

Métodos de evaluación con usuarios

Amaia Calvo-Fernández Rodríguez
Sergio Ortega Santamaría
Alicia Valls Saez

PID_00176614



Universitat Oberta
de Catalunya

www.uoc.edu



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-Compartir igual (BY-SA) v.3.0 España de Creative Commons. Se puede modificar la obra, reproducirla, distribuirla o comunicarla públicamente siempre que se cite el autor y la fuente (FUOC. Fundació per a la Universitat Oberta de Catalunya), y siempre que la obra derivada quede sujeta a la misma licencia que el material original. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-sa/3.0/es/legalcode.ca>

Índice

Introducción.....	5
1. La evaluación de la usabilidad con usuarios.....	9
1.1. Evolución del concepto de usuario y de las interfaces	9
1.2. Primera aproximación a los tests con usuarios	13
1.2.1. Trabajo con modelos mentales	14
1.2.2. Qué es un test con usuarios	17
1.2.3. Cuándo llevar a cabo un test con usuarios	18
1.2.4. Personas necesarias en un test con usuarios	21
1.2.5. Laboratorios y herramientas para la evaluación de la usabilidad	25
1.3. Proceso del test	32
1.3.1. Definición y objetivos del test	33
1.3.2. Preparación del test	34
1.3.3. Captación de los participantes	43
1.3.4. Ejecución del test: ejemplo práctico	44
1.3.5. Consejos para ser un buen facilitador de test	47
1.4. Clasificación de los tests con usuarios	52
1.4.1. Presencial / remoto	52
1.4.2. Formal / de guerrilla	56
1.4.3. Tests con prototipos de alta/baja fidelidad	57
2. Variantes del test con usuarios.....	59
2.1. Test de codescubrimiento	59
2.2. Seguimiento ocular	60
2.3. Lápiz y papel	61
3. Otros métodos de evaluación con usuarios.....	64
3.1. Indagación contextual	65
3.2. Entrevistas	67
3.3. Encuestas	68
3.4. Cuestionarios	68
3.4.1. Software usability measurement inventory (SUMI)	69
3.4.2. Measuring the usability of multi-media systems (MUMMS)	69
3.4.3. Website analysis and measurement inventory (WAMMI)	70
3.4.4. System usability scale (SUS)	70
3.4.5. Otros cuestionarios estandarizados	74
3.5. Dinámicas de grupo (<i>focus groups</i>)	74
3.6. Clasificación de tarjetas (<i>card sorting</i>)	75

3.7.	Registros de usuario	75
3.8.	Sesiones guiadas (<i>journalled sessions</i>)	76
3.9.	Registro informático (<i>logging</i>)	77
4.	Análisis de resultados	79
4.1.	Organización de la información	79
4.1.1.	Organización sobre un criterio preestablecido	80
4.1.2.	Organización sin un criterio preestablecido	81
4.2.	Tratamiento de los datos: métricas	81
4.2.1.	Eficiencia y efectividad	82
4.2.2.	Tratamiento de los datos numéricos	83
4.2.3.	Métricas en un test con usuarios: Ejemplo de aplicación	87
4.3.	Interpretación de los datos	88
4.3.1.	Interpretación de los datos	88
	Resumen	92
	Bibliografía	93

Introducción

En el 2008, la revista *Fortune* entrevistaba a Steve Jobs, CEO de Apple, y le preguntaba sobre las claves del éxito de su compañía. Entre las numerosas respuestas de Jobs, una de ellas venía acompañada de una celebre cita de Henry Ford, diseñador y fabricante de coches y fundador de la Ford Motor Company.

“Si hubiera preguntado a mis clientes qué es lo que querían, me habrían dicho que un caballo más rápido”.

Henry Ford

Esta afirmación constituye una de las claves de las siguientes páginas. Apple no suele recurrir a la investigación de mercados para lograr el desarrollo de productos, pero sí realiza encuestas a sus usuarios y obtiene continuamente información de ellos para saber qué es lo que les gusta o qué es lo que pueden necesitar en cada momento.

Sin embargo, ni Apple ni otras muchas empresas que trabajan con productos y servicios interactivos necesitan, en el sentido más literal de la frase, conocer directamente la opinión de los consumidores. Debemos reconocer que las personas que utilizamos dichos productos no somos buenos articulando y verbalizando una respuesta precisa y clara sobre nuestras necesidades.

En definitiva, no son los usuarios quienes tienen que ofrecer soluciones concretas –pedir caballos veloces o máquinas potentes–, pero sí deben ser la fuente de información más fiable que permita que valoremos su participación en los procesos de construcción.

Para descubrir el mejor diseño y para ir optimizando progresivamente nuestras interfaces web, necesitamos ver cómo los usuarios trabajan, cómo ejecutan las tareas que hemos diseñado, y todo ello siguiendo tres “normas” ineludibles, que Nielsen (2001) destaca siguiendo la línea que hemos marcado hasta ahora:

- 1) Observar qué hace y cómo se comporta realmente la gente.
- 2) No creer aquello que dicen que hacen.
- 3) No creer aquello que la gente predice del futuro y de lo que pueden llegar a hacer.

Insistiremos en esta filosofía de trabajo en los próximos apartados, pero veamos qué sucede del lado de los diseñadores y desarrolladores.

Lectura complementaria

Entrevista a Steve Jobs, CEO de Apple:

- “Steve Jobs speaks out”, *Fortune* (07 de marzo de 2008)

Lectura complementaria

J. Nielsen (2001). “First rule of usability? Don’t listen to users”. *Alertbox* (núm. 5, agosto).

Cuando un equipo de desarrollo está creando un producto o servicio interactivo, piensa que éste será útil y fácil de usar. Sus miembros creen que han colocado el botón en el lugar más visible de la página, que han diseñado un formulario claro y agradable y que han ordenado los contenidos de un sitio de la manera más lógica posible.

Todas las ideas preconcebidas por parte de este equipo desaparecen en el momento en que un usuario se encuentra por primera vez con este producto o servicio. Quizá no encuentre el botón porque pensaba que debía estar en un lugar determinado de la interfaz, no logra entender el formulario ni finalizarlo o no encuentra conceptos básicos dentro del árbol de contenidos, que han sido estratégicamente pensados y analizados.

Es entonces cuando nos damos cuenta de que evaluar nuestros productos o servicios con usuarios reales puede ser una buena manera de saber si vamos por el buen camino, es decir, si el producto o el servicio que desarrollamos es fácil, eficiente y agradable de usar.

“Ser abogado no es fácil cuando no te dejan ver a tu cliente. Como profesionales de la usabilidad, aplicar normas reconocidas de diseño de interfaz nos ayudarán a hacer mejor nuestro trabajo y a velar por los intereses de nuestro defendido: el usuario.”

“La interfaz perfecta no se logra al primer intento.”

“No hay usabilidad sin contexto.”

“No hay usabilidad sin usuarios.”

“Aplicando correctamente la metodología de diseño centrado en el usuario con el usuario nos ponemos en situación de obtener la mejor interfaz para un producto concreto destinado a usuarios específicos.”

Luis Villa, grancomo.com

La evaluación de la usabilidad no siempre requiere la participación de los usuarios. Algunos métodos de evaluación facilitan la localización de problemas de usabilidad y el logro de mejoras en el diseño a partir del cumplimiento de principios, la consulta a expertos o el análisis de métricas, por señalar algunos ejemplos. En ocasiones, valorar el uso de estas metodologías antes de implicar al usuario nos puede ayudar a reducir una cantidad relevante de problemas de usabilidad que con una evaluación con usuarios podrían suponer un coste elevado.

Por eso, el principal objetivo de una evaluación con usuarios, también conocido como *test de usabilidad con usuarios*, será verificar la existencia de problemas mayores de usabilidad en nuestro producto o servicio interactivo (aquellos que generan un mayor impacto en el usuario) y así poder ofrecer de forma eficaz soluciones para los problemas encontrados.

Hay buenas razones para llevar a cabo evaluaciones con usuarios durante lo que viene a llamarse *ciclo de vida* completo del producto o servicio. Podemos tomar diferentes momentos para efectuar la prueba:

- Antes de comenzar a crearlo, para detectar de forma temprana errores de diseño.
- Durante su construcción, para no desviarnos de nuestros objetivos o tareas.
- Después de su lanzamiento, para comprender qué tendremos que mejorar en una siguiente versión o para trabajar sobre un rediseño.

Ahora bien, cuando hablamos de diseño centrado en el usuario (DCU) hablamos de un enfoque de investigación y un desarrollo de principios guiados por el conocimiento constante de los usuarios y sus modelos mentales a partir de sus procesos de interacción. Esto implica que si conseguimos obtener una buena información en etapas tempranas, las probabilidades de asegurar la consecución de un producto y su funcionalidad adecuada para usuarios concretos serán mayores.

No debemos olvidar que la búsqueda de una mejora progresiva del producto nos invita a conocer cuanto antes nuestros errores de diseño.

1. La evaluación de la usabilidad con usuarios

La historia del método de evaluación con usuarios está íntimamente ligada al desarrollo de la interacción persona-ordenador (IPO¹), que a su vez no se entendería sin analizar la aparición de la tecnología como elemento de gran consumo (en especial, el ordenador personal, a partir de 1980) y la creación de las interfaces gráficas de usuario (GUI²).

⁽¹⁾IPO es la sigla de *interacción persona-ordenador*.

⁽²⁾GUI es la sigla de la expresión inglesa *graphical user interface*, en español *interfaz gráfica de usuario*.

Pero también es importante atender a otros aspectos cruciales que surgieron a finales de la década de los setenta y principios de los ochenta, y que también nos van a permitir entender el comienzo de los tests de usabilidad, a saber:

- Evolución de la psicología cognitiva y una mayor valoración de los procesos cognitivos humanos que nos han permitido conocer cómo las personas sienten, perciben, actúan, almacenan información, recuperan...
- Proliferación de estudios, investigaciones y publicaciones relacionadas con el procesamiento humano de la información.
- Crecimiento de modelos sobre sensación, percepción, memoria... y, en consecuencia, aumento en la realización de experimentos parciales y académicos centrados en la ejecución de tareas.
- Desarrollo de las interfaces gráficas, que suponen una mejora en la comunicación con los usuarios a partir de avances y mejoras en la interacción, en la visualización o en la organización.

1.1. Evolución del concepto de usuario y de las interfaces

Los primeros tests de usabilidad basaban sus propósitos en obtener validaciones estadísticas o en evaluar al final del proceso para demostrar que todo estaba correcto.

Pero, atendiendo a los diferentes aspectos nombrados anteriormente, se comenzó a observar que ese tipo de actuación provocaba demasiadas frustraciones en la creación de productos y servicios dirigidos a un público determinado. Ya no se trataba de saber y confirmar que todo funcionaba correctamente o de validar un nuevo modelo. El mayor beneficio estaba en descubrir y averiguar dónde estaban los errores, cómo se producían o qué es lo que no funcionaba bien.

De todo esto se desprende que, gracias a los avances en campos como la comunicación, las ciencias de la información, la psicología cognitiva y la interacción persona-ordenador, ahora sabemos más sobre la forma en que los seres humanos procesan la información.

La consideración de estos avances nos ha llevado a incluir conceptos como evaluación temprana, continuada e iterativa en los enfoques dados a las metodologías, así como a abandonar su exclusividad científica o académica para convertirse en una valiosa, ágil y eficiente forma de obtener la información adecuada en la construcción de los productos interactivos.

De estos avances logrados en las últimas décadas, cabe mencionar la importancia del progreso en las interfaces de usuario. Su evolución ha facilitado el aumento de la productividad que se ha estado prometiéndolo durante años gracias al uso de la tecnología.

Hemos aprendido que la interfaz de usuario es uno de los elementos críticos en la aceptación y apropiación de tecnologías, productos y servicios por parte del usuario. Su diseño ha requerido planteamientos cuidadosos, de tal manera que el modelo creado por el usuario fuera correcto, y en parte coincidente con el que tenía el creador.

El diseño de una interfaz debe tomar en consideración cómo las personas adquieren, interiorizan, procesan o exteriorizan información a partir de las sensaciones, percepciones e interacciones para lograr mayor productividad, precisión y satisfacción.

Neal Stephenson nos habla en su libro *En el principio... fue la línea de comandos* sobre la aparición de las GUI, no sólo para entender el impacto de las interfaces en nuestra vida, sino también sobre las contradicciones que se han creado con su aparición.

“Si el vídeo se hubiera inventado hace cien años, tendría una ruedecita para la sintonización y una palanca para avanzar y rebobinar, y una gran asa de hierro forjado para cargar o expulsar los casetes. Llevaría un gran reloj analógico delante, y habría que ajustar la hora moviendo las manillas en la esfera. Pero debido a que el vídeo se inventó cuando se inventó –durante una especie de incómodo periodo de transición entre la era de las interfaces mecánicas y las GUI– tiene sólo unos cuantos botones delante, y para fijar la hora hay que pulsar los botones de modo correcto. Esto le debe de haber parecido bastante razonable a los ingenieros responsables, pero para muchos usuarios es sencillamente imposible. De ahí el famoso 12:00 que parpadea en tantos vídeos. Los informáticos lo llaman el síndrome del 12 parpadeante. Cuando hablan de ello, empero, no suelen estar hablando de vídeos.

