

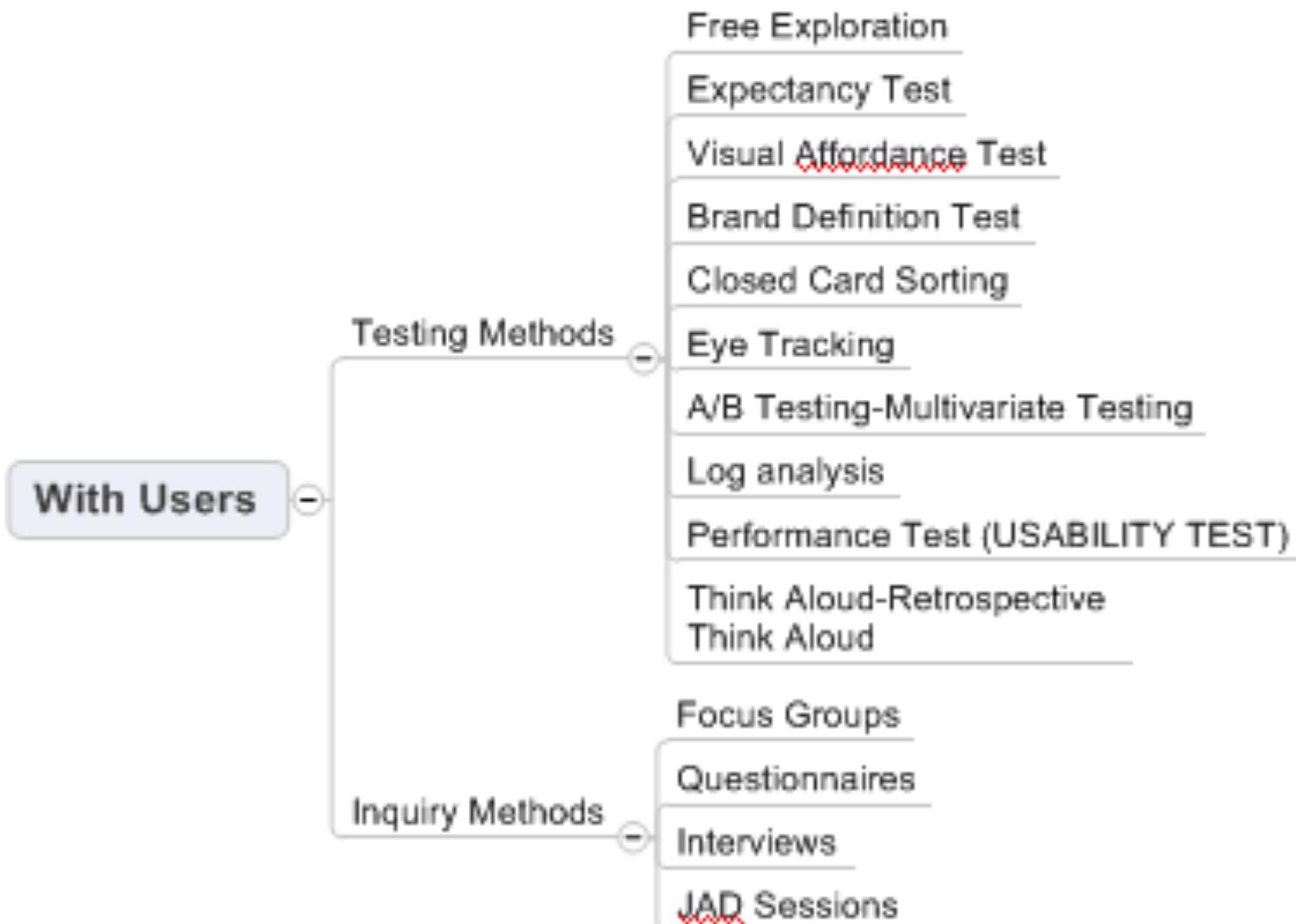
Gestión Calidad Software



GRADO II
ITINERARIO ING. SW
OPT. 4º CURSO 2º C.

UT con usuarios

Técnicas



UT: Test de Rendimiento

Introducción

- Un test de rendimiento es una técnica para testar la usabilidad de un sitio o aplicación en el que el testeador recoge datos empíricos mientras observa o graba a usuarios representativos durante su uso del producto para realizar tareas significativas



UT: Test de Rendimiento

Introducción

- **Facilita:**
 - Comprobar si el diseño funciona
 - Diagnosticar problemas
 - Comparar alternativas
 - Verificar que los objetivos de diseño se han cumplido
- **Beneficios directos:**
 - Proporciona feedback directamente de los usuarios (no interm)
 - Proporciona datos para decisiones de diseño (no opiniones)
 - Ahorra tiempo de desarrollo total (ahorra mucho mantenimiento)
- **Beneficios indirectos:**
 - Da datos para convencer al jefe de que hay que cambiar
 - Sus medidas son la base para cálculos de ROI

UT: Test de Rendimiento

Introducción



¿Por qué es importante realizar un test de rendimiento?



Do users accurately report how they will determine the credibility of a Web site?

No. Users often act in ways that deviate from how they say they will act. It is necessary to study the user's behavior.

Eysenbach, G., & Kohler, C. (2002). "How do consumers search for and appraise health information on the world wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews." *British Medical Journal*.

Tipos de Tests de Rendimiento

UT: Test de Rendimiento

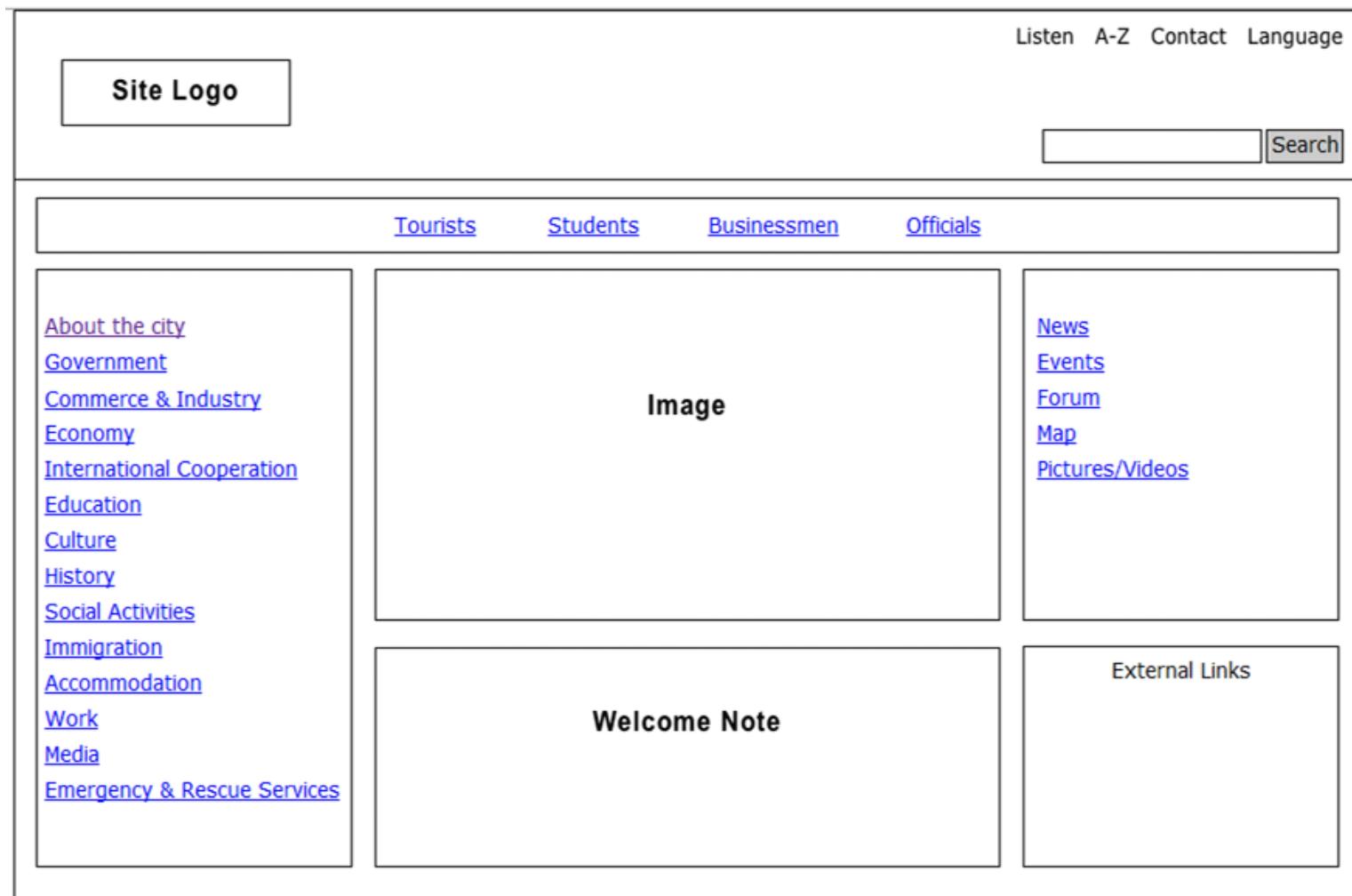
Introducción

- Pueden ser formativos o sumativos, cuantitativos o cualitativos, y ser aplicados a prototipos tempranos o prototipos avanzados
 - Test formativo: encuentra y arregla problemas de usabilidad
 - Test sumativo: cuantifica la usabilidad de la aplicación mediante métricas
- En HCI predominan los tests formativos y cualitativos sobre prototipos avanzados:
 - Muestra pequeña sobre prototipos que arroja resultados cualitativos, es decir, datos en forma de descripciones de problemas y recomendaciones de diseño

UT: Test de Rendimiento

Ejemplo: PT sobre prototipo temprano

- Ejemplo: Testeo de dos prototipos de una web turística realizado por The Media Department en una universidad sueca.
 - <https://www.loop11.com/wireframe-usability-testing/>
- PROTOTIPO 1:



UT: Test de Rendimiento

Ejemplo: PT sobre prototipo temprano

- Ejemplo: Testeo de dos prototipos de una web turística realizado por The Media Department en una universidad sueca.
 - <https://www.loop11.com/wireframe-usability-testing/>
- PROTOTIPO 2:

The wireframe shows a header with 'Site Logo' and navigation links for 'Listen', 'A-Z', 'Contact', and 'Language'. The main content area contains a sidebar with links like 'About the city', 'Government', 'Commerce & Industry', etc., and a central area labeled 'Image'. A search bar and 'Search' button are in the top right. A 'External Links' box is on the right.

