Universitat de Lleida

EVALUATION TECHNIQUES AND USABILITY TESTING

Theme I – Introduction to the Evaluation of the Usability

Montserrat Sendín
DIEI - UdL

Contents

- What is the evaluation in the HCI scope?
- Objectives of evaluation and aspects to consider
- Taxonomy in evaluation methods
- Integration of the evaluation techniques to the development life cycle
- Conclusions

What is the evaluation in the HCI scope?

What is usability ?

Extent to which an interactive system can be used with **effectiveness, efficiency, and satisfaction** by specified users to achieve specific objectives in specific contexts of use

[ISO 9241-11, 98]

What is the evaluation in the HCI scope?

What is usability evaluation ?

- It refers to **evaluate the usability**, as the **main objective** in an interactive application
- **Basic activity** along the development of an interactive system, and one of the **essential parts** of the User Centered Design (UCD)
- Hence, it is a **¡ need !**

Typical attitudes to be avoided:

- ***Focus in functionality without paying attention to other aspects***
- ***Usable? If I can use it, it's ok ...***

What is the evaluation in the HCI scope?

What implies ?

- Applying **iterative life cycle** and carrying out **usability evaluation methods** in a continued manner
- **Analyzing users and the environment** where users are going to use the product, **testing prototypes, designs** and so on, achieving to **integrate definitely users in the development life cycle**

Evaluation covers a set of methods and techniques that analyze the usability of an interactive system in different stages of the life cycle

What is the evaluation in the HCI scope?

How contributes ?

- Obtaining **better products in a contrastable way**
- To the extent that users are able to carry out tasks in a more **effective, efficient and satisfactory** way
  - **The User eXperience improves**
  - Allowing to know if a design or a system
    - Satisfies or not the **users' expectations**
    - Conforms to users' **social, physical and organizational context**
- The **development efficiency** and **cost** also improves

Objectives of evaluation

- Check the system functionality
- Verify the **impact** of the user interface in future or potential users, trying to **improve** it
- Identify any specific problem related to the system **usage** and its **interaction**

[A. Dix et al., Human-Computer Interaction, 97]

➔ **Guarantees the usability** of the system

Aspects to consider for designing evaluation

- **Cost**
  (both, in terms of time and material needed)
- **People involved**
  (evaluators and/or users)
- **Place where is going to be carried out**
  (laboratory or real workplace environment)
- **Automation degree**
  (level of human intervention during the deployment)
- **Life cycle stages**
  (the sooner, the better)