Los vídeos modernos habitualmente tienen algún tipo de programación en pantalla, lo cual significa que se puede fijar la hora y controlar las demás propiedades mediante una especie de GUI primitiva. Las GUI también tienen botones virtuales, claro, pero también tienen otros tipos de controles virtuales, como botones de radio, casillas que tachar, espacios para introducir textos, esferas, y barras. Las interfaces compuestas de estos elementos parecen ser mucho más fáciles para muchas personas que pulsar esos botoncitos en la máquina, y así el propio 12:00 parpadeante está desapareciendo lentamente de los salones de Estados Unidos. El problema del doce parpadeante ha pasado a otras tecnologías.

Así que la GUI ha pasado de ser una interfaz para ordenadores personales a convertirse en una especie de metainterfaz, que se emplea en cualquier nueva tecnología de consumo. Raramente es ideal, pero tener una interfaz ideal o incluso buena ya no es la prioridad; lo importante ahora es tener algún tipo de interfaz que los clientes usen realmente, de

tal modo que los fabricantes puedan afirmar con toda seriedad que ofrecen nuevas posibilidades.”

Neal Stephenson, *En el principio... fue la línea de comandos* (1999)

En el ámbito de los ordenadores personales, el salto inicial hacia esta interfaz “que los clientes utilicen realmente” se dio con la aparición de conceptos como Ventana, Icono, Menú, Puntero (Window, Icon, Menú, Pointing Device, WIMP), aplicado por el equipo de Xerox Palo Alto Research Center (PARC) en el modelo Xerox Alto, que era experimental. Le seguiría el Xerox Star, que se comercializó en 1981.

El intenso trabajo en las pruebas de usabilidad que los investigadores del PARC llevaron a cabo para el desarrollo del Star y el deseo de ayudar al usuario a comprender y aceptar nuevos planteamientos, desarrollos, metáforas... en las interfaces gráficas de usuario condujo a una forma de interacción persona-ordenador más enriquecedora y beneficiosa.

Los miembros del equipo eran conscientes de estos cambios y así lo hacían saber en sus informes y presentaciones públicas.

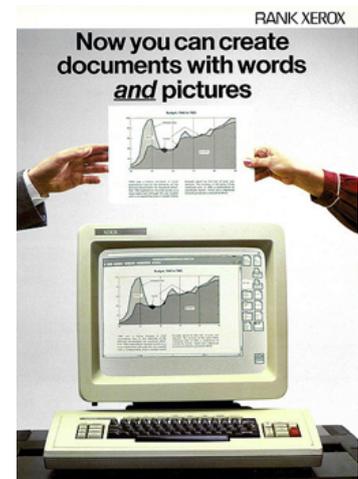
“La grabación en vídeo ha sido una herramienta muy importante. Para empezar, la cámara nos permitía ver y escuchar al sujeto sin estar en la misma habitación. En segundo lugar, era un registro de todas las actividades, por lo que no teníamos la necesidad de tomar notas perfectas durante el experimento. En tercer lugar, los diseñadores se convenían mejor con las cintas que con nuestros fríos números de que la gente tenía problemas con su sistema.”

“Human factors testing in the design of Xerox’s 8010 Star office workstation”. Conferencia CHI de 1983

Como decíamos al comienzo, la IPO vivía en los años ochenta un período de grandes avances, y la introducción de la psicología cognitiva, especialmente a partir de las aportaciones de Card, Moran y Newell (1983) y de Norman y Draper (1986) ha facilitado la búsqueda de un objetivo muy claro: conocer y representar cómo interactúa el ser humano con el ordenador.

De esta manera, la anterior conclusión de los investigadores de PARC, obvia desde una perspectiva actual, resume perfectamente ese deseo por conocer y evaluar involucrando al usuario real en los procesos de diseño. Las virtudes de los tests con usuarios comenzaban a aflorar.

Apple se basó en la interfaz del Xerox Star para crear la interfaz de Apple Lisa, y también llevó a cabo tests con usuarios, reclutando voluntarios entre su familia y amigos para probar el nuevo sistema. Recogiendo nuevamente otra de las ideas de Steve Jobs en su entrevista en Fortune, la idea no es tratar de engañar a la gente, no se trata de convencerles de la necesidad de algo que ellos no quieren. Tenemos que averiguar qué es lo que nosotros queremos y tener el control adecuado para poder pensar si una gran cantidad de personas también van a quererlo.



Fuente: Wikipedia (CC BY-SA 3.0)

Lecturas complementarias

S. K. Card; T. P. Moran; A. Newell (1983). *The psychology of human-computer interaction*. Hillsdale: Erlbaum.

D. Norman; S. Draper (1986). *User centered system design: new perspectives on human-computer interaction*. Hillsdale: Erlbaum.

Lectura complementaria

“Steve Jobs speaks out”, *Fortune* (07 de marzo de 2008)

Quizás lo difícil de este proceso sea obtener el significado en relación con otros. Dicho de otro modo, nuestras percepciones y significados de los productos que construimos no pueden ser incompatibles con la posibilidad de convertirse en percepciones y significados compartidos.

Durante las pruebas realizadas por el equipo de Xerox PARC, los miembros recogieron resultados que vienen a confirmar esta idea de la percepción y el significado individual frente a la percepción y el significado de los otros.

“Cuando el software requería confirmación del usuario, mostraba una pequeña ventana llamada ‘caja de diálogo’, que contenía una pregunta y presentaba dos botones, para la confirmación positiva o negativa. Los botones tenían las etiquetas ‘Do It’ y ‘Cancel’. Los diseñadores observaron durante los tests que los usuarios tendían a atascarse en esta caja de diálogo y acababan clicando sobre ‘Cancel’, cuando se esperaba que lo hicieran sobre ‘Do It’.

Finalmente, el facilitador del test interrumpió la sesión con un usuario que parecía particularmente molesto por la ventana para preguntarle cuál era el problema. El usuario respondió ‘I’m not a dolt, why is the software calling me a dolt?’ (No soy un idiota, ¿por qué el software me llama idiota?; ‘dolt’ es la palabra inglesa para “idiota”).

Resultó que el usuario no percibía el espacio entre la “o” y la “I” de “Do It”, y la tipografía del sistema hacía que la “I” pareciera una letra ele. Por tanto, los usuarios llegaban a la caja de diálogo y veían que tenían que elegir entre “Idiota” o “Cancelar”.

El equipo decidió sustituir el “Do It” por el más directo “OK”, que conocemos bien, pero que en un principio habían descartado por parecerles excesivamente coloquial.”

Fuente: Andy Hertzfeld, “Do It”, *folklore.org* (junio de 1982)

En cualquier caso, tanto Xerox Star como Apple Lisa fueron fracasos comerciales, también motivados por el alto precio de los equipos y cierto desencanto del público con respecto a las prometedoras interfaces gráficas.

Se transformó también la IPO, al cuestionar las teorías cognitivas e incorporar otras disciplinas como la antropología para tener en cuenta también el entorno y las situaciones de uso de los sistemas (Suchman 1987; Löwgren, 1995) y valorar el creciente interés que estaba teniendo la usabilidad.

En la siguiente década, será Nielsen (1994) quien introducirá los *discount usability methods* a partir de la ingeniería de la usabilidad (IU) como una forma más barata y cualitativa –pero igualmente operativa– de efectuar pruebas de usabilidad, especialmente el test con usuarios.

El desarrollo y la evolución de Internet han consolidado una nueva trayectoria donde la competitividad, la productividad y el cumplimiento de objetivos han mejorado los esfuerzos profesionales y han favorecido la agilidad de las pruebas rápidas pero continuas durante el proceso de desarrollo del producto o servicio.

Lecturas complementarias

S. Card; T. Moran; A. Newell (1983). *The psychology of human-computer interaction*. Hillsdale: Lawrence Erlbaum Associates.

J. Nielsen (1994). *Usability engineering*. San Francisco: Morgan Kaufmann.

J. Löwgren (1995). *Perspectives on usability*. Linköping: Department of Computer and Information Science, Linköping University. IDA Technical Report.

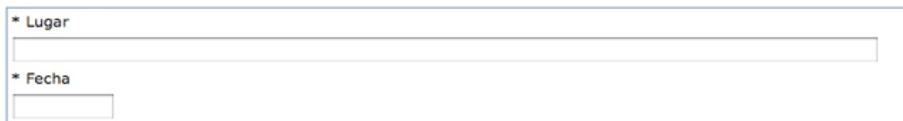
D. Norman; S. Draper (1986). *User centered system design: new perspectives on human-computer interaction*. Hillsdale: Erlbaum.

Lucy A. Suchman (1987). *Plans and situated actions: the problem of human-machine communication*. Nueva York: Cambridge University Press.

M. Ribera Turró (2005). "Evolución y tendencias en la interacción persona-ordenador". En: *El profesional de la información* (vol. 15, nov.-dic., núm. 6, págs. 414-422).

1.2. Primera aproximación a los tests con usuarios

Observad los siguientes campos de un formulario recogido de una página en Internet:



The image shows a screenshot of a web form with two input fields. The first field is labeled '* Lugar' and is a wide, empty text box. The second field is labeled '* Fecha' and is a smaller, empty text box. Both fields are enclosed in a light blue border.

Sin tener una contextualización clara, tan solo observando la posición, la forma y los tamaños, ¿qué y cómo escribiríais en cada caja?

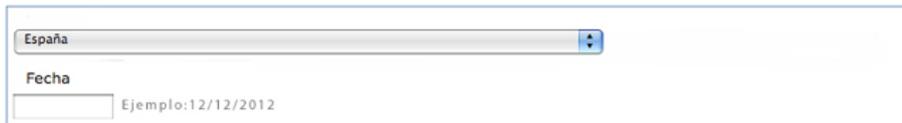
Podemos deducir que el campo "Lugar" me permite introducir una gran cantidad de caracteres, dando por hecho que algunos lugares pueden estar formados por palabras extensas. Por ejemplo: Villafranca del Penedés (Barcelona).

El campo "Fecha" por su parte, no aclara cómo debo escribir la fecha y parece que no es importante para recoger la información. Eso sí, quizás podamos deducir que podemos escribirla con números porque de otra manera no entraría en la caja. Entonces, podríamos escribir una fecha con este formato (12/12/12) o con este otro (12-12-12), o incluso con este último (12 12 2012).

Un detalle más. Observamos en ambos campos un asterisco junto al nombre. No se indica nada pero podemos hacer dos suposiciones. Una sería que el campo es obligatorio y no puedo dejarlo vacío. Otra sería que existe una anotación en la parte inferior de la página que aclara algún aspecto de estos campos marcados.

Con este ejemplo notoriamente confuso, podemos comprobar la cantidad de suposiciones que se pueden llegar a hacer cuando alguien nos solicita unos simples datos y la forma de introducirlos no es clara. Incluso, si pidiéramos más opiniones sobre la forma de introducir los datos, encontraríamos más variantes posibles.

La solución a esta tarea parece sencilla. Basta con no conceder tanta libertad y margen de error al usuario y evitar al máximo la ambigüedad.



Un formulario con un menú desplegable que muestra 'España' y un botón de flecha hacia abajo. Debajo, un campo de texto etiquetado 'Fecha' con el ejemplo 'Ejemplo:12/12/2012'.

Si colocamos los elementos donde se espera que se encuentren, nos aseguraremos de que estos serán percibidos. Pero además, si evitamos la confusión y generamos interacciones acordes con las expectativas, crearemos experiencias que el usuario encontrará satisfactorias. Esto significa estar en sintonía con los modelos mentales de nuestros usuarios.

Ahora bien, en un caso tan sencillo como el que hemos explicado podemos intuir un diseño equilibrado y adecuado que vaya acorde con las acciones que realizamos en otras muchas páginas. Pero ¿qué sucedería si la respuesta no fuera tan sencilla?

Pensad en algo más complejo, como por ejemplo un proceso de compra, una configuración de una cuenta o, por ir a algo más cercano para todos, pensad en las opciones de configuración de una política de privacidad en una red social. ¿Cómo la percibirán los usuarios? ¿Qué conceptos pueden ser introducidos? ¿Qué interacciones serán necesarias?

1.2.1. Trabajo con modelos mentales

Construimos modelos mentales a partir de nuestra experiencia y los almacenamos en nuestra mente. Esta afirmación que parece tan evidente nos recuerda, tal y como veíamos en el subapartado anterior, que en todo lo que hacemos participan nuestras ideas, perspectivas, presuposiciones o estrategias. Es decir, participan nuestras ideas generales, que dan forma a nuestros pensamientos y a todos nuestros actos.

Las estrategias que la gente emplea para resolver un problema están fundamentadas en lo que cada individuo ya conoce o cree conocer.

Ejemplo de modelo mental

Lindgaard (1994) propone un ejemplo que da mayor claridad al tema de los modelos mentales.

Imaginemos que estamos haciendo un pan y cuando lo sacamos del horno, no ha crecido como esperábamos. ¿Qué hemos hecho mal?

El horno estaba demasiado caliente cuando introducimos el pastel, lo que mató las células de levadura, por lo que el pan no subió.

No se mezclaron bien los ingredientes y la levadura no pudo hacer su función.

La cantidad de levadura era insuficiente para la cantidad de masa.

No se dejó fermentar suficientemente la masa antes de meterla en el horno.

Alguien abrió la puerta del horno antes de tiempo e interrumpió la cocción.

Dependiendo de cuál de estas explicaciones decidamos que es la correcta, adoptaremos una estrategia u otra la próxima vez que hagamos un pan: mezclaremos mejor, variaremos las cantidades de ingredientes, controlaremos mejor la temperatura del horno. Y la elección de la explicación dependerá de nuestra experiencia anterior en hacer panes; si somos expertos panaderos, podremos analizar el color, el nivel de cocción de la masa o la textura, para llegar a una conclusión.

