

EVALUATION TECHNIQUES AND USABILITY TESTING

Theme II – Usability Testing Study (2nd Part – Test Plan Elaboration)

Montserrat Sendín
DIEI - UdL

Contenido

- ◉ Technique definition
- ◉ Test typologies
- ◉ More common protocols
- ◉ Usability laboratories
- ◉ **Test Plan design and elaboration**
 - > Overview
 - > Main structure and documents associated
 - > Other aspects to be considered

Test Plan – Overview

Test plan document

collects all decisions taken by the working group

describes all the elements and **necessary resources** that compose the test

documents all the steps to be carried out before, during and after the test

▫ Why is crucial to elaborate a Test Plan ?

- Allows to **coordinate efforts**
- **Makes easier** for all members **the tracking** of the different phases for the test deployment, and in particular of its staging

Result: Test plan serves as a **communication vehicle** among working group members and also with the client

Indispensable for the success of the test

‘failing to plan is planning to fail’ [C. Myhill, 2014]

Test Plan – Overview

- Test Plan rules of thumb:

- **Work in group:** *progress over the base of **consensus***
- **Prevention better than cure**
- **Revise as much as you can** material generated and also test tracking:
 - it is necessary to **test the test** in advance, in a collaborative way

- **Elaborate a good test plan implies:**

- **Decide** and **state main aspects**
- **Take** some **actions**
- **Prepare** necessary **documentation**

Test Plan – Documents to prepare

- Test Plan main document structure:
 - Product definition
 - Product specification
 - Product state and problems in which to incise
 - **Purpose and test objectives**
 - Test type and technology involved
 - Target audience, user profile and target user sample
 - Participants recruitment process
 - Methodology
 - Procedure and protocol
 - Number of participants
 - Description of the tasks to be evaluated
 - Test environment and equipment
 - Metrics to be handled (data to be registered)
 - List of complementary documents

Test Plan – Documents to prepare

□ Documents to prepare:

- Test Plan main document
- **Annexes:**
 - Screener
 - Questionnaires and/or interviews
 - List of scenarios for tasks
 - Informed consent form about participation and recording session
 - Test calendar
 - Observation document
 - Script for the moderator
 - Additional documents to be used in the test:
 - » Briefing guide
 - » Checklist

Test Plan – Documents to prepare

▫ Test Plan main document structure:

- Product definition
- **Purpose and test objectives**
 - Test type
- Target audience, user profile and target user sample
- Participants recruitment process
- Methodology
 - Procedure and protocol
 - Number of participants
 - Description of the tasks to be evaluated
- Test environment and equipment
- Metrics to be handled (data to be registered)
- List of complementary documents

Test Plan – Setting the purpose and objectives

Test purpose and Objectives

Critical aspect. Do not pretend to evaluate everything related to the product/prototype. It is **impracticable**

*The purpose, as a **general objective**, describes **the client's needs** and/or **development worries** regarding the current system prototype or product, in order to **justify the necessity to carry out a test***

In order to **bound** it, it is ideal to count on **specific problems** from previous activities:

- Field study reports, Focus Group sessions, logs, calls to customer service, polls, competence analysis, etc.

Remember to take into account some aspects, such as:

- Test **realization place** and conditions
- **Who** is interested in the results (client) and their **needs**
- **Budget** and **time** available

Purpose and objectives must be aligned with the client

Test Plan – Setting the purpose

Establishing main Test purpose

A) Informal test

- Nature of data to be registered: **only** qualitative data
- User sample size: no more than **5 users**
- Purpose: Collecting **feedback** for a **general diagnosis**
Diagnose the **UI** (User Interface) **impact** over users
 - Are users able to successfully use all or parts of the system ?
 - Can users successfully navigate through the system ?
 - Is the information logically organized and grouped so that new users can easily locate it ?
 - Possible future market acceptance according to interest observed

B) Formal test (statistical studies)

- Nature of data to be registered: **especially** quantitative data
- User sample size: about **25 users** (if eyetracking about 40 users + reserve)
- Possible purposes:
 - Comparison to a previous version, or to a competence product

Ex: TMB Mobile

Test Plan – Setting the purpose

Establishing main Test purpose

C) Semi-Formal test

Nature of data to be registered: **both**, qualitative and quantitative data

User sample size: between **10** and **20 users**

Different purposes:

1. As in an informal test, collecting **feedback** for a **general diagnosis**

Ex: Chinese HolidayInn.com (**ISO** usability properties), Reciclame (**ISO** usability properties), Janair0 (Efficiency & Engaging & Easy to learn)

2. There is a previous/known concern, so **concrete questions to be addressed** are formulated, which are focused on **specific problems**

- Concern-oriented, problem-oriented or task-oriented

Ex: MealpayPlus (**ISO** usability properties)

- *Is the user able to add more than one child to their profile ?*

- ...