Listen A-Z Contact Language

Site Logo

About the city
Government
Commerce & Industry
Economy
International Cooperation
Education
Culture
History
Social Activities
Immigration
Accommodation
Work
Media
News
Events
Forum
Emergency & Rescue Services

Image

Search

External Links

UT: Test de Rendimiento

Ejemplo: PT sobre prototipo temprano

- Estas dos versiones se presentaron a 60 usuarios
- Todos los usuarios realizaron seis tareas: encontrar la lista de eventos, encontrar los mapas de la ciudad, saber más sobre cursos de idiomas, etc. El proyecto Loop11 guardó si las tareas eran completadas de manera exitosa y cuánto tiempo se tardaba en realizar cada tarea.
- Medidas: Ratio compleción de la tarea, Tiempo
- Resultados generales:

Dashboard - Prototype_1

Average task completion rate

58%

■ Success	58%
■ Fail	37%
■ Abandon	4%



Dashboard - Prototype_2

Average task completion rate

51%

■ Success	51%
■ Fail	45%
■ Abandon	4%



UT: Test de Rendimiento

Ejemplo: PT sobre prototipo temprano

- EL PROTOTIPO 1 arrojó mejores resultados: cuatro de seis tareas mostraron sobre el prototipo 1 ratios de compleción mayores que sobre el prototipo 2.
- Sin embargo, la mayoría de esas tareas tomaron bastante más tiempo a los usuarios en el prototipo 1
- Por otro lado, el prototipo 2 sólo ganó en cuanto a ratio de compleción en dos tareas
- Sin embargo, en cinco de las seis tareas el prototipo 2 demostró ser más rápido.
- Por último, sólo una tarea se ejecutó más rápido sobre el prototipo 1 que sobre el prototipo 2: Buscar alojamiento de estudiante.



UT: Test de Rendimiento

Ejemplo: PT sobre prototipo temprano

- ¿Qué haríais vosotros con esos resultados?



UT: Test de Rendimiento

Clasificación

- En Persona/Remoto
- Moderado/No Moderado



¿Modalidad
de estas dos
fotografías?



- Normalmente cuando pensamos en tests de rendimiento pensamos en los **tándems moderado y en persona o no moderado y remoto.**
- **¿Cuándo pensáis que es más conveniente utilizar cada uno?**
 - Aspectos a considerar: tipos de tareas, preguntas de investigación, test formativo/sumativo, presupuesto, inmediatez de los datos, grado de conocimiento de tu usuario objetivo, ...

UT: Test de Rendimiento

Ventajas

En persona + Moderado	Remoto + No moderado
<p>El testeo moderado está especialmente indicado para productos con múltiples facetas o tareas complejas que NO tienen una secuencia estructurada de pasos.</p> <ul style="list-style-type: none">+ Capacidad de observar lenguaje no verbal+ Think aloud protocol+ Probing questions adaptadas a lo que el usuario acaba de hacer (por qué...)+ Pueden tener hasta 10-12 tareas (más o menos 45 min-1hora) sin cansar demasiado al usuario.+ Drop-off muy bajo	<p>El testeo no moderado está especialmente indicado para situaciones en las que tienes preguntas muy específicas acerca de cómo la gente utiliza una interfaz de usuario para tareas relativamente simples y directas.</p> <ul style="list-style-type: none">+ Las sesiones de test pueden incluir cientos de personas simultáneamente, lo que puede permitir justificar de manera muy convincente la importancia de los problemas encontrados.+ El usuario interactúa en su medio natural (por lo que aumenta la validez externa del estudio)+ Se pueden realizar sesiones de test con muchas webs de manera simultánea (e.g. mi web vs web de competidores, de otros países, etc.)+ Los premios para los usuarios de tests no moderados pueden ser mucho menores.+ Algunas herramientas de testeo no moderado pueden reclutar usuarios interceptándolos en tu sitio web. Esto permite recoger información clave sobre la motivación real y el propósito de los visitantes de tu web.+ Es posible recoger datos en tiempo real

UT: Test de Rendimiento

Desventajas

En persona + Moderado	Remoto + No moderado
<ul style="list-style-type: none">- Posibilidad de No-shows o replacements de última hora- Social Desirability Bias: cuando preguntas a alguien directamente dan opiniones más positivas que cuando se usa una encuesta anónima.	<ul style="list-style-type: none">-No pueden ser largos (los tests no moderados deberían tomar entre 15 y 30 minutos máximo, e incluir entre 3 y 5 tareas) porque el porcentaje de tests no respondidos aumenta de manera drástica a partir de dicha duración.-No puedes saber si los participantes están haciendo el test en serio o solo quieren la recompensa, ni si están usando algún tipo de dispositivo periférico.- Los participantes deben tener una conexión a Internet con cierta velocidad (y no liarse con el uso del entorno remoto)-El análisis de datos sin ningún tipo de información cualitativa (e.g. comentarios, expresiones) puede dar una impresión errónea de la experiencia real del usuario, sus necesidades y deseos, mientras navegaba por la web. Por ello, se deben incluir siempre preguntas cualitativas en los estudios, e incluso rastrear después a los usuarios para discutir su feedback.-Lo que los participantes dicen en encuestas puede ser muy distinto a lo que hacen en realidad.- Para continuar con la siguiente tarea, los participantes deben ser capaces de decidir si han completado la tarea actual. Por tanto, es posible que los participantes piensen que ya han completado de manera exitosa una tarea cuando en realidad no lo han hecho. Para evitarlo, es necesario definir tareas directas que tengan estados finales bien definidos.

UT: Test de Rendimiento

Clasificación

- Sin embargo, otras las modalidades son posibles
 - E.g. Test remoto moderado sobre prototipo en papel con GoToMeeting o WebEx o incluso a través del teléfono

The image consists of two side-by-side screenshots. The left screenshot shows a paper prototype of a user interface on a wooden table. The prototype has a white rectangular card with the text "Shuffle Selections?" in a handwritten style. Below it are two small rectangular boxes labeled "YES" and "NO". At the bottom of the card is a small checkbox with the text "set as default". To the right of the card, there are some handwritten initials: XM, SH, CB, and SP. The right screenshot is a screenshot of a computer screen showing a software application titled "Mailings". The interface includes a navigation bar with links for "Home", "Mailings", "Email Lists", and "Reports". Below the navigation is a search bar labeled "Search mailing". A main table lists several items, each with a checkbox, a name, a status, and a date. Red arrows from the paper prototype point to the "Create New Mailing" button at the top of the software interface and the "Search mailing" input field.

UT: Test de Rendimiento

Clasificación: En persona vs remoto

- Consejos para testeo remoto
 - Deshabilita cualquier programa que pueda distraer o interferir
 - Graba la sesión con los dispositivos necesarios (e.g. si moderas por teléfono, utiliza un grabador de teléfono)
 - Utiliza tan pocas herramientas como sea posible
 - Haz copias de seguridad!!!
- Unmoderated remote usability testing: good or evil?
 - <http://www.slideshare.net/usableinterface/unmoderated-remote-usability-testing-good-or-evil-4446847>
 - Nota: el testeo remoto no debería usarse como sustituto para el testeo moderado, pero sí es útil como complemento de tu investigación cualitativa

- Herramientas para testar interfaces móviles en remoto:
 - <https://get.gotoassist.com/es-es>
 - <http://www.airsquirrels.com/reflector>
 - Wiser?
- Nota: tened en cuenta que realizar un testeo de usabilidad remoto y moderado para apps móviles es un poco más complicado que para web o escritorio.

Protocolo Think-aloud

UT: Test de Rendimiento

Protocolo Think-aloud

- Thinking-aloud protocol: ¿QUÉ Y POR QUÉ?
 - Consiste en pedir al usuario que verbalice lo que piensa (o ha pensado) mientras está (estaba) usando el sistema. Esta verbalización permite al evaluador entender cómo ven los usuarios el sistema, y qué partes del diálogo causan los mayores problemas.
 - Ventajas: Algunos usuarios se hacen más conscientes cuando tienen que verbalizar. Los datos vienen acompañados de comentarios explícitos y vívidos que son muy convincentes de cara a los jefes.
 - Inconvenientes: Las mediciones de rendimiento pueden (en función del tipo de protocolo) no ser tan exactas. Además, algunas tareas (e.g. Calcular algo) se llevan mal con el protocolo. Poco natural. Analizar los datos obtenidos es costoso.