Si nunca habíamos hecho un pan, tendremos menos herramientas para analizar las “evidencias del delito” y llegar a una conclusión.

Lo que el ejemplo quiere ilustrar es que, ante una elección, el razonamiento difiere enormemente entre unas personas y otras, entre usuarios noveles y expertos de un sistema (panaderos o manazas).

Si nos trasladamos al terreno de las interfaces, sucede que en una evaluación, ante un problema o elección, nuestros usuarios adoptan una estrategia relacionada con la forma en la que entienden el problema que se les presenta: algunos piensan que el horno está demasiado caliente, otros añaden más levadura, otros mezclan con más vigor. Es decir, actúan según su modelo mental de “formas de hacer un pan perfecto”.

Por tanto, los modelos mentales no son “patrimonio” único de las etapas de recogida de datos sobre nuestro usuario o del diseño y conceptualización del producto. Llegan a nosotros durante el test con usuarios y hemos de prestarles atención en el momento del análisis.

Un modelo mental es la representación de la realidad que cada persona se construye para entender fenómenos específicos. Don Norman los definió así: “En la interacción con el entorno, con los demás, con los artefactos tecnológicos, la gente se forma modelos mentales internos de ellos mismos y de las cosas con las que están interactuando. Estos modelos proporcionan capacidad de prever y explicar la interacción” (Getner y Stevens, 1983).

Siguiendo la perspectiva de la evaluación, que es la que abordamos en estas páginas, se exponen algunas características importantes de los modelos mentales que conviene resaltar antes de dar paso a las explicaciones sobre los tests con usuarios.

a) Los modelos mentales son incompletos

Imaginemos que queremos evaluar un nuevo buscador. ¿Qué pasaría si el modelo mental de todos nuestros usuarios fuera “esto es como Google”? ¿Podemos haber influido en parte en la creación de este modelo incompleto a través del diseño de nuestro test?

A menudo las evaluaciones también son parciales, por lo tanto, hay que juzgar críticamente si el modelo que adoptaron los usuarios estaba motivado por las características de nuestro test:

- Si evaluamos un prototipo o no evaluamos en profundidad todas sus funcionalidades, es natural que den lugar a un modelo mental incompleto.
- Si las tareas o escenarios no abarcan el conjunto del producto o servicio, es natural que los usuarios no tengan un modelo mental completo.

b) Los modelos mentales no tienen límites precisos: diferentes dispositivos e interacciones se confunden entre sí

Durante la evaluación ¿los usuarios han utilizado como referencia otros dispositivos para interactuar? Cuando el usuario ha expresado verbalmente lo que piensa, ¿se han registrado frases como “Esto funciona como esto o aquello” o “Probaré a hacer lo que haría con esto/aquello, para ver si funciona igual”?

c) Los modelos mentales tienen un componente “a-científico” y “supersticioso”

Podemos descubrir que nuestros usuarios han desarrollado “rituales” de interacción durante las pruebas de evaluación. Definiríamos los rituales como patrones de interacción que los usuarios adoptan, que saben racionalmente que son innecesarios, pero que les permite pensar que el sistema responde con mayor eficacia.

El comportamiento supersticioso se refuerza en la interacción con sistemas en los que hay transacción económica. Por ejemplo, en unas pruebas con máquinas recreativas de azar (popularmente, “tragaperras”), comprobamos que los usuarios pensaban que la manera de apoyarse en la máquina o de pulsar los botones influía sobre el resultado.

Estos comportamientos son difíciles de racionalizar, pero es necesario que, como equipo que diseña o evalúa la interacción con el producto, registremos su existencia y los tengamos en cuenta a la hora de analizar los resultados de una evaluación.

Don Norman apunta, “acertadísimamente³” (Getner y Stevens, 1983), que algunas de estas supersticiones tienen su origen en que la persona ha encontrado algún tipo de problema durante el uso del producto o servicio y cree que una particular secuencia de acciones resolverá el problema o evitará que se repita. Por tanto, debemos estar atentos también al origen de las supersticiones de nuestros usuarios: ¿están enmascarando algún problema que debamos conocer?

d) Los modelos mentales son ahorrativos

Los usuarios prefieren hacer más interacciones que dedicar energía mental a pensar el mejor camino para resolver su problema. Esto es un indicador doble:

- Por un lado, del nivel de conocimiento que tienen nuestros usuarios del producto o servicio, ya que el comportamiento más “certero” o el uso de los atajos denotan un conocimiento más profundo del sistema. Si nadie utiliza atajos, ¿qué puede significar?
- Por el otro, de que los modelos mentales tienden a simplificar lo complejo ¿qué aspectos detectamos que se simplifican?

Ejemplo

Mediante el análisis de registros informáticos (*logs*) podemos detectar, por ejemplo, que los usuarios de una intranet dan una serie de pasos innecesarios para llegar a una funcionalidad concreta. Ello no es suficiente para afirmar que existe un modelo mental determinado, pero es una buena pista para investigar mediante métodos como entrevistas, dinámicas de grupo (*focus groups*) o tests de tareas.

1.2.2. Qué es un test con usuarios

El test con usuarios es una prueba de usabilidad enmarcada en el enfoque de diseño centrado en el usuario. Consiste en la evaluación de un sitio web a partir de la observación, expresión verbal e interacción de los usuarios mientras ejecutan las tareas propias de un producto o servicio interactivo.

En un test de usabilidad se analizan los puntos fuertes y débiles de un producto o servicio con usuarios reales y representativos bajo condiciones reales de uso. Se fijan unos objetivos de test y se determina un público objetivo que reclutar para dicho test. Se crean las tareas y los escenarios que estructuran la prueba.

Otras “supersticiones” de usuario

Otras “supersticiones” de usuario que hemos recogido en diferentes proyectos: “Cuando navego por una entidad bancaria, cierro el resto de las ventanas”. “Cuando un botón falla, vuelvo a pulsarlo pero durante más rato”.

⁽³⁾Nunca se ensalza suficientemente a Don Norman, os animamos a hacerlo.

Una persona experta –el facilitador o moderador– sería la encargada de moderar el test mientras observa el comportamiento del usuario; uno o varios observadores tomarían nota de las acciones que aquél ejecuta. Una vez finalizados los tests, se analizarían e interpretarían los resultados para detectar los aspectos que se tienen que corregir.

Aun así, la naturaleza de los tests con usuarios es iterativa. El proceso puede comenzar una y otra vez e ir refinando los resultados. De hecho, nunca se agotarían los errores posibles, sino que su especificidad iría en aumento.

Los tests de usabilidad permiten contrastar todas las ideas del equipo que desarrolló el producto o servicio con la realidad del usuario que lo tiene que consumir. A menudo dentro de un departamento de desarrollo se generan discusiones sobre si un diseño es adecuado o no y muchas veces el mejor recurso para poner fin a estas discusiones es llevar a cabo un test; de esta manera sabremos qué es lo que realmente hace el usuario. Es una manera de corroborar que las decisiones que se toman tienen de verdad un impacto positivo sobre el producto o servicio mediante una demostración válida con hechos tangibles.

En este sentido, los tests se parecen en cierta medida a los experimentos que se llevan a cabo como parte del método científico: se formulan unas hipótesis (¿puede mi usuario cumplimentar este formulario?), se seleccionan unos sujetos de test (los usuarios), se efectúan mediciones durante la prueba, se analizan los resultados y se intenta demostrar que la hipótesis queda validada.

1.2.3. Cuándo llevar a cabo un test con usuarios

Un test con usuarios puede ser una de las técnicas de evaluación más costosas a nivel económico, por eso se recomienda utilizar previamente otras técnicas para descubrir problemas de uso básicos. Por ejemplo, podemos analizar con anterioridad el producto o servicio llevando a cabo un análisis heurístico basado en los principios generales de usabilidad. De esta manera encontraremos los errores de diseño más básicos y fácilmente detectables sin necesidad de la participación de usuarios.

Antes de realizar un test con usuarios, debemos asegurarnos de que el servicio o producto cumple los requisitos básicos de usabilidad para evitar desperdiciar tiempo y dinero (Hassan Montero; Martín Fernández, 2003).

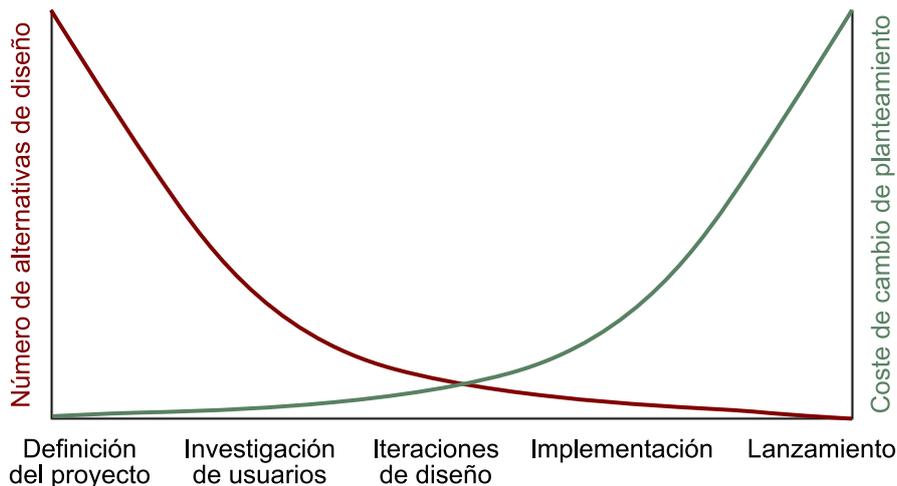
Es importante, además, que el test con usuarios no se deje para el final del proceso de creación del producto o servicio; a menudo, se realiza el test demasiado tarde, como para arreglar los fallos que se detectan. Cuanto más rápido detectemos un problema, menos costosa será su reparación.

El siguiente gráfico refleja esta realidad: la línea descendente representa el número de alternativas de diseño a medida que avanza el ciclo de vida del proyecto; la línea ascendente representa el impacto económico de “cambiar de

Lectura complementaria

Y. Hassan Montero; F. J. Martín Fernández (2003). “Método de test con usuarios”. *No Solo Usabilidad* (núm. 2).

idea”, es decir, de corregir el planteamiento del producto o servicio. A medida que el proyecto avanza en el tiempo y va atravesando diferentes fases –definición, investigación de usuarios, iteraciones de diseño, implementación y lanzamiento–, el número de alternativas de diseño se va reduciendo y aumenta el coste de cambiar de planteamiento, ya que “se van cerrando puertas”. Por esta razón el test con usuarios, como cualquier otro método de evaluación, no debería dejarse para el final (la fase de implementación, por ejemplo) ya que, si nos hemos equivocado en alguna de las decisiones anteriores, será muy costoso volver atrás.



Fuente: De la conferencia de Jeffrey Veen en Web Essentials 05: “Beyond Usability: Designing the Complete User Experience”

Procuremos no empezar la casa por el tejado

Imaginemos que queremos construir un edificio de pisos. Si empezamos por el tejado, no tendremos nada que sustente el edificio. Siempre empezaremos por los cimientos, por el suelo, basándonos en los proyectos que hayan desarrollado los arquitectos responsables (es decir, aplicar técnicas de usabilidad como una evaluación heurística).

Mediante evaluaciones previas sabremos si el edificio cumple las normativas legales de construcción. Para averiguar si los materiales y productos que se emplean para su construcción son los adecuados, deberemos ir piso por piso a comprobar su calidad.

Aunque no sea muy común en construcción, invitaremos a nuestros futuros inquilinos a pasearse por cada piso, para que nos ayuden a detectar si el edificio cumple con las necesidades, las características y las expectativas de éstos. Si esperamos a que el edificio esté terminado para que se paseen, quizá sea demasiado tarde para cambiar los errores. Si, por el contrario, efectuamos evaluaciones a medida que se van construyendo las plantas, seguramente evitaremos repetir errores y nos aseguraremos de que los pisos cumplen con los proyectos desarrollados por los arquitectos (requisitos básicos de usabilidad) y con las necesidades de los inquilinos-usuarios.

Así pues, cuanto antes se realice un test con usuarios, mejor será el resultado y menos costosa su realización.

Otro punto importante del “cuándo” es que no hace falta tener el producto completamente terminado para realizar un test con usuarios. Podemos realizarlo mediante prototipos, de alta o baja fidelidad.

Krug (2001) hace sobre esta cuestión las siguientes observaciones importantes:

Ved también

Podéis ver los prototipos de alta o baja fidelidad en el subapartado 1.4.3 de este módulo didáctico.

- Probar un usuario es 100% mejor que no probar ninguno; probar siempre funciona. Incluso la peor prueba con el usuario equivocado mostrará cosas que podemos hacer para mejorar el producto o servicio.
- Probar un usuario al principio del proyecto es mejor que probar 50 casi al final; una prueba al principio de un proyecto, cuando todavía tenemos tiempo de utilizar y cambiar los errores que detectamos, es casi siempre más valiosa que una sofisticada prueba más tarde. Es importante evitar construir mal desde el principio.
- La importancia de reclutar usuarios representativos está sobrestimada; es importante realizar los tests con usuarios que van a utilizar el producto o servicio, aunque es más importante probarlo pronto y con frecuencia.

La vehemencia de Steve Krug

Steve Krug es un ante todo un comunicador del área de la usabilidad. En este sentido, da más importancia a transmitir su pasión por su trabajo que a ser metodológicamente correcto en cada frase. ¿Por qué decimos esto? Porque decir que “la importancia de reclutar usuarios representativos está sobrestimada” es una frase provocadora. Como veremos más adelante, elegir bien a los usuarios es un ingrediente fundamental para que los resultados de una evaluación sean fiables. No obstante, estamos de acuerdo con Krug en que hacer un test con cualquier usuario, por inadecuado o poco común que sea su perfil al producto o servicio que evaluamos (ejemplo: una persona mayor y un teléfono móvil Blackberry), es mejor que no hacer ningún test con ningún usuario.

