Human-Computer Interaction

2024/2025

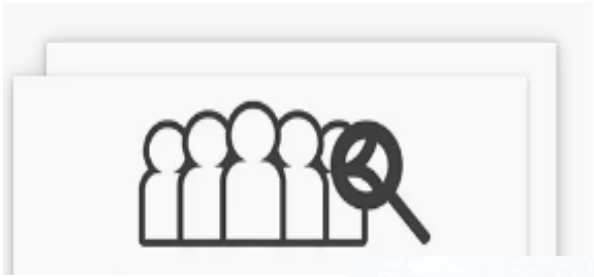
Lecture 10

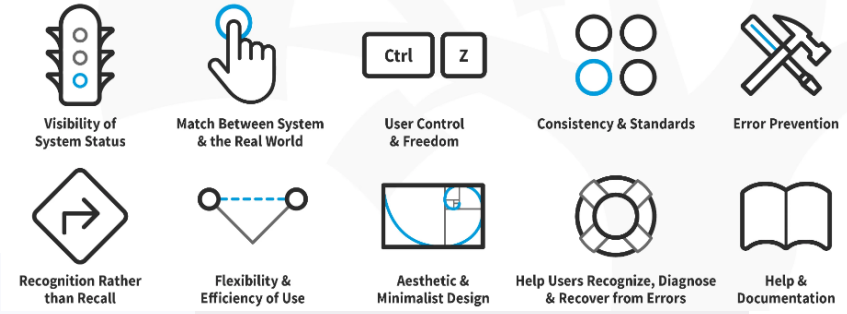
Usability Evaluation

universidade de aveiro
deti departamento de electrónica, telecomunicações e informática

Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Usability
Evaluation Methods - overview





What is Usability Evaluation? | IxDF (interaction-design.org)

Beatriz Sousa Santos

How to measure usability ??

- ISO 9241-11's three factors of usability have become five in ISO 25010's quality in use factors:
- Effectiveness
- Efficiency
- Satisfaction
- Freedom from risk
- Context coverage

ISO/IEC 25019:2023 - Systems and software engineering — Systems and software Quality Requirements and Evaluation
Usability Evaluation | The Encyclopedia of Human-Computer Interaction, 2nd Ed.

“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, and reliable metrics

- Evaluation methods and metrics are thoroughly documented …

Usability Evaluation | The Encyclopedia of Human-Computer Interaction, 2nd Ed.
When to Use Which User-Experience Research Methods - NN/g

Evaluation Methods
(widely used classification)

- **Analytical** (without users)
  - Heuristic Evaluation ✓
  - Cognitive Walkthrough ✓
  - Model based methods
  - Review methods
- **Empirical** (involving users)
  - Observation → usability tests ✓
  - Query
  - Controlled Experiments ↙

(Dix et al.)

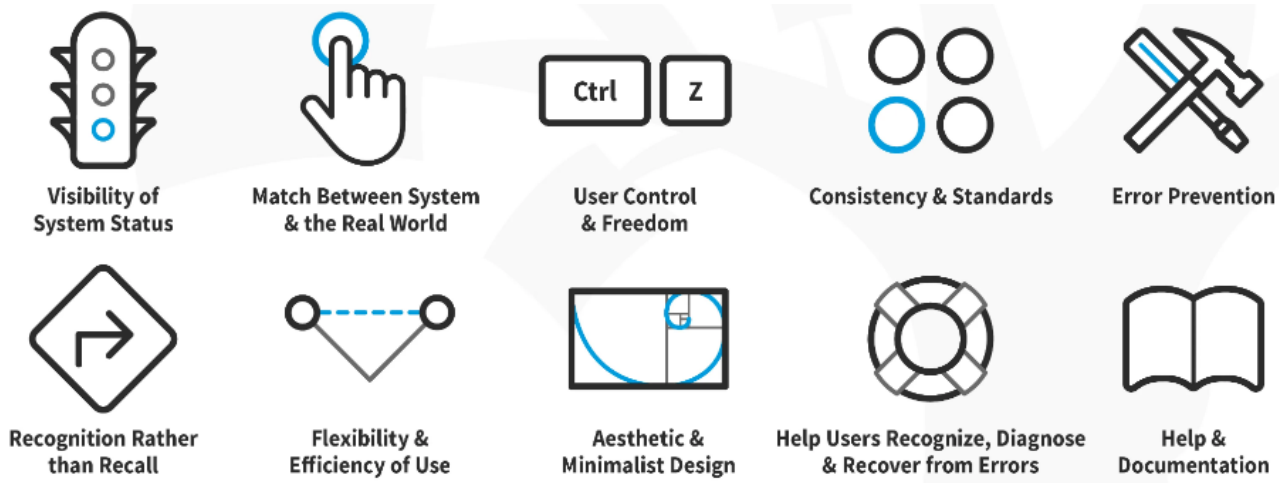
✓ - have used in Lab classes

↘ - have seen in papers

Analytical methods

Do not involve users, but analysts/evaluators

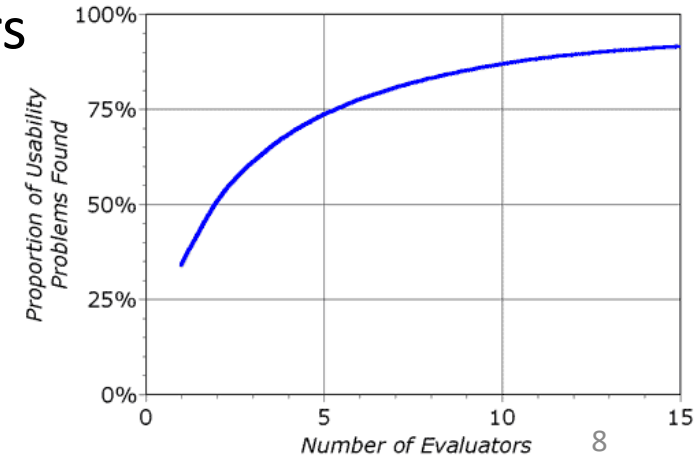
- Heuristic Evaluation
- Cognitive Walkthrough
- Model based methods
- Review methods



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 to gain by using larger numbers



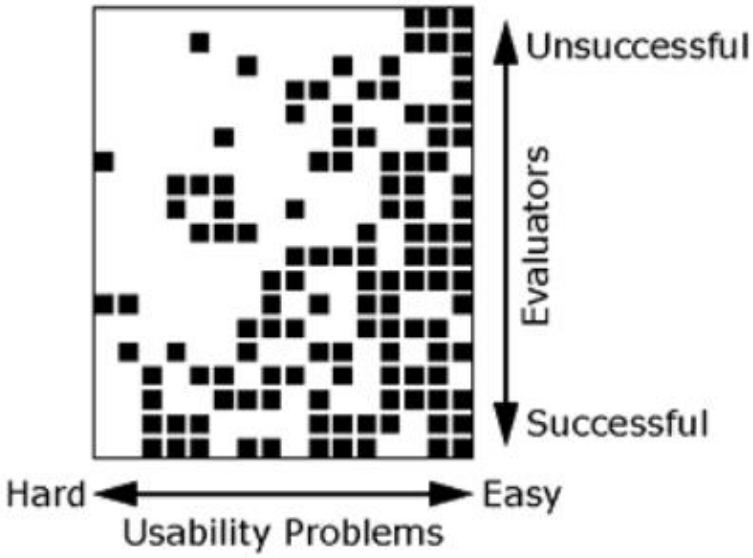
Heuristic Evaluations: How to Conduct - NN/g

Example:

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

black square - problem found

white square – not found



[Heuristic Evaluations: How to Conduct - NN/g]

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
  - **Criticality** of the system (cost of user errors)
  - …

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!

Heuristic Evaluations: How to Conduct - NN/g

- Nielsen proposed **10 general usability heuristics**,

- yet **there are other sets, e.g. for**

  - different types of applications
    (web, mobile, visualization … applications)
  - different types of users
    (for seniors, children…)

Usability Evaluation | The Encyclopedia of Human-Computer Interaction, 2nd Ed.