UT: Test de Rendimiento

Protocolo Think-aloud

■ Thinking-aloud protocol: Tipos

Concurrent (CTA)



Retrospective (RTA)



Vs.

Un solo usuario (CTA)

Más de un usuario (CI)

■ CTA (Concurrent Think Aloud)

- Variación: Más de un usuario: 'Interacción constructiva' (CI).
 - Dos usuarios usando el sistema juntos.
 - Ventaja: situación de test más natural, hace que los usuarios realicen más comentarios.

Sin apoyo visual

Con apoyo visual (e.c. Vídeos)

■ RTA (Retrospective Think Aloud)

- Ventaja: evita que el hablar interfiera con el rendimiento
- Desventaja: se pierde información. Se puede paliar enseñándole al usuario lo que hizo en cada tarea (vídeo) o incluso una grabación en eyetracking de lo que miró (el usuario recuerda más información).

Número de usuarios de un test de rendimiento cuantitativo

UT: Test de Rendimiento

Número de usuarios

¿Cuántos usuarios pensáis que
deberíais reclutar para vuestro test
de rendimiento?
¿En conjunto o por perfil de
usuario?

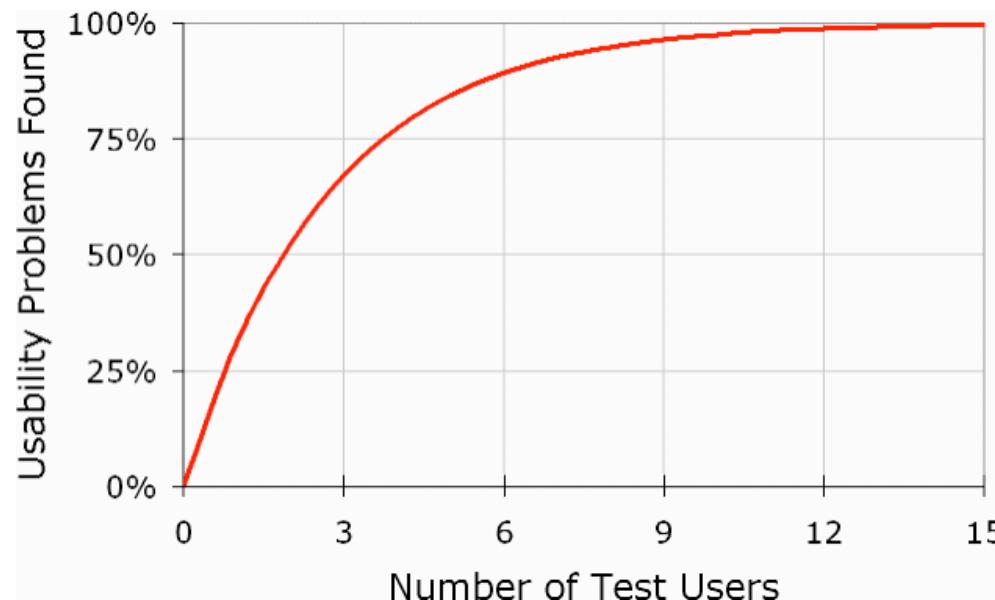


UT: Test de Rendimiento

Número de usuarios

■ Número de usuarios

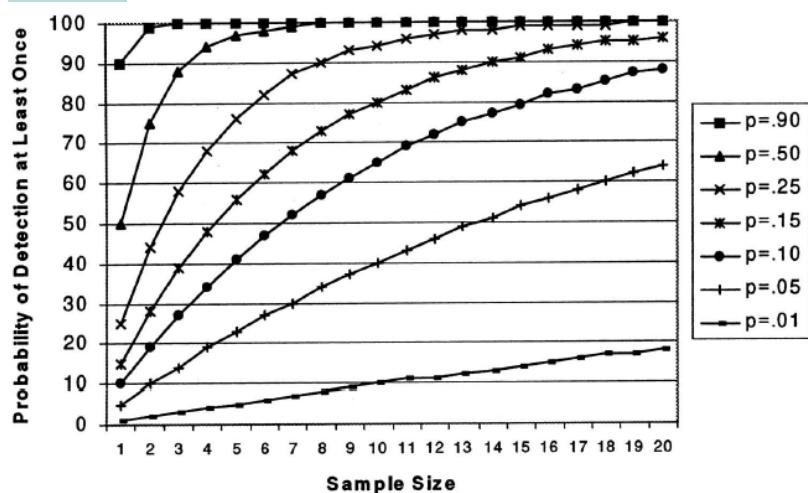
- Cinco usuarios descubrirán aproximadamente el 80% de los problemas de usabilidad de un sitio (Virzil, 1992 y Nielsen, 1993)
- Sin embargo Spool & Schroeder (2001) reportaron que con 5 usuarios sólo encontraron el 35% de los problemas de usabilidad
- Faulkner (2003) encontró que 5 usuarios encontraban alrededor del 85% de los problemas pero con mucha variabilidad (55% - 100%), mientras que 10 usuarios encontraban el 95% de los problemas con rango 85%-100%



UT: Test de Rendimiento

Número de usuarios

- RATIO DE DESCUBRIMIENTO DE PROBLEMAS DE USABILIDAD (Lewis): Dado un método de descubrimiento de problemas M, y un número total de problemas detectables mediante ese método P(M), el ratio de descubrimiento de problemas p se define como (a) la media de la proporción de problemas observado para cada participante o (b) la media de la proporción de participantes que experimentan cada problema observado.
 - El número de problemas detectables depende del método empleado y del estado del prototipo sobre el que se realiza la evaluación (su nivel de evolución)
 - En métodos observacionales, además, depende del número de observadores, la pericia de los observadores, la experiencia de los participantes, los escenarios utilizados
 - En evaluaciones heurísticas, además, depende de la pericia de los evaluadores y el conjunto específico de heurísticas.



Prob de detectar un problema al menos una vez en el test:
$$1 - (1 - \rho)^n$$

ρ =estimación del ratio de descubrimiento de problemas
 n =num usuarios

FIGURE 1 Projected problem-discovery curves as a function of n for various values of p .

UT: Test de Rendimiento

Número de usuarios

- RATIO DE DESCUBRIMIENTO DE PROBLEMAS DE USABILIDAD (Lewis): Ejemplo
 - Test con tres usuarios. Vemos que de las 21 celdas posibles (7 problemas identificados, tres participantes) , 16 tienen una cruz: $p=0.76$ (ratio de descubrimiento de problemas, independiente del número de usuarios)

Table 2: Hypothetical Results for a Problem-Discovery Usability Study: First 3 Participants

Participant	Prob 1	Prob 2	Prob 4	Prob 5	Prob 6	Prob 8	Prob 10	Count	Proportion
1	x	x	x		x	x	x	6	0.86
2	x	x	x		x	x		5	0.71
3	x	x	x	x	x			5	0.71
Count	3	3	3	1	3	2	1	16	
Proportion	1.00	1.00	1.00	0.33	1.00	0.67	0.33		0.76

Note. Prob = problem; x = specified participant experienced specified problem.

- Con este ratio, el investigador puede ajustar el número de usuarios necesarios para un determinado umbral de probabilidad de detectar un problema al menos una vez en el test
 - $1-(1-0.76)^3 = 1-(0.24)^3 = 1-0.014=0.986$
 - $1-(1-0.76)^4 = 1-(0.24)^4 = 1-0.0033=0.9966$

UT: Test de Rendimiento

Número de usuarios

- Vemos que un ratio de descubrimiento de problemas de 0.76 es altísimo. Veamos ahora qué ocurre con diez participantes

Table 1: Hypothetical Results for a Problem-Discovery Usability Study

Participant	Prob 1	Prob 2	Prob 3	Prob 4	Prob 5	Prob 6	Prob 7	Prob 8	Prob 9	Prob 10	Count	Proportion
1	x	x		x		x		x		x	6	0.6
2	x	x		x		x		x			5	0.5
3	x	x		x	x	x					5	0.5
4	x	x		x			x				4	0.4
5	x	x	x	x		x			x		6	0.6
6	x	x	x					x			4	0.4
7	x	x	x		x						4	0.4
8	x	x	x		x		x				5	0.5
9	x		x		x		x		x		5	0.5
10	x		x		x		x		x	x	6	0.6
Count	10	8	6	5	5	4	4	3	3	2	50	
Proportion	1.0	0.8	0.6	0.5	0.5	0.4	0.4	0.3	0.3	0.2		0.50

Note. Prob = problem; x = specified participant experienced specified problem.

- En realidad $p=0.5$ (cincuenta celdas cubiertas de las 100 posibles), por lo que con tres usuarios la estimación ($p=0.76$) estaba claramente sobredimensionada en más de un 50%
 - Con el nuevo p , la probabilidad de detectar un problema con 3 usuarios se puede estimar como $1-(1-0.5)^3 = 1-0.125 = 0.875$
 - Incluso este valor es demasiado alto (en realidad habían descubierto un 60%)

UT: Test de Rendimiento

Número de usuarios

- El problema es la sobreestimación de p para pequeñas muestras.
- Para solventarlo, se aplica una fórmula que consiste en una combinación de normalización y estimador de Good-Turing

$$truep = \frac{1}{2}[(estp - 1/n)(1 - 1/n)] + \frac{1}{2}[estp / (1 + GT_{adj})]$$

- $GT_{adj} = E(N_1)/N$, es el ajuste al espacio de probabilidad, que debe entenderse como la proporción del número de problemas que han sido detectados por un solo usuario/número de problemas diferentes detectados.

UT: Test de Rendimiento

Número de usuarios

Table 10.4 Usability Problems by Participant

Problem Number	Description	Task	P1	P2	P3	P4	P5	P6
1	System did not recognize zip code.	1		2				
2	Asked for store rather than service location.	1			1	1		
3	Phone numbers play a little fast.	1	4	4			4	
4	Had to repeat phone number.	2		3				
5	Picked store from list, didn't wait for prompt to request zip code.	2			4			
6	Experienced "Was that <silence>?"	2			3			
7	System did not recognize an ungrammatical utterance.	2				3		
8	Tried to enter zip with keypad.	2				3		
9	Went back to main menu to order accessory.	3		3				
10	Participant said entire model number, then rejected 4-digit "Was that" response.	3				3		
11	A main menu option was confusing.	4	3	3	3	3	3	
12	Participant said ungrammatical variant of prompted option, had to listen to help and retry.	4				3	3	

$$truep = \frac{1}{2} [(estp - 1/n)(1 - 1/n)] + \frac{1}{2} [estp / (1 + GTadj)]$$

- Problemas detectados durante el testeo de cuatro tareas (columna Task)
- 1,2,3,4 indican impacto(grave,moderado,leve,menor)
- Estp (ratio observado de descubrimiento de problemas) = 20/72 = 0.278 (ver columnas P1..P6)
- n (participantes) = 6
- $GT_{adj} = E(N_1)/N = 8/12 = 0,66$
- truep (formula)=0,134
- Probl descubiertos=1-(1-0,134)⁶ = 1-0,422= 57,8% de problemas descubiertos
- Número de sujetos necesarios para descubrir un 95% de problemas : 21 (habría que seguir aumentando n hasta que $(1-0,134)^n < 0.05$)
- ¿Nº total de problemas que podemos esperar que haya en la aplicación?
 - Si 12 son un 57.8, un 100 serían aproximadamente 21 problemas.

UT: Test de Rendimiento

Ejemplo: PT sobre prototipo temprano

- ¿Qué indica el bajo ratio de descubrimiento de problemas? ¿Y si lo combinamos con la baja severidad de la mayoría de los mismos (números 3 y 4 de la tabla?)



UT: Test de Rendimiento

Número de usuarios

- Con este método, mientras más usuarios se utilicen para calcular el p inicial, más exacto será el número de participantes calculado para asegurar cierto nivel de cobertura de errores.
 - En Lewis (2001), el error cuando se estimó p a partir de 6 participantes nunca superó 1 participante.
- En la práctica, se puede obtener un tamaño de muestra ajustado para un objetivo de descubrimiento de problemas entre el 70 y el 95% haciendo una estimación inicial de la muestra tras 2 tests, y luego ajustando el estimador tras obtener datos de otros dos (total de 4 participantes)

Roles

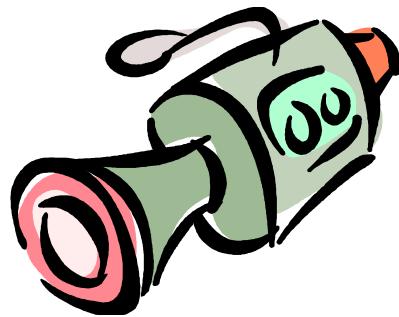
UT: Test de Rendimiento

Roles

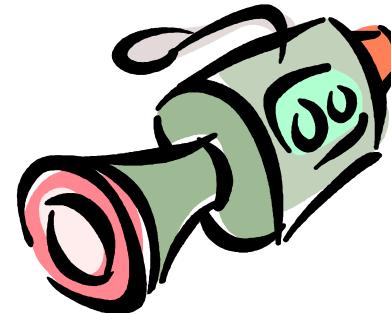
- Facilitador: supervisor de todo el proceso: planificación, testeo y reporte de resultados
- Participante: usuario; cliente actual o potencial. No se deberían usar representantes de usuario
- Observador: persona que graba los eventos según van ocurriendo. Tiene una interacción limitada con el cliente/usuario. Contribuye al reporte de resultados.

UT: Test de Rendimiento

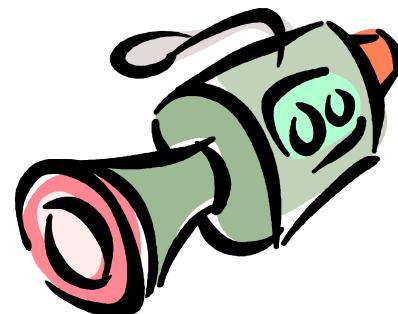
Ejemplo



Performance test sobre Paper prototype:
Daikin (En castellano)



Performance test over Paper prototype: WebPoint (English)
<http://www.youtube.com/watch?v=ppnRQD06ggY>



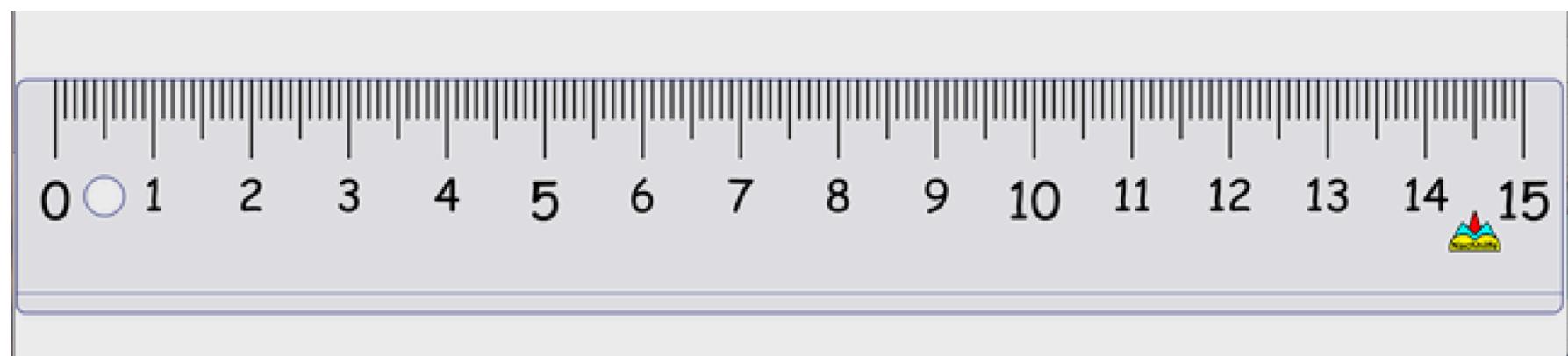
Performance test over Paper prototype: Application for kids
(Friendly Animals English) (7:40)
<http://www.youtube.com/watch?v=9wQkLthhHKA>

Medidas de Usabilidad

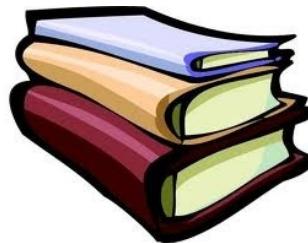
UT: Test de Rendimiento

Medidas de usabilidad: Intro

- El tipo de test más adecuado y su estructura viene en parte determinado por el tipo de información que se desea obtener
- ¿Qué es importante medir para mi proyecto?



- Además de la ISO 25010, se han propuesto varios modelos de usabilidad que caracterizan el concepto de usabilidad e incluyen medidas
 - McGee, Rich and Dumas 2004: Understanding the usability Construct
 - QUIM Seffah, Donyaee, Kline and Padda: Usability Mesurement and Metrics, a consolidated model
 - Winter, Wagner and Deissenboeck (2008): A comprehensive model of usability



Seffah, Donyaee, Kline and Padda (2006):
Usability Mesurement and Metrics, a consolidated model

Resumid los principales modelos de usabilidad propuestos
y sus principales características

- QUIM (Seffah et al. 2006)
 - 10 factores correlacionados: efficiency, effectiveness, productivity, satisfaction, Learnability, Safety, Trustfulness, Accessibility, Universality, usefulness
 - Puede ser extendido
 - Cada factor dividido en criterios medibles (sub-factores). E.g. time behaviour, operability, user guidance, consistency, ... donde cada criterio medible puede afectar a más de un factor (de ahí la correlación de los factores)
 - Cada criterio lleva asociada al menos una medida (en total 127)
 - Las medidas se clasifican en medidas de testeo o predictivas (de diseño), en función de la fase del ciclo de vida en la que se apliquen.

UT: Test de Rendimiento

Medidas de usabilidad: Modelos de calidad

- QUIM: Factores y criterios medibles relacionados (un criterio medible puede afectar más de un factor)

Table 4 Relations between factors and criteria in QUIM.