- El hecho de probar no es aprobar o desaprobar algo. Es informar de su decisión; a menudo a la gente le gusta pensar que puede utilizar los tests para comprobar si un sistema de navegación “a” es mejor que uno “b”. Nadie tiene los recursos para establecer el tipo de experimento controlado que necesita. La prueba nos puede proporcionar inversiones incalculables que, junto a la experiencia, decisión profesional y sentido común, harán que podamos elegir con confianza entre “a” o “b”.
- Probar es un proceso repetitivo; probar no es algo que se haga solamente una vez. Tenemos que probar, solucionar y probarlo de nuevo.
- Nada puede con una reacción viva del público; lo importante es conocer lo que quieren nuestros usuarios. Debemos probar y probar con ellos hasta llegar a la decisión definitiva.

Cuando empezamos a diseñar un producto o servicio nunca es pronto para efectuar un test de usabilidad. Aunque únicamente tengamos un prototipo en papel, podemos enseñarlo a los usuarios para conocer qué piensan de él. Además, al no tener diseño, no se distraerán con detalles, sino que se centrarán más en el producto o servicio en sí mismo.

Lectura complementaria

S. Krug (2001). *No me hagas pensar* (pág. 142). Madrid: Prentice Hall.

Algo que se deduce de estos puntos es que tenemos que actuar como científicos y aprender de nuestros experimentos, sean cuales sean los resultados. Si éstos contradicen nuestros modelos mentales, aportarán igualmente información valiosa, si les prestamos la atención adecuada.

A continuación presentaremos los recursos humanos y materiales necesarios para poder llevar a cabo una prueba con usuarios.

1.2.4. Personas necesarias en un test con usuarios

En un test con usuarios participan distintas personas que desempeñan diferentes roles durante la prueba:

- el **usuario**, para llevar a cabo la prueba;
- un **facilitador**, que modera el test, y
- **observadores**, que recogen diferentes aspectos de lo que sucede durante el test.

El usuario

Según la RAE, *usuario* es aquella persona que “usa ordinariamente algo”.

El usuario es el consumidor potencial y habitual del producto o servicio que evaluamos.

La selección de los usuarios adecuados constituye gran parte del éxito final del test, ya que las observaciones que hagamos de los mismos, así como sus opiniones, comentarios y reacciones, serán la materia prima con la que se elaborarán los resultados. Por esta razón, la muestra de usuarios ha de ser lo más representativa posible de quien esperamos que utilice finalmente el producto y servicio que desarrollamos.

¿En qué consiste el rol de usuario? Consiste en:

- Tomar parte en el test, ya sea remoto o presencial.
- Firmar los acuerdos de confidencialidad y cesión vinculados al proyecto.
- Realizar las tareas del test, siguiendo las pautas marcadas por el facilitador.
- Responder a los cuestionarios pre-test y post-test.
- Verbalizar sus pensamientos y reflexiones acerca de la tarea realizada.

El facilitador o moderador

El facilitador o moderador es la persona encargada de conducir el test y acompañar al usuario durante todo el proceso.

El rol de facilitador no es fácil, ya que debe conjugar tres funciones principales:

1) Salvaguardar la integridad física y psicológica de los participantes. Puede parecer exagerado, pero en el fondo, un test de tareas se basa en que un usuario cometa errores delante de un experto; esta situación no es agradable para nadie, por lo que el facilitador tiene que velar por los usuarios:

- Para que se sientan cómodos. Un usuario incómodo no reacciona como lo haría en una situación real, lo que invalida los datos.
- Estar atentos a signos de estrés o cansancio, incluso decidir cuándo se hacen las pausas para que el usuario descanse.
- Aliviar los agobios y tensiones del usuario ante la tarea. Lo cierto es que obtenemos información muy valiosa de los momentos duros por los que pasa el usuario. El facilitador tiene la obligación de aliviar al usuario, transmitirle serenidad en todo momento; transmitirle la idea de que no es un “torpe” por no saber utilizar algo, sino que el producto necesita mejorar y que su ayuda es imprescindible.

2) Maximizar el flujo de información hacia los observadores. A pesar de que los observadores están cerca, suele ser el facilitador el que más información recibe por parte del usuario. El facilitador debe animar al usuario a manifestar lo que piensa y cuidarse de que esta información llegue hasta los observadores. El facilitador:

- Es el responsable de que se cumpla el protocolo de pensamiento manifiesto⁴, también llamado *pensamiento en voz alta*.
- Tiene que subrayar las acciones relevantes del usuario, para que los observadores puedan registrarlas.

Ejemplo

En un test con una página web, el usuario elige una búsqueda avanzada en lugar de la búsqueda simple; el facilitador puede subrayar este hecho simplemente diciendo “Has elegido búsqueda avanzada”. Esto servirá para que los observadores estén atentos y para que el usuario probablemente explique por qué.

- Debe animar a que los usuarios pregunten pero no dar ninguna respuesta: las preguntas del usuario ayudan a determinar qué funcionalidades tienen que mejorarse.

Ejemplo

En una web de compra:

USUARIO: Si toco este botón de aquí ¿voy directo al carrito?

FACILITADOR: ¿Probamos?

⁽⁴⁾En inglés, *thinking aloud*.

Ved también

Podéis ver el concepto de pensamiento manifiesto en el subapartado 1.3.4 de este módulo didáctico.

- Ha de fomentar que el usuario exprese lo que piensa, formulando preguntas abiertas: “¿es lo que esperabas?” “¿qué es lo que intentas hacer?”.
- Debe prestar atención al lenguaje no verbal. Las expresiones del rostro y la postura de los usuarios dicen mucho de lo que pasa por su mente; en ocasiones, hay un observador dedicado a registrar las reacciones emocionales del usuario que recoge este lenguaje no verbal, pero es conveniente ayudarlo a que verbalice lo que pasa.

Ejemplo

Un resoplido del usuario a menudo expresa disgusto; no se trata de preguntar “¿qué te ha hecho resoplar?”, ya que probablemente el usuario ni siquiera es consciente de ello; la pregunta abierta es una buena opción: “¿qué ha pasado?”.

- Tiene que evitar que el usuario se desvíe del tema. Hay usuarios que se expresan en sus propias opiniones sobre lo que se evalúa; hay que recordar que no nos interesan las opiniones del usuario, nos interesa lo que éste consigue hacer durante el test.

Ejemplo

Los usuarios tienen tendencia a generalizar desde lo que ellos opinan hacia el mundo: “Creo que la gente tendrá problemas con esto”; el facilitador debe reaccionar: “¿Tú tienes problemas?”.

- Debe decidir cuándo se acaba una tarea. Si tenemos una tarea en la que el usuario se atasca, el facilitador puede sugerirle pasar a la siguiente. A menudo todos los usuarios de un test se atascan en un mismo punto. Si el equipo ya tiene información suficiente para demostrar que existe un problema, el criterio del facilitador puede ser que no hace falta más.

Ejemplo

Los usuarios más voluntariosos insisten en retomar tareas que dejaron inacabadas si durante el test han comprendido cómo hacerlas; también responde al criterio del facilitador permitirlo. En todo caso, de cara al análisis es muy importante que se recoja que la tarea pudo resolverse en un segundo intento.



O, simplemente, recordarle que puede abandonar la tarea cuando lo crea conveniente.

Lectura recomendada

En relación con el lenguaje no verbal, es interesante revisar el proyecto siguiente:

Eva de Lera; Muriel Garreta-Domingo (2006). “Ten emotion heuristics”.

3) Mantener el máximo grado de integridad en la información. El facilitador debe mantener siempre en mente que un test con usuarios es una prueba orientada a obtener datos fiables. Por tanto, debe velar porque estos datos se contaminen lo menos posible; el facilitador puede decidir que un usuario no cumple con el perfil y que no debe realizar un test, o que una tarea ha quedado invalidada por alguna razón.

La pirámide de la imagen muestra, de la base a la cúspide, una serie de preguntas que el facilitador puede hacerle a un usuario “atascado”. Se trata de no llegar a preguntar “Para qué crees que puede servir...”, ya que es una intervención directa del facilitador en el test. Aun así, queda a criterio de cada facilitador decidir si interviene o no.

Los observadores

Tan importante como el facilitador o el usuario es la labor del equipo de observación. Los observadores son personas encargadas de registrar datos de test para su análisis posterior; el número y el papel de los observadores dependerán de los objetivos que nos hemos fijado en el test, ya que tendrán que recoger unos datos u otros; por ejemplo, puede interesar medir el número de clics que hacen los usuarios, el número de veces que usan el buscador o el número de veces que consultan el manual de ayuda.

Una distribución básica de observadores podría ser:

a) Observador encargado de los tiempos y del resultado de las tareas:

- ¿consigue el usuario llevar a cabo la tarea?
- ¿cuánto tiempo tarda?

b) Observador encargado de las manifestaciones verbales de los usuarios.

c) Observador especializado A: por ejemplo, reacciones emocionales de los usuarios.

d) Observador especializado B: por ejemplo, en una prueba de móviles, qué teclas pulsa por tarea.

Reflexión

Se corta la conexión a Internet en la mitad de la realización de una tarea con una web y se pierden los datos de un formulario ¿qué debemos hacer?, ¿permitir que el usuario vuelva a comenzar?, ¿pedir que pase a la siguiente tarea y desestimar ésta?

Es importante planificar antes de pasar el test y estandarizar la labor de los observadores, así como decidir con antelación los datos que se recogen y en qué formato. Resulta útil celebrar una reunión previa con todo el equipo implicado en el test y plantear preguntas como las siguientes:

- ¿Cuándo consideraremos que el usuario ha realizado con éxito la tarea?
- ¿Cuándo lo consideraremos fracaso?
- ¿Qué pasa si abandona la tarea? ¿Es un fracaso en la tarea?
- ¿En qué formato recogeremos los tiempos de los usuarios?

En un equipo de test llegamos a definir también el “falso éxito” y el “falso fracaso” como datos que pueden ser recogidos, ya que a veces los usuarios resuelven adecuadamente las tareas sin saberlo, o erróneamente sin darse cuenta.

Ejemplo

- ¿Cómo recogeremos los datos?
- ¿Apuntamos la hora del test al lado de cada expresión, declaración o manifestación verbal del usuario?

En cualquier caso, los observadores no son siempre imprescindibles. La participación de los observadores dependerá de los objetivos y nivel de complejidad del test y de los recursos disponibles para su realización.

1.2.5. Laboratorios y herramientas para la evaluación de la usabilidad

Podemos distinguir dos clases de **laboratorios de usabilidad** según el tipo de instalaciones:

- 1) laboratorios fijos o estables, con instalaciones pensadas y adaptadas a la evaluación, y
- 2) laboratorios portátiles, pensados para desplazarse.

Por otra parte, los **componentes del mobiliario, hardware y software** más habituales en un laboratorio de usabilidad suelen ser:

- 1) Elementos de sala, mobiliario y ambiente: mesas y sillas, espejo de observación unidireccional, luces regulables e insonorización.
- 2) Elementos de apoyo para la recogida de datos: sistema de grabación en vídeo, micrófonos, monitores, equipamiento para la grabación en vídeo, ordenador, y software para el apoyo a la recogida de datos.

Reflexión

En una ocasión, al terminar un test, nos encontramos que los observadores encargados de recoger los tiempos habían hecho su tarea de manera no estandarizada: uno ponía a cero su cronómetro en cada tarea, otro anotaba los minutos de grabación del vídeo y otro las horas en que ocurría cada registro. Todos los datos eran válidos, pero hubo que hacer trabajo extra de homogeneización.

Laboratorios de usabilidad

1) Laboratorios de usabilidad estables

Este tipo de laboratorios de usabilidad son, como su nombre indica, espacios diseñados y adaptados especialmente para el desarrollo de pruebas de evaluación de usabilidad. Esta adaptación consiste fundamentalmente en conseguir reproducir de manera realista uno o varios entornos de uso y en dotar al equipo de evaluadores de las herramientas necesarias para llevar a cabo su trabajo.

Reproducir de manera realista significa adaptar el espacio al contexto de uso del producto o servicio interactivo, tal y como interactuaría el usuario con el mismo.