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 and discuss with other evaluators

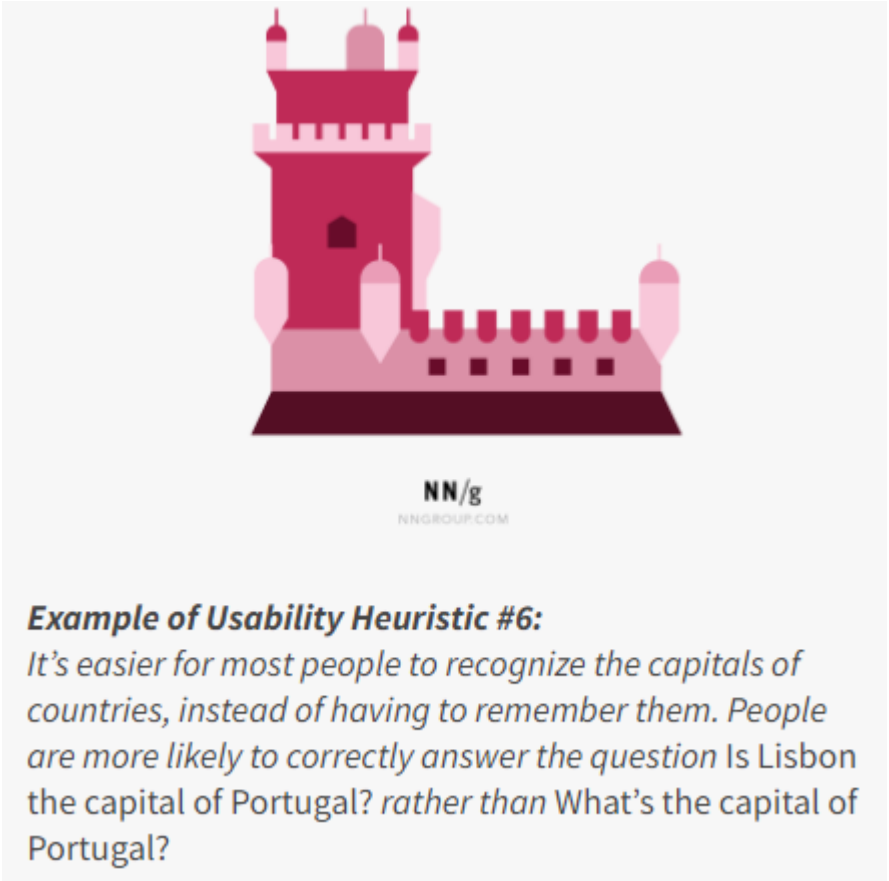
Heuristic Evaluations: How to Conduct - NN/g

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

[10 Usability Heuristics for User Interface Design - NN/g]

Example:
Heuristic #6 - Recognition rather than recall



Tips

- Let people recognize information in the interface, rather than having to remember ("recall") it.
- Offer help in context, instead of giving users a long tutorial to memorize.
- Reduce the information that users have to remember.

Learn more:

Memory Recognition and Recall in User Interfaces - NN/g

Severity rating of usability problems

Is a combination of **three factors**:

- The **frequency** with which the problem occurs
- The **impact** of the problem if it occurs
- The **persistence** of the problem

The following 0 to 4 **rating scale** can be used to rate the severity of usability problems:

0 = I don't agree that this is a usability problem at all (to be used in the discussion)

1 = **Cosmetic problem**

2 = **Minor usability problem**

3 = **Major usability problem**

4 = **Usability catastrophe**

Severity Ratings for Usability Problems: Article by Jakob Nielsen - NN/g

- **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**

Evaluate Interface Learnability with Cognitive Walkthroughs - NN/g

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:

comprises the 3 first questions of CW

- 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

According to Spencer the method can be applied successfully if the usability specialist:

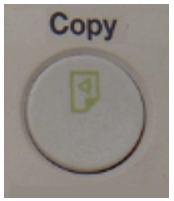
- takes care to prepare the team for the walkthrough,
- avoids design discussions during the walkthrough,
- explicitly neutralizes defensiveness among team members,
- streamlines the procedure by collapsing the first three questions into one question,
- and captures data selectively

How to Conduct a Cognitive Walkthrough | IxDF (interaction-design.org)

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

[Task-Centered User Interface Design : 4. Evaluating the Design Without Users]

- 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.

Another example: 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!

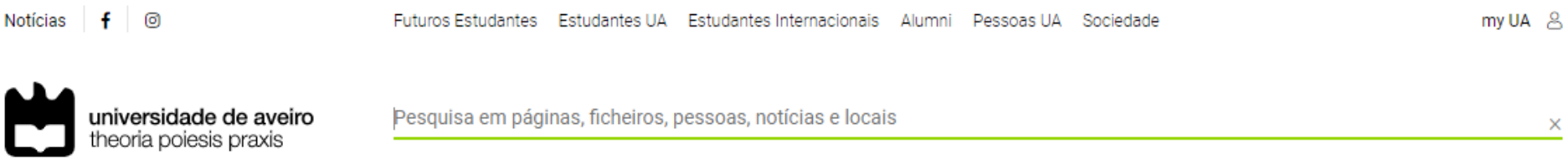
Previous action for foreign students: Select the English version seems easy (it is a "standard" way to do it in sites)

First action in the Portuguese version: find the icon 

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

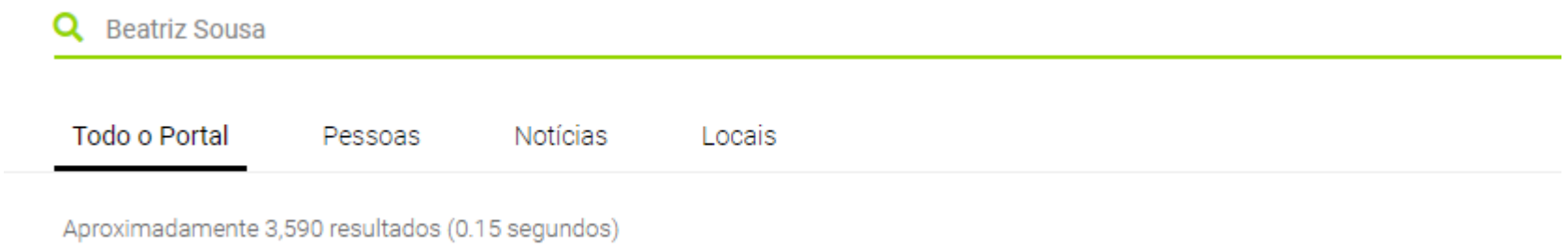
Even without tooltip the correct icon seems recognizable (it is "standard") 

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?



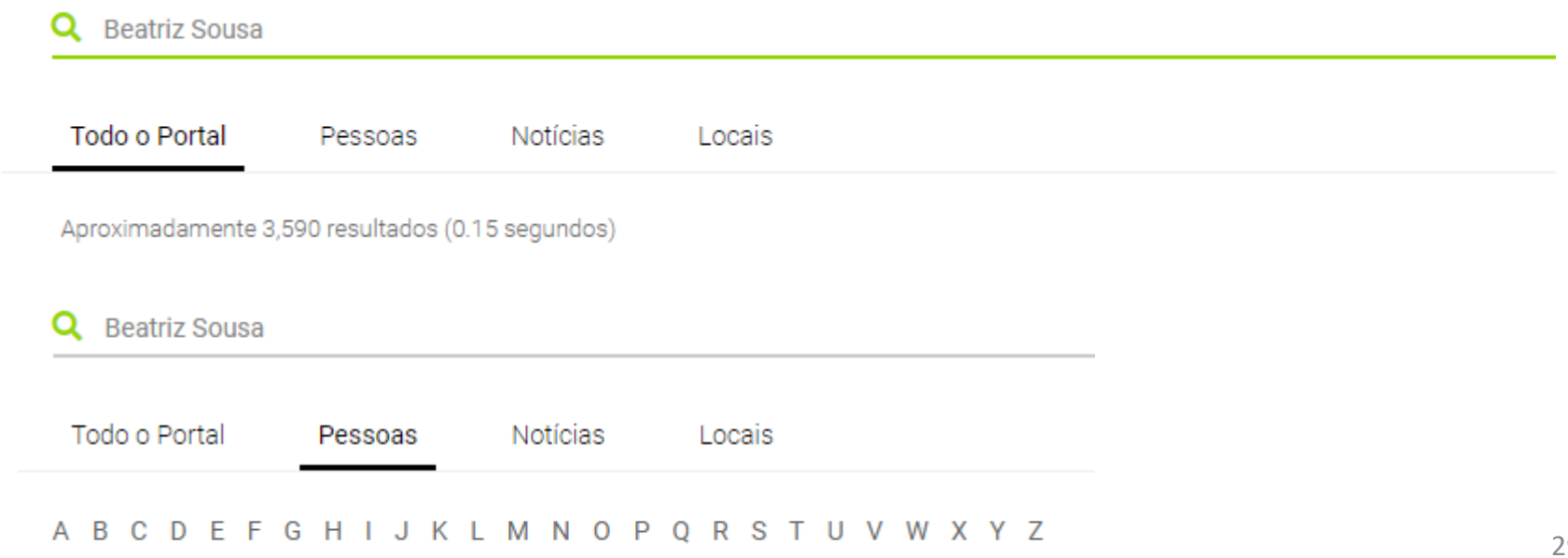
Probably yes; while it may not look a search bar, it is adequately labeled (Pesquisa em páginas, ...)