Criteria	Factors									
	Efficiency	Effectiveness	Satisfaction	Productivity	Learnability	Safety	Trustfulness	Accessibility	Universality	Usefulness
Time behavior	+			+						
Resource utilization	+			+						+
Attractiveness			+						+	
Likeability			+							
Flexibility		+	+				+	+	+	+
Minimal action	+		+		+		+	+		
Minimal memory load	+		+		+		+	+		
Operability	+		+			+	+	+		
User guidance			+		+		+	+		
Consistency		+			+		+	+		
Self-descriptiveness					+		+	+		
Feedback	+	+							+	+
Accuracy		+				+				+
Completeness		+				+				
Fault-tolerance						+	+			
Resource safety						+				+
Readability							+	+		
Controllability							+	+		
Navigability	+	+					+	+		
Simplicity					+		+	+		
Privacy						+	+	+		
Security						+	+			
Insurance						+	+			
Familiarity						+	+			
Loading time	+			+				+	+	

UT: Test de Rendimiento

Medidas de usabilidad: Modelos de calidad

■ Ejemplos de medidas:

Metric	Definition
Essential Efficiency (EE; Constantine and Lockwood, 1999) Essential Efficiency (EE; Constantine and Lockwood, 1999) Estimates how closely a given user interface design approximates the ideal expressed in the use case model	$EE = 100 \times S_{\text{essential}} / S_{\text{enacted}}$ $S_{\text{enacted}} S_{\text{essential}} = \text{The number of user steps in the essential use case narrative}$, $S_{\text{enacted}} = \text{The number of steps needed to perform the use case with the user interface design (rules for counting the number of enacted steps has come in the reference)}$
Layout Appropriateness (LA; Sears, 1995), Favors arrangements where visual components that are most frequently used in succession are closer together, reducing the expected time (cost) of completing a mix of tasks	$LA = 100 \times C_{\text{optimal}} / C_{\text{designed}}$, $C = \sum P_{i,j} \times D_{i,j} V_{i,j}$ $P_{i,j} = \text{Frequency of transition between visual components } i \text{ and } j$ $D_{i,j} = \text{Distance between visual components } i \text{ and } j$
Task Concordance (TC; Constantine and Lockwood, 1999), Measures how well the expected frequencies of tasks match their difficulty, favors a design where more frequent tasks easier are made easier (e.g., fewer steps)	$TC = 100 \times D/P$ $P = N(N - 1)/2$ $N = \text{The number of tasks being ranked}$, $D = \text{Discordance score, i.e., the number of pairs of tasks whose difficulties are in the right order minus those pairs whose difficulties are not in right order}$
Task Visibility (TV; Constantine and Lockwood, 1999),	$TV = 100 \times (1/S_{\text{total}} \times \sum V_i) V_i$ $S_{\text{total}} = \text{Total number of enacted steps to complete the use case}$ $V_i = \text{Feature visibility (0 or 1) of enacted step } i \text{ (i.e., how to count enacted steps and allocate a visibility value to them is defined by some rules in the reference)}$
The proportion of interface objects or elements necessary to complete a task that are visible to the user Horizontal or Vertical Balance (Sears, 1995). These metrics evaluate how well balanced the screen is both vertically and horizontally (a score of 100 indicates perfect balance)	$\text{Balance} = 200 \times W_1 / (W_1 + W_2)$ $W_1 = \text{Weight of side one}$ $W_2 = \text{Weight of side two}$ $\text{Weight of a side} = \text{Number of pixels used} \times \text{side's distance from the center}$ $\text{Center} = \text{Halfway between the left edge of the left-most visual element and the right edge of the right-most element}$

UT: Test de Rendimiento

Medidas de usabilidad: Tipos de medida

- Medidas típicas de un test de usabilidad (Sauro & Lewis 2009):
 - Tiempo de tarea (99% estudios): eficiencia
 - Ratio de compleción (98% studies): eficacia (completeness)
 - Errores (58%): eficacia (accuracy)
 - Satisfacción post-test (48%): satisfacción
 - Satisfacción post-tarea (40%): satisfacción
- En algunos casos se utilizan medidas combinadas que intenta resumir la usabilidad de la aplicación, aunque suele ser más útil considerar cada aspecto de la usabilidad por separado.
- Fijaos en la visión CUANTITATIVA que subyace en un test de rendimiento

UT: Test de Rendimiento

Medidas de usabilidad: Tipos de medida

- Sin embargo, datos obtenibles de dos tipos:
 - **Cuantitativos:** tiempo, clicks de ratón, errores, ...
 - Medidas de éxito: número de tareas completadas, número de pasos exitosos dentro de cada tarea
 - Medidas de error: número de paths erróneos intentados, número de reintentos, tasa de errores por tarea
 - Medidas de eficiencia: tiempo de compleción de la tarea, número de pasos requeridos, número de tareas realizadas en una unidad de tiempo
 - Medidas de satisfacción: satisfacción a nivel de tarea, satisfacción a nivel de test
 - Medidas de acceso a ayuda
 - PERO TAMBIÉN **cualitativos:** impresiones acerca de cómo se han sentido los usuarios, listas de problemas de usabilidad, incluyendo frecuencia y severidad, etc.
 - Esta visión se ha fortalecido con la aparición del concepto del UX

- Recopilación de medidas:
 - Modelo QUIM (127 medidas): Seffah, Donyaee, Kline & Padda. Usability measurement and Measures: a consolidated model.
 - **Hornbaek: Current practice in measuring usability: challenges to usability studies and research: 54 tipos de medida**
 - En gran parte vuestra selección de medidas dependerá del tipo de test de rendimiento (en persona/remoto, moderado/no moderado, formativo/sumativo, sobre prototipos tempranos/avanzados, ...) y de las herramientas de apoyo de las que dispongáis.
 - VEAMOS ALGUNAS DE ELLAS...

UT: Test de Rendimiento

Medidas de usabilidad: Tipos de medida

- Tiempo de tarea (eficiencia): cuánto tiempo pasa un usuario en una actividad
 - Tiempo medio de compleción de tarea: Tiempo de los usuarios que completaron con éxito la tarea
 - Tiempo hasta fallo: Tiempo medio que los usuarios que fallaron pasaron hasta que abandonaron o completaron la tarea de manera incorrecta.
 - Tiempo total de tarea: Duración media de todos los usuarios realizando la tarea.



UT: Test de Rendimiento

Medidas de usabilidad: Tipos de medida

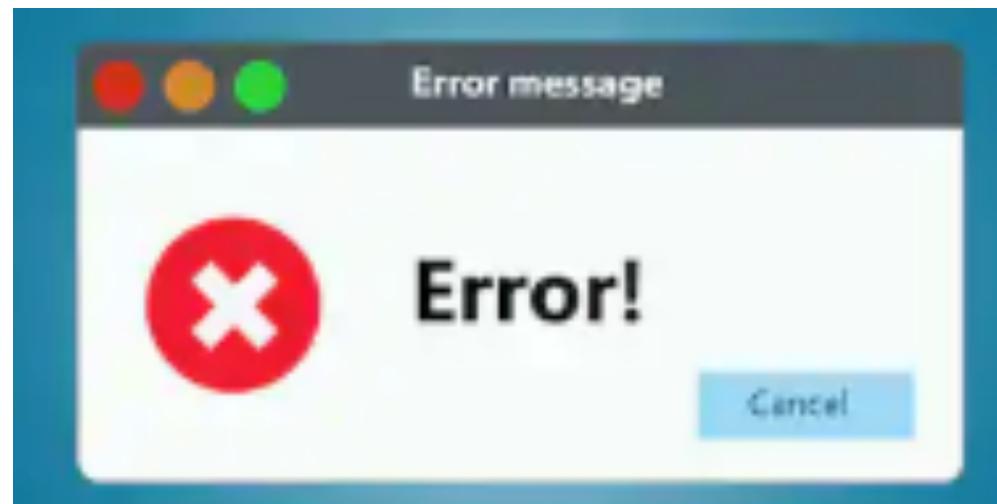
- Ratios de compleción (ratios de éxito: eficacia).
 - Tareas muy sencillas: Medida binaria: 1 (éxito), 0 (fracaso). E.g. Checkout en una
 - E.g. Si de 10 usuarios, ocho completan la tarea de manera exitosa, ratio de éxito = 0.8
 - Tareas más complejas: Escala ordinal. E.g. Completar una lección en una app de aprendizaje de idiomas.
 - En función de la tarea, los investigadores pueden también definir criterios para éxitos parciales, e.g. Éxito usando la ayuda, éxito tras más de un intento,
- ...



UT: Test de Rendimiento

Medidas de usabilidad: Tipos de medida

- Errores: acciones no intencionadas, despistes, fallos u omisiones que el usuario comete mientras está intentando realizar una tarea (eficacia)
 - Medida de tipo ratio, de 0 a infinito
 - Medida binaria: 1 (el usuario tuvo algún fallo) or 0 (no tuvo ningún fallo)



UT: Test de Rendimiento

Medidas de usabilidad: Tipos de medida

- Ratios de satisfacción (satisfacción): escalas que miden la percepción de la facilidad de uso de un sistema
 - Cuestionarios Post-task
 - Cuestionarios Post-test
 - Cuestionarios Off-test



UT: Test de Rendimiento

Medidas de usabilidad: Tipos de medida

- Aunque no son tan comunes, también se pueden incluir medidas de eficiencia y eficacia SUBJETIVA (basadas en opiniones de los usuarios acerca de su eficiencia/eficacia)
 - Esta visión subjetiva es a menudo tan importante como los datos objetivos
- Del mismo modo, se pueden incluir medidas OBJETIVAS de satisfacción
 - ¿Se os ocurren cuáles?
 - Observación de la postura del cuerpo, reconocimiento de expresiones faciales, sensores de sudoración para detectar estrés, ...

UT: Test de Rendimiento

Medidas de usabilidad: Tipos de medida

- Las medidas más apropiadas pueden variar en función del objetivo del test

Table 3.1 Ten Common Usability Study Scenarios and Their Most Appropriate Metrics

Usability Study Scenario	Task Success	Task Time	Errors	Efficiency	Learnability	Issues-Based Metrics	Self-Reported Metrics	Behavioral and Physiological Metrics	Combined and Comparative Metrics	Live Website Metrics	Card-Sorting Data
1. Completing a transaction	X			X		X	X				X
2. Comparing products	X			X			X			X	
3. Evaluating frequent use of the same product	X	X		X	X		X				
4. Evaluating navigation and/or information architecture	X		X	X							X
5. Increasing awareness							X	X			X
6. Problem discovery						X	X				
7. Maximizing usability for a critical product	X		X	X							
8. Creating an overall positive user experience							X	X			
9. Evaluating the impact of subtle changes											X
10. Comparing alternative designs	X	X				X	X				

UT: Test de Rendimiento

Medidas de usabilidad: Medidas agregadas

- Algunos autores proponen combinar medidas en una medida global de usabilidad
 - SUM: Single Usability Measure [Kindlburg and Sauro 2005]: derived from completion, time, error and satisfaction
 - MUS: Master Usability Scaling [McGee 2004]: derived from subjective perceptions

- Sin embargo, esta práctica está desaconsejada por otros autores, basándose en metaanálisis:
 - Hornbaek 2007: Meta-analysis of correlations among usability measures
 - Medidas de eficiencia, efectividad y satisfacción están poco relacionadas
 - Esto es independiente de la complejidad de la tarea (en contra de la opinión de algunos autores que decían que la correlación disminuía según la tarea se hacía más compleja)
 - Sin embargo la correlación sí disminuye a medida que las medidas utilizadas para medir estos constructos se hacen más complejas.

Entornos de testeo De usabilidad

UT: Test de Rendimiento

UTEs

- Existen muchas herramientas que proporcionan información sobre la actividad del usuario en tu web, pero no todas se pueden incluir en la categoría de Usability Testing Environments (UTE's)
 - Herramientas de análisis web: crazyegg, clickdensity
 - Herramientas de análisis web con capacidades de vídeo: Userfly, clicktale, Tealeaf, Clixpy
 - Ninguna de estas herramientas permite definir tareas que guíen el test.
 - Info típica que debería capturar un UTE incluye task-completion rate, time on task, time on page, clickstream paths, satisfaction ratings or opinion rankings, Web analytics data—such as browser, operating system, and screen resolution

UT: Test de Rendimiento UTEs

	Moderado (síncrono)	No Moderado (asíncrono)
En persona	Morae (\$\$) UTE Observer (Noldus) Ovo studios	
Remoto		Loop11(\$) Keynote Web Effective (\$\$\$ \$) UserVue UserZoom (\$\$\$\$\$) RelevantView(\$\$\$\$) Webnographer(\$\$\$\$) OpenHallway(\$) UserTesting.com(\$) EasyUsability.com(\$) Usabilla(\$) Chalkmark(\$) Treejack(\$)

Estructura genérica de un Test de Rendimiento

■ Estructura genérica de un Test de Rendimiento

- Formulario de confidencialidad y consentimiento informado (Nondisclosure/Informed consent form)
- Cuestionario de antecedentes (Background questionnaire)
- Cuestionarios pre-test (e.g. Primera impresión)
- Escenarios y cuestionarios de tarea
- Cuestionarios post-test
- Cuestionario de aclaraciones (debriefing)
- Otros artefactos

UT: Test de Rendimiento

Formulario de consentimiento informado

- Formulario de confidencialidad y consentimiento informado incluye:
 - Acuerdo de confidencialidad de información de producto propietaria
 - Permiso de grabación
 - Permiso para utilizar los datos de manera agregada
 - Guías/regulaciones legales
 - <http://www.socialpsychology.org/consent.html>



UT: Test de Rendimiento

Formulario de consentimiento informado

- Ejemplos: OHIM (papel), HUT (papel), SMonkey (online)
- Más plantillas:
 - http://www.who.int/rpc/research_ethics/informed Consent/en/

Consent Form

The purpose of this research project is....This is a research project being conducted by..... at SurveyMonkey University. You are invited to participate in this research project because you are.....

Your participation in this research study is voluntary. You may choose not to participate. If you decide to participate in this research survey, you may withdraw at any time. If you decide not to participate in this study or if you withdraw from participating at any time, you will not be penalized.

The procedure involves filling an online survey that will take approximately 30 minutes. Your responses will be confidential and we do not collect identifying information such as your name, email address or IP address. The survey questions will be about.....

We will do our best to keep your information confidential. All data is stored in a password protected electronic format. To help protect your confidentiality, the surveys will not contain information that will personally identify you. The results of this study will be used for scholarly purposes only and may be shared with SurveyMonkey University representatives.

If you have any questions about the research study, please contact....This research has been reviewed according to SurveyMonkey University IRB procedures for research involving human subjects.

ELECTRONIC CONSENT: Please select your choice below.

Clicking on the "agree" button below indicates that:

- you have ready the above information
- you voluntarily agree to participate
- you are at least 18 years of age

If you do not wish to participate in the research study, please decline participation by clicking on the "disagree" button.

agree
 disagree

UT: Test de Rendimiento

Cuestionario de experiencia previa

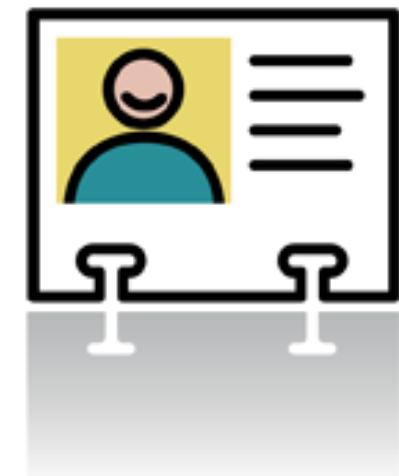
- Experiencias del participante, actitudes y preferencias
 - Datos que complementan los que ya se recogieron durante el proceso de screening (selección de participantes).



UT: Test de Rendimiento

Cuestionario pre-test

- Permite entender el rendimiento, confirmar que los sujetos tienen el perfil adecuado y ayudar al observador
- Trata objetivos de test específicos:
 - Descubre actitudes y primeras impresiones
 - Cualifica a los participantes para su inclusión en un determinado grupo
 - Establece el conocimiento que el participante debe tener como pre-requisito.



UT: Test de Rendimiento

Cuestionario pre-test

■ Ejemplo: OHIM

The screenshot shows the homepage of the OHIM (Office for Harmonization in the Internal Market) website. The header includes the OHIM logo, a European Union flag, and links for Home, Trade Marks, Designs, Quality plus, About OHIM, and Observatory. A banner at the top right promotes a conference: "8 - 9 april 2013 - Alicante - Spain" and "10 years community design ADDA - Auditorio de la Diputación de Alicante". The left sidebar features sections for Legal texts (Trade marks, Designs), Case-Law (Trade marks, Designs), and Most Requested (Press Room, SMEs and newcomers section, Newsletter: Alicante News, Public procurement, How to obtain protection for the whole of the European Union, Fees and payments (CTM), Fees and payment (RCD), The registration process - Classification (CTM)). The main content area displays news items: "Coming up this week in Luxembourg" (January 21, 2013), "Bulgaria database harmonisation" (January 14, 2013), and "Coming up this week in Luxembourg" (January 14, 2013). The right sidebar contains links for eSearch Plus (Beta), Databases, Forms/Filings, and MyPage.

UT: Test de Rendimiento

Cuestionario pre-test

■ Ejemplo: Norman Nielsen

Participant Background

Participant name:

Intranet/Company:

City/State:

Date and time:

Facilitator(s):

1. How long have you been working at this company?
2. What is your job title? What department do you work in?
3. Have you ever used the company intranet?
 - a. If so, what kind of things have you used it for?
 - b. If so, about how often do you use the company intranet?
 - c. If so, when was the last time you used the company intranet?
4. For how long have you been using the Web (the Internet, not the company intranet)?
5. About how frequently do you go online?
6. What kinds of things do you do online?

UT: Test de Rendimiento

Escenarios de tarea

- Escenario de tarea: representación realista de las actividades que estamos razonablemente seguros que los participantes querrían realizar con nuestro producto



UT: Test de Rendimiento

Escenarios de tarea

Ejemplo:

- Vives en Alicante, pero tu novia vive en Madrid. Quieres sorprenderla, y comprar un billete de tren de ida y vuelta que salga de Alicante el próximo viernes después de las 6 p.m y vuelva el Domingo por la tarde. Tu presupuesto es ajustado, así que te gustaría comprar el ticket más barato posible que cumpla tus condiciones de horario.



- Encuentra un ticket de ida y vuelta a Madrid

UT: Test de Rendimiento

Escenarios de tarea

Ejemplo:

- Vives en Alicante, pero tu novia vive en Madrid. Quieres sorprenderla, y comprar un billete de tren de ida y vuelta que salga de Alicante el próximo viernes después de las 6 p.m y vuelva el Domingo por la tarde. Tu presupuesto es ajustado, así que te gustaría comprar el ticket más barato posible que cumpla tus condiciones de horario.



- Encuentra un ticket de ida y vuelta a Madrid



UT: Test de Rendimiento

Escenarios de tarea

- Existen distintas opiniones sobre la longitud y detalle que deberían tener las tareas en un test de usabilidad
- Mi consejo: los escenarios de tarea deberían incluir...
 - El resultado que el participante debe conseguir
 - Motivo plausible para querer realizar la tarea
 - Datos reales y nombres en lugar de generalidad
 - Estado del sistema cuando se inicia la tarea



UT: Test de Rendimiento

Escenarios de tarea

- ¿Vocabulario?
 - No proporciones claves de cómo realizar la tarea incluyendo vocabulario de interfaz: nombres de opciones de menú o de botones, etc.
- ¿Número de tareas?
 - Test in-situ: no más de 10-12 tareas **no-triviales** (menos de 1 hora)
 - Test remoto: no más de 15 minutos (max 5 tareas).
- ¿Orden de tareas?
 - Si es posible, secuencia los escenario de tarea de la manera en que un usuario real las ejecutaría.
 - Si existe duda, opta por poner primero los escenarios más sencillos, de manera que el usuario pueda ir ganando confianza con la interfaz poco a poco.

UT: Test de Rendimiento

Escenarios de tarea

Ejemplo: OHIM: Lawyers

The screenshot shows the OHIM (Office for Harmonization in the Internal Market) website. The header includes the OHIM logo, a search bar, and links for FAQ, Contact us, Site map, and Links. The main navigation menu has tabs for Home, Trade Marks, Designs, Quality plus, About OHIM, and Observatory. On the left, there are sidebar menus for Legal texts (Trade marks, Designs), Case-Law (Trade marks, Designs), and Most Requested (Press Room, SMEs and newcomers section, Newsletter: Alicante News, Public procurement, How to obtain protection for the whole of the European Union, Fees and payments (CTM), Fees and payment (RCD), The registration process - Classification (CTM)). The central content area features a banner for '10 years community design' (8 - 9 april 2013 - Alicante - Spain) and news items about upcoming hearings in Luxembourg and Bulgaria database harmonisation. To the right, there are links for eSearch Plus (Beta), Databases, Forms/Filings, and MyPage.

UT: Test de Rendimiento

Escenarios de tarea

- Los escenarios de tarea necesitan tener medidas y procedimientos de recogida de datos bien definidos.
- **Medidas de rendimiento (objetivas)**
 - Medidas de éxito: número de tareas completadas con éxito, número de pasos exitosos dentro de cada tarea, etc.
 - Medidas de error: número de intentos fallidos (paths fallidos), número de veces que el usuario ha intentado realizar la tarea, número de errores (clicks erróneos) por tarea, ...
 - Medidas de eficiencia: tiempo que se tarda en realizar la tarea, número de pasos necesarios para completar la tarea, ratio de tareas por unidad de tiempo, ...
- **Medidas de satisfacción(subjetivas): expresiones faciales, comentarios, etc.**

UT: Test de Rendimiento

Escenarios de tarea

- Los escenarios de tarea necesitan tener procedimientos de recopilación de datos bien definidos:
- Por tanto durante la fase de diseño del estudio hay que pensar en varias cosas:
 - ¿Qué significa que una tarea ha sido completada con éxito? ¿Es conveniente definir distintos grados de éxito? ¿éxito parcial?

EXAMPLE OF TASK GIVEN TO A USER	SUCCESS SCORE NOTES
What is John Doe's telephone extension?	<ul style="list-style-type: none">○ Finds employee address book: ½ point○ Types the name in the correct place: 1 point○ Invokes the search: ½ point○ Finds the right user's phone extension in results or in profile doc: 1 point○ Finds the extension: 1 point

Source: OHIM Report

Pass/Fail	Menu label	Description
Pass	Easy	1 st try - no problem
Pass	Medium	2 nd /3 rd try- observed difficulty
Fail	Hard	3 rd /4 th try- expressed difficulty
Fail	Assist	Succeeded with assistance
Fail	Fail	Failed or gave up

Table 3.1 - Effectiveness scoring criteria

Source: Norman Nielsen:
Intranet Usability Report (2007)

UT: Test de Rendimiento

Escenarios de tarea

- Además, tendremos que decidir...
 - ¿Tipos de errores que voy a controlar? ¿Caracterización de la severidad de los errores? ¿Niveles de complejidad de corrección de errores?

Type of nature:

- Graphical
- Structure
- Navigation
- Labelling
- Users' satisfaction

Severity:

- Low
- Medium
- High

Complexity:

- Low
- Medium
- High

The categorisation by severity was defined by the usability team based on the business goals set by the project owner at the beginning of the process

The categorisation by complexity was defined by the usability team after talking to each responsible of fixing the problem/bug

UT: Test de Rendimiento

Escenarios de tarea

- En general, siempre deberías comprobar...
 - ¿Fueron los usuarios capaces de terminar la tarea?
 - ¿Encontraron la información que necesitaban en cada momento?
 - ¿Encontraron un path eficiente?
 - ¿Entendían lo que estaban haciendo?
 - ¿Qué tipo de problemas encontraron? Desorientación, no saber cómo continuar, ...
 - ¿Se pudieron recuperar de los errores? ¿Por sí solos, o necesitaron ayuda?

UT: Test de Rendimiento

Escenarios de tarea

- Es posible que necesites diseñar algún tipo de formulario de recolección de datos para acelerar el proceso
- Ejemplo: OHIM

The screenshot shows the OHIM (Office for Harmonization in the Internal Market) website. At the top, there's a banner for the "10 years community design" event held on April 8-9, 2013, in Alicante, Spain. Below the banner, there's a "News" section with three articles: "Coming up this week in Luxembourg", "Bulgaria database harmonisation", and another "Coming up this week in Luxembourg". On the left side, there's a sidebar with links for "Legal texts" (Trade marks, Designs), "Case-Law" (Trade marks, Designs), and "Most Requested" (Press Room, SMEs and newcomers section, Newsletter: Alicante News, Public procurement, How to obtain protection for the whole of the European Union, Fees and payments (CTM), Fees and payment (RCD), The registration process – Classification (CTM)). On the right side, there are sections for "eSearch Plus (Beta)" (Search Community trade marks, designs and case-law), "Databases" (Search Community trade marks, Search trade marks in the EU - TMview, Search classification for trade marks - EuroClass, Search Community designs, Verify certified copies), "Forms/Filings" (File a Community trade mark, File a Community design), and "MyPage".

UT: Test de Rendimiento

Escenarios de tarea

- Si estás haciendo benchmarking o quieres comparar tu producto con otros, es necesario definir tipos de tareas genéricas y luego personalizarlas para cada aplicación particular

TASK TYPE	EXAMPLE OF TASK GIVEN TO A USER
Find an employee's phone number or email address	What is John Doe's telephone extension?
Find information about a group	Name a few of the services that the Copy Department offers.
Find information about a high-level manager	What was John Doe's job just prior to becoming president of this organization?
Find information about a policy	You have jury duty next week. Are you granted any paid days off, and if so, how many?
Make a request or order	Get yourself some business cards.
Apply for another job	See if there are any available positions at Acme that you might actually be interested in. If so, apply.
Sign up for a training course	Enroll in an enrichment class that would help you in your work or personal life.
Find information via search	How much money did the organization donate to UNICEF in 2008?
Update some personal information (HR or status)	Your mobile number has changed to 212.555.5555. Update this.
Find past news	When will the cafeteria begin accepting credit card payments?
Go to the intranet and log in	<i>We did not specifically ask users to log in. We simply presented them with the first task—usually, "Read the latest news." The logging in was implied.</i>

Extracted from Norman Nielsen:
Intranet Usability Report (2007)

UT: Test de Rendimiento

Escenarios de tareas

- Tareas predefinidas vs Tareas basadas en Entrevistas
 - Una alternativa a predefinir tareas es entrevistar al usuario acerca de cómo usa el producto y utilizar las respuestas para crear tareas en el momento para él
 - Ejemplo: ¿Cuál es la última cosa que hiciste cuando usaste este producto? ¿Me lo puedes enseñar?

UT: Test de Rendimiento

Escenarios de tareas

- Los escenarios de tarea pueden venir acompañados de cuestionarios de tarea
 - Escala Expectativa-Experiencia
 - ASQ
 - Cuestionarios personalizados
 - Sugerencias:
 - Evita realizar preguntas cuya respuesta puedes directamente observar
 - Diseña las preguntas y respuestas de manera que sean simples y breves.
- ¿Los recordamos?

UT: Test de Rendimiento

Escenarios de tareas

- Escala Expectativa-Experiencia: ítems de 7 puntos (muy fácil..muy difícil), antes y después de la tarea
 - Globalmente, creo que esta tarea va a ser Muy difícil* * * * * Muy fácil
 - Globalmente, esta tarea fue Muy difícil* * * * * Muy fácil



UT: Test de Rendimiento

Escenarios de tareas

- ASQ (After-Scenario-Questionnaire): Tres ítems de 7 puntos (totalmente en desacuerdo... totalmente de acuerdo), una vez que la tarea se ha completado
 - Estoy satisfecha con la facilidad de compleción de esta tarea
 - Estoy satisfecho con la cantidad de tiempo que me llevó completar esta tarea
 - Estoy satisfecha con la información de soporte (ayuda on-line, mensajes, documentación) que se me proporcionó mientras estaba completando la tarea

UT: Test de Rendimiento

Escenarios de tareas

- Cuestionarios privados (e.g. Desarrollados dentro de la empresa)

TASK ONE:										
1	How confident are you that you completed the task?	Not Confident	Very Confident	1	2	3	4	5	6	7
2	Was doing this task satisfying?	Very Unsatisfying	Very Satisfying	1	2	3	4	5	6	7
3	Was doing this task frustrating?	Very Frustrating	Not Frustrating	1	2	3	4	5	6	7

Extracted from Norman Nielsen:
Intranet Usability Report (2007)

UT: Test de Rendimiento

Cuestionario Post-test

- Dentro de un test de rendimiento también se usan Cuestionarios Post-test: usados para recopilar información que profundice nuestro entendimiento de las fortalezas y debilidades del diseño de nuestro producto
 - Suelen centrarse en la facilidad de uso y la facilidad de aprendizaje

UT: Test de Rendimiento

Cuestionario Post-test

■ Cuestionarios estandarizados. E.g. SUS: System Usability Scale

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

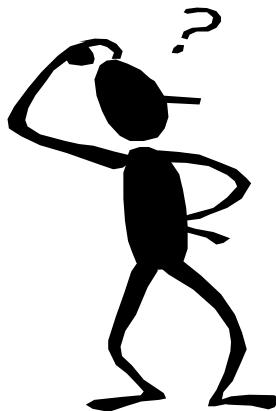
Strongly Disagree					Strongly Agree				
<input type="text"/>									
1	2	3	4	5	1	2	3	4	5
<input type="text"/>									
1	2	3	4	5	1	2	3	4	5
<input type="text"/>									
1	2	3	4	5	1	2	3	4	5
<input type="text"/>									
1	2	3	4	5	1	2	3	4	5
<input type="text"/>									
1	2	3	4	5	1	2	3	4	5
<input type="text"/>									
1	2	3	4	5	1	2	3	4	5
<input type="text"/>									
1	2	3	4	5	1	2	3	4	5
<input type="text"/>									
1	2	3	4	5	1	2	3	4	5

UT: Test de Rendimiento

Cuestionario Post-test

- Valores base de SUS:

- Media: 66
- 25%: 57
- 75%: 77



¿Cuál queréis que sea
vuestro objetivo de SUS?

- El SUS se puede adaptar a la idiosincrasia del sistema particular que estéis testando.
 - Ejemplo: OHIM

UT: Test de Rendimiento

Escenarios de tareas

- NPS: Net Promoter Score
 - ¿Cómo de probable es que recomiendes esta aplicación a un amigo o colega? (0..10)
 - Promotores: 9-10
 - Pasivos: 7-8
 - Detractores: 0-6
 - $NPS = \% \text{ promotores} - \% \text{ detractores}$

UT: Test de Rendimiento

Escenarios de tareas

- ACSI:

- ¿Cuál es tu satisfacción global con respecto a esta aplicación? (Escala 10 puntos: muy insatisfecho... muy satisfecho)
- ¿Hasta qué punto ha cumplido tus expectativas esta aplicación? (No ha llegado a cumplir mis expectativas.. Ha excedido mis expectativas)
- ¿Hasta qué punto se acerca esta aplicación a tu ideal de aplicación para... (el propósito de tu red social)? (No se acerca mucho a mi ideal... Se acerca mucho a mi ideal)
 - Fórmula: (Satisfacción (de 0 a 9) *0,3885 + Expectativa (de 0 a 9) * 0,3190 + Rendimiento (de 0 a 9)*0,2925) /9*100
 - Media de Computer Software: 71

UT: Test de Rendimiento

Cuestionario Post-test

- Otros temas que podrías querer tratar:
 - Aspectos generales de la página o pantalla software
 - Aspectos generales de documentación
 - Aspectos generales sobre el panel de control de hardware

- Más ejemplos de preguntas
 - ¿Qué tres cosas te gustaron más?
 - ¿Qué tres cosas te gustaron menos?
 - Si pudieras cambiar una cosa, ¿cuál sería?
 - ¿Usarías esta aplicación?
 - ¿Piensas que alguien puede tener problemas usando esta aplicación?
 - ¿Tienes alguna otra pregunta o comentario?



UT: Test de Rendimiento

Cuestionario Post-test

■ Ejemplo cuestionario post-test personalizado: OHIM

- What are the three things you like best and least about the Web site?

1. _____

2. _____

3. _____

- If you could make one significant change to this Web site, what change would you make?
- (Ask the participant if he is using/knows other IP websites) How do you find the site in comparison to other IP web sites?
- Do you feel this site is current? Why?
- If you were to describe this site to a colleague in a sentence or two, what would you say?

UT: Test de Rendimiento

Cuestionario de reunión posterior al test

- Cuestionario de reunión posterior al test: “Debriefing” con participantes: entender por qué cada error, dificultad u omisión ocurrió con cada participante de cada sesión.
- “Debriefing” con observadores: obtener distintos puntos de vista con respecto a lo que está pasando en las sesiones.



UT: Test de Rendimiento

Materiales adicionales

- MATERIALES ADICIONALES
- Script de Orientación (script de introducción): herramienta de comunicación pensada para ser leída *verbatim* a los participantes
 - Qué ocurrirá durante la sesión de test
 - Establece el tono de la sesión en la mente de los participantes
 - Objetivo: tranquilizar y relajar a los participantes.
- Algunos consejos:
 - Lee el script
 - Mantén el tono del script profesional pero amigable
 - Hazlo breve

UT: Test de Rendimiento

Materiales adicionales

- Secciones de un script de orientación (Ejemplo)
 - Introducción y no relación con el producto (si es así)
 - Ofrecer algún tipo de refrigerio
 - Explicar por qué está ahí el participante
 - Explicar el setup del testing
 - Explicar lo que se espera de los participantes
 - **Asegurar a los participantes que ellos no están siendo evaluados.**
 - Explicar cualquier requisito inusual
 - Mencionar que está bien preguntar en cualquier momento, incluso si el moderador no puede contestarlas!
 - Preguntar si tiene alguna pregunta
 - Referirse a cualquier formulario que deba ser llenado por el usuario y pasárselo
 - Indicar el tiempo que se espera sea necesario para acabar el test

UT: Test de Rendimiento

Materiales adicionales

- Materiales de entrenamiento
 - Para poder testar funciones avanzadas
 - Para hacer testing con usuarios avanzados: facilidad de uso en lugar de facilidad de aprendizaje (que es lo que se suele testar en realidad con usuarios noveles)



Recomendaciones adicionales

UT: Test de Rendimiento

Recomendaciones: Test Piloto

- Todos los materiales deberían pasar por un **test piloto** para evitar ambigüedades y sesgos
 - Pregunta a tus usuarios piloto si se sintieron confundidos por alguna pregunta o tarea
 - Comprueba que las tareas no están sesgadas, y que no dirigen al usuario a la respuesta correcta.
 - Solicita ideas para nuevas tareas
 - Pregunta si sienten que alguna tarea/pregunta es superflua
 - Aseguráos de que los límites de tiempo marcados para el test se cumplen
 - Los cuestionarios deberían ser diseñados de manera que sean simples y breves.

UT: Test de Rendimiento

Recomendaciones: Stakeholders

- Diseñadores, desarrolladores y otros interesados que puedan estar interesados de cualquier manera en los resultados del test deberían ser incluidos desde las fases de planificación.
- Incluye a todos los interesados tan pronto como sea posible en las discusiones sobre metas, usuarios y tareas.
 - De este modo se facilitará el poder trabajar juntos en un futuro para mejorar los diseños.

UT: Test de Rendimiento

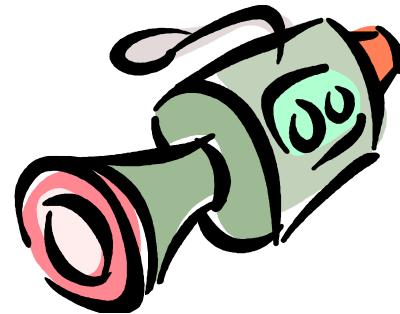
Recomendaciones: Observadores

- Las personas que siempre deberías invitar a observar los tests son aquellos que usarán la información para mejorar el diseño (disminuirás sus resistencias si ven ‘in situ’ los problemas que causa el diseño actual).

UT: Test de Rendimiento

Ejemplo

- Testeo de la usabilidad en acción



Usability Testing Example: Steve Krug. Zipcar
<http://www.youtube.com/watch?v=QckIzHC99Xc>

Mientras lo escucháis, responded al quiz

UT: Test de Rendimiento

Ejemplo

1.- ¿Cuál de los siguientes elementos no se especifican en la introducción de un test de usuario?

- Duración del test
- El hecho de que estamos testando la aplicación, no el usuario
- Un resumen de las distintas tareas que el usuario va a realizar

2.-¿Utiliza este test el protocolo think aloud?

- Sí
- No

3.-¿Utiliza el probing el moderador del test?

- Sí
- No

4.-¿Sobre qué pagina introduce el moderador del test un test de exploración?

- Página home
- Página de producto
- Página de contacto

5.-¿Qué tipo de escenarios se definen en el test?

- Tareas cortas, con el objetivo básico
- Escenarios más largos, con detalles de contexto

6.-¿Interacciona el moderador del test con el usuario para ayudarle a realizar las tareas?

- Sí
- No

UT: Test de Rendimiento

Ejemplo

7.-¿Cuáles de los siguientes NO son problemas del sitio de Zipcar según el usuario testado?

- Al usuario no le gusta no saber qué significan determinados conceptos
- El usuario se siente frustrado y no confía en el contenido del sitio
- El usuario cree que el vocabulario del sitio es muy complicado.

8.-¿Cuáles son los tres principales problemas de usabilidad detectados con este test?

9.-¿Cómo los arreglarías?

UT: Test de Rendimiento

Referencias

- *Handbook of usability testing. How to plan, Design and Conduct Effective Tests.* Second Edition. Jeffrey Rubin and Dana Chisnell. Wiley (2008)
- *Measuring the user experience: Collecting, Analyzing, and Presenting Usability Metrics.* Thomas Tullis, William Albert. Morgan Kauffman (2008)