Es por esto por lo que a menudo se divide el espacio del laboratorio en dos salas:

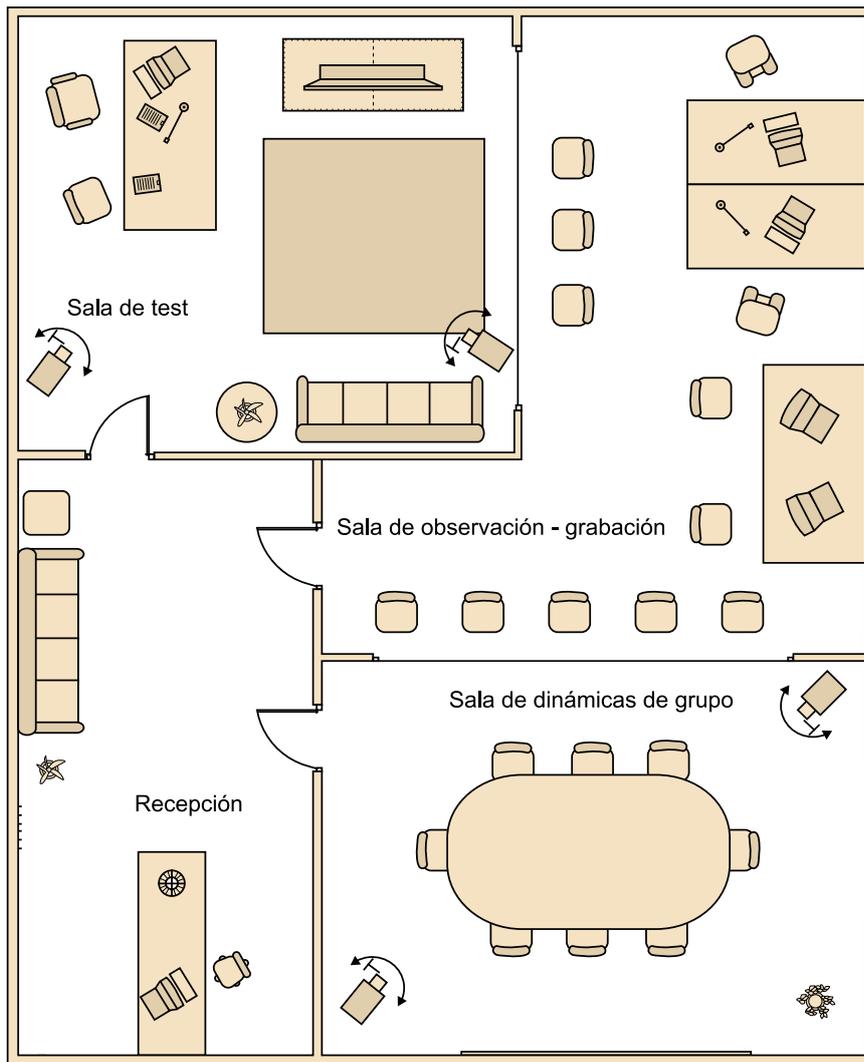
- a) Sala de observación: en ella se sitúa el equipo de observadores, el equipo de grabación y, en ocasiones, un cliente que pueda estar viendo el test.
- b) Sala de test, donde tiene lugar la prueba. A menudo está ambientada como un entorno doméstico –por ejemplo, una sala de estar– o un entorno de oficina.

Entre ambas habitaciones se encuentra instalado un cristal de observación unidireccional o de visión unilateral que permite ver a los usuarios desde la sala de observación mientras efectúan la prueba, aunque no desde el otro lado. Se completa con cámaras de grabación controladas remotamente, que sirven para grabar la prueba o tener planos concretos del usuario.

En algunos laboratorios de usabilidad existen salas adicionales donde se pueden hacer otro tipo de pruebas de usabilidad que requieren más infraestructura.

Cámaras de Gesell

En el ámbito científico estas salas donde se desarrollan los tests de usabilidad se denominan *cámara de observación* o *cámara de Gesell*. El psicólogo y pediatra Arnold Gesell fue pionero en el uso de estas salas en sus investigaciones en el campo de la educación.



Los beneficios de un laboratorio de usabilidad son los siguientes:

- Disponer de una habitación separada nos permite disponer de más gente, aparte del facilitador, que visualice el test, y por tanto, más ojos centrados en los movimientos, los comportamientos y las interacciones del usuario sin que éstos sean conscientes de ello.
- En este tipo de laboratorios fijos o estables el facilitador no tiene que preocuparse por temas técnicos, tales como la grabación de audio y vídeo, o de colocar una cámara enfocada al usuario. Estos laboratorios ya disponen de un equipo técnico montado y permiten tener técnicos especialistas que se encarguen de los detalles de grabación.

Aunque los resultados de realizar un test en un laboratorio fijo o portátil pueden ser muy parecidos, a veces es clave dar a entender a la organización que la usabilidad juega un papel primordial en el desarrollo y diseño y que por ello se construyen laboratorios específicos para su estudio. A veces podemos encontrarnos con clientes que pueden darle importancia a este tipo de aspectos.

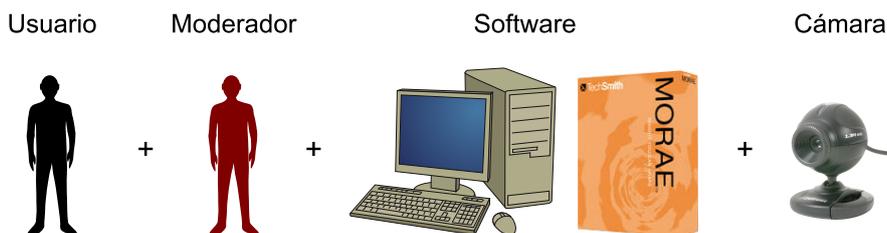
- Al disponer de observadores que toman notas acerca de cómo el usuario se desenvuelve con el producto o servicio evaluado y observan su comportamiento, dejan más libertad al facilitador para poderse centrar en otros aspectos, lo cual quizá no fuera posible en un laboratorio portátil, al encontrarse solos con el usuario, sin nadie que pueda tomar notas.

2) Laboratorios de usabilidad portátiles

Convertir una oficina o una habitación en un laboratorio es muy sencillo. Únicamente necesitamos un ordenador, una cámara y el software disponible para grabar la interacción, el audio y el vídeo del test de usabilidad. De esta manera podemos realizar pruebas donde queramos.

Los beneficios de un laboratorio portátil son los siguientes:

- Mayor autonomía; es más fácil organizar un test de usabilidad portátil, ya que no dependemos de una sala específica para hacer el test. Podemos llevarlo a cabo en cualquier habitación de nuestra oficina, en casa del usuario, etc.
- Ahorro de costes; es más caro hacer un test de usabilidad en un laboratorio fijo que montar un laboratorio portátil.
- Comodidad para el usuario; a veces puede parecer más frío y artificial hacer el test en un laboratorio de usabilidad fijo.
- Adaptación al entorno y contexto de uso; en ocasiones, es importante hacer el test en el entorno real de uso del producto o servicio, que no siempre puede reproducirse en una sala de laboratorio fijo.



Mobiliario, herramientas de hardware y software

1) Elementos de la sala, mobiliario y ambiente

Los elementos de sala, mobiliario y ambiente son los que nos permiten crear el espacio que da acogida a los usuarios en cada prueba.

a) Mobiliario. En la elección de mobiliario, las premisas son:

- **Polivalencia:** es poco probable que dispongamos de una sala donde desarrollar cada método; lo más habitual es que las salas de tests sean polivalentes y su estructura se adapte al tipo de método que aplicamos. Por ello, la elección de mobiliario tiene que permitir que podamos trabajar con un único usuario o con un grupo de usuarios. En este sentido, son útiles las mesas modulares con ruedas o las sillas apilables.
- **Comodidad:** cualquiera de las tipologías de test puede llegar a durar una hora, por lo que el mobiliario debe de ser cómodo y ergonómico.
- **Inmersión:** si las instalaciones lo permiten, podemos disponer de una sala con un entorno inmersivo que reproduzca un hogar o una oficina. Esto implica la inclusión de elementos decorativos: cuadros, plantas, alfombras, etc. La ventaja de un espacio inmersivo es que ayuda a los usuarios a sentirse más “en casa” y propicia que se comporten con naturalidad durante las pruebas.

b) Espejos de observación unidireccional. Hemos hablado de la separación entre dos ambientes por un cristal de visión unilateral. Permiten a un observador mirar sin ser visto, desde una sala a oscuras, lo que ocurre en la sala adyacente iluminada. Son un lujo que no muchos laboratorios pueden permitirse, ya que supone una adaptación ex profeso de la estructura de las salas. Estos cristales espejo no sólo permiten que los observadores de un test estén fuera de la sala, sino que el posible cliente o parte del equipo técnico que desarrollan el producto también puedan asistir a la sesión de evaluación en directo, sin tener que mirar las cintas. No olvidemos que este tipo de grabación debe acogerse a unas normas legales y a unas políticas de uso y privacidad que deben respetarse.

c) Luces regulables. Son muy recomendables si hemos instalado los cristales de observación unidireccional, ya que éstos funcionan iluminando más una sala que la otra; las luces regulables nos permiten no tener que trabajar en la absoluta oscuridad. También son un valioso elemento para reproducir diferentes ambientes: iluminación total para oficinas, luces más tenues para reproducir un hogar, etc.

d) Insonorización. Si tenemos la suerte de contar con más de una sala, es recomendable insonorizarlas. Esto facilitará la presencia de un equipo de observadores y de invitados externos en cada prueba que en sus lugares correspondientes podrán hacer comentarios o tener conversaciones que no afecten al desarrollo de la prueba.

2) Elementos de apoyo para la recogida de datos

Los elementos de apoyo para la recogida de datos sirven para facilitar la recogida de datos durante el test para su análisis posterior, y aumentar así la riqueza de los datos y la documentación.

a) Sistema de grabación en vídeo. Las cámaras de vídeo permiten registrar las sesiones y filmar en detalle aspectos como la cara de los usuarios. Podemos llegar a utilizar dos cámaras, una dentro de la sala donde está el usuario y otra en la sala de observación. La primera podrá registrar la pantalla, la postura del usuario o los movimientos del ratón. La segunda cámara puede tomar planos detalle de la cara, las manos o de cualquier anotación que el usuario tenga que hacer sobre un cuaderno situado en la mesa. Las características técnicas del sistema de grabación en vídeo van a depender del presupuesto con que se cuente, del uso que se les pretenda dar y del nivel de competencia de nuestro equipo humano. Un sistema analógico o digital de grabación necesita un técnico que entienda cómo mezclar las imágenes de diferentes cámaras (utilizando un mezclador, por ejemplo) y sea capaz de hacer un montaje o una posproducción. Si así fuera, no debemos olvidar dar las instrucciones necesarias a dicha persona sobre aquello que queremos recoger.

b) Micrófonos. Pueden instalarse micrófonos omnidireccionales en las salas, tal y como se utilizan en las dinámicas de grupo⁵. Pero para el test con usuarios, es recomendable utilizar micrófonos “de corbata”, que captan perfectamente las palabras del usuario. También es útil que el facilitador cuente con un sistema de comunicación con el resto del equipo de observación (comúnmente, un auricular) para recibir instrucciones o conocer si hay algún problema técnico.

⁽⁵⁾En inglés, *focus group*.

c) Monitores. Disponer de varios monitores conectados al sistema de vídeo permitirá a los observadores ver con detalle lo que sucede en el test, tanto en la pantalla del usuario como en sus reacciones.

d) Ordenador. Un ordenador o varios son necesarios para que el usuario lleve a cabo el test, recoger los datos, grabar digitalmente la sesión, etc.

e) Software de apoyo para la recogida de datos. Este tipo de software posibilita que lo que sucede durante el test pueda ser recogido de manera ordenada y vinculado a las grabaciones de vídeo. Esto tiene varias ventajas:

- Homogeniza la recogida de datos entre diferentes observadores.

- Permite que los datos del test estén disponibles de manera inmediata, sin que haya que “pasar a limpio” las notas.
- Posibilita hacer un seguimiento de los tiempos de grabación de audio y vídeo y una posterior extracción de clips de vídeo o secuencias.
- Ofrece una automatización de las métricas y estadísticas.
- Permite la creación de “marcas” vinculadas a eventos del test. Por ejemplo, una marca que represente cada vez que el usuario suelta una grosería o un improperio; esta marca se configura de manera que se registra vinculada a un tiempo y un momento de la grabación.

Son muchas las soluciones gratuitas o comerciales que se ajustan a éstas y otras características. Algunas de ellas incluso facilitan la grabación a partir de webcam de forma presencial o remota, con lo que podríamos ahorrar costes.