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

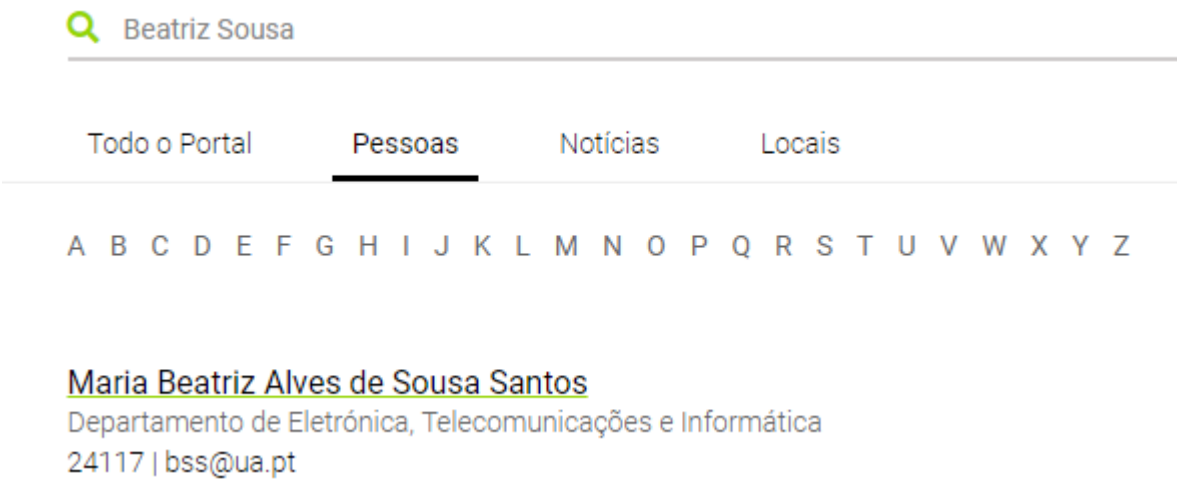


Q1 - Will the user know what to do at this step?

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:

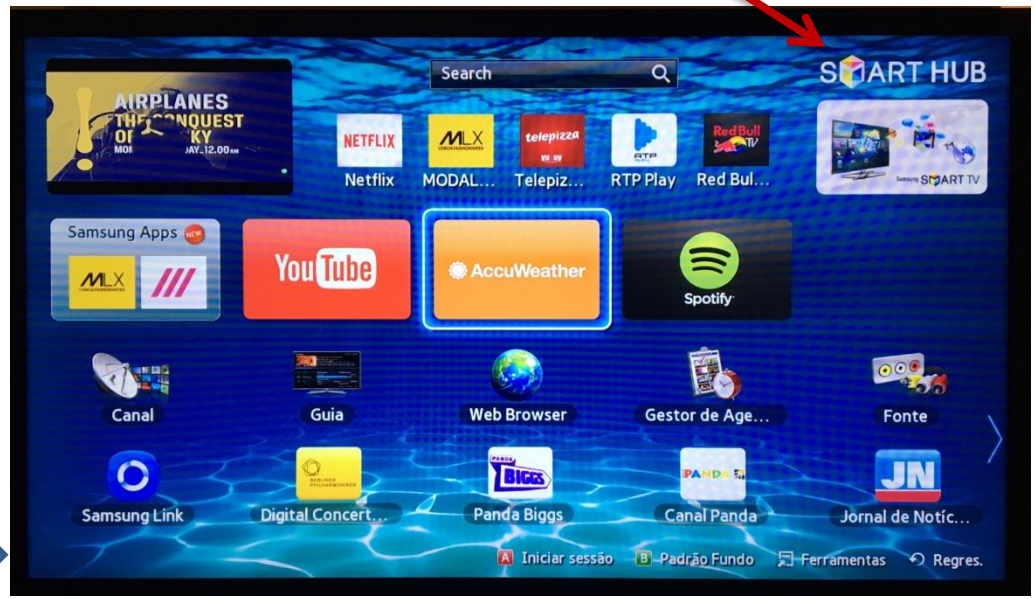
- it seems easy for the target users to reach the phone number and email address;
- however, the phone number may be not recognized as such

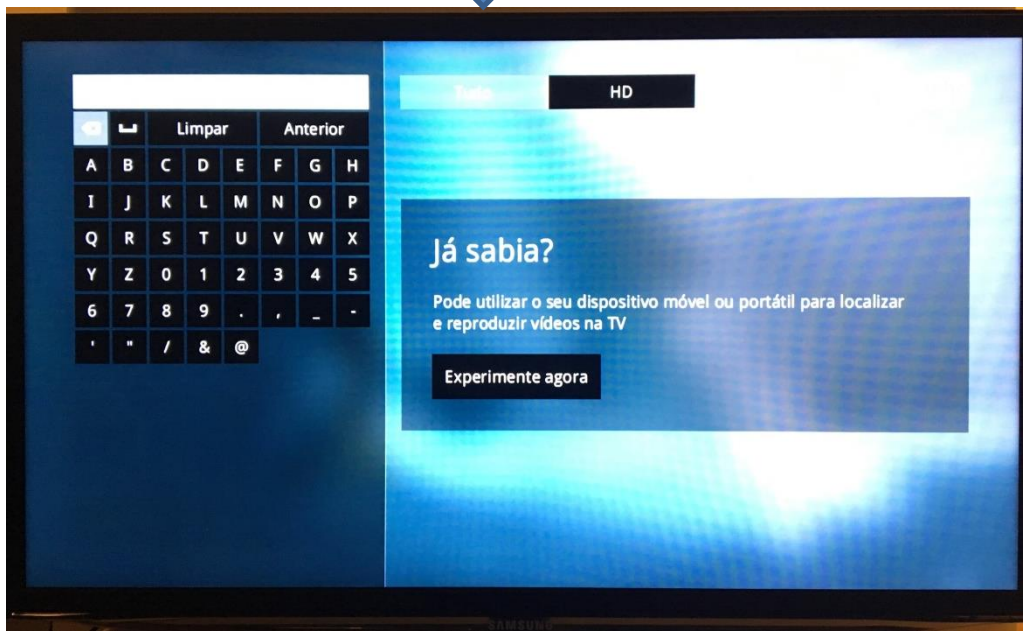
Another example: Smart TV

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:

Analyzing interactive systems/applications that should be very intuitive (e.g. consumer electronics):

- 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

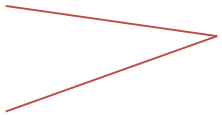
- Change the Channel using the box of your TV service (not the remote control)
  user: anyone having a TV box

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)

Empirical Methods

- Observation
- Query
- Controlled Experiments

usability tests

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!

Anyone doing research with people should have a good grasp of research ethics!

Ethical Maturity in User Research - NN/g

Empirical evaluation styles

These methods may be performed:

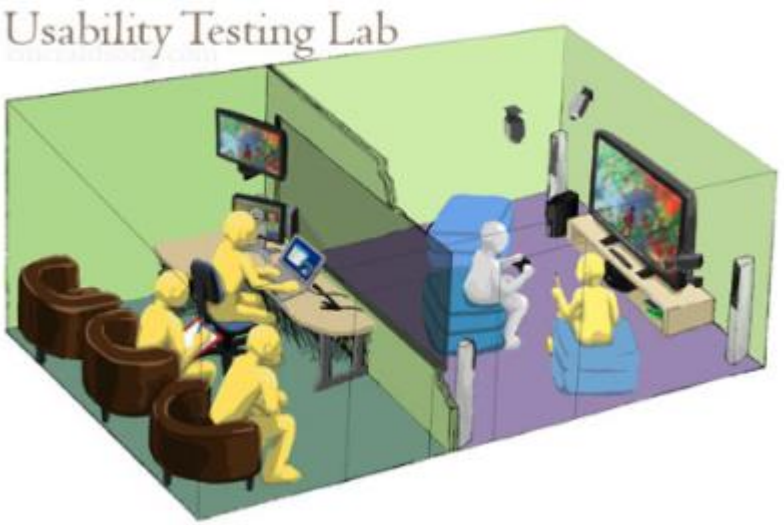
- In the laboratory (more controlled)
- In the field (more realistic)

They produce complementary information;

if possible use both!

Field Studies - NN/g

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

Usability Testing with Users | Usability Body of Knowledge

Think aloud Observation

Participants are asked to use the system while continuously thinking out loud (verbalizing their thoughts as they use the system)

Benefits:

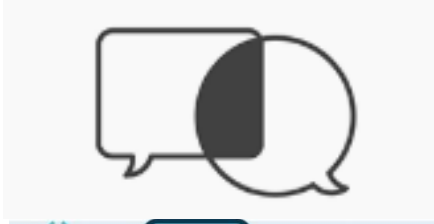
- Inexpensive
- Flexible
- Easy to learn and apply

Limitations:

- Unnatural situation
- Filtered statements
- Changing user behavior

Thinking Aloud: The #1 Usability Tool - NN/g /
When Thinking Aloud Fails (Video) - NN/g
Usability Testing with Users | Usability Body of Knowledge

Query



- Two main variants:
  - Questionnaire
    (reach more people; less flexible)
  - Interview
- **Should always be carefully prepared and tested**
- Collected data should be carefully analyzed

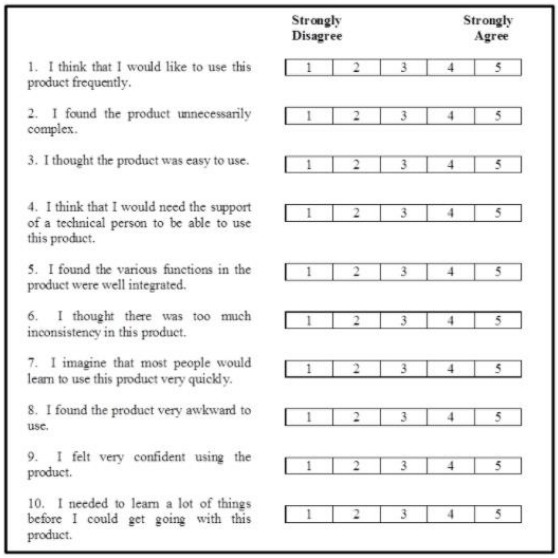
Useful Survey Questions for User Feedback Surveys | IxDF

How to Conduct User Interviews | IxDF

Well-known usability questionnaires

- **System Usability Scale** (SUS)

- Questionnaire for User Interface Satisfaction (QUIS)



	Strongly Disagree				Strongly Agree
1. I think that I would like to use this product frequently.	1	2	3	4	5
2. I found the product unnecessarily complex.	1	2	3	4	5
3. I thought the product was easy to use.	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this product.	1	2	3	4	5
5. I found the various functions in the product were well integrated.	1	2	3	4	5
6. I thought there was too much inconsistency in this product.	1	2	3	4	5
7. I imagine that most people would learn to use this product very quickly.	1	2	3	4	5
8. I found the product very awkward to use.	1	2	3	4	5
9. I felt very confident using the product.	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this product.	1	2	3	4	5

- 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 many 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

Usability – Digital.gov
Beyond the NPS: Measuring Perceived Usability with the SUS, NASA-TLX, and the Single Ease Question After Tasks and Usability Tests - NN/g

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.

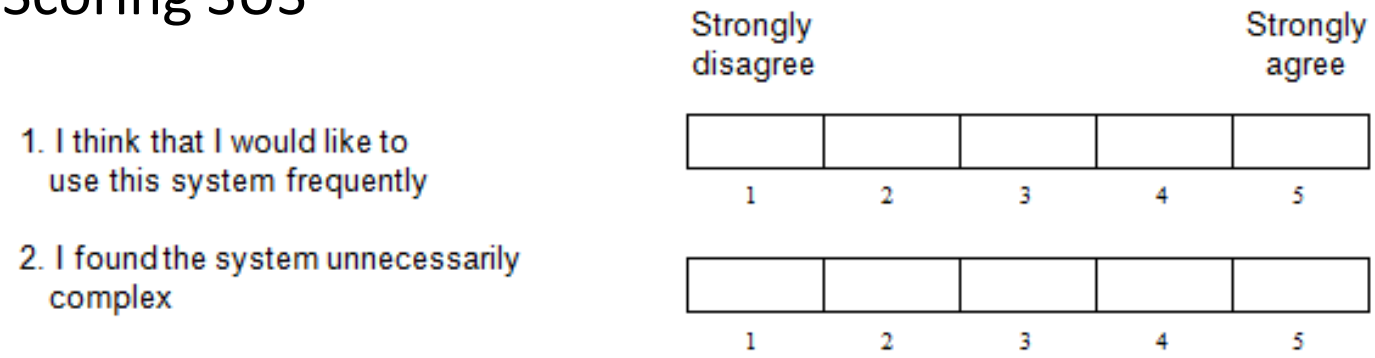
Usability Starter Kit – Digital.gov
SUS-Universal.docx

Scoring SUS

- SUS provides a value in **[0-100]**
- To obtain the value:
  - Add the scores of all questions:
    - odd numbered questions - subtracting 1 from the score
    - even numbered questions - subtracting their value from 5
  - Multiply the sum by 2.5.

SUS > **68 would be considered above average**

Scoring SUS



Step 1: Convert the user ratings from the 10 questions into points:

Strongly Disagree: 1 point
Disagree: 2 points
Neutral: 3 points
Agree: 4 points
Strongly Agree: 5 points

Step 2:

x= Sum of points of all odd-numbered - 5
y= 25 - Sum of points from all even numbered questions

Step 3: SUS Score = (x+y) * 2.5 **SUS [** 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.

