Universitat de Lleida

EVALUATION TECHNIQUES AND USABILITY TESTING

Theme II – Usability Testing Study

(1rst Part – Technique definition. Typologies)

Montserrat Sendín
DIEI - UdL

Superior Polytechnic School – Master's Degree in Informatics Engineering

Contents

- **Technique definition**
  - Genesis
  - General description
  - Clasification within the general taxonomy
- Test typologies
- More common protocols
- Usability laboratorios
- Test Plan design and elaboration

MIPO

Taxonomy in evaluation methods

- Technique Typology -

- **Inspection**
  - **Heuristic Evaluation**
  - **Cognitive walkthrough**
  - Pluralistic Walkthrough
  - Standards inspection
  - Model-based evaluation

- **Inquiry**
  - Field observation
  - Proactive field study
  - *Focus Group*
  - Card sorting
  - Question techniques
    - Interviews
    - Questionnaires
  - Log-based techniques
    - Logging actual use
    - Monitoring physiological answers (*eye-tracking, GSR, ECG, EEG, etc.*)

- **Test**
  - Thinking aloud
  - Constructive interaction or Co-discovery Learning
  - Coaching method
  - Performance Measurement
    - Performance measures
    - Subjective measures
  - Retrospective testing
  - Remote testing

Taxonomy in evaluation methods

- Technique Typology -

- **Inspection**
  - **Heuristic Evaluation**
  - **Cognitive walkthrough**
  - Pluralistic Walkthrough
  - Standards inspection
  - Model-based evaluation

- **Inquiry**
  - Field observation
  - Proactive field study
  - *Focus Group*
  - Card sorting
  - Question techniques
    - Interviews
    - Questionnaires
  - Log-based techniques
    - Logging actual use
    - Monitoring physiological answers (*eye-tracking, GSR, ECG, EEG, etc.)*

- **Test**

Usability Tests

Variants

Definition

▫ **Process** that intends to know from users hand **how usable is** a product, **observing them while they try to solve concrete tasks,** as well as **registering certain measures** *(how fast, easy and pleasant is the use)*

▫ **It is intended to** *"see the interface through participants' eyes"*

UT technique genesis

Scientific origin

▫ Usability measurement is **rooted** in concepts and techniques from:

 – Experimental Phycology

 – Statistical Analysis

Its application to HCI

▫ Originally were used strictly **experimental design** methods, which are based on user samples among 30 and 50 participants

 – Applicable for *formal tests* (also called experiment or rigorous test)

▫ In the early 90s, scientist studies showed that **effective tests** also can be realized with a **minor effort**

 – Applicable for *informal tests* (also called quick tests), as it is shown next

[Nielsen & Landauer, 93]

General description (I)

- **Tipology**: **Quantitative technique** based on the **users behavior analysis** while the realization of <u>concrete tasks</u>

  Qualitative aspects are always also analyzed (comments or non verbal expressions)

- **Life cycle stage**:

  – From **design** onwards, once existing a functional enough prototype

- **Basic reference**: **[Nielsen & Landauer, 93]**

General description (II)

- **Objectives**:

    - **Observe and register** what each participant **says and/or does**, **gathering** quantitative and/or qualitative information

    - **Ask opinion** to participants during and after the test (questionnaires)

    - **Analyze collected data**

    GOALS:

    - **Diagnose usability problems** in the context of a **representative tasks set**

    - **Elaborate recommendations** and **propose improvements** to the user experience with concrete changes

        To be synthetized and defended in a final report and corresponding presentation upon clients

General description (III)

□ **Participants type:**

– **Users: target users**, that is, potential, real users, <u>representative enough</u>, with experience and motivation in similar systems usage

– **Moderator:** responsible of **conducting the test** and of **guiding** participants

– **Highly recommended**: participation of **observers**

□ **Recommended number of users:**

– **It depends** on the <u>type of test</u> (the level of rigor)

– Not less than 4 participants nor much more than 5, *for quick tests*

– Must be taken into account if **different user profiles** and other kind of factors

– *Optimal number*: this is a unsolved question in the scientist scope (formal test)

General description (IV)

Mathematical model that established that **the ideal participants number** is a function of the total number of usability problems detected in a UT for *quick tests*

$$U \left(1 - (1 - p)^n \right)$$

- p – percentage of usability problems detected by an only user
- **n – number of participants**
- U - total number of usability problems detected

[Nielsen&Landauer, 93] *https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/*
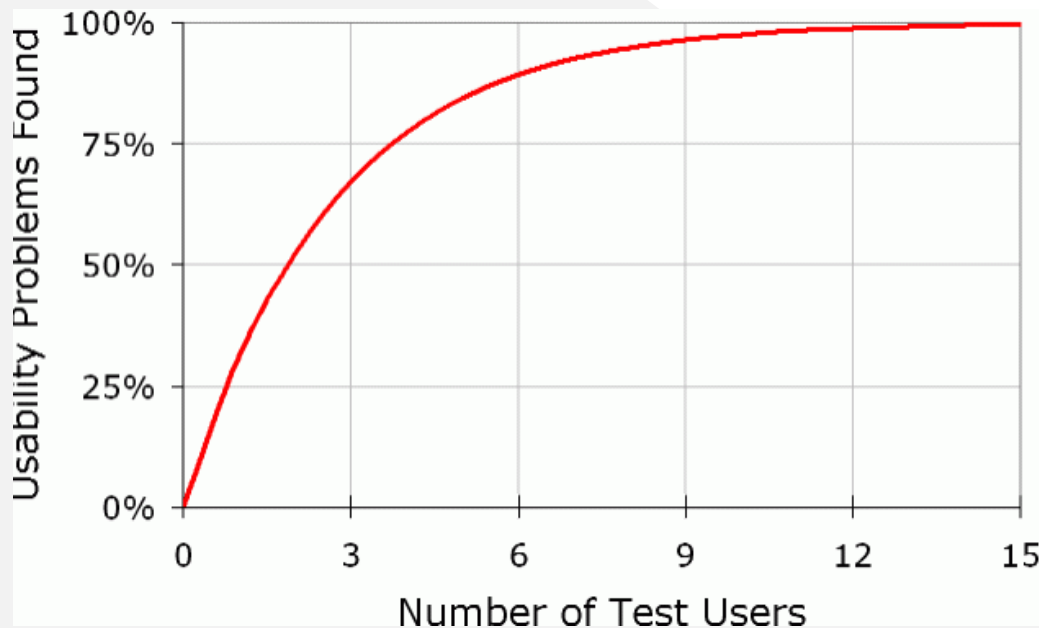
Clasification within the general taxonomy

▫ **Realization place:**

A) Quick or informal test

Not needed a laboratory, but recommended. Any space wide enough with the appropriate equipment can be useful.

B) Semi-formal or formal test (experiment)

Laboratory (controlled environment)

When evaluating an *improved version*: recommended the real environment, because it helps to understand the **use context** (*field study*)

Clasification within the general taxonomy

- **Automation degree:**

  – **Data registration in controlled laboratories is automated**

  – However, the rest of the process is **manual** (*qualitative* data collection, users recruitment, etc.)

    - Only can be considered completely automated in **asynchronous remote tests**, where not even the moderator presence is required (*unmoderated tests*)

      They allow a large volume of participants (*benchmarking*)

Contents

- Technique definition

- Test typologies

- More common protocols

- Usability laboratorios

- Test Plan design and elaboration

MIPO

Test typologies (I)

- **Variants** according to level of rigor:

  - **Quick or informal test.** Focused on collecting opinions or comments from users along the test (**subjective measures** about '*what it is said*')

    Affordable and effective alternative to traditional approach

  - **Semi-Formal test.** Collects and analyzes a combination of **subjective** and **objective measures** about users **behavior** (focused on '*what is done*')

  - **Formal test** (*experiment* or *rigorous test*). Apart from collecting **subjective** and **objective measures**, it applies a statistical analysis applying some sort of experimental design

    - It throws significant results, but it is **costly**

Test typologies (II)

- Quick or informal test **Characteristics**:

  - Protocol: informal, flexible, interaction with user is allowed.
    Drawing conclusions **in situ**

  - Users sample size: bounded (**between 4 or 5** participants is worthy)

  - Realization place: Not necessarily needed a laboratory (real environment is the usual place)
    Field studies are recommended in improved versions of products

  - Other known names: '*guerrilla*' test, *discount usability testing*, *usability check, do-it-yourself, "Low Fidelity" usability testing, "paper prototype" testing, Napkin tests,* **R**apid **I**terative **T**esting and **E**valuation (RITE)

  The Art of Guerilla Usability Testing [D. P. Simon, 2013]

  http://www.uxbooth.com/articles/the-art-of-guerilla-usability-testing/

Test typologies (III)

- <u>Quick or informal test **Characteristics**</u>:

  - <u>Other methodological aspects</u>:

    Formative test

    - Involves only **qualitative data**
    - **Focused on the process**
      - **Short and frequent** life cycle **Iterations**
    - **Immediate results**, such as a **diagnosis**
      - Conceived for **diagnosing preliminary problems**, starting as soon as possible (early stages)
    - **Non significant results**

      ➢ *Perfectly suited for **UCD** methodology*

Test typologies (IV)

▫ <u>Semi-Formal test **Characteristics**</u>:

    – Protocol: quite formal, minimal interaction with participants

    – Users sample size: **between 10 and 20** participants

    – Realization place: Mainly laboratory but also real environment

Test typologies (V)

▫ Semi-Formal test **Characteristics**:

– Other methodological aspects:

▪ Combines **qualitative and quantitative data**

▪ **Focused on the process**

– Generally applied over a **prototype**

– Conceived for **diagnosing and solving concrete problems** throughout the developing process, starting when having a prototype (intermediate stages)

▪ **Non immediate results**, they require to be analyzed

➢ **It does not involves a statistic analysis**

▪ **Results** are **more or less significant**, and can be used to improve the product evolution

Test typologies (VI)

- <u>Formal test **Characteristics**</u>:

  - – Protocol: very formal, invariable, without no interaction with participants (avoiding influencing results)

  - – Users sample size: a **<u>significant</u>** <u>enough sample</u> is required (*recommended* **25** participants minimal)

  - – Realization place: Laboratory

  - – <u>Other known names</u>: *"High Fidelity" usability testing*, controlled study, performance measures

Test typologies (VII)

▫ Formal test **Characteristics**:

  – Other methodological aspects:

  Summative test

  ▪ Combines **qualitative and quantitative data**

  ▪ **Focused on the result**
    – Applied over the **product final version** (deploying stages)
    – Results can be used to improve later product versions

  ▪ **Non immediate results**

    – **Results require to be exploited and statistically analyzed**

  ▪ **Involves a statistic analysis.** The goal is to reach contrastable results and conclusions in order to certain premises regarding usability can be proven

    – **Significant results**

Test typologies (VIII)

▫ <u>Formal test **typical scenarios**</u>:

   – **Comparative between products/versions**

      » Ex: *benchmarking* studies, competence analysis, consulting project

   – Objective audience composed of **different users groups** (user profiles) or destined for the **general public**

      » Interesting approach: **remote tests** (place and/or time asynchrony)

   – **Complex system** or that involves some risk

Test typologies - *Conclusions* (IX)

– In general, **frequent quick tests** are recommendable

– In any case, **do not limit** to an only formal test

Decisive factors to decide level of rigor**:**

- Available time and budget

- Developing product state

- Test objective

- Objective users and localization

- Realization place

NO! That's not how you're supposed to use it!

Developer watching videotape of usability test.

Both variants are common and dominate in UCD methodologies

Contents

- Technique definition

- Test typologies

- More common protocols

- Usability laboratorios

- Test Plan design and elaboration

MIPO

More common protocols

Thinking aloud

- Consistence
  - Users are asked to **express aloud** their thoughts, feelings and opinions while interacting with the system

Application

- **Quick tests** or **semi-formal tests** ➔ users can be prompted to verbalize
- **Formal tests** ➔ without interferences by the moderator

Results

- **Qualitative data**

- Added value: Very useful in capturing of

- Mental model
- Terminology
- Cognitive activities

More common protocols

Retrospective Thinking Aloud

▫ Consistence

– Test session is **video-recorded** and afterwards **revised with the user**

– Evaluator has the chance of asking users questions regarding their behavior during the test and participants can explain what they were doing and why

– There exist analysis tools for video-recordings

▪ For example **Tobii Studio** (RTA)

▫ Application

– **Quick** or **semi-formal tests**

▫ Results

– **Qualitative data** (not a high number of users)

▫ Added value:

– Allows collecting more information from the user. **User comments** while reviewing the video are more **extensive**, although it takes at least twice as long with each one

More common protocols

Coaching method

▫ Consistence

- **Variant** of previous protocols where **explicit interaction** between user and moderator is produced
- Moderator 'conducts' participants to the right direction during the system usage

▫ Application

- **Quick tests**

▫ Results

- **Qualitative data** to obtain from a reduced number of users

▫ Added value:

- Useful **with inexpert users**, in order to discover their information needs

More common protocols

Constructive interaction (co-discovery learning)

- Consistence
  - **Variant** of *Thinking aloud* protocol
  - Instead of one, **two users** perform tasks together, trying to help each other
  - They are **encouraged to explain** what they are thinking about in the meantime, **dialoguing** between them

- Application
  - **Quick tests**

- Results
  - **Qualitative data** to obtain from a reduced number of users

- Added value: **It makes more natural** for the test users to verbalize their thoughts during the test

- Disadvantages regarding *Thinking aloud*:
  - Users can have different learning strategies
  - It is required twice number of participants

Contents

- Technique definition

- Test typologies

- More common protocols

- Usability laboratories

- Test Plan design and elaboration

MIPO

Usability laboratories

Especially adapted spaces for usability testing and other usability evaluation methods realization

Spaces:

- Generally composed of two rooms

    - **Test room** (especially adapted space to carrying out tests)

    - **Observation room** (especially adapted space for evaluators)

- It is common that both rooms are separated with a crystal with **visibility in one direction**: from the observation room to the test room
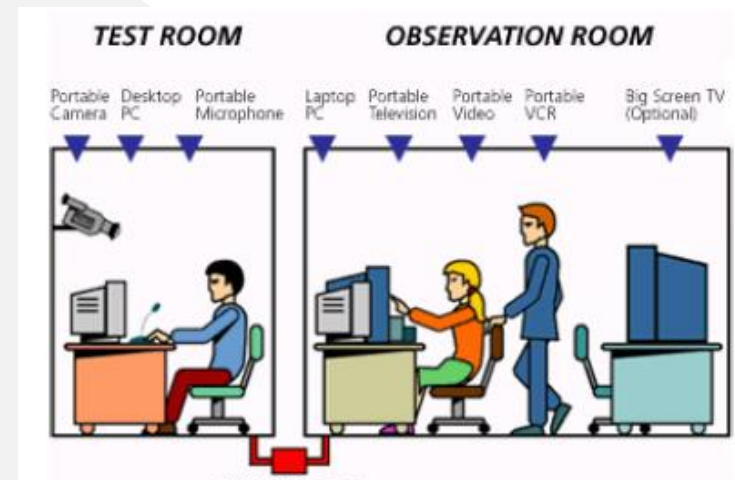
Usability laboratories

Spaces:

- Some laboratories have an adjoining room to the observation room that allows <u>additional observers</u> or **developers** to be located
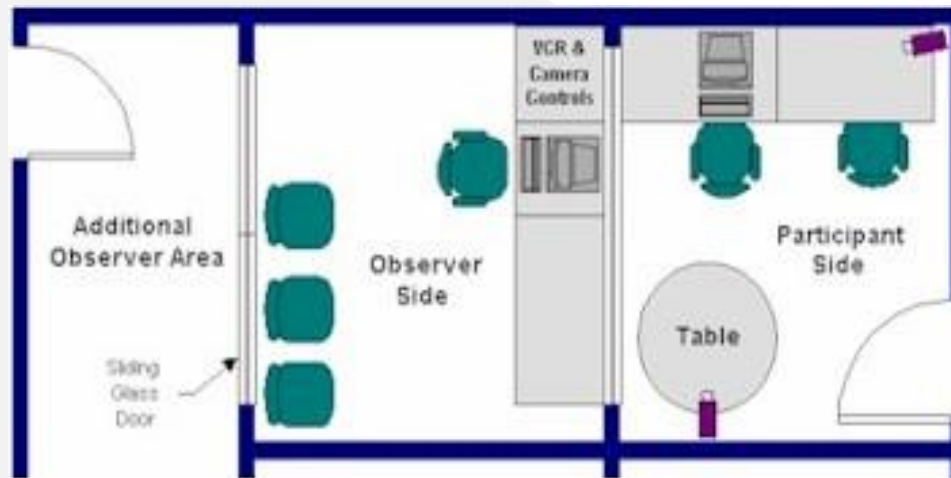
Usability laboratories

Equipment for the test room:

- Terminal with a **specific software** to **register the interaction** of the user with the system in a specific usage scenario

- Video cameras (remotely controlled form the observation room)

- Microphones

- *OPTional*) Required material to **reproduce the workplace user environment** where the user is going to use the interactive product
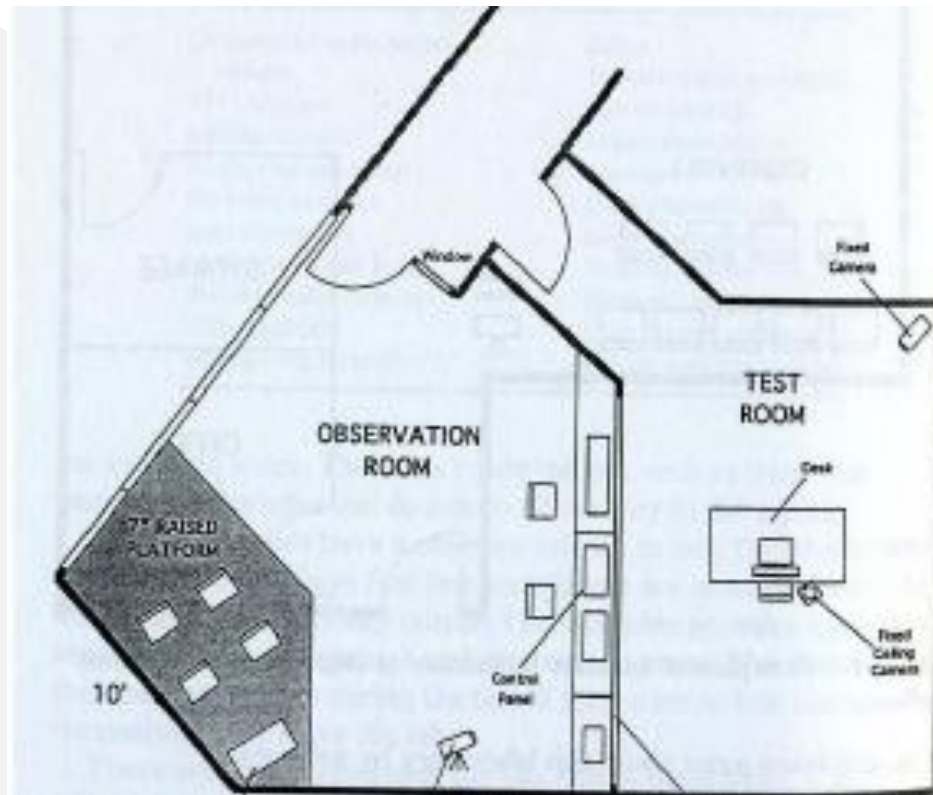
Usability laboratories

Equipment for the observation room:

- Terminal with a **specific software** to **monitor the user interaction** and **analyze** the information afterwards

- Video cameras **control**
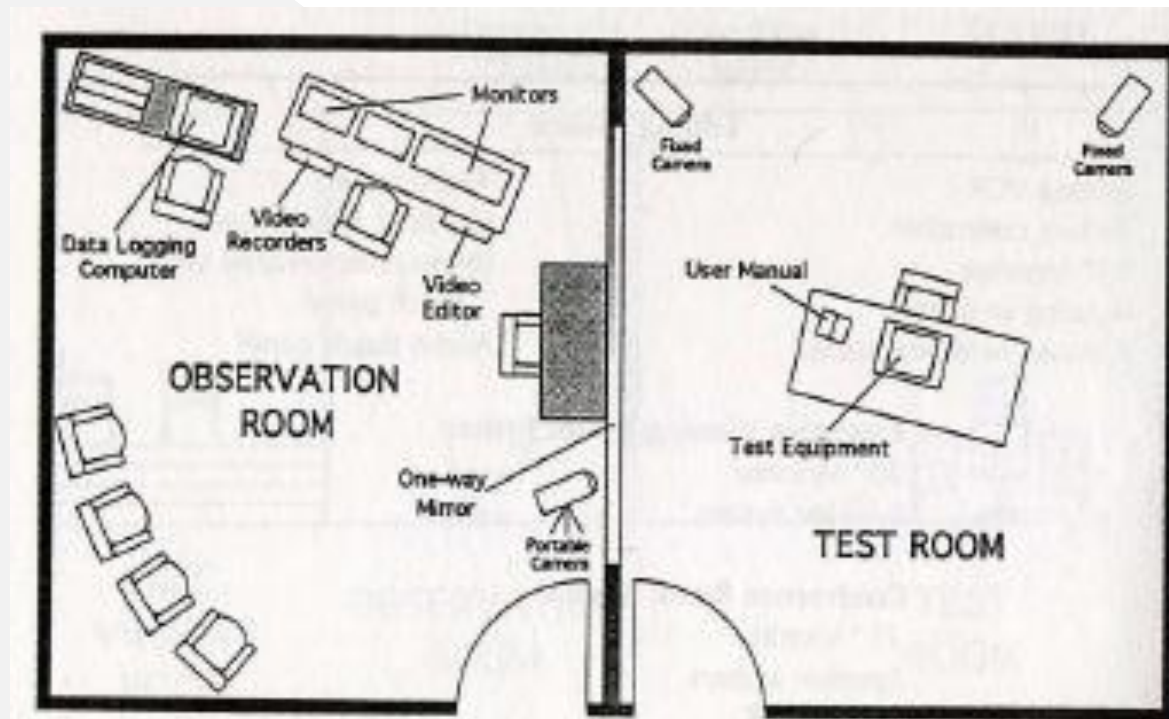
- Digital mixer

- Input scanner

- etc.

Usability laboratories

American Institutes for Research

Usability laboratories

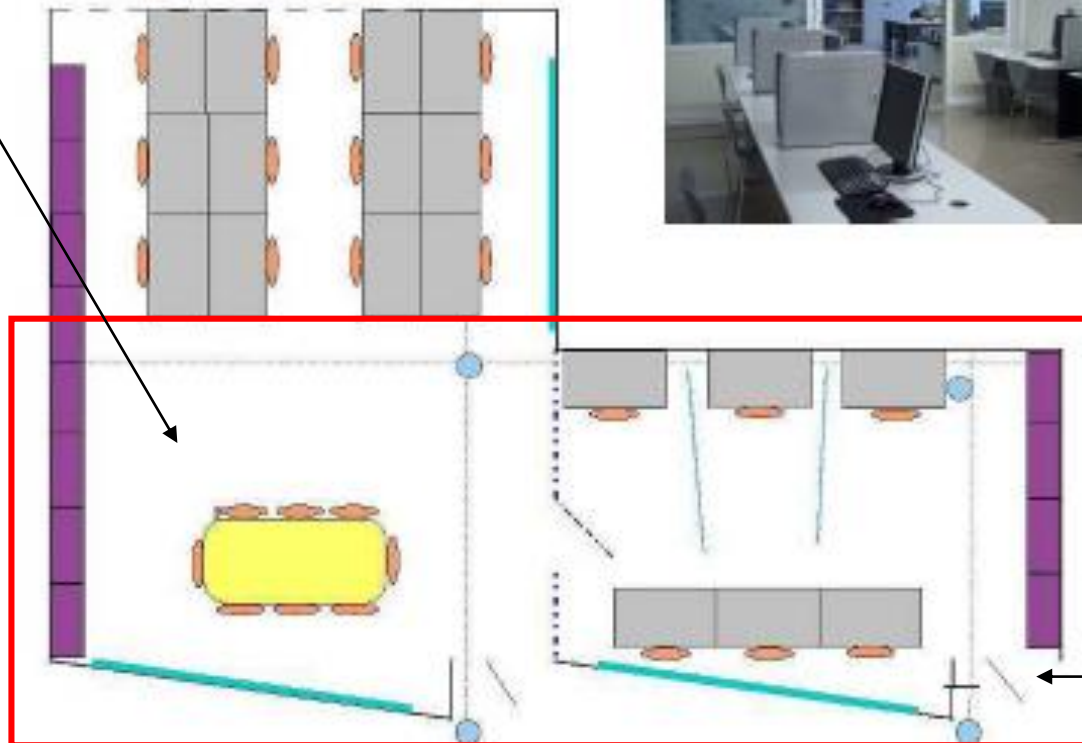**Lotus Development Corporation**

Usability laboratories

La Salle

Usability laboratories

UsabiliLab

http://www.griho.udl.cat/about/infrastructure.html

Additional observer area

Test and
observation room

Usability laboratories

UsabiliLab

- **Specific equipment for the test and observation room:**

    - 6 work stations

    - Specific software for usability testing **Morae** (*Manager*, *Recorder* and *Observer* modules)

    - 3 web cameras and microphones for the user observation

    - The 6 work stations are <u>arranged in pairs</u> (*Recorder/Observer*), so that each pair is prepared to realize a usability testing

        Recorder terminals are **in front of** observer terminals

        Morae Manager module allows to analyze, afterwards, the information (audio, video and data) registered during the usability test sessions
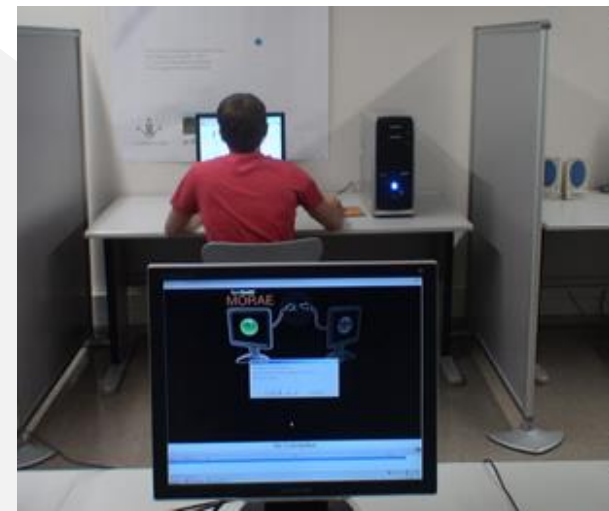
Usability laboratories

UsabiliLab

- **Additional equipment:**

  - *Eye-tracking* equipment for both, desktop and mobile environments
  - Multi-touch screen attachable to ET monitor
  - Video camera
  - Camera tripod
  - *Tablets* and *Smartphones*
  - Multimedia centers
  - Specific software for accessibility evaluations, such as **Jaws**