Tabla

Algunas soluciones de apoyo para la recogida de datos

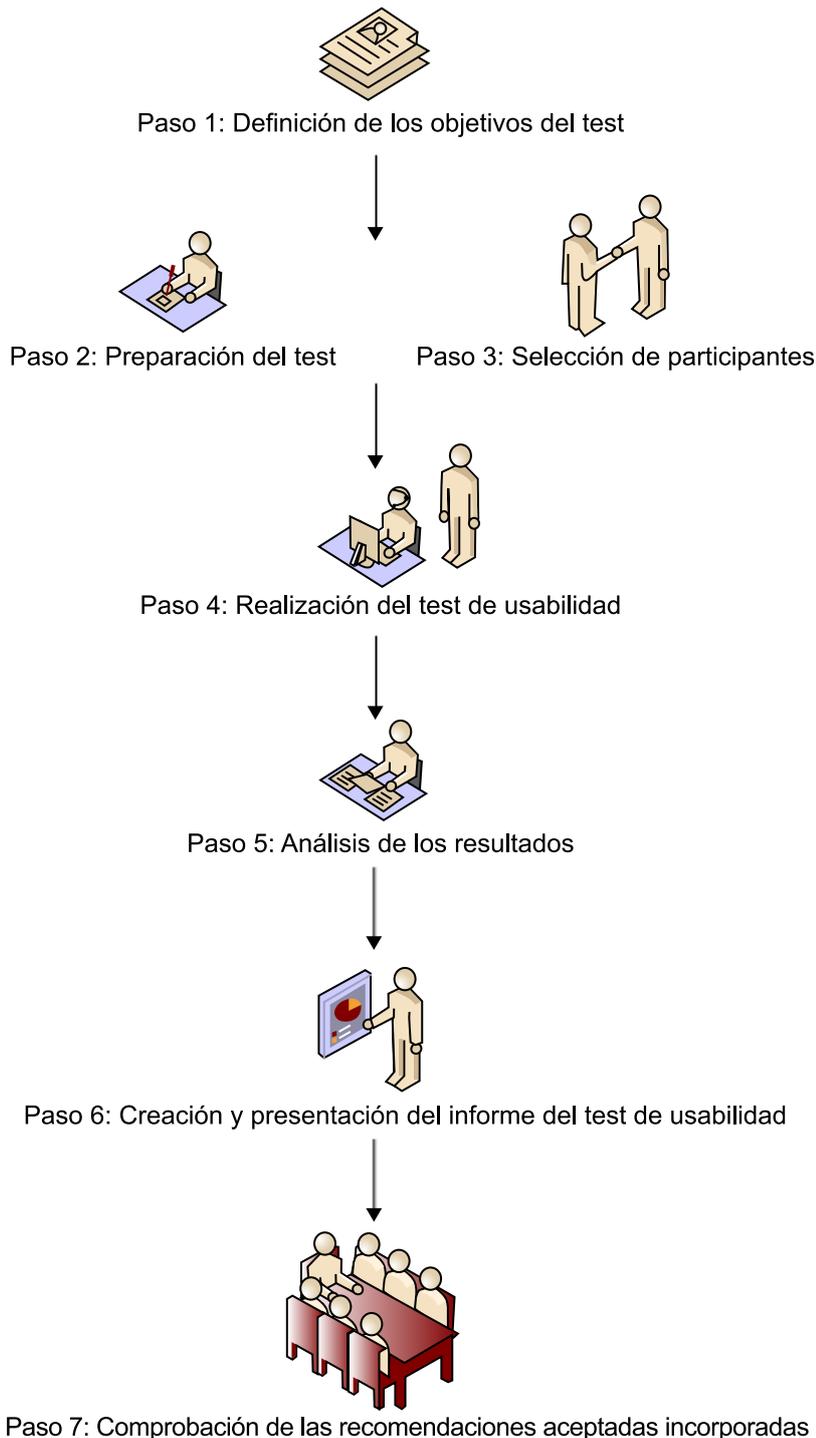
Nombre	Web	Tipo
Silverback 2.0. (Clearleft)	silverbackapp.com	Shareware (30 días)
Morae (Techsmith)	techsmith.com/morae.asp	Shareware (30 días)
ObServant (User Works)	observant.com.au/	Comercial
Usability Activity Log (Bit Debris Solutions)	bitdebris.com	Comercial
The Observer XT (Noldus)	noldus.com	Comercial
Userfly	userfly.com/	Versión Free limitada y versión Premium
UserTesting	usertesting.com	Comercial
Crazyegg	crazyegg.com	Comercial
Loop11	loop11.com	Comercial
Chalkmark (Optimal Workshop)	optimalworkshop.com/ chalkmark.htm	Comercial
Usability Test	usabilitytest.com/	Freeware

No obstante, la ausencia de este tipo de software no es un obstáculo para la realización de un test con usuarios. Puede ser sustituido por una simple hoja de cálculo, un editor de texto y un poco de organización en la recogida de datos.

1.3. Proceso del test

Lo primero que debemos hacer es definir los objetivos del test, es decir, tener claro lo que queremos lograr con la consecución de la prueba de usabilidad. Una vez definidos los objetivos, podemos empezar a preparar el test, planificar todo el procedimiento del test (las tareas que los usuarios deberán realizar, los escenarios ficticios y los cuestionarios) según los objetivos especificados.

Ejemplo de etapas de proceso de un test



Cuando tengamos claro qué queremos evaluar y cómo será el proceso del test, podremos empezar a seleccionar a los participantes, siempre teniendo en cuenta los objetivos definidos y las características del producto o servicio que evaluamos.

Cuando tengamos a los participantes seleccionados, podremos empezar el test. Una vez realizado, deberemos recoger todos los datos recopilados durante el test y analizar los resultados.

1.3.1. Definición y objetivos del test

Para llevar a cabo un test con usuarios es necesario definir previamente los objetivos que se quieren conseguir con la finalidad de aclarar y definir aquello que queremos evaluar del producto o servicio.

El objetivo general de una prueba o test de usabilidad es conocer, con un elevado nivel de detalle, la facilidad de uso que presenta un producto o servicio y el nivel de eficacia, eficiencia y satisfacción que produce en el usuario.

El punto principal de un test de usabilidad es proveer de información durante el proceso de diseño y desarrollo para asegurarnos que el producto o servicio que probamos sea de uso fácil e intuitivo.

A partir de aquí, se pueden proponer objetivos específicos. Los objetivos específicos se traducirán en tareas que los usuarios participantes en el test realizarán. Los objetivos del test han de ser específicos y razonables para que las tareas que se desprendan de ellos sean cercanas a los usuarios a los que se dirige, factibles y con una duración adecuada. En resumen, en primer lugar hay que definir claramente los objetivos del test, una vez definidos, se definirán las tareas a realizar por los usuarios.

A continuación pasamos a explicar ámbitos de actuación que cabe considerar a la hora de marcar unos objetivos para el test y la posterior definición de tareas precisas.

a) Eficacia y eficiencia en la navegación

En este punto nos interesa saber si los usuarios encuentran lo que necesitan fácilmente. Se trata de descubrir si los usuarios encuentran un patrón de navegación que encaja con su modelo mental, es decir, si las opciones de navegación son lo suficientemente claras para encontrar lo que buscan, si las categorías son oportunas, si el lenguaje utilizado es cercano al usuario, etc.

Reflexión

Nótese que buena parte de los ejemplos que se proponen a continuación corresponden a la evaluación de la usabilidad de sitios web, por ser el entorno más evolucionado y propicio para efectuarlas. No obstante, otros entornos, tecnologías, dispositivos, productos y servicios también están sujetos a evaluaciones de usabilidad.

Debemos preguntarnos lo siguiente:

¿Los usuarios encuentran lo que necesitan de manera intuitiva, fácil y rápida?

b) Contenido útil y práctico

Es necesario conocer la información que el usuario necesita realmente y organizarla de manera clara para que la encuentre fácilmente. Un usuario no tiene la necesidad de leer todo el contenido de un sitio para encontrar la información, les debemos ayudar a encontrarla de la manera más útil y cómoda posible.

Debemos preguntarnos lo siguiente:

¿Qué información es la que quiere el usuario?

¿Hemos organizado la información de manera clara para que los usuarios la encuentren rápidamente?

c) Claridad en la presentación

El diseño gráfico, la fuente, el tamaño de la navegación así como el contenido son elementos que pueden ayudar al usuario o, por el contrario, pueden distraerlo y crear una barrera entre este y la información.

Debemos preguntarnos lo siguiente:

El diseño, los colores, la fuente... ¿ayudan al usuario o, por el contrario, lo distraen y hacen que se sienta incómodo?

d) Ratios de las tareas

En un test con usuarios se crean una serie de tareas para comprobar el grado de eficiencia del producto. Conseguir que los usuarios realicen las tareas de manera satisfactoria nos aportará información válida para el desarrollo del producto o servicio y nos ayudará a saber cómo se han sentido los usuarios al realizarlas.

Debemos preguntarnos lo siguiente:

¿Los usuarios han sido capaces de realizar las tareas planteadas? Si es así, ¿cómo se han sentido?

Es importante tener estos objetivos en mente durante todo el proceso para no desviar el foco del proyecto. Teniendo claros los objetivos del test, los resultados serán mucho más precisos.

1.3.2. Preparación del test

En este subapartado definiremos y organizaremos todos los elementos que se utilizarán durante el test. Estos elementos formarán el guión de nuestra planificación.

Cuestionario pre-test

Un primer cuestionario ayuda a recopilar especialmente datos sociodemográficos que nos van a ayudar en el proceso de análisis de los resultados. Las consideraciones sociodemográficas están siempre presentes en el cuestionario pre-test tales como nombre, edad, sexo, empresa, cargo... El resto de preguntas del cuestionario dependerán de la aplicación que evaluemos.

Usabilidad de una tienda de ropa en línea

Imaginemos que efectuamos un test de usabilidad de una tienda de ropa en línea. Si no tenemos información suficiente sobre el grupo de participantes, podríamos formular algunas preguntas sencillas relacionadas con sus datos personales y situación laboral. A continuación podremos trabajar sobre cuestiones específicas para el test de usabilidad.

Algunos ejemplos:

- Frecuencia de conexión a Internet:
 - ¿Cuántas horas a la semana te conectas a Internet? 1-2 horas/semana;
 - ¿Desde dónde te conectas habitualmente?
- Compras en línea:
 - ¿Has comprado alguna vez en línea?
 - ¿Me puedes especificar las webs donde has comprado?
 - ¿Cada cuánto sueles comprar?
 - ¿Con qué frecuencia accedes a Internet?
 - ¿Cuándo fue la última vez que compraste algo?
 - ¿Has efectuado alguna compra mediante tu teléfono móvil?

Antes del inicio de la prueba se suele dejar que cada usuario haga una pequeña exploración (durnate unos minutos) y empiece a familiarizarse con el producto o servicio. Una vez finalizados estos minutos podemos preguntarle sobre esta primera exploración.

Escenarios y tareas

El **escenario** es la descripción de un personaje en una situación y contexto de uso con unas tareas y actividades asignadas. Colocamos a los usuarios en situación y les damos los datos necesarios para que asuman el papel y actúen en consecuencia.

No olvidemos que el contexto hace referencia a las condiciones bajo las cuales un producto interactivo va a ser usado. De esta manera debemos ofrecer la suficiente información al usuario como para lograr que se sitúe y se adecue. Por ejemplo, si vamos a efectuar un test con usuarios para una compañía telefónica con usuarios que no pertenecen a la compañía debemos pedirles que actúen como si lo fueran, y aportarles la información necesaria para ello.

Según la aplicación

Si preparamos un test de usabilidad de una página de compras en línea, haremos preguntas tales como: frecuencia de conexión a la página, número de compras que se efectúan por mes...

Ejemplo de escenario

Imaginad que sois clientes de la compañía X y un amigo os comenta que desde la web de la compañía podéis llevar a cabo muchas de las acciones que sois haciendo por teléfono.

Las **tareas**, por su parte, son el punto principal en el guión de un test con usuarios. Son las acciones que se le pide al usuario que realice con el producto o servicio a lo largo del test.

Ejemplo de tarea

Accede a la página web de la compañía telefónica X e intenta obtener datos sobre vuestro consumo durante el mes de enero.

¿Cómo elaborar escenarios?

Para elaborar los escenarios es imprescindible conocer y tener bien detectados los objetivos que se pretenden extraer del test de usabilidad, así como las funcionalidades o características que queremos analizar del producto o servicio.

Los tests de usabilidad son pruebas que nos ayudan a evaluar la calidad y la medida en que un producto o servicio es usable para el público al que se dirige. El usuario que realiza el test puede ser una persona que conozca o no el producto o servicio, pero sea como sea debemos ponerlo en situación con el fin de que los resultados del test sean lo más fiables posibles. Por ello, al pedirle que haga una serie de acciones que quizá no realiza en su día a día, debemos dotarlas de un sentido completo.

Por lo general los escenarios se utilizan para introducir las tareas y crear un ambiente imaginario que de sentido a la acción.

Por ejemplo, al introducir al usuario una tarea hay una gran diferencia entre:

- Sin escenario: “Accede a la página principal de una web de compras en línea y encuentra un libro que quieras comprar” (tarea).
- Con escenario: “Imagina que el martes de la semana que viene es el cumpleaños de tu madre y sabes que le haría ilusión tener la última novela de un gran escritor. Además, sabes que a tu madre le gustan los libros de bolsillo porque acostumbra a leerlos en la playa o en el autobús. ¿Podrías acceder al sitio web de compras en línea X y tratar de encontrar el libro que quiere?” (tarea).

Como vemos, en el ejemplo sin escenario, el usuario no se ve envuelto en ningún contexto, por lo que llevará a cabo la tarea sin ningún tipo de motivación y sin tener información suficiente que le permita dar sentido a dicha tarea. Por el contrario, como vemos en el segundo ejemplo, que incorpora un escenario, el usuario se ve implicado en un contexto y ya dispone de una motivación suficiente para realizar la acción.

Cuando tengamos las características que queramos evaluar y las tareas que en consecuencia queramos que los usuarios lleven a cabo, podremos dotar a la situación de un contexto, de un escenario.

¿Cómo elaborar tareas?

A la hora de elegir qué evaluar, el propio cliente para el que trabajemos nos puede proporcionar pistas que nos ayuden a generar las tareas y marcar los aspectos que pueden ser críticos en su servicio o producto.

En otras ocasiones, es importante contar con la opinión de un usuario experto o celebrar reuniones de equipo previas para saber qué tipo de tareas son más habituales y qué tipo de tareas son críticas.

Las tareas de un test de usabilidad tienen que basarse en las características siguientes:

- **Razonables:** la tarea debe ser razonable, es decir, debe reflejar algo que es normal hacer para el usuario, nunca algo que no se haría nunca. De lo contrario el usuario no podrá sentirse lo suficientemente integrado en el papel y no realizará correctamente la tarea.

Ejemplo

Razonable: "Comprar dos entradas para un concierto."

No razonable: "Comprar dos entradas para seis conciertos."

- **Orientadas a la tarea:** cuando un usuario está solo en su casa o trabajo haciendo uso de un producto o servicio, lo hace por algún motivo. Para el usuario ese producto o servicio es una herramienta con la que interactuar, ya que le proporciona algo a cambio. Al poner al usuario en un contexto irreal, en un laboratorio de usabilidad, ajeno a su vida cotidiana, debemos lograr que sienta que la acción que está realizando tiene un objetivo. La motivación del usuario cuando realiza una tarea es esencial. Por tanto, debemos dar un escenario lo más realista posible.