Questionnaire for User Interface Satisfaction | Digital Healthcare Research (ahrq.gov)

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

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

Sequence of screens

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

Usability tests

- Involve **observation and query**

- Main aspects:
  - Participants
  - Tasks
  - Test facilities and systems
  - Protocol
  - Usability measures
  - Data analysis

- May have a **complex logistics**

- Standard: **Common Industry Format (CIF)** for usability test reports

What is Usability Testing? | IxDF
Usability – Digital.gov

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, profession, 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**

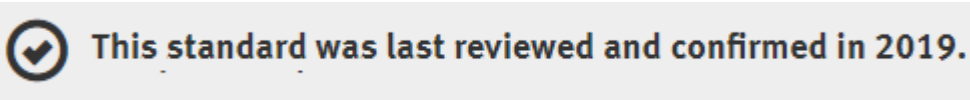
Common Industry Format (CIF) for usability test reports ISO/IEC 25062:2025

- 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

ISO 25062:2025 - Systems and software engineering - Common Industry Format (CIF) for reporting usability evaluations

Usability Standards: ANSI CIF Usability Test Reporting

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



The format includes the following elements:

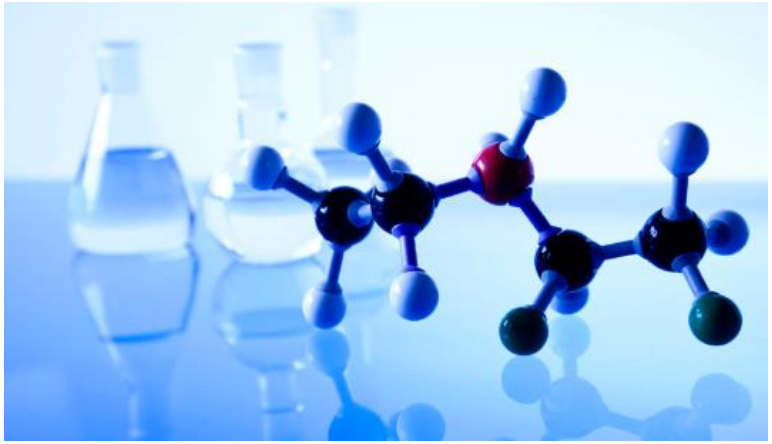
- the description of the product,
- the goals of the test,
- the test participants,
- the tasks the users were asked to perform,
- the experimental design of the test,
- the method or process by which the test was conducted,
- the usability measures and data collection methods, and
- the numerical results.

Evaluation Methods

- Analytical (without users)
  - Heuristic Evaluation ✓
  - Cognitive Walkthrough ✓
  - Model based methods
  - Review methods
- Empirical (involving users)
  - Observation > usability tests ✓
  - Query
  - **Controlled Experiments** ↙

(✓ - have used in Lab classes
↘ - have seen in papers/participated)

Controlled experiments



- The **“work horse” of experimental science** ...

- Important issues to consider:
  - Hypothesis
  - Variables (input or independent; output or dependent, secondary)
  - Experimental design (within-groups; between-groups)
  - Protocol
  - Participants (number, profile)
  - Statistics

Experimental Methods in Human-Computer Interaction | The Encyclopedia of Human-Computer Interaction, 2nd Ed.

Controlled experiment

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

  To the user
  - list of tasks for the experimenter to take notes

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

Controlled experiment

Variables:

- **Independent or input variables** – what is controlled
  (e.g. interaction method)

- **Dependent or output variables** – what is measured
  (e.g. times and errors)

- **Secondary variables –** not controlled but may influence the result
  (e.g. age, previous experience)

Controlled experiment

Experimental design:

- **Within-groups or within-subjects** – all participants use the same conditions (usually in randomized order to avoid bias)

  advantages – a smaller number of participants
  same profile
  disadvantages – prone to fatigue or learning bias

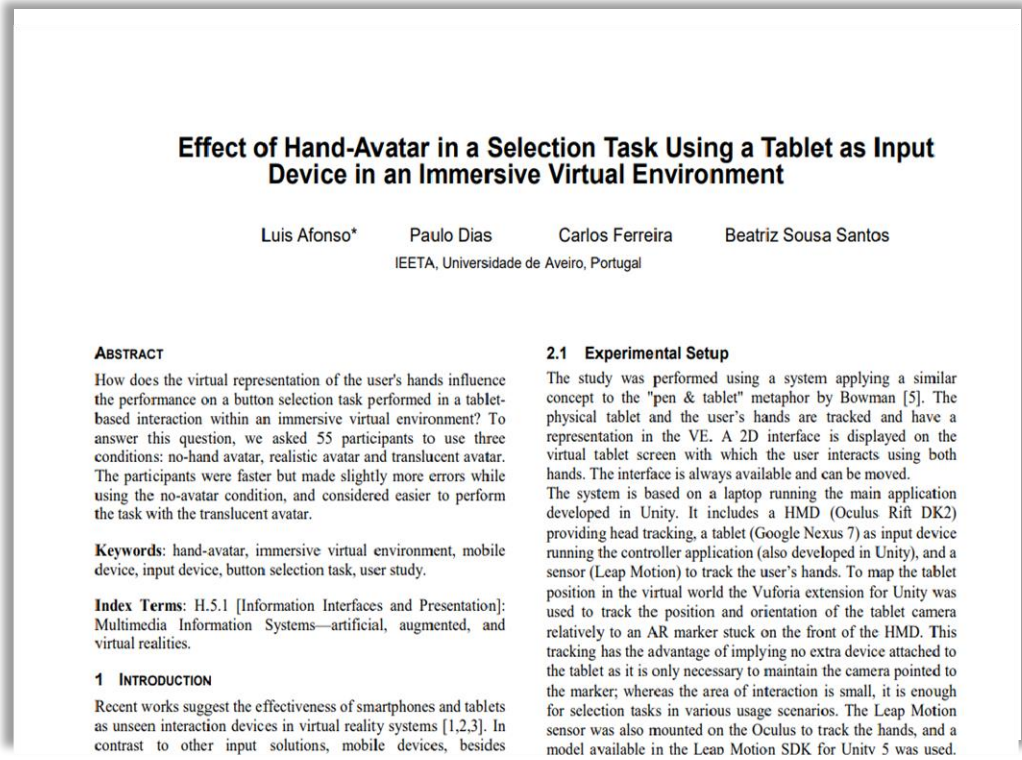
- **Between-groups or between-subjects** – each participant uses only one condition

  advantages – less fatigue or learning bias
  disadvantages – higher number of participants needed
  different participants' profile

Example of a controlled experiment:

Effect of Hand Avatar Using a Tablet as Input in a IVE

- **Laboratory Study**



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

Luis Afonso* Paulo Dias Carlos Ferreira Beatriz Sousa Santos
IEETA, Universidade de Aveiro, Portugal

ABSTRACT

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? To answer this question, we asked 55 participants to use three conditions: no-hand avatar, realistic avatar and translucent avatar. The participants were faster but made slightly more errors while using the no-avatar condition, and considered easier to perform the task with the translucent avatar.

Keywords: hand-avatar, immersive virtual environment, mobile device, input device, button selection task, user study.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—artificial, augmented, and virtual realities.

1 INTRODUCTION

Recent works suggest the effectiveness of smartphones and tablets as unseen interaction devices in virtual reality systems [1,2,3]. In contrast to other input solutions, mobile devices, besides

2.1 Experimental Setup

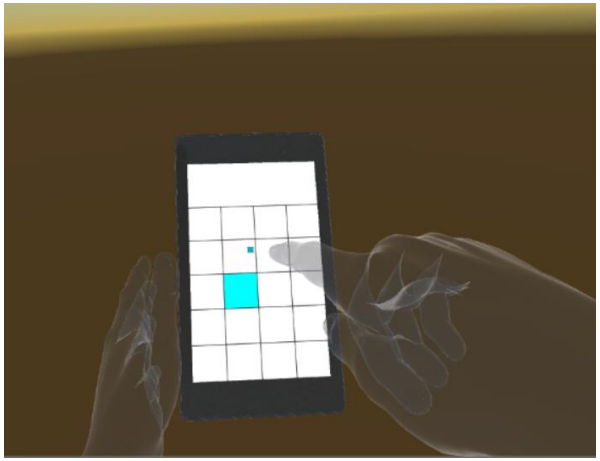
The study was performed using a system applying a similar concept to the "pen & tablet" metaphor by Bowman [5]. The physical tablet and the user's hands are tracked and have a representation in the VE. A 2D interface is displayed on the virtual tablet screen with which the user interacts using both hands. The interface is always available and can be moved.

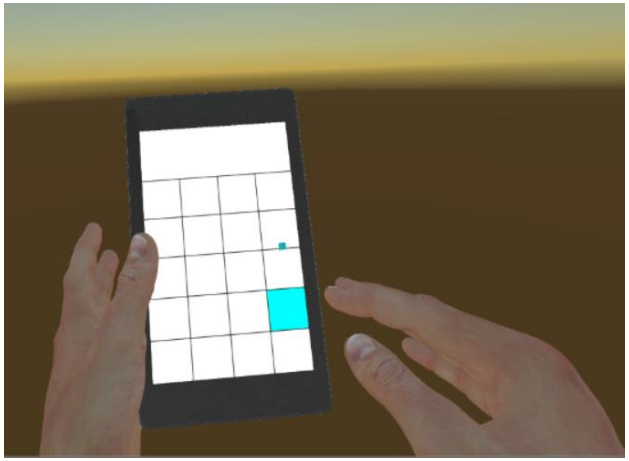
The system is based on a laptop running the main application developed in Unity. It includes a HMD (Oculus Rift DK2) providing head tracking, a tablet (Google Nexus 7) as input device running the controller application (also developed in Unity), and a sensor (Leap Motion) to track the user's hands. To map the tablet position in the virtual world the Vuforia extension for Unity was used to track the position and orientation of the tablet camera relatively to an AR marker stuck on the front of the HMD. This tracking has the advantage of implying no extra device attached to the tablet as it is only necessary to maintain the camera pointed to the marker; whereas the area of interaction is small, it is enough for selection tasks in various usage scenarios. The Leap Motion sensor was also mounted on the Oculus to track the hands, and a model available in the Leap Motion SDK for Unity 5 was used.

Example of a Controlled Experiment performed @ HCI - DETI

- How should the user's hands be represented in a specific Virtual Environment?

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





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

Effect of hand-avatar in a selection task using a tablet as input device in an immersive virtual environment | IEEE Xplore

- **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 slightly faster but made more errors with no-avatar

- Considered easier to perform the task with the translucent avatar

Experimental Design

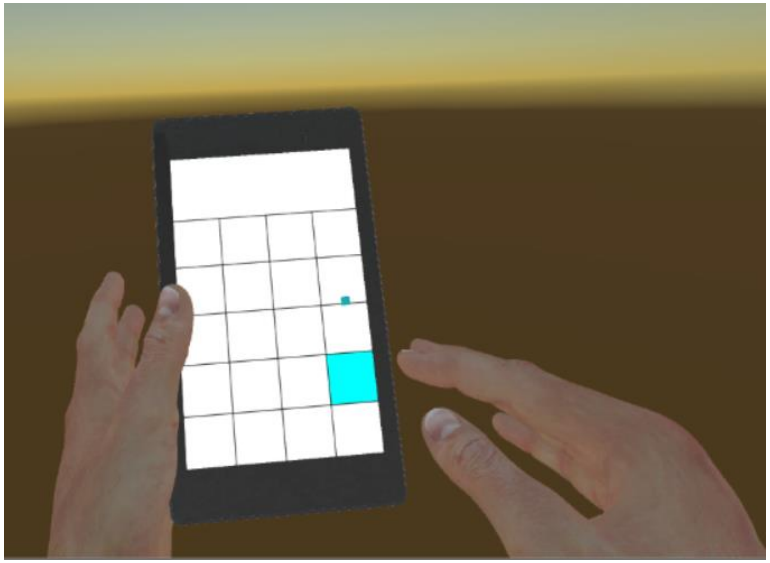
Null Hypothesis: usability is independent of the hands representation

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 learning or fatigue bias)

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

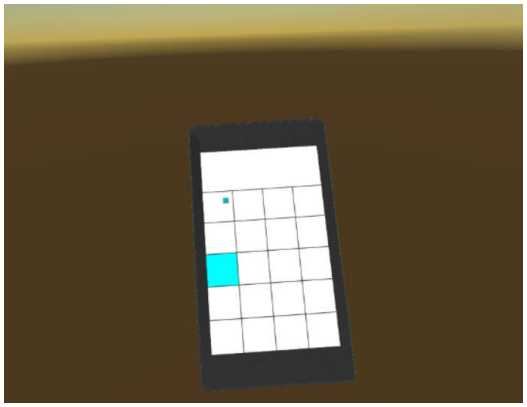


Experimental Conditions

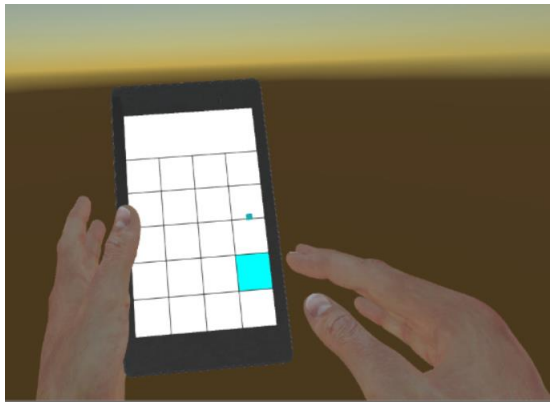
1- **No avatar**: the user only sees the virtual tablet;

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

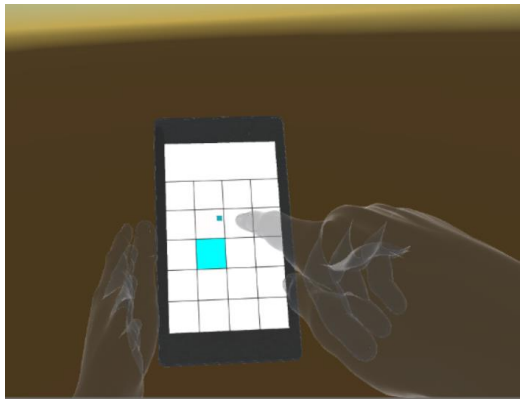
3- **Translucent avatar**: a translucent hand model is used (to alleviate occlusion)



No-avatar



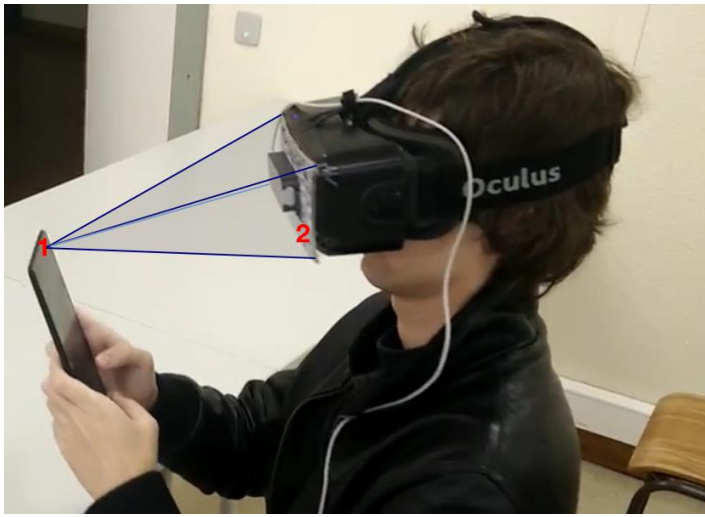
Realistic avatar

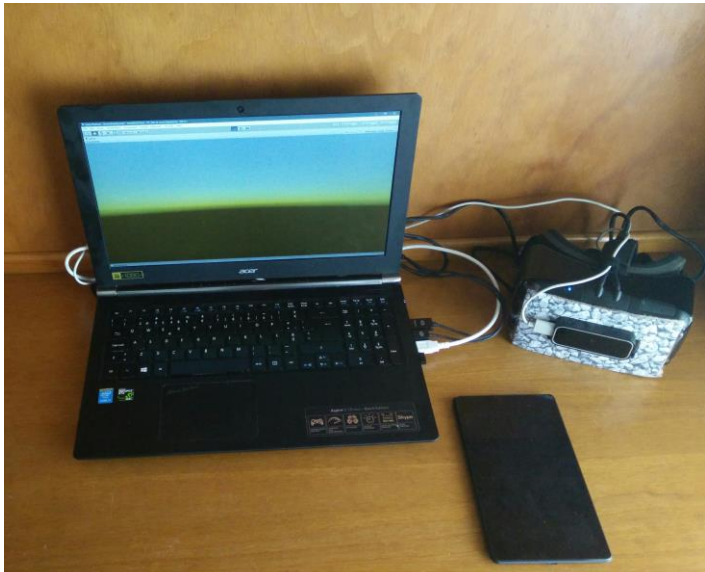


Translucent-avatar

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:

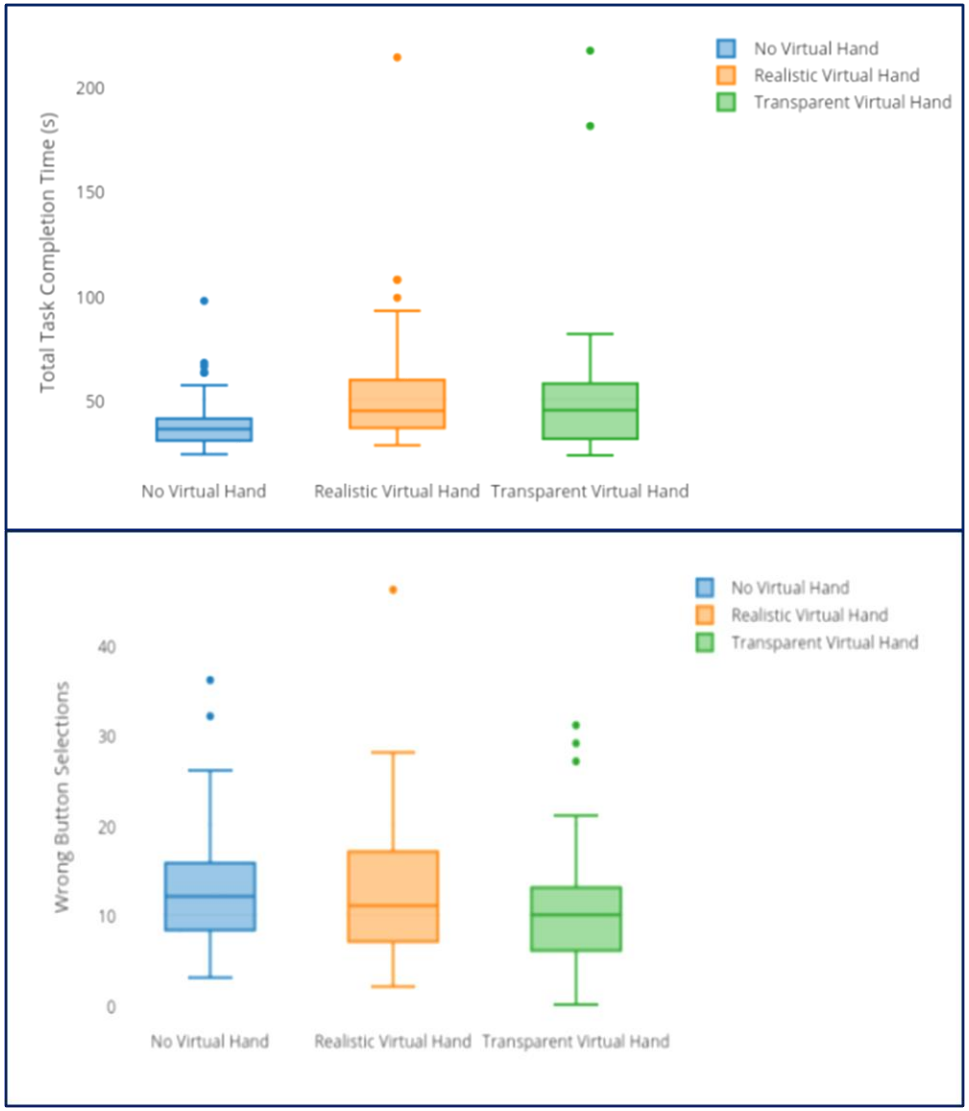
Participants made slightly less errors with avatar - realistic or translucent- (**statistically significant**)

Participants' opinion:

The translucent avatar:

- was more preferred

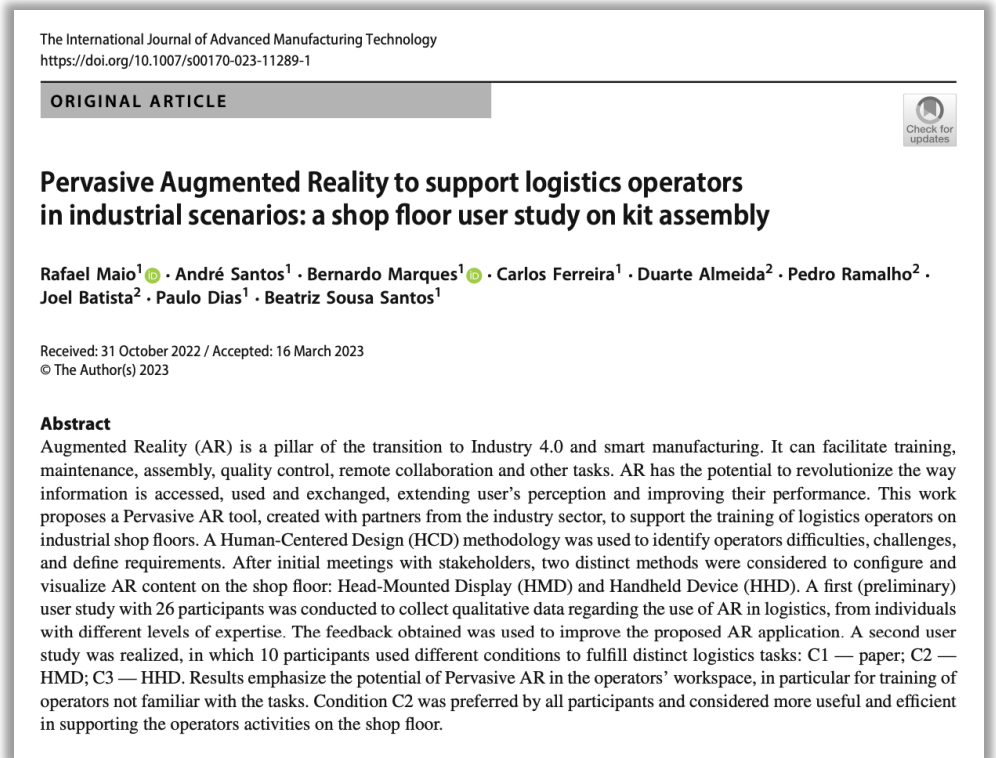
- was considered as better than the realistic avatar (statistically significant)



Example of a controlled experiment:

Pervasive AR to support logistics operators in Industrial scenario: a shop floor user study on kit assembly

➢ **Field Study**



The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-023-11289-1

ORIGINAL ARTICLE

Check for updates

Pervasive Augmented Reality to support logistics operators in industrial scenarios: a shop floor user study on kit assembly

Rafael Maio[1] · André Santos[1] · Bernardo Marques[1] · Carlos Ferreira[1] · Duarte Almeida[2] · Pedro Ramalho[2] · Joel Batista[2] · Paulo Dias[1] · Beatriz Sousa Santos[1]

Received: 31 October 2022 / Accepted: 16 March 2023

Abstract
Augmented Reality (AR) is a pillar of the transition to Industry 4.0 and smart manufacturing. It can facilitate training, maintenance, assembly, quality control, remote collaboration and other tasks. AR has the potential to revolutionize the way information is accessed, used and exchanged, extending user's perception and improving their performance. This work proposes a Pervasive AR tool, created with partners from the industry sector, to support the training of logistics operators on industrial shop floors. A Human-Centered Design (HCD) methodology was used to identify operators difficulties, challenges, and define requirements. After initial meetings with stakeholders, two distinct methods were considered to configure and visualize AR content on the shop floor: Head-Mounted Display (HMD) and Handheld Device (HHD). A first (preliminary) user study with 26 participants was conducted to collect qualitative data regarding the use of AR in logistics, from individuals with different levels of expertise. The feedback obtained was used to improve the proposed AR application. A second user study was realized, in which 10 participants used different conditions to fulfill distinct logistics tasks: C1 — paper; C2 — HMD; C3 — HHD. Results emphasize the potential of Pervasive AR in the operators' workspace, in particular for training of operators not familiar with the tasks. Condition C2 was preferred by all participants and considered more useful and efficient in supporting the operators activities on the shop floor.

“Pervasive AR to support logistics operators in Industrial scenario: a shop floor user study on kit assembly”

- **Industry 4.0** proposes integrating digital and physical worlds into the industrial procedures;
- **Human operators** remain an essential part of any industrial processes;
- During **picking procedures**, operators must identify the necessary components and their location, so that they can be picked up correctly;
- This is mostly done through traditional printed **paper manuals**;
- Using paper manuals is prone to mistakes, decreasing efficiency;
- One **possible solution is using AR** (one of the nine pillars of Industry 4.0).

Study Goal: compare the performance of distinct audiences using 3 conditions

Method: Controlled experiment; field study

C1 – Paper
C2 – HMD (Head Mounted Display)
C3 – HDD (Hand Held Device)

[Pervasive Augmented Reality to support logistics operators in industrial scenarios: a shop floor user study on kit assembly | The International Journal of Advanced Manufacturing Technology]

Experimental Design

Null Hypothesis: all conditions are equally usable

Independent (input) variable (with 3 levels): device

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

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

Task: Real-life picking task: participant had to collect components based on a pre-existing list

Adaptation period was provided.

- **Data collected: Dependent and secondary variables**:
  - Demographic information, task duration, preferences and opinion;
  - Dimensions considered (1—Low;7—High):
    - Level of confusion or distraction about the content used;
    - Level of physical effort;
    - Level of mental effort;
    - Level of satisfaction.

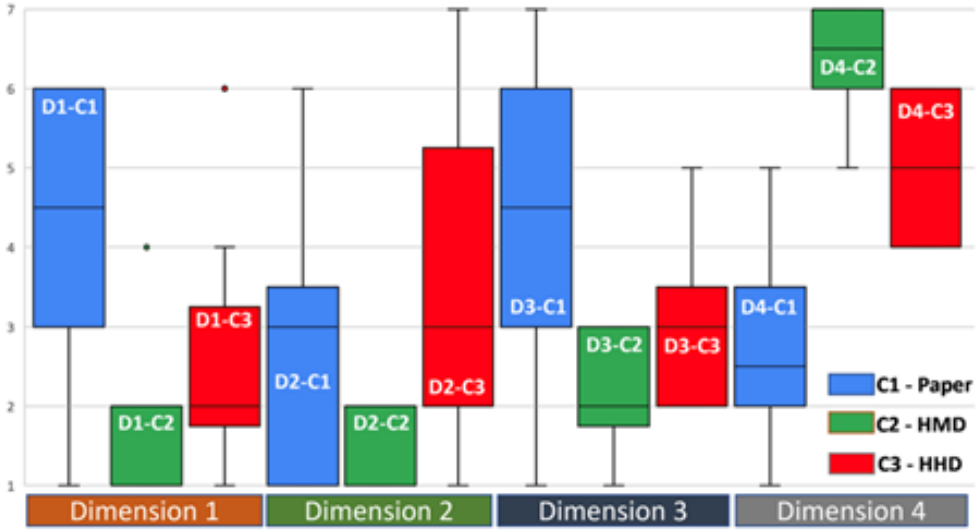
- 10 Participants (1 female) – 22 to 39 years old (M= 26.6, SD= 4.9);
- Data engineer, automation engineer, industrial engineer, maintenance engineer, mechanical engineer, line manager, line worker and manufacturing digitalization engineer;
- 6 participants had previous experience with AR;
- Participants had no experience at the procedures of the task.

- **NOTE:** previous Lab user study with 12 participants (simulated environment);
  two visits to the shop floor – 26 people experienced the proposed methods.

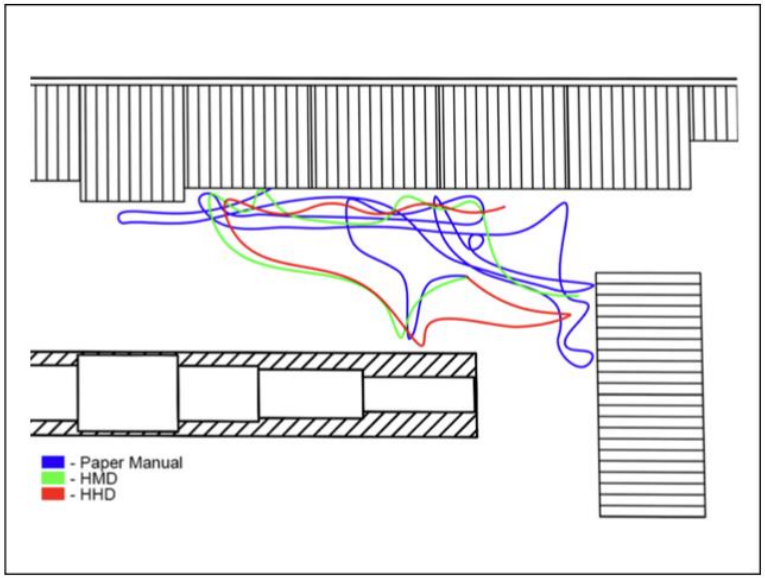
Main results

- D1 – level of confusion about information;
- D2 – level of physical effort;
- D3 – level of mental effort;
- D4 – level of satisfaction.

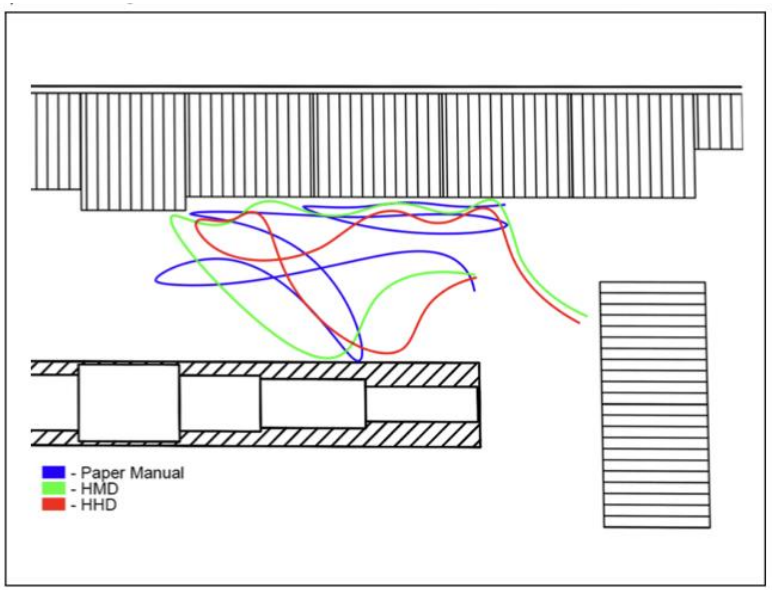
Relevant dimensions considered based on the focus of the study



Main results



(b) Picking routes for the first inexperienced participant.



(c) Picking routes for the second inexperienced participant.

Results emphasize the **potential of Pervasive AR** in the operators' workspace, in particular for training of operators not familiar with the tasks.

C2-HMD was preferred by all participants
considered more **useful and efficient** to support operators on the shop floor

Some observations
(Regarding AR usage)

- Gloves affected hand tracking of human operators;
- Some colors used were easily confused with the environment;
- The percentage of components gathered should be displayed to constantly provide the status of the task;
- The virtual information should not instantly disappear after validating the picking of a specific material;
- The HMD requires an adjustment for ergonomic purposes before being used;
- Having a hands-free setting facilitates accomplishing the components picking tasks.

Future Work

- Longitudinal study with more complex tasks - to better comprehend how AR may affect operators' health, motivation, and productivity;
- Verify how easily the proposed AR solutions can expand to support other use cases at the shop floor;
- Move into co-located settings, allowing various users to visualize the same AR content.

Usability evaluation take away:

Evaluation is fundamental

Usability evaluation take away

- Usability is a fundamental and 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 reliable methods and measures
- There are several methods thoroughly documented
- Should be selected and used according to the guidelines …

Bibliography for Usability evaluation – Books and links

Y. Rogers, H. Sharp, J. Preece, Interaction Design, 6th Edition, Wiley, 2023
(chaps 14-16)

Explore other books at this playlist: Interactive Systems Playlist

UX & Usability Articles from Nielsen Norman Group - NN/g

Usability – Digital.gov

Usability articles and resources