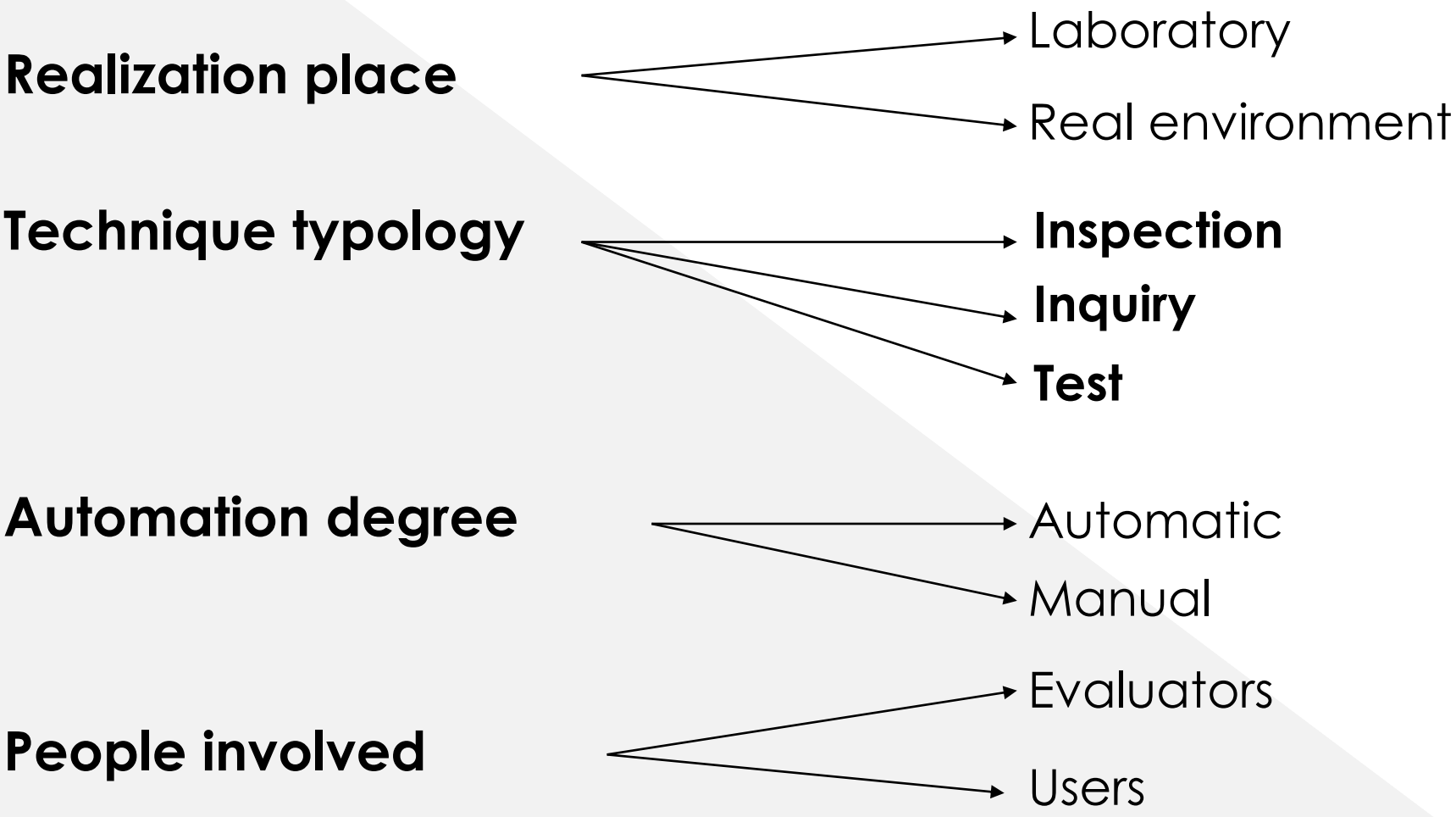
Aspects to consider for designing evaluation

- **Evaluation Plan**:
  - **Make a clear identification of objectives** to obtain, before starting evaluation
  - Each evaluation must have a designated **responsible** person
  - Evaluation must be integrated regarding **planning** and **economic cost** inside the project

Contents

- What is the evaluation in the HCI scope?
- Objectives of evaluation and aspects to consider
- Taxonomy in evaluation methods
- Integration of the evaluation techniques to the development life cycle
- Conclusions

Taxonomy in evaluation methods



Taxonomy in evaluation methods

- Realization place

- **Laboratory**

  + Appropriate for a **controlled and objective evaluation development** with **thorough observation**

    - **Sophisticated equipment facilities**, which make easier monitoring, observation and registration of specific and diverse aspects in an automated or semi-automated and precise way, **minimizes users distractions** and **reproducing the same conditions** for the whole user sample

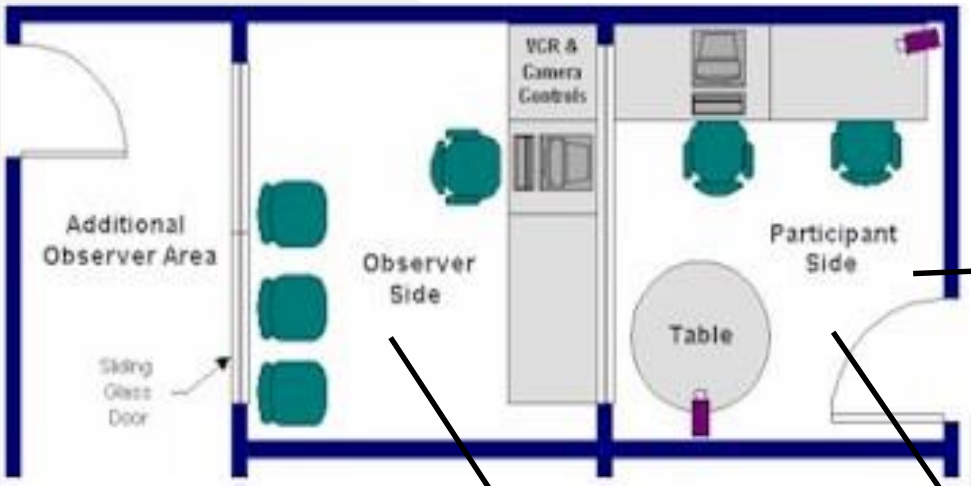
  - Participants can feel **influenced** by the evaluator presence (*intrusiveness*)

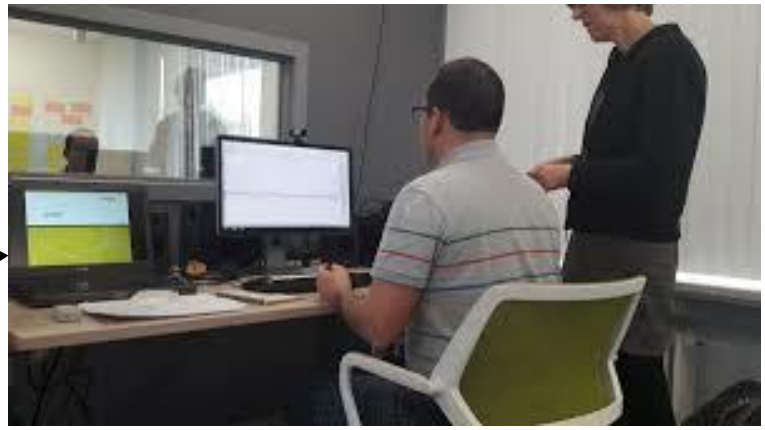
  - **Natural user environment conditions** get lost

➢ **Not realistic environment** regarding **work conditions**

Realization place – *Usability laboratory*

Especially adapted spaces for usability evaluation methods realization





Realization place – *Usability laboratory*

Desktop Lab for mobile devices

Portable Lab for mobile devices

Portable Lab provided with *Eyetracking* and mobile support

Taxonomy in evaluation methods

- Realization place

- **Real environment** (field studies)

  - **There is no control** over the user activity
    - **Common workplace distractions** can make observation difficult
  - \+ **Intrusiveness** feeling gets **reduced**
  - \+ **Natural user environment conditions** and real context **reproduction**
    - Great opportunity to observe how users cooperate with each other
    - Helps to an exhaustive **requirements specification**

Control vs. Naturality

Ideally, development process should include **both styles**

It is claimed that laboratory studies should dominate in the *early stages*, and field studies in the *implementation stages*

Taxonomy in evaluation methods

- Automation degree

- **Automatic methods:**
  - **Consume** little resources, but high **time-consuming preparation**
  - They are highly **efficient**
    - They can be carried out **quickly**
  - They need a **laboratory**
    - They make possible to carry out **remote tests**
  - Results are based on the same parameters, **without subjective appreciations**

➢ Generally applied in **formal** evaluation methods, focused on statistic results (experiments)

Taxonomy in evaluation methods

- Automation degree

- **Manual methods:**
  - **Consume** more resources, but less **time-consuming preparation**
  - They don't need a **laboratory**
  - They **allow to improvise** and to evaluate aspects than ***'fall outside the pattern'***

➢ Appropriate for **informal** evaluation methods, not focused on statistic results

Taxonomy in evaluation methods

- People involved

- **Methods without users**

  Only expert evaluators with the support of **scripts, guidelines or specific documents**

  + They can be carried out **quickly** and are **quite cheap**

  - They do not evaluate the **real system usage**
  - **They lack of feedback** from an evaluation with users
  - **Subjectivity**: they are subject to evaluators point of view

  Imprescindible to combine both kind of methods

Taxonomy in evaluation methods

- People involved

- **Methods with users and/or stakeholders**

  Direct participation of **representative** users or *stakeholders*

  - **People or organisms implied** in the system specification or development, having a direct or indirect influence in the system requirements

  + They provide **feedback** from users, evaluating the **real system usage**
  + Tests can be initiated in **early development stages**

  - Difficulties related to **users recruitment**
  - They are **expensive** and **time-consuming**

Taxonomy in evaluation methods

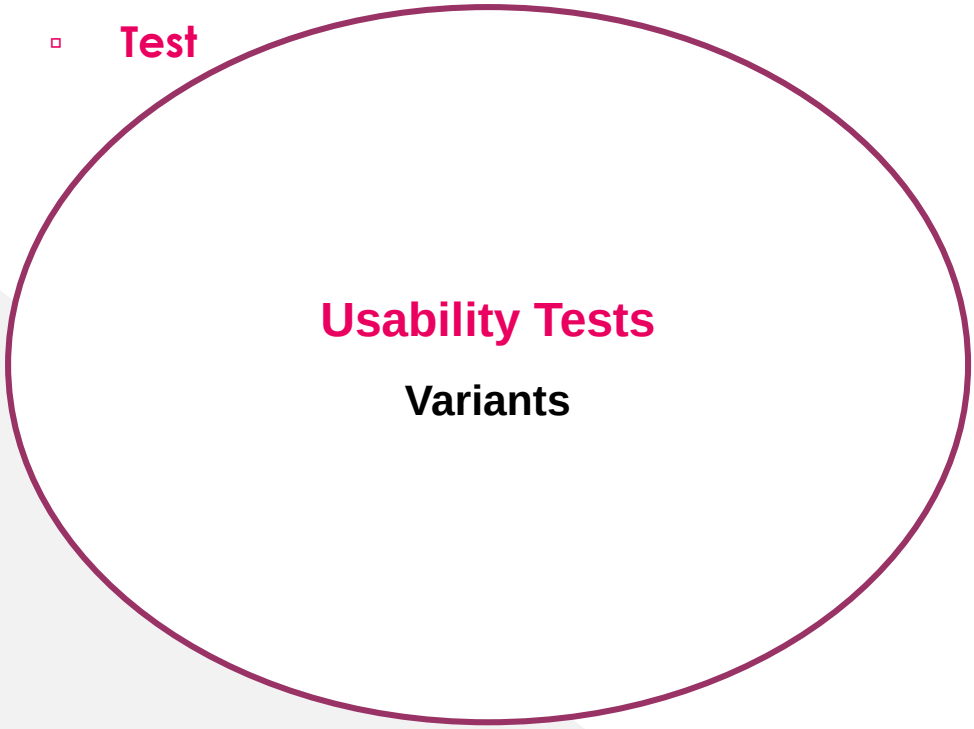
- Technique Typology -

- **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**



Technique Typology

- ***Inspection***
- ***Inquiry***
- ***Test***

Inspection

- Method category where some **usability specialists examine thoroughly** (inspection) diverse **usability-related aspects** of the user interface
- They produce a **fundamental factor** for usability evaluation:
  - **Opinions** and **judgements from expert evaluators** regarding the **accordance** of the user interface to certain **general usability principles** universally recognized (e.g. heuristic principles)

- **Low cost** (users no required), because **only evaluators**
- They **allow to detect a lot of** potential usability **problems**
- Able to be applied in **diverse development stages**

Evaluation Methods based on Inspection

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

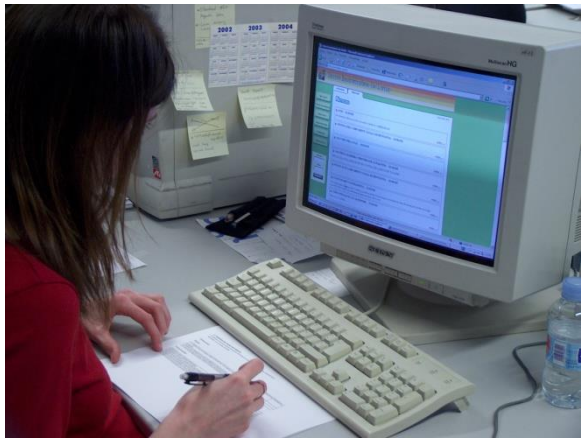
Inspection: **Heuristic Evaluation**

developed by **Jakob Nielsen and Rolf Molich**

- **Several evaluators** examine **independently** the interface and then **analyze its accordance to a set of** specifically selected **heuristic principles**[1] (*heuristics* and *sub-heuristics*)
  - Recommended between 3 and 5 evaluators
  - Each evaluator issues an individual report
  - Finally, results are analyzed on the whole

[1] **Guideline**, **general principle** or **rule of thumb** that describes the common properties of an usable user interface

Ejemplo de aplicación de la EH



Proyecto: WEB de Els Infants de la Paeria de Lleida

evaluador: Marta Gonzalez

4. Control y libertad para el usuario

Los usuarios eligen a veces funciones del sistema por error y necesitan a menudo una salida de emergencia claramente marcada, esto es, salir del estado indeseado sin tener que pasar por un diálogo extendido. Es importante disponer de deshacer y rehacer

Sub heurísticos

	Impacto	Frecuencia	Persistencia
a) Es posible deshacer una acción siempre que sea una operación funcional y operativa.	3	MEDIA	
b) En caso de un proceso de diversos pasos, es posible volver a pasos anteriores del proceso para modificarlos.	3	MEDIA	
c) Existe una salida de la página, del proceso o de la estructura de información mediante acciones tipo "Desconectar" o "Cancelar".	1		
d) Se inician de manera automática acciones que el usuario no ha solicitado explícitamente.	0		
e) Se utilizan animaciones no controladas por el usuario.	1		
f) El scroll no ocupa más de dos pantallas.	0		
g) Es posible guardar información.	1		
h) Es posible imprimir la información sin perder información.			
i) Existe un vínculo que permite volver al inicio de la aplicación.	0		
j) Es posible aumentar y disminuir el tamaño de la letra.	1	ALTA	
k) El sistema se visualiza perfectamente utilizando diferentes resoluciones de pantalla.	0		
l) La interfaz no introduce tecnologías que requieren versiones actualizadas de elementos externos (navegadores, plugs-ins, DLL's,…).	1		

notas del evaluador:

a) cuando pulsas en la opción de enviar no hay la posibilidad de cancelar. Lo envia y ja está.
b) cuando pulsas en la opción de enviar no hay la posibilidad de cancelar. Lo envia y ja está.
c)
d)
e) si, en el banner informativo
f)
g)
h) cuando he impreso toda la banda inferior se ha girado!!
i)
j)
k)
l) Si necesita el plug-in del Macromedia Flash, pero es muy habitual.

Inspection: **Cognitive Walkthrough**

- **Evaluators explore** the user interface design
  - In the context of one or diverse previously established tasks
  - **Evaluators** assume the user role, according to a pre- established profile

- Interface to evaluate needs to be in a certain advanced state (prototype, design or system)

  ➢ **Ideal** in the **design stage**

- Variation of traditional CW: **Pluralistic Walkthrough**

  - **Participants:** developers and experts

Inspection: **Standards inspection**

Set of methods that are all based on having evaluators inspect a user interface

- An expert on an interface standard **inspects** the interface **for compliance**
  - *Standard*: for example, the style guides for a specific mobile platform (Android style guide, iOS style guide, etc.,)

Inquiry

- Category of methods destined to obtain **information about users** in **early development stages**, such as:
  - users' needs
  - users' likes, dislikes
  - users' understanding of the system
  - possible existing problems (*current product version*)

- Intended to **uncover**, **learn**, **generate design ideas** about the product that is going to be developed (*early stages*)
- Apart from in early stages, it also can be applied **once the product is finalized** (*deployment stages*)

Inquiry

- **Ways to carry out inquiry methods**:
  - **talking** to them
  - **observing** them using the system in real working place
  - letting them **answer** questions verbally or in written form

Evaluation Methods based on 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)

Inquiry: **Field observation**

- Evaluators **move to the users working place** and they **observe** in order to understand:
  - **How** users make uso of the system in their daily tasks
  - The users **mental model** about the product to be evaluated

  ➢ *Adequate in **early developing stages** and also in test and **deployment stages***

Inquiry: **Focus Group**

- Data collecting technique where about 6 to 9 users are **brought together to discuss** certain issues related to the system
  - Generally related to initial requirements and context of use
- A **human factors** expert plays the role of a **moderator**, in order to:
  - lead the group
- Or the moderator himself, or either an observer:
  - prepare the list of issues to be discussed beforehand
  - seek to gather the needed information from the discussion

➢ *Useful to capture **spontaneous user reactions** and **ideas** that evolve in the **dynamic group process***

FG put in practice

Inquiry: Question techniques

Direct and structured way to collect information

- **Ques.:** *How to know if a system satisfies the user requirements?*
- **Answ.:** *¡ask to the own user!*
  - ➢ Obtain from users hand point of view
- Advantages:
  - **Extract information** over the user preferences, feelings and attitudes
  - Help to find options not considered in the design, as well as certain design problems
- Problems:
  - Subjective information

Inquiry: **Question techniques**

Interviews

- *Top-down* approach
- **Context adaptable:** the interview can be adapted on the progress in order to obtain the maximum benefit

Inquiry: **Question techniques**

Questionnaires

- **Less flexible** than interviews
- Can be applied to a **largest group**
- Can be analysed more rigorously than an interview

Questionnaire types:

- **Pre-test**
  - Information and participants profiles
- **Post-task**
  - Collect opinions and ratings of each task
- **Post-test**
  - Collect opinions and ratings once participants finish the tasks

Inquiry: Logging actual use

- **Automatically collects statistics about the system usage** (e.g. web analytics)
- Focused on registering information about the **user actions** while system interaction
  - Information provided by **peripheral** system devices (mouse, keyboard), or the corresponding operating system
  - Specific code injection in order to specialize the system in order to register **usage data**
- Different human aspects registration
  - *Eye-tracking*
  - Physiologic (GSR, ECG, EEG, etc.)

Test

- **Representative users make use** of the user interface
  - solving **concrete typical tasks,** interacting **with the own system or prototype to evaluate**
- **Evaluators observe, collect and analyze results** in order to see how the user interface supports the users when carrying out designed tasks

Usability Testing

Characteristics:

- Participants represent **real users**
- Participants have to solve **real and representative tasks**
- What participants do and say **is observed and registered**
- **Data is analyzed, real problems are diagnosed** and finally **some changes are recommended**

Two data types can be can managed:

- *Quantitative* ➔ *Performance measures*
- *Qualitative* ➔ *Subjective measures*

Typical Protocols applied on Tests

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

Typical Protocols applied on Tests

- Thin
-

Usability Tests

Variants

Test: Thinking aloud

Concrete **protocol** that is usually applied in usability testing

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

- **Very useful in capturing** a wide range of cognitive activities
  - Mental model
  - Terminology
  - Cognitive activities

Test: Constructive interaction or Co-discovery Learning

Variant of *Thinking aloud* protocol

- **Two users** perform tasks together, trying to help each other
- They are **encouraged to explain** what they are thinking about in the meantime
- **It makes more natural** for the test users to verbalize their thoughts during the test

Disadvantages:

- Users can have different learning strategies
- It is required twice number of participants

Test: Coaching method

Variant of test techniques in which **interaction between user and evaluator** is more outstanding

- Participants are **allowed to ask** any system-related questions
- **Evaluator serves as the coach**, conducting the user in the right way while using the interface

- **Focused to discover the information needs of users** in order to provide better design, training and documentation

➔ *Suitable for **inexpert users** and in **qualitative tests***

Test: Retrospective Testing

- 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
  - User can describe what he/she is doing and why
- There exist **analysis tools** for video-recordings
  - ➢ For example **Tobii**
    - ➔ *Allows collecting more information from the user*

Disadvantages:

- It takes at least twice as long with each user

Test: Remote Testing

- Evaluators **are separated in space and/or time** from participants
  - Evaluator cannot observe the testing process directly
  - Participants are usually not in a formal usability laboratory
- There are different types of remote testing
- Specialized tools and a videoconferencing environment are required
- **Remote test types**:
  - **Moderated**. Same-time but different-place
    - Tester can observe the test user's screen through computer network, and may be able to hear what the test user says during the test through speaker
  - **Unmoderated**. Different-time & different-place
    - User's test session are guided and logged
- **Advantages**:
  - **Availability**
  - **Geography**
  - **Speed**

Test: Performance measures

Quantitative nature. Destined to know **how the user *'behaves'*** during test

- Based on a detailed quantification related **to user performance**
- They require careful observations, to collect in a **controlled way and laboratory**

Some examples:

Objective measures related to

Effectiveness and

Efficiency

• **related to successes and failures in task performance**

number of successful completed tasks
number of completed tasks in a certain time
number of completed tasks with or without assistance
number of uncompleted tasks (blockade) ➔ **failure**

• **related to time**

time to complete a task
time consumed in navigation menus
time consumed in online help
time in searching information in the manual
time in errors recovering

• **related to committed errors or assistances received**

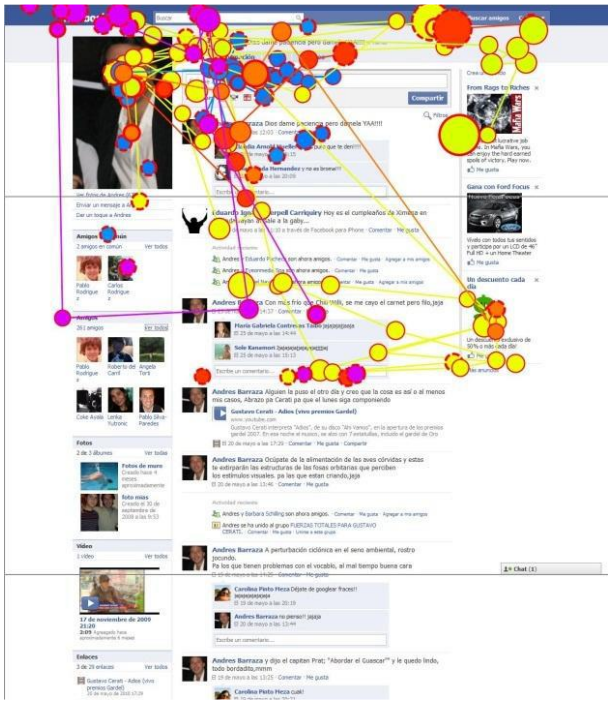
number of wrong menu options
number of wrong dialog box options
number of wrong selection icons
number of functional keys incorrectly selected

number of help calls
number of online help screens
number of consultations to manual

Test: Performance measures

Quantitative nature

- **<u>Some examples related to *Eyetracking*</u>:**
  - **Fixation duration** (*gaze duration*)
  - **Fixation count**
  - **Time to first fixation**
  - **First fixation duration**
  - **Fixations before**



Test: Subjective measures

Qualitative nature. Destined to know **how the user '*feels*' ➔ Satisfaction**

- **Some examples:**

Subjective measures related to **satisfaction**

- Appreciations about
  - usage easy
  - learning easy
  - easy in solving a concrete task
  - Installation easy
  - easy in finding information
  - help usefulness
- Preferences and related reasoning
  - regading a previous version
  - about competitor product
- Behavior predictions
  - product will be bought?
- Spontaneous comments
  - *I was totally lost*
  - *It was easy*
  - *I do not understand the message*

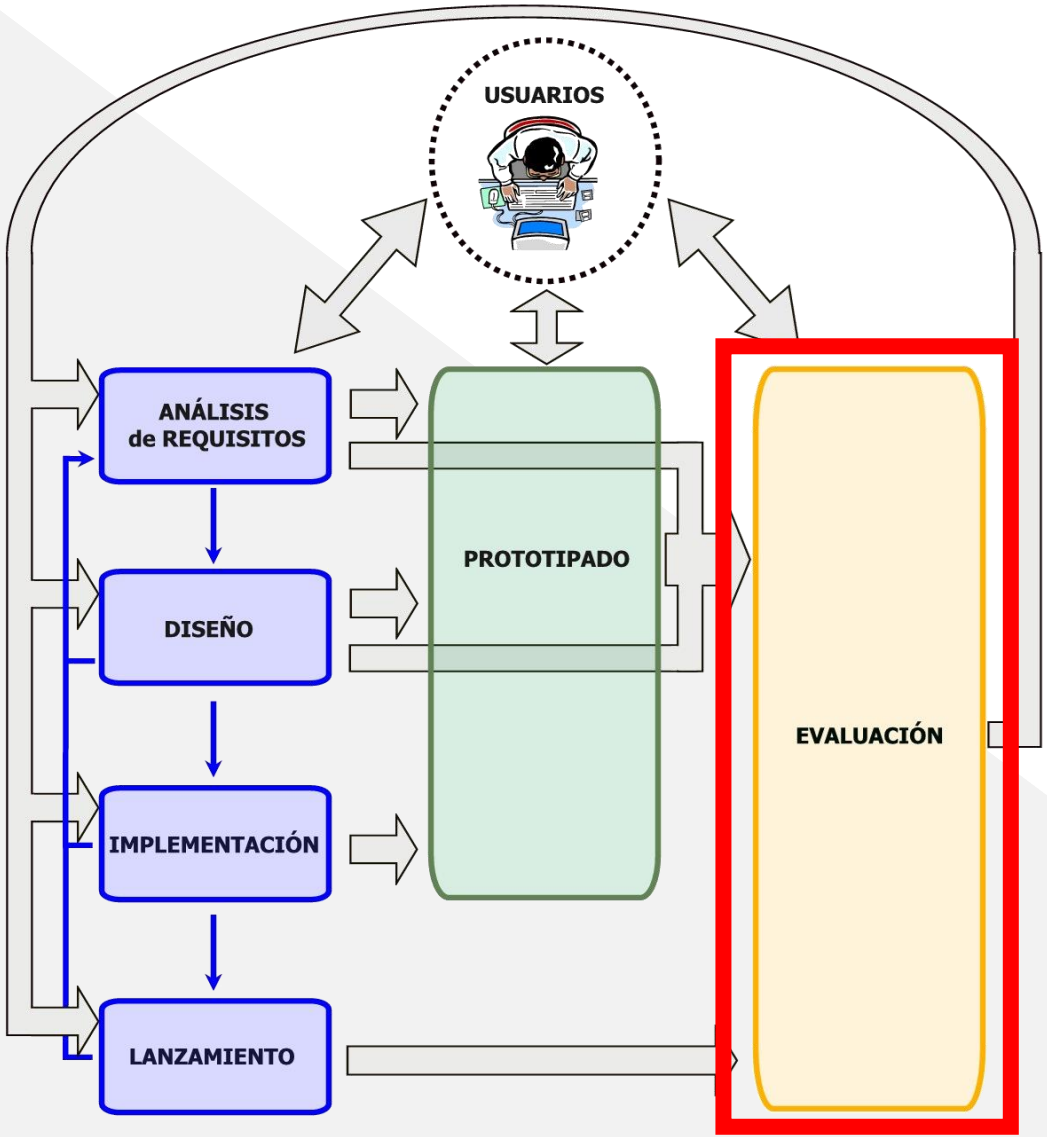
Apart form these,

standardized opinion questionnaires produce **objective** measures regarding **satisfaction**

Contents

- What is the evaluation in the HCI scope?
- Objectives of evaluation and aspects to consider
- Taxonomy in evaluation methods
- **Integration of the evaluation techniques to the development life cycle**
- Conclusions

Integration of the evaluation techniques to the development life cycle



Different techniques in the life cycle

Método	Etapas del ciclo de vida				
	Requisitos	Diseño	Codificación	Test	Despliegue
Recorridos plurales		X			
Chequeo de un sistema de escenario		X	X	X	X
Evaluación heurística		X	X	X	X
Pensando en voz alta		X	X	X	X
Recorrido cognitivo		X	X	X	X
Medida de prestaciones				X	X
Entrevistas		X	X	X	X
Focus group	X			X	X
Cuestionarios				X	X
Observación de campo	X				X
Inspección de estándares				X	X
Grabación del uso				X	X

Contents

- What is the evaluation in the HCI scope?
- Objectives of evaluation and aspects to consider
- Taxonomy in evaluation methods
- Integration of the evaluation techniques to the development life cycle
- Conclusions

Conclusions

- Evaluation is **an essential part in the design of interactive systems** and must be carried out along the whole life cycle
- It **is in charge of** proving not simply the functionality, but the **usability**, to identify and correct problems
- **Evaluation can be done** in the laboratory or in the user workplace, and **it is crucial active and continuous user participation**

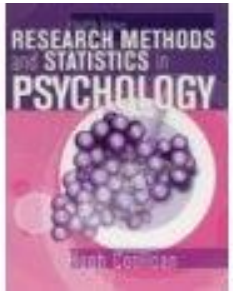
Conclusions

- Evaluation **Advantages**
  - Production costs **reduction**
  - Redesign, maintenance and support costs **reduction**
  - Learning curve for users **reduction**
  - **Product quality improvement**, in particular, **higher user satisfaction**, because it will be easier to use

Bibliography

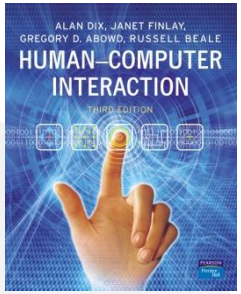
Books

- Hugh Coolican,
  Research Methods and Statistics in Psychology
  Hodder Arnold (2004)



- Toni Granollers, Jesús Lorés, José J. Cañas
  Diseño de sistemas interactivos centrados en el usuario
  Editorial UOC (2005)
  Capítulo 5

- Alan Dix, Janet Finlay, Gregory D. Abowd, R. Beale
  Human-Computer Interaction
  Pearson Prentice Hall – 3ª edició (2004)
  Capítulo 9



Bibliography

- **Support material**
  - Toni Granollers

    Sitio Web MPIu+a: *Modelo de Proceso de la Ingeniería de la usabilidad y de la accesibilidad*

    http://mpiua.invid.udl.cat/category/ipo-hci/evaluacion/tecnicas-de-evaluacion/

- **Other links**
  - No solo usabilidad
    - http://www.nosolousabilidad.com/manual/3.htm
  - Nielsen Norman Group
    - https://www.nngroup.com/articles/
  - Usability Body of Knowledge
    - http://www.usabilitybok.org/usability-evaluation-methods
  - Usability Evaluation
    - http://www.usabilityhome.com/