Ejemplo

Orientada a tarea: "Comprar un CD nuevo para un familiar cercano porque te has dado cuenta de que el que se compró el año pasado está estropeado y no se oye bien."

No orientada a tarea: "Comprar un nuevo CD."

- **Específicas:** para ser consistentes y centrarnos en la tarea, ésta debe tener un objetivo final específico. De esta manera el usuario entenderá mejor la acción que debe realizar y tendrá una motivación específica. Debemos ir con cuidado para no utilizar términos o nombres incluidos en el producto o servicio analizado, ya que podemos, sin querer, generar pistas al usuario y alterar los resultados del test. Por ejemplo, si la acción que el usuario debe realizar es acceder a la pestaña llamada "Contratos", debemos intentar no decir la palabra contratos, sino formular la tarea de manera que no incluya esta palabra.

Ejemplo

Específica: “Compra el último libro de Pérez-Reverte.”

No específica: “Compra un novela de aventuras.”

- **Factibles:** no se debe proponer una tarea imposible al usuario ya que no ayuda a mejorar el diseño y puede causarle frustración. Un caso ligeramente exagerado sería pedirle que busque el precio de un piso en un pueblo de Madrid aun sabiendo que no va a encontrar ninguno porque la base de datos no refleja esa población. Estamos evaluando el diseño, no al usuario.
- **Secuencias de tareas realistas:** las tareas deben reflejar el orden de trabajo natural con la interfaz. No debemos inventarnos secuencias que no se correspondan con la realidad del producto o servicio.

Ejemplo

Si pensamos en una web de vehículos de ocasión, la secuencia realista sería navegar por las categorías ofrecidas o efectuar una búsqueda para después filtrar y seleccionar. Pedirle que encuentre un vehículo de baja cilindrada con tubos de escape cromados sería salirse de la secuencia –de los objetivos– y no conseguiremos analizar la fluidez del usuario con el producto o servicio.

- **No especializadas:** al proponer tareas debemos probar algo que todo el mundo pueda hacer o conocer, no algo sobre lo que alguien sepa mucho. Una tarea debe ser algo que el usuario conozca pero hay que procurar que no sea un experto en ese campo. Para ello debemos evitar elegir usuarios que se adapten con excesiva rapidez a los esquemas y modelos mentales ofrecidos por el producto o servicio que se evalúa, utilizando las entrevistas previas.

Ejemplo

Si estamos realizando un test sobre motores de aviones, en la fase de reclutamiento podemos preguntar: ¿cuál es tu nivel de conocimiento sobre aviones?

- **Duración razonable:** generalmente las tareas que deben realizarse no son muy complejas, por tanto el usuario no debe tardar más de diez minutos en completar cada una de ellas. Debemos repartir bien el tiempo por tarea. Si el test dura 90 minutos, las tareas deberían ser de 10 o 12 minutos de duración. De esta manera las tareas se pueden dividir en subtareas y podemos incluso tener algunas opciones en reserva por si viéramos que la prueba va bien de tiempo y los usuarios se encuentran cómodos.

Para la realización de las pruebas de usabilidad es recomendable crear una tabla donde se colocarán todas las tareas. Así, los observadores pueden ir escribiendo al lado de cada tarea los comentarios que suscitan las acciones que el usuario lleva a cabo y otras notas.

Para cada característica que queramos evaluar, debemos generar una tarea y con todas ellas obtendremos el guión de tareas para la prueba.

Ejemplo

Objetivo de la tarea: “Queremos comprobar si los usuarios utilizan el buscador y si su ubicación es la adecuada.”

Formulación de la tarea: “Se te ha roto el aspirador de casa y necesitas comprar uno nuevo. ¿Cómo harías para obtener el mismo modelo del listado de todos los aspiradores disponibles?”

Quizás algo que haya que valorar acerca de las tareas son las circunstancias de uso de una interfaz. Vamos a hablar del tiempo de dedicación, de la motivación del usuario y de otros aspectos cruciales. Pero no debemos olvidar que la tarea viene dada por unas circunstancias específicas que también debemos definir. Por ejemplo, comprar un billete con tres meses de antelación no es lo mismo que encontrarse en la necesidad de comprar un billete de un día para el otro. El comportamiento no será el mismo y se acercará más o menos a una situación real dependiendo del sitio, la página o las tareas que evaluemos.

Una vez tengamos seleccionadas las tareas, es recomendable efectuar una o varias pruebas con algún compañero de equipo para comprobar si el test es demasiado largo o si las tareas son adecuadas y comprensibles.

Ejemplo de documento para un test de evaluación

A continuación, se muestra un ejemplo del tipo de documento que se puede utilizar en una prueba donde se recogen las tareas-escenarios, las métricas y posibles comentarios.

Escenario	Tarea	Éxito / Fracaso	Tiempo	Comentarios
Imagina que la semana que viene es el cumpleaños de tu madre y sabes que le haría ilusión tener la novela de Pérez Reverte de la que tanto habla.	“¿Podrías acceder a la web X y tratar de encontrar la última novela de Arturo Pérez Reverte?”	Se trata de anotar si el usuario fue capaz de realizar la tarea encomendada.	[hh:mm]	Para anotaciones que puedan ser interesantes respecto a la tarea; por ejemplo: encuentra una novela de Reverte pero no es la última.
	“¿Puedes comprar el libro?”			
	“¿Puedes comprobar en qué fecha se entregará el libro?”			
Hace poco has leído <i>Los pilares de la Tierra</i> y quieres dejar un comentario para compartir tu opinión con otros lectores.	“Busca el libro en la web.”			
	“Da una puntuación al libro.”			
	“Contesta al comentario de otro usuario.”			

Métricas

Las **métricas** son datos cuantitativos y cualitativos que recogemos durante el test. Algunos ejemplos de métricas posibles pueden ser:

- La velocidad con la que el usuario realiza una tarea (tiempo).
- La cantidad de errores que comete (número de errores).
- La frecuencia con que rectifica el error (número de rectificaciones).
- La cantidad de usuarios que completan la tarea satisfactoriamente (número de usuarios).

Los objetivos numéricos no son la meta del diseño. Uno de los grandes errores es pensar que si hay algo que no se puede medir, entonces no importa o no cuenta.

No hay un conjunto universal de métricas que sirvan para todas las pruebas, sino que éstas dependen de los objetivos concretos que se hayan definido y del tipo de producto o servicio que se evalúa.

A diferentes dispositivos, diferentes métricas. Por ejemplo, una web o un dispositivo móvil pueden compartir la métrica de tiempo por tarea, pero mientras que en una web puede interesar la observación del uso del botón de regreso (*back*), en un dispositivo móvil esta opción no puede existir como tal; en cambio, podemos observar y registrar si el usuario emplea la marcación numérica o por voz.

Apuntamos aquí una división simple de las métricas y algunos ejemplos habituales:

a) Cuantitativas

- Éxito: si el usuario es capaz de realizar la tarea
- Fracaso: si el usuario no es capaz de realizar la tarea
- Tiempo: tiempos de cada usuario por tarea

La definición de “éxito” y “fracaso” es más difusa de lo que parece. Si un usuario completa todos los pasos de una tarea excepto el último ¿podemos hablar de fracaso?

Algunos profesionales otorgan porcentajes de éxito a fracaso a cada usuario y tarea. Aun así, el grado de refinamiento con respecto a cualquier métrica que queramos recoger es una decisión que el equipo evaluador debe contemplar para cada test.

Refinando un poco más el éxito y el fracaso, podemos intentar recoger métricas de los casos en los que los usuarios creen haber completado correcta o incorrectamente la tarea. Se pueden recoger métricas como “falso éxito” y “falso fracaso” a partir del pensamiento manifiesto del usuario, donde verbaliza si ha terminado o no la tarea.

b) Cualitativas

- **Observaciones:** anotaciones sobre las dificultades, los comportamientos inusuales o las situaciones en las que la causa del error no es obvia.
- **Comportamiento:** acciones efectuadas por el usuario que permiten entender la ejecución de la tarea. Por ejemplo, en un test con una web, la ruta que un usuario toma para resolverla: Home > opción 1 > opción 2...
- **Literales:** opiniones subjetivas sobre la experiencia y la interfaz expresadas por los participantes, o comentarios que acompañan la acción y forman parte del protocolo de pensamiento manifiesto.
- **Análisis de la experiencia emocional:** anotaciones sobre las reacciones emocionales de los usuarios, tales como exclamaciones de sorpresa, disgusto, expresiones de frustración, de alegría, etcétera. Este tipo de observaciones emocionales son profundamente subjetivas y conviene que el observador que las recoja tenga bien tipificados los rasgos emocionales que debe registrar antes de concluir que se ha producido una emoción u otra.

Cuestionario post-test

Al final de la sesión de test debemos dejar un cierto tiempo (como mínimo 10 minutos) para que el usuario exprese su opinión sobre el producto o servicio. Para ello es necesario disponer de un cuestionario que nos ayudará a guiar estas preguntas y conocer así el grado de satisfacción del usuario. Preguntas del tipo “¿Qué es lo que más te ha gustado?”, “¿Qué es lo que menos te ha gustado?”, “¿Qué te ha parecido confuso?”... ayudan a conocer la opinión global del usuario. Es decir, este último tipo de preguntas están centradas especialmente en evaluar la usabilidad subjetiva, la percepción que ha tenido el usuario del sitio web y de la tarea.

El cuestionario post-test puede realizarse a modo de preguntas cerradas o preguntas abiertas. Por lo general, en ambos casos, al final siempre se deja un tiempo para que el usuario exprese observaciones o comentarios abiertos.

Ejemplos de cuestionarios post-test

Preguntas cerradas:

- A nivel general, ¿crees que es fácil utilizar el sitio web?

Lectura recomendada

Leed el artículo siguiente:
Eva de Lera; Muriel Garreta-Domingo (2007). “Ten emotion heuristics: guidelines for assessing the user's affective dimension easily and cost-effectively”.

Poco 0 1 2 3 4 5 6 7 Mucho

- Tu impresión general sobre el uso del sitio web es:

Pésima 0 1 2 3 4 5 6 7 Excelente

- ¿Has sabido en qué página estabas en todo momento?

Nunca 0 1 2 3 4 5 6 7 Siempre

- ¿Consideras que el diseño gráfico te ha ayudado a encontrar lo que buscabas?

Nunca 0 1 2 3 4 5 6 7 Siempre

- ¿El sitio web te ha permitido cumplir con tus tareas?

Poco 0 1 2 3 4 5 6 7 Mucho

Preguntas abiertas:

- ¿Qué es lo que más te ha gustado del sitio web?
- ¿Qué es lo que menos te ha gustado del sitio web?
- ¿Qué te ha parecido el diseño gráfico?
- ¿Crees que utilizarías esta página en otras ocasiones?

Carta de autorización

La **carta de autorización** es un documento en materia de protección de datos y privacidad del usuario, donde se requiere su consentimiento previo para capturar y almacenar información, y que aclara la finalidad de su tratamiento y el uso que se le va a dar.

Cada usuario debe firmar esta carta, que autoriza al facilitador a tomar notas, grabar audio y vídeo, y confirma que la información es confidencial para la empresa encargada de llevar a cabo el test.

Ejemplo de carta de autorización

Yo _____, con DNI número _____ autorizo la captación de la imagen, voz y testimonio, a _____, con domicilio en _____ y NIF/CIF _____, las cuales podrán ser utilizadas con la única y exclusiva finalidad de análisis del estudio del que he sido informado.

En caso de revocación del consentimiento, se deberá avisar por escrito a la siguiente dirección postal:

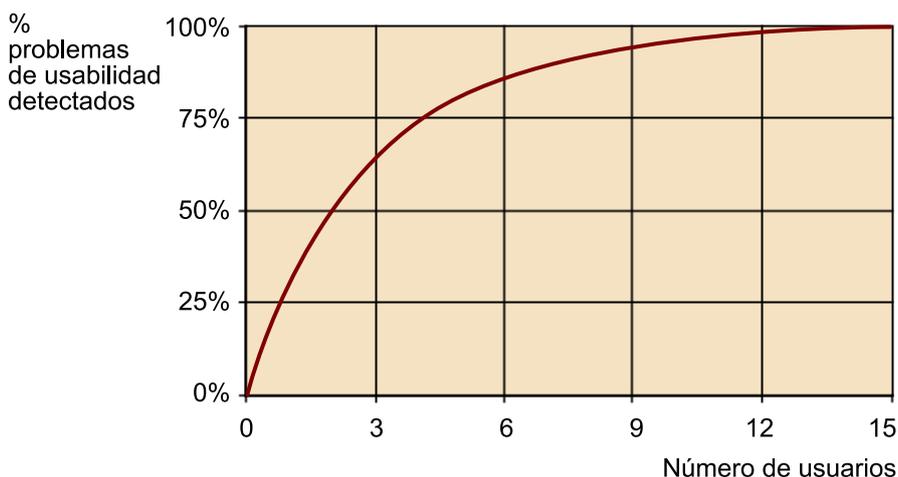
XXXXXX

También acepto que toda la información intercambiada es propiedad exclusiva de XXXX.

Firmado:

1.3.3. Captación de los participantes

Se ha conjeturado mucho sobre el número adecuado de usuarios al realizar un test de usabilidad, pero los resultados obtenidos en diferentes pruebas concluyen que un número elevado no se corresponde con mejores resultados. Nielsen (2000) considera que un número óptimo de participantes en un test con usuarios debe ser al menos de 5. Incluso recomienda que en lugar de trabajar con 15 usuarios en un solo test, es mejor llevar a cabo tres pruebas en diferentes momentos del proceso con 5 usuarios.



No se trata de dar solución al 100% de los problemas o de buscar todas las debilidades del sistema. El propósito es mejorar el diseño y de forma iterativa ir dando respuestas a la mejora de la usabilidad.

Para empezar a reclutar usuarios para un test con perfiles adecuados, debemos tener muy claro quién es el público objetivo del sitio web.

Aunque nosotros mismos podemos encargarnos del reclutamiento, poniendo anuncios en páginas concretas, hay empresas especializadas en este tipo de servicios que agilizarán el proceso. Una vez seleccionados los participantes debemos ponernos en contacto con ellos para acordar un día, un lugar y una hora para realizar el test.

Generalmente, en los tests con usuarios se ofrece una gratificación por el tiempo destinado. Estas gratificaciones por colaboración ayudan y motivan a los usuarios, algo que no debemos perder de vista en ningún momento.

1.3.4. Ejecución del test: ejemplo práctico

Según las explicaciones anteriores, una prueba de usabilidad podríamos dividirla en seis pasos o fases de desarrollo.

Primer paso: Bienvenida y explicación del test

Antes de empezar con el test de usabilidad, con el cuestionario y las tareas, debemos crear un ambiente cercano y confortable. El usuario se enfrentará sólo a la tarea aunque haya gente a su alrededor y se sentirá observado ante una persona desconocida (facilitador) y una cámara o webcam que estará grabándole. Es por ello que debemos romper el hielo para que se sienta cómodo e intente actuar con toda normalidad. De lo contrario los resultados podrían salir alterados.

Debemos empezar por explicar que el objetivo de la prueba es evaluar la calidad de uso del producto o servicio, nunca se evalúa al usuario. Tenemos que reiterar que si el usuario comete algún error durante el test no será culpa suya, sino del diseño. Es de suma importancia que esto le quede muy claro al usuario.

También se le tiene que insistir en la aplicación del protocolo de pensamiento manifiesto durante el test. Esta técnica es la que permite al equipo de evaluadores conocer (o por lo menos, acercarse un poco a) lo que el usuario piensa mientras ejecuta las tareas (usabilidad percibida); por tanto, es útil para detectar qué le llama la atención, qué le motiva a ejecutar cada acción, qué es lo que no comprende, etc.

Se debe explicar al usuario que no estamos en su cabeza y, por tanto, no sabemos qué piensa en cada momento. Escuchar sus opiniones sobre nuestro producto o servicio durante todo el test implica decir en voz alta todo lo que se le pase por la cabeza. A lo largo del test podemos ayudarlo con frases como: ¿qué opinas?; ¿qué ves ahora?; ¿qué te parece esta nueva pantalla?...

Ejemplo de público objetivo

Si vamos a desarrollar una intranet para una empresa, deberemos preguntar a nuestro cliente por usuarios que hagan uso de la misma, ya que éstos serán los usuarios potenciales de nuestro test.

Una buena forma de explicar el pensamiento manifiesto al usuario es mostrarle un ejemplo práctico de lo que esperamos de él.

Además, debe quedar muy claro que la función del facilitador es la de observar, en ningún caso ayudar, al usuario. Por tanto, el usuario debe saber que el facilitador no podrá apoyar sus decisiones en la consecución de las tareas.

Por último, es importante dejar claro el tiempo que durará el test y mantenerlo. Debemos respetar los horarios de los participantes.

Segundo paso: Agenda

Explicar al usuario los pasos que vamos a seguir durante el test para situarlo.

Ejemplo

“Primeramente te haré una serie de preguntas antes de empezar con el test para conocerte un poco mejor. A continuación te pediré que realices una serie de tareas.

Te iré poniendo un conjunto de escenarios ficticios para ponerte en situación y tú deberás resolver las tareas que te vaya proponiendo. Recuerda que yo no puedo ayudarte.

Si no sabes cómo resolver la tarea, me lo comentas y pasamos a la siguiente. No te preocupes si no la haces bien o no la acabas, ya que, como te he comentado anteriormente, no te estoy evaluando a ti, sino a la aplicación.

Al acabar con las tareas te haré unas preguntas relacionadas con el test para tener una opinión global.

Este proceso suele durar una hora aproximadamente así que cuando quieras, empezamos.”

Tercer paso: Cuestionario pre-test

El cuestionario pre-test puede ser rellenado por el usuario o preguntado por el facilitador.

Recomendamos que sea el facilitador quien formule las preguntas e intente conocer lo máximo posible al usuario, de esta manera también puede aprovechar este momento para acercarse al usuario y romper el hielo, si todavía no se siente cómodo.

Cuarto paso: Tareas

Indicamos al usuario que vamos a empezar, realizando una serie de tareas que deberá llevar a cabo.

Primeramente explicaremos al usuario el producto o servicio que vamos a analizar y dejaremos que éste acceda a él y dedique unos minutos a observarlo, tocarlo..., en definitiva, a familiarizarse con él. Al cabo de unos minutos le preguntaremos qué es lo que ve, de qué cree que se trata, cuál cree que es su objetivo y su opinión sobre el primer vistazo.

En este sentido se puede aprovechar este primer contacto del usuario con la página para efectuar una breve tarea previa denominada “test de los 5 segundos”. Consiste en un sencillo ejercicio que puede aportar información sobre la página del sitio web que mostremos inicialmente al usuario. Durante 5 segundos podrá ver dicha página y, pasado ese breve tiempo, tendrá que minimizarla, ocultarla o apagar la pantalla. Después le pediremos que nos cuente todo lo que recuerda, su primera impresión y qué tipo de contenido cree que contiene el sitio web.

Toda la información que nos ofrezca el usuario nos permitirá conocer aspectos relacionados con el diseño de la página y con la jerarquía visual del contenido. Si es capaz de dar una explicación coherente del propósito de la misma, destacar lo que resulta más relevante, y ello se corresponde con los objetivos de la página, tendremos un primer dato sobre la comunicación de nuestro sitio y sobre la ordenación visual y la claridad de los contenidos.

Se trata de una técnica rápida y fácil de aplicar que podemos llevar a cabo antes o después de situar a los usuarios en un contexto determinado.

Una vez realizada esta primera parte procederemos a analizar la facilidad de uso del producto o servicio mediante tareas y escenarios concretos.

Durante el transcurso de las tareas podemos interrumpir al usuario, si lo creemos necesario. De la misma manera, si percibimos que el usuario no sabe finalizar una tarea, le diremos amablemente que no hay problema y pasaremos a la siguiente. Debemos recordar que un usuario nunca debe pensar que hace mal una tarea. Es el producto o el servicio el que genera el error, o las personas que lo diseñaron.

Si vemos que no se expresa, debemos recordarle que es importante que nos cuente sus impresiones con tal de saber lo que piensa.

Quinto paso: Cuestionario post-test

Una vez los usuarios han completado las tareas, podemos pasar al cuestionario post-test para evaluar la opinión que tienen del producto o servicio y para que expliquen cuáles han sido sus percepciones y sensaciones durante la prueba. Podemos aprovechar para solicitarles cualquier otro comentario u observación que deseen añadir.

Sexto paso: Firma de la autorización y entrega de la gratificación

Una vez concluido el test, les entregaremos la carta de autorización y, en los casos que proceda, se hace entrega de la gratificación o de un pequeño obsequio o regalo que tenga relación con la página, con la empresa, o que simplemente permita recordar al usuario su participación en la prueba.

Lectura complementaria

Sobre el test de los 5 segundos, podéis consultar la obra siguiente:

C. Perfetti (2005). “5-Second Tests: Measuring Your Site's Content Pages”. *User interface engineering*.

Ejemplo

“Por último, te pido que leas y, si estás de acuerdo, firmes este papel conforme los datos obtenidos en esta prueba, así como el audio y el vídeo. Es nuestra política de privacidad, que permite hacerte saber que no podemos hacer uso de la información para otra cosa que no sea cubrir los objetivos y propósitos del proyecto en cuestión.

También te hacemos entrega de un obsequio como agradecimiento por habernos dedicado tu tiempo.

Muchas gracias”

1.3.5. Consejos para ser un buen facilitador de test

Ya que el rol de facilitador es fundamental en un test con usuarios, a continuación presentamos una serie de consejos prácticos para desempeñarlo.

1) Ser consciente de los prejuicios y proyecciones

Un facilitador tiene dos mantras: “no estamos juzgando al usuario” y “no hay buenas ni malas respuestas”. Creer en estas dos premisas constituye la mitad de la pieza clave de su labor; actuar en consecuencia es la otra mitad.

El comportamiento del facilitador, implícito o explícito, influye sobre el usuario.

La dedicación profesional plena a la usabilidad tiene, como en otras muchas profesiones, un trato directo con clientes y usuarios. Es natural que haya gente con la que no exista afinidad, incluso hacia la que se pueda sentir cierto rechazo por su actitud. Esto, que es un hecho, no debe evidenciarse en ningún momento. Todos los usuarios son iguales ante el facilitador.

Es tan difícil controlar el agrado como el desagrado; ésta es la razón por la cual también se desaconseja que un facilitador modere un test con familiares o amigos. Entre la gente que se conoce existen todo tipo de condicionantes emocionales que contaminan el test.

De la misma forma, el sentimiento no puede ser mutuo. No hace falta convertirse en el mejor amigo del usuario, pero tampoco hay que dejar que nos rechace, ya que el éxito del test radica en que confíe en nosotros como facilitadores y en que se esfuerce con nuestra sola presencia en la ejecución de las tareas.

Un test es una evaluación, al fin y al cabo. Cuando miramos con “ojo crítico” lo que otra persona hace, desde el punto de vista psicológico se produce una proyección de nuestras percepciones: vemos en el otro aspectos que nosotros podríamos evitar o que haríamos de otra manera.

¿Qué se puede hacer al respecto? Ser consciente de esta realidad es la única forma de tomar medidas para evitarlo y comportarnos de la forma más natural y neutra posible.

En términos prácticos, una forma de frenar los prejuicios del facilitador es evitar respuestas que implican valoración: comentarios como “Excelente” o “Muy bien”, como reacción al protocolo de pensamiento manifiesto, pueden sustituirse por otros más neutros como “Entiendo” o “De acuerdo”. Incluso las preguntas que incidan en las respuestas del usuario pueden ayudar a que siga sus comentarios: “¿Entonces, decís que el menú de navegación no os funciona?”

2) Ser transparente y ganarse la confianza del usuario

Imaginad que en pleno test con usuarios se produce un error técnico en la grabación. Pánico en la sala de observación. Al facilitador, que está en comunicación directa con su equipo de observadores, le gritan que hay un problema técnico con la cinta y que no se está grabando el test. El facilitador se remueve inquieto desde su puesto junto al usuario, ¿qué hacer?, ¿detener la prueba?, ¿seguir?; el usuario nota que el facilitador está preocupado y se pregunta ¿tan mal lo estoy haciendo?, ¿le estoy aburriendo tanto?

El usuario no es ajeno al comportamiento del facilitador, por lo que conviene que éste sea sincero y transparente. Por ejemplo, cuando sucede algo poco habitual, como es el corte de una grabación o un corte en la conexión a Internet, por poner dos ejemplos, pueden ser transmitidos con naturalidad al usuario, indicando que vamos a hacer una breve pausa para solucionar el problema.

Incluso puede capitalizarse un comportamiento no previsto. Si la conexión a Internet falla, podemos recopilar más información y preguntarle al usuario qué haría en un caso así y cómo afectaría a su interacción con el producto o servicio. No hace falta fingir que hemos producido el fallo en la conexión a propósito.

El axioma de la transparencia también se aplica a las reacciones del facilitador ante lo que hace el usuario. Aunque las palabras del facilitador sean neutras, si su comportamiento no verbal indica sorpresa, desagrado o aburrimiento, el usuario se sentirá inseguro.

Resoplar, removerse en la silla si se está sentado, mover excesivamente las manos, tamborilear con los dedos... la lista de comportamientos involuntarios y no verbales que ejercen una impresión negativa sobre el usuario es larga.

3) El último test, como el primero

Muy en relación con lo expuesto en el punto anterior, el facilitador no debe dejar que el cansancio o la monotonía afecten a su comportamiento durante cualquiera de las sesiones de test. En ocasiones la tarea de facilitación se hace repetitiva, ya que en un bloque de varias pruebas, los usuarios caen recurrentemente en los mismos problemas de diseño del producto o servicio. Este hecho no debe desvirtuar el último test con el último usuario. Si un error se repite una vez más, también es un dato valioso que debemos a ese último usuario.

4) Adaptarse al estilo de comunicación del usuario

A menudo después de un test se cae en la tentación de “comentar la jugada” del usuario, especialmente sobre su capacidad de llevar a cabo el protocolo de pensamiento manifiesto. El usuario es catalogado como demasiado hablador o demasiado callado, demasiado tímido, demasiado exagerado, etc.

En vez de juzgar, cabe recordar que es responsabilidad del facilitador que el usuario realice un buen trabajo. Hay que adaptarse a las características del usuario, no esperar que él se adapte a nuestro protocolo:

- Los usuarios extravertidos tienden a pensar y hablar al mismo tiempo. Es el caso del usuario que hila en una misma frase “no-lo-encuentro-aquí-está-ay-esto-no-era-dónde-estoy”. Su peligro es que vuelcan durante el test un exceso de información no siempre relevante. Por lo tanto, el papel del facilitador es centrar a este usuario y ser capaz de separar los comentarