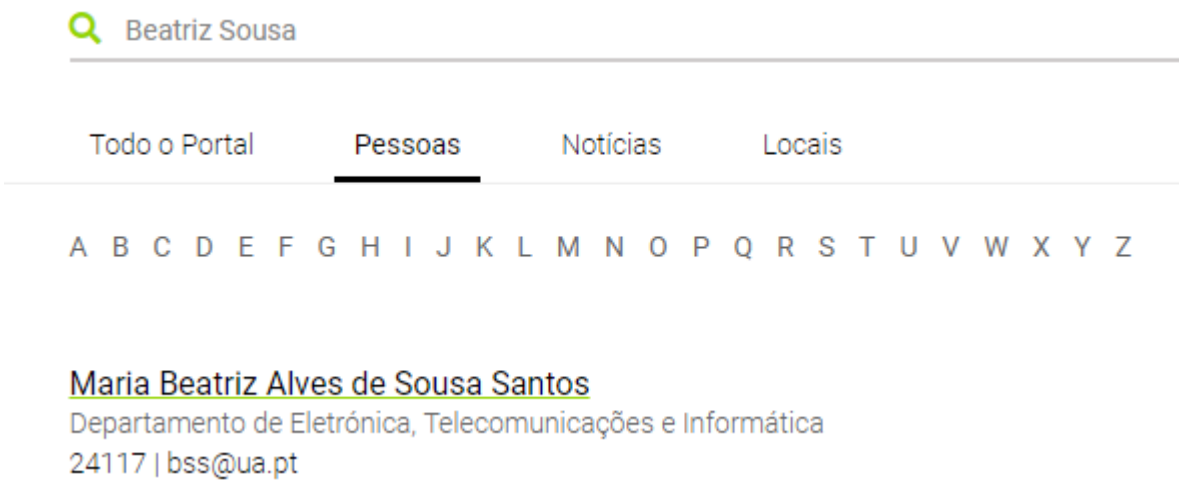


Q1 - Will the user know what to do at this step?

Probably yes; it is easy to recognize that s/he should input the person's name and select "Pessoas"



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



Probably yes; however, some users may not recognize 24117 as a phone number (it only has 5 digits, as it is internal, and not 9 as possibly expected)

In conclusion:

- it seems easy for the target users to reach the phone number and email address;
- however, the phone number may be not recognized as such

Previous version of the site:

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 adequatly 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?

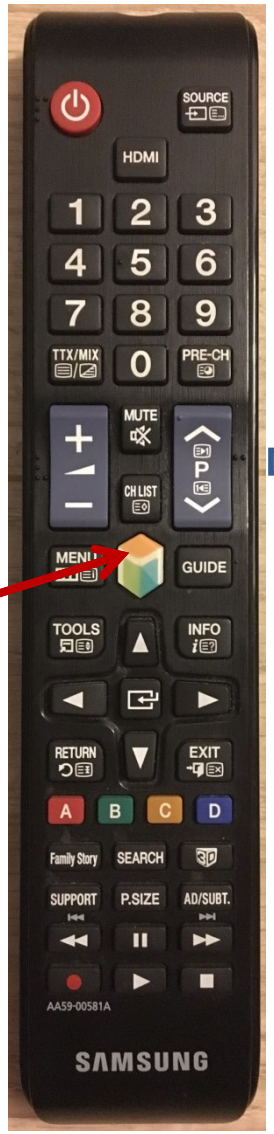
Probably yes; however, some users might not recogize 24117 as a phone number (it only has 5 digits, as it is internal, and not 9 as possibly expected)

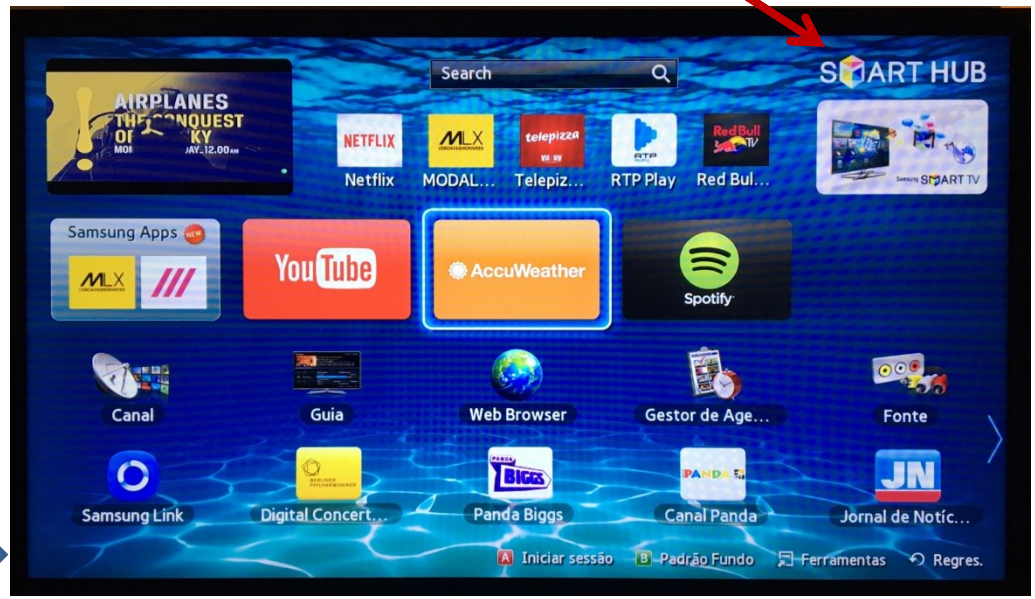
Another example: Smart TV

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

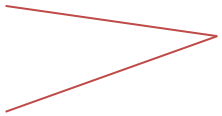
- Change the Channel using the box of your TV service (not the remote control)
  user: anyone having a TV box

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)

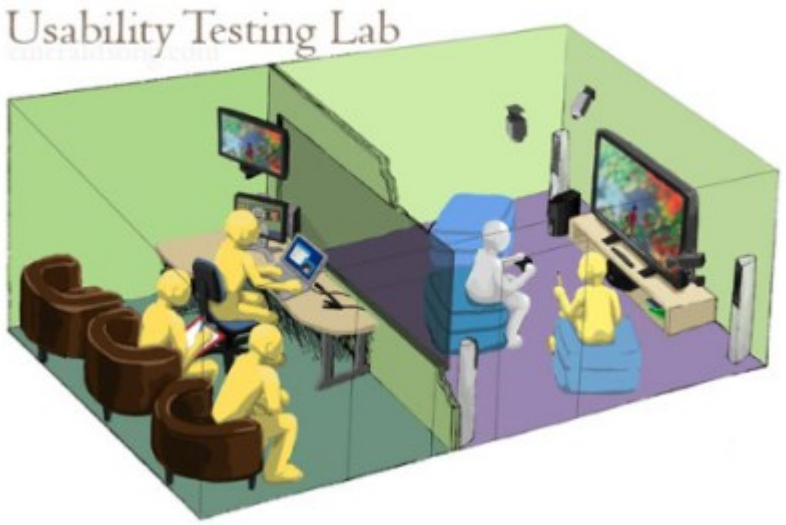
Ethics in applying empirical methods

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/ vídeo – 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

Usability tests

- Involve observation and query

- Main aspects:
  - Participants
  - Tasks
  - Test facilities and systems
  - Protocol
  - 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)

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

Protocol

- Procedure: the logical design of the test
- Participant general instructions and task instructions
- The usability measures to be used:
  a) for effectiveness (completeness rate, errors, assists)
  b) for efficiency (times)
  c) for satisfaction

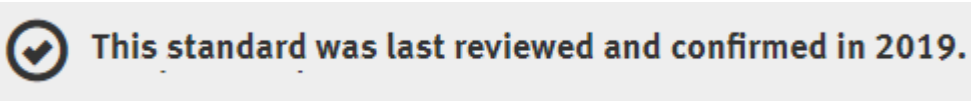
Common Industry Format (CIF) for usability test reports ISO/IEC 25062:2006

- 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

https://www.iso.org/standard/43046.html

https://www.userfocus.co.uk/articles/cif.html

Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- **Common Industry Format (CIF) for usability test reports**



The format includes the following elements:

- the description of the product,
- the goals of the test,
- the test participants,
- the tasks the users were asked to perform,
- the experimental design of the test,
- the method or process by which the test was conducted,
- the usability measures and data collection methods, and
- the numerical results.

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

Controlled experiment

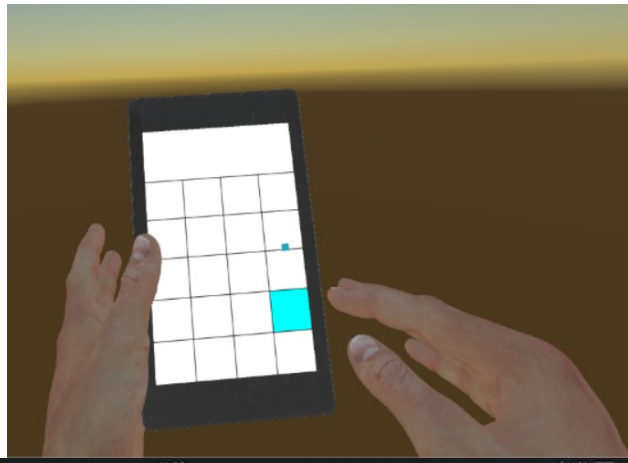
- Define an hypothesis
- Define input (independent), output (dependent) and secondary variables
- Define experimental design (within-groups / between groups)
- Select the participants
- Prepare all the documentation:
  - list of tasks and perceived difficulty
  - final questionnaire

  → To the user
  - list of tasks for the observer to take notes

  → To the observer
- Run a pilot test
- Take care of the logistics … and after the experiment analyze data

Examples of Controlled Experiments performed @ HCI - DETI

- Study of the Effect of Hand-Avatar in a Selection Task using a Tablet as Input Device in an Immersive Virtual Environment



- Comparing two alternative versions of Meo Go

“Effect of Hand-Avatar in a Selection Task using a Tablet as Input Device in an Immersive Virtual Environment”

L. Afonso, P. Dias, C. Ferreira, B. Sousa Santos
IEEE 3D UI, Los Angeles, March 2017

- Research question: How does the virtual representation of the user's hands influence the performance on a button selection task performed in a tablet-based interaction within an immersive virtual environment?

- Method: Controlled experiment

- 55 participants used three conditions:
- no-hand avatar,
- realistic avatar,
- translucent avatar.

- Participants were faster but made slightly more errors with no-avatar

- Considered easier to perform the task with the translucent avatar

Experimental Design

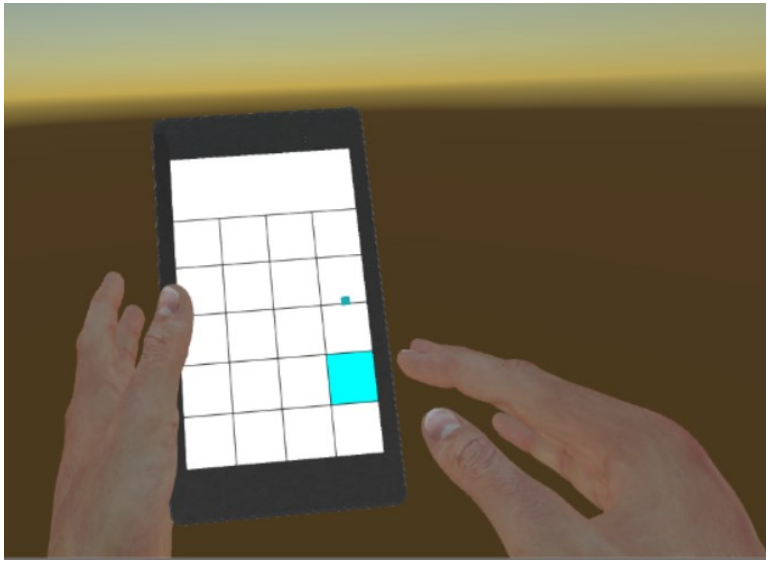
Null Hypothesis: usability is independent of the hands representation

Independent (input) variable (with 3 levels): representation of the hands

Dependent (output) variable: usability (performance + satisfaction)

Within-groups: all participants used all experimental conditions (in different sequences to avoid bias)

Task: selecting as fast as possible a highlighted button from a group of twenty buttons (repeated measures)

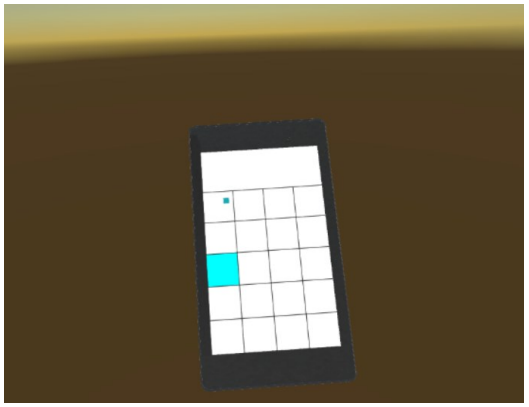


Experimental Conditions

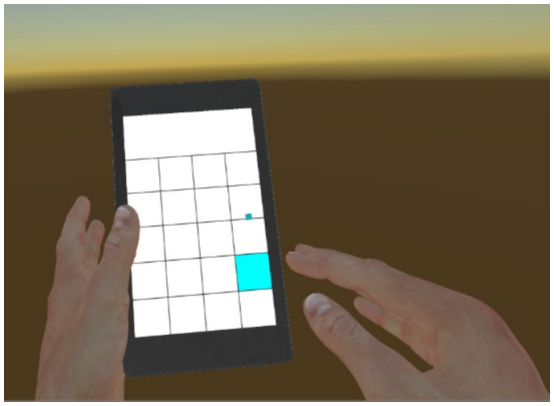
1- No avatar: the user only sees the virtual tablet;

2- Realistic avatar: a realistic representation of the hands movement is shown

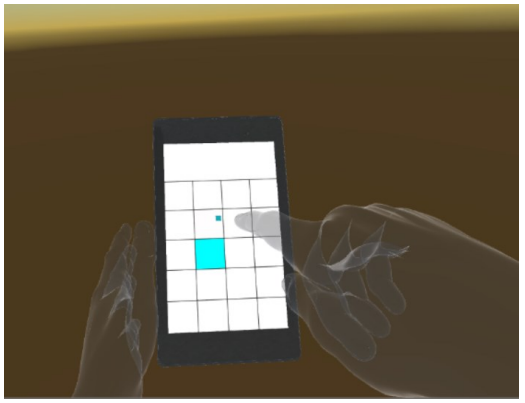
3- Translucent avatar: a translucent hand model is used (to alleviate occlusion)



No-avatar



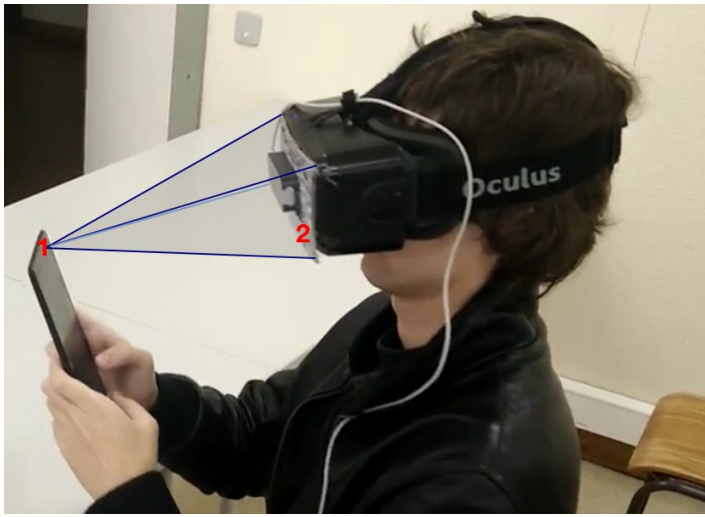
Realistic avatar

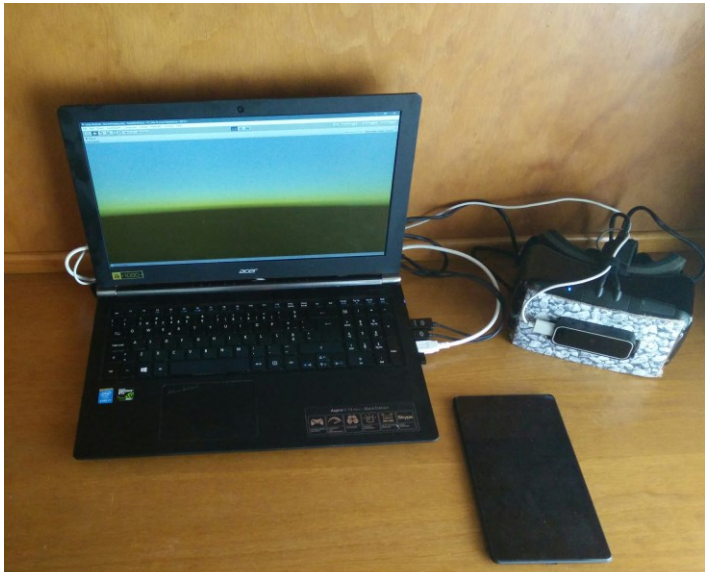


Translucent-avatar

Experimental Set-up

- Laptop running the main application (in Unity)
- HMD (Oculus Rift DK2) providing head tracking
- Tablet (Google Nexus 7) as input device running the controller application (in Unity)
- Leap Motion (mounted on the HMD) to track the user's hands
- Tablet camera tracking the position and orientation of an AR marker on the HMD to map tablet position in the virtual world (using Vuforia)



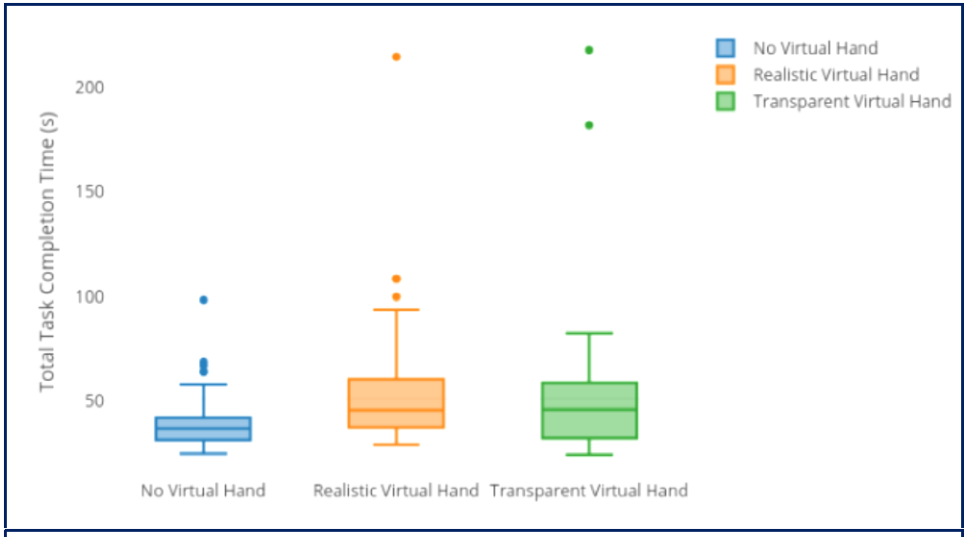


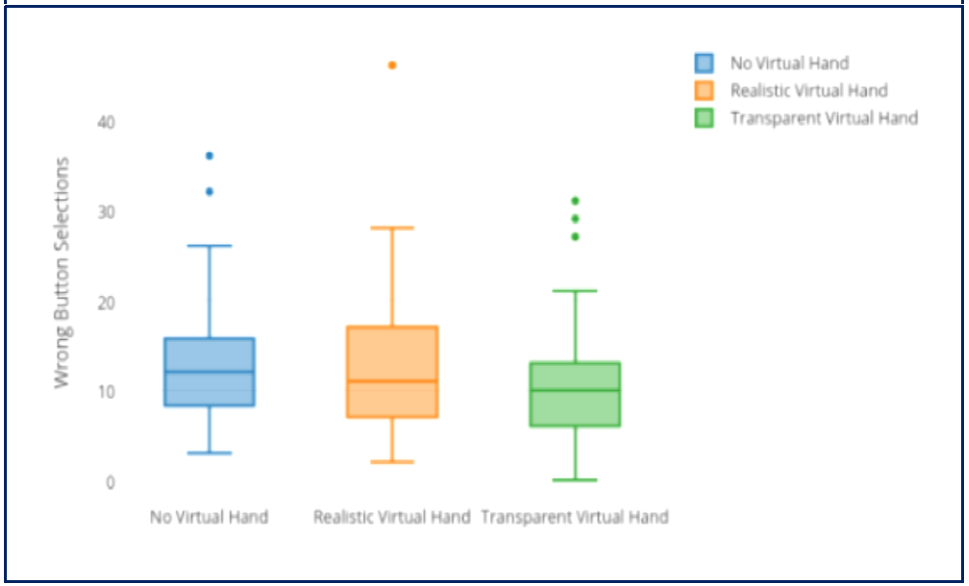
Main Results

Selection time:
Participants completed the button selections in average faster with no-avatar (**statistically significant**)

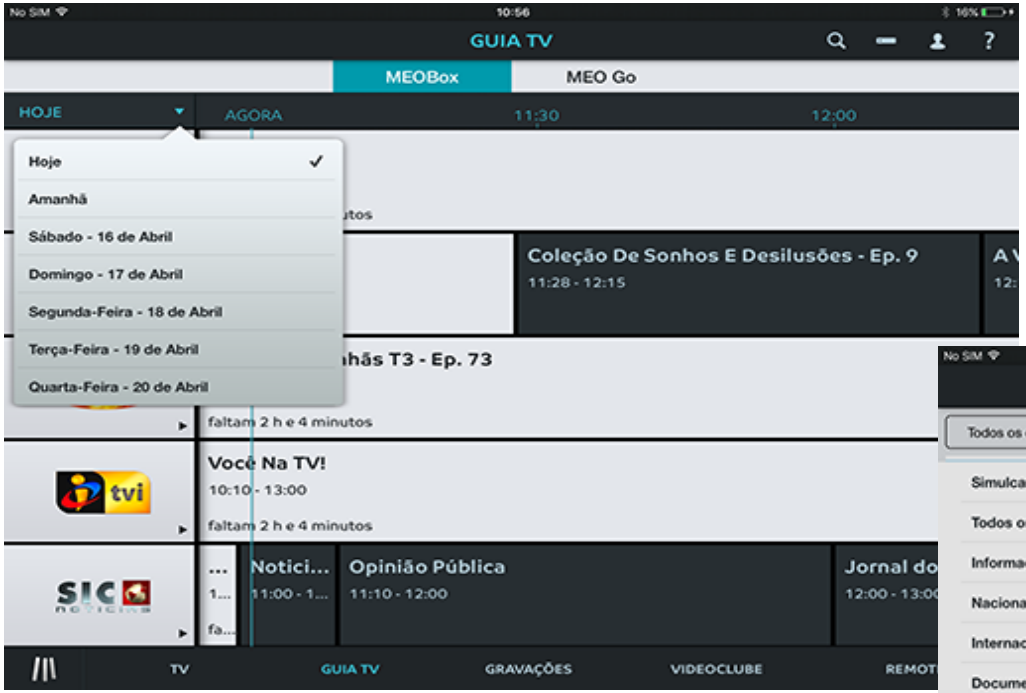
Selection errors:
Participants made slightly less errors with avatar - realistic or translucent- (**statistically significant**)

Participants' opinion:
The translucent avatar:
- was more often preferred
- was considered as better than the realistic avatar (statistically significant)





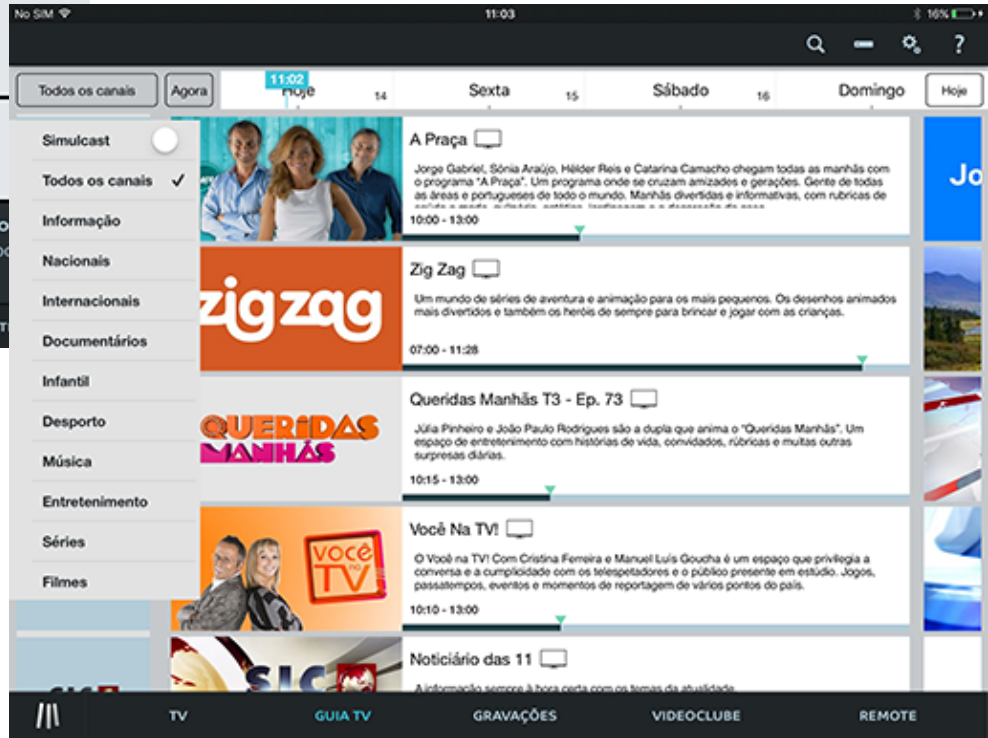
Comparing two versions of Meo Go



Current version

Altisse Labs User - UX Group
DETI – HCI couse
April/2016

New version



Controlled experiment: comparing two versions of Meo Go: Current version *vs* Version to be deployed

- Null Hypothesis: both are equally usable
- Input (independent) variable: version
- Output (dependent) variables: performance, opinion and satisfaction
- Secondary variable: participant profile
- Participants: 66 volunteer HCI -2016 students (12 female)
- Experimental design: between groups (one version per participant)
- Exploratory Data Analysis and non-parametric tests (ordinal variables)

Epilogue of this course

- We will have the Human in the loop for long in most situations …
- And even when/if they are no longer in the loop

Technology shall serve the Human

(and not the other way around…)

And don’t forget

“The interface between humans and computers is harder than ever to define, we can interact with computers just by walking through a public space.”

Sellen, A., Rogers, Y., Harper, R., & Rodden, T., “Human Values in the Digital Age”, *Communications of the ACM*, *52*(3), March 2009, pp. 58–66

And thus we need more and more to know the right methods

Bibliography for Usability evaluation – 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

- Jenny Preece, Yvonne Rogers, Helen Sharp, D. Benyon, S. Holland, T. Carey, *Human-Computer Interaction*, Addison Wesley, 1994

- **Peter Mitchell, *A Step-by-step Guide to Usability Testing*, iUniverse, 2007**

- Gilbert Cockton, Usability Evaluation. In: Soegaard, Mads and Dam, Rikke Friis (eds.), The Encyclopedia of Human-Computer Interaction, 2nd Ed, 2013, Aarhus, Denmark: The Interaction Design Foundation. (June, 10, 2018)

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

- Norman/ Nielsen Group site - http://www.nngroup.com/articles/

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

- Standard ISO/IEC 25062:2006 – *CIF - Common Industry format for Usability Tests Report*

- What is CIF? - http://www.usabilitynet.org/prue/cif.htm

Bibliography for usability evaluation - Papers

- Spencer, R. The Streamlined Cognitive Walkthrough Method, Working Around Social Constraints Encountered in a Software Development Company CHI 2000 vol.2 issue 1 pp353–359

- Wharton, C. Bradford, J. Jeffries, J. Franzke, M. Applying Cognitive Walkthroughs to more Complex User Interfaces: Experiences, Issues and Recommendations CHI ’92, 1992, pp381–388

http://www.usabilitynet.org/prue/cif.htm

Preparing the Exam

- Study the Slides
- Study the mandatory readings
- Answer the Exam preparation questions available at Moodle

Mandatory Readings for the Exam

- Slides available at Moodle and at the course web page:

  http://sweet.ua.pt/bss/disciplinas/IHC-ECT/IHC-ECT-home.htm

- Alan Dix t al., *Human-Computer Interaction*, 3rd ed., Prentice Hall, 2004
  (at the Library)
  - Chapters: 1 to 4 (for the topics addressed in the slides)
  - Chapter 9 (for the topics addressed in the slides)
  - Chapters 12, 14 and 16 (for the topics addressed in the slides)

- Randolph Bias, Deborah Meyhew, *Cost Justifying usability*, Morgan Kaufmann, 2005
  (Chapter 3, pag. 42-55 available at Moodle)

- Ian Sommerville, *Software Engineering*, 9. Ed., Addison Wesley, 2009
  (Chapter 29, available at Moodle)