3. A combination of 1. and 2.

Ex: Bookify (the **5Es**), Commed (**ISO** usability properties & Easy to learn)

Test Plan – Setting the purpose and objectives

Test purpose and Objectives

It consists of establishing:

a) **A general objective** (the purpose)

Ex1: Compare a product to the one's from the competence (B)

Ex2: The development is on the good line ? (A)

Ex3: The app is able to have a good acceptance in the market ? (A)

Ex4: Can the user create a new profile easily ? (C.2)

b) **Concrete this general objective in specific and measurable terms** (the objectives)

Ex: Efficiency: collect time metrics and amount of errors committed

Giving thus a response to:

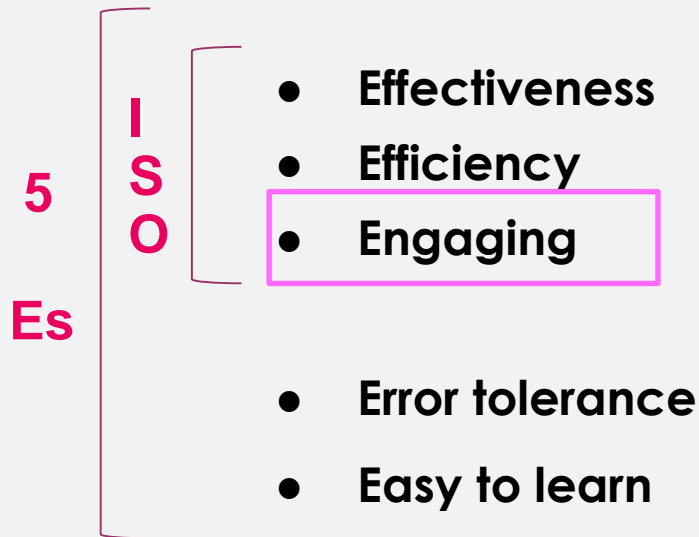
- **What information is going to be collected**
- **How test results are going to be measured → Setting of metrics**
 - Qualitative (observable), quantitative (quantifiable), or both types

Test Plan – Defining objectives

Test objectives

Select any combination of objectives from

- the **three standardized ISO usability properties** or
- the complete **typified usability goals set** (the 5Es):



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 information is going to be obtained ?

Test Plan: Defining objectives - *Effectiveness*

Effectiveness: Capacity to achieve an objective (Ex: complete a concrete task), taking into account accuracy and plenitude with which users achieve it



Successes and failures

Accuracy and completeness with which users achieve specific goals

[9241-11, 08]

- **Task by task:** What we consider successes and failures?
 - *Users achieve the last step and make the contract for purchasing the service ?*
 - *Users achieve the last step, in spite of they do not make the contract ?*
 - *Users are not able to achieve the last step ?*
- **There is a inconsistency** between effectiveness results and the level of satisfaction expressed by participants ?
 - **“False success”**
 - *Have we obtained false successes ?*

Test Plan: Setting of metrics - *Effectiveness*

- **Degrees** of **success** and/or **failures** handled:



- Completeness of a task **successfully in the expected time**
- Completeness of a task **successfully beyond the expected time**
- Completeness of a task correctly **with some help**
- Failure in the completeness of a task **with the user acknowledgment**
- Failure in the completeness of a task **without the user acknowledgment**

➔ “**False success**”

- **Success** and **failures METRICS** typically handled:

- **Degree** of **success** or **failure obtained in each task**
- **Number of assistances** received or number of built-in help accesses
- **Percentage** of successfully completed tasks, % of non completed tasks
- **Completion rate** (also *success rate*)

- **Recommendation:**

- **Prioritize the analysis** of possible **causes** for success and failure

Test Plan: Defining objectives - Efficiency

Efficiency: Measure of if the effort employed to reach established goals is reasonable and fulfills the expectation



They are performance measures related to **costs** and **human effort** (usually **errors** and **time**)

It is required a very **precise observation**

Resources used in relation to the results achieved

[9241-11, 08]

– Some examples:

- *Users are able to choose the right menu option ?*

→ count the number of failed chosen options from menu

- *It is intended to check out the ease for understanding the icons*

→ count the number of failed chosen icons

- *It is intended to check out the ease for finding what users search on built-in help*

→ count the employed time in looking up the built-in help

□ **Recommendation:**

- As they are **relative measures** (they depend on what are compared to), consider to apply **thresholds** and/or **reference measures**

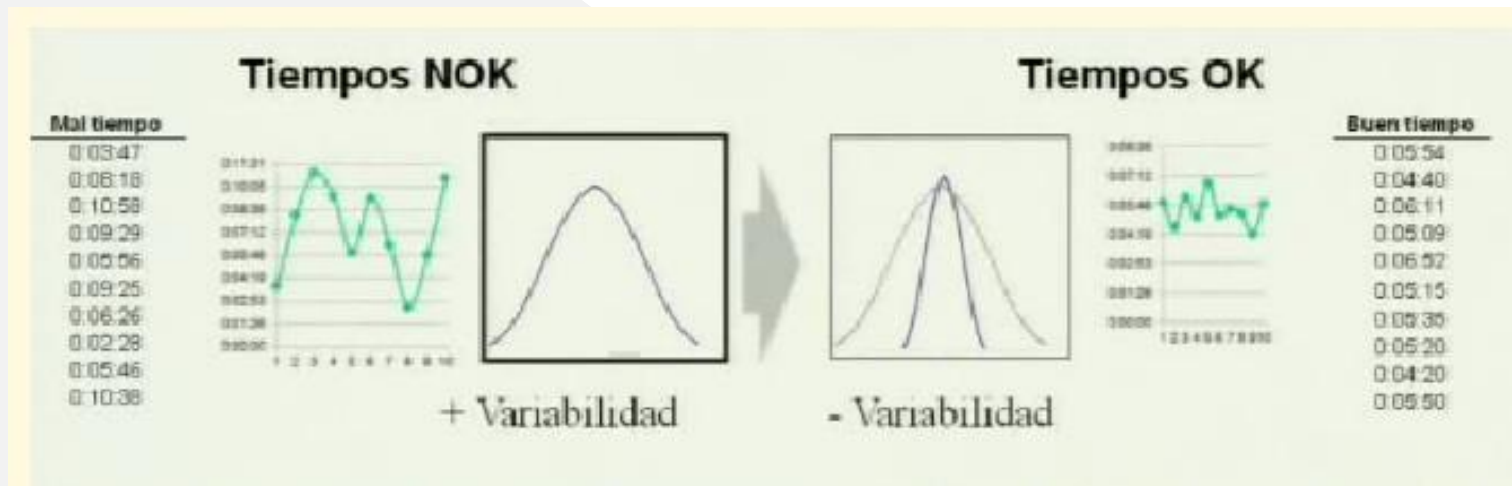
Test Plan: Setting of metrics - *Efficiency*



Time

How long a user spends on doing a task

- *How we measure the time ?*
- *Is the time registered for the task the real time ?*
- *How to interpret inputs regarding time as a whole ?*



Recommendations:

- **Adjust the time** taken to the time actually destined to complete the task
- **Specify** the measure unit used for time
- **The dispersion** in time values needs to be analyzed on the whole

Test Plan: Setting of metrics - Efficiency



Errors

Any unintended action, mistake or omission a user makes while doing a task

Some clarifications must be performed:

- What will be considered an error ?
 - *Deviations with respect to the natural way ?*
 - *After deviations, the own user has led back successfully, or he/she has needed some help ? Is in this case considered an error ?*
- Errors must be categorized
 - **Critical** error (blocking errors) vs **non critical** errors
 - **Occasional** error vs **recurrent** error
- Errors observed in the test correspond with those we were expecting ?

○ **Error METRICS** typically handled:

- **Number of errors**
- **As a binary measure:** the user either encountered an error (1 = yes) or did not (0 = no)

▪ **Recommendations:**

- **Categorize errors** by its level of **criticality** (blocking or not blocking errors)
- Give the necessary relevance to the **frequency** of an error (occasional or recurrent)

Test Plan: Defining objectives - Satisfaction



Satisfaction: Absence of discomfort and positive attitude by the user towards the use of the product

Extent to which the user's physical, cognitive and emotional responses that result from the use of an interactive system meet the user's needs and expectations

[9241-11, 08]

User satisfaction has an **outstanding importance**

Composed by **a)** subjective information (**Qualitative data**), **opinions regarding** satisfaction, but also **is quantified with b) questionnaire responses (Quantitative)**

a-QI) listen and **observe** for things that show if users were satisfied with the experience

- **Qualitative data** and **Post-Tasks** questionnaires



b-Qf) ask questions after the test to register in which extend users liked the product

- **Standardized Post-Test** questionnaires



Satisfaction = Qualitative data + Quantitative data from questionnaires

Test Plan: Defining objectives - Satisfaction



Satisfaction: Absence of discomfort and positive attitude by the user towards the use of the product



Some typical questions:

- Easy to use
 - Easy to navigate
 - Utility
 - Utility of the built-in help and other supports
 - Number of positive aspects that can remember
- **There is a correspondence** between the level of satisfaction expressed by participants and the number of successes and failures (effectiveness results) ?
- “Do you think you could have completed the tasks without any help ?”*

➔ **“False success”**

Setting of metrics

Nature of data to be registered

Cualitativo	Cuantitativo
El objetivo es una descripción completa, detallada.	El objetivo es clasificar rasgos, contarlos, y construir modelos estadísticos en una tentativa de explicar lo que es observado.
El investigador sólo puede saber aproximadamente de antemano lo que él/ella busca	El investigador sabe claramente de antemano lo que él/ella busca
Recomendado durante fases más tempranas de proyectos de investigación	Recomendado durante fases últimas de proyectos de investigación.
El diseño surge cuando el estudio se despliega.	Todos los aspectos del estudio son diseñados con cuidado antes de que los datos sean coleccionados.
El investigador es el instrumento de recopilación de datos.	El investigador usa instrumentos, como cuestionarios o equipos para coleccionar datos numéricos.
Los datos están en la forma de palabras, cuadros u objetos	Los datos están en la forma de números y estadística.
Subjetivo - la interpretación de los acontecimientos de los individuos es importante	Objetivo - busca la medida precisa y el análisis de conceptos objetivos
Los datos cualitativos son 'más ricos', tiempo consumido, y menos capaz de ser generalizado.	Los datos cuantitativos son más eficientes, capaces de probar hipótesis, pero pueden perder el detalle contextual.
El investigador tiende a hacerse subjetivamente sumergido en la materia.	El investigador tiende a permanecer objetivamente separado de la materia.

Setting and Collecting Qualitative data



Measuring satisfaction qualitatively

Qualitative data or “qualitative feedback”: Opinions and subjective appreciations that **strengthen the perception about satisfaction**

– *What is considered qualitative feedback ?*

- **Spontaneous reactions**
- **Subjective appreciations** of users
- **Comparative appreciations** between design proposals
- **Preferences** declarations
- Answers regarding **intention of purchase/contract**
- **Comments** or **representative reactions** (not only during ‘Thinking aloud’)
- **Answers** to post-task questionnaires
- **Answers** to interviews (in informal tests they can substitute satisfaction Q.)
- **Answers** to open questions in questionnaires

– **There is a correspondence between these data and numerical data ?**

➔ **“False success”**

Setting and Collecting Qualitative data



Measuring satisfaction qualitatively 

Qualitative data or “qualitative feedback”

– *When are they obtained ?*

Apart from specifically registered in **post-task questionnaires answers**:

- **Along the task realization**, during the *Thinking aloud*
- **Between tasks** (time space for interchanging of impressions by means of improvised and/or *informal interviews*, and *post-task questionnaires*)
- **Especially interesting at the end of the test**, taking profit of user “post-test” relax (useful for obtained added value information)

Setting and Collecting Qualitative data



Measuring satisfaction qualitatively

Qualitative data or “qualitative feedback”

– How are they obtained ?

- **Video recording**
 - **Register of the “best-moments”** (specific codes on Observer module)
- **Literals transcriptions, verbalizations, comments or reactions**
during the *Thinking aloud* or the *Retrospective Thinking aloud*.
They can also be taken by the observer/s during the test
- **Annotations** during informal interviews by means of the moderator
- **Answers to open questions** in questionnaires

Test Plan – Defining objectives

Test objectives

Select objectives from the standardized **ISO** or the typified usability goals set (the 5Es):

- **Effectiveness** — *Can users realize an order successfully or get registered to obtain a service easily ?*
Can users find the information they need to complete tasks without assistance ?
- **Efficiency** — *Can users realize tasks within a certain period of time ?*
Can users realize an order or sign up for a service without committing errors ?
- **Engaging** — *Do users consider the experience satisfactory ?*
Do users rate their experience as satisfying or enjoyable ?
Users comments or body expression suggest they are having a positive experience ?
- **Error tolerant**
- **Easy to learn**

<http://www.nngroup.com/articles/usability-101-introduction-to-usability/>

Test Plan – Defining objectives – Error tolerance

Test objectives

Select objectives from the standardized ISO or the typified usability goals set (the 5Es):

- The three ISO usability properties

- **Error tolerance:** How well the system reacts to errors (user's and/or system's errors), and **helps** the user recover from mistakes

Reaction: The system **informs** the user and allows **easy error correction/recovering**, while supporting the user as far as possible

*When errors occur, is the system able to **recover** successfully ?*

*Do users **understand error messages** ?*

Prevention: If possible, the product **avoids occurring errors** autonomously

If this is not feasible, **redo** options can help

*System has been **designed** trying to restrict doing the wrong action ?*

*System offers the opportunity to "**redo**", allowing to go back and start again ?*

- Easy to learn

<https://www.interaction-design.org/literature/article/an-introduction-to-usability>

Test Plan – Defining objectives – *Easy to learn*

Test objectives

Select objectives from the standardized ISO or the typified usability goals set (the 5Es):

- The three ISO usability properties

- Error tolerance

- **Easy to learn:** How well the system supports both the **initial orientation** and **continued learning** throughout the complete lifetime of use
The best way to support ease of learning is to design systems that match an existing user's **mental model**

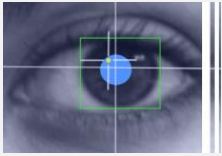
Can users **get started** right away ?

Do users **improve** their abilities as far they **become familiar** with the system ?

Does the system architecture agrees with **users' mental model** ?

<https://www.interaction-design.org/literature/article/an-introduction-to-usability>

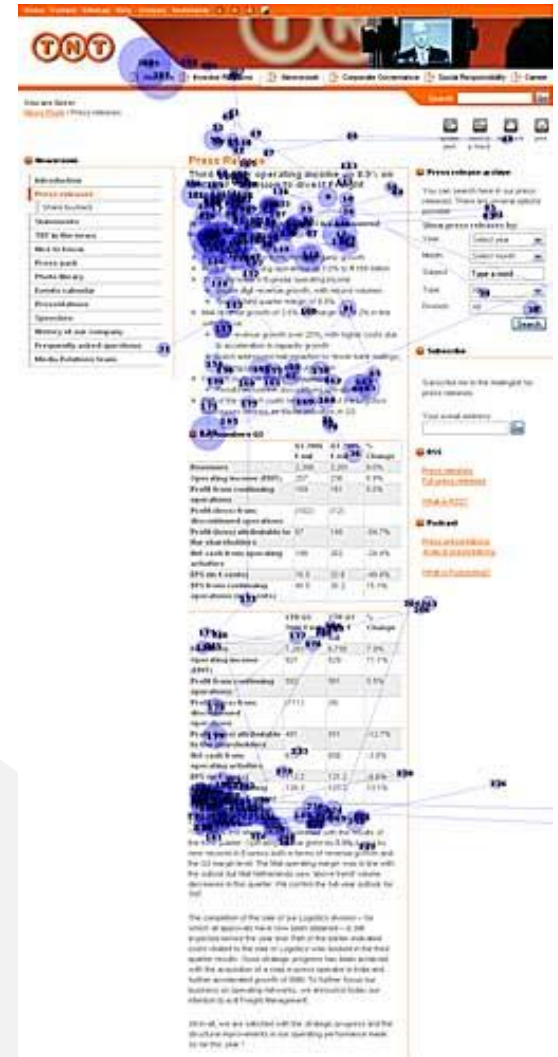
Test Plan: *Specific metrics for eyetracking*



Specific metrics for eyetracking

A tool for determining precisely **how participants moved and focused their gaze**

- **Visual fixation:** maintaining of visual gaze at a single location
- **Duration of fixations**
 - Fixation duration mean
- **Number of fixations** (*fixation count*): amount of visual fixations registered
 - Average fixation count
- **Saccade movements:** quick eye movements that occur between fixations (**saccades**)
- **Saccade path:** spatial graph that shows the sequence of fixations and saccade movements between them



Recommendable: <https://www.youtube.com/watch?v=FrSWaoA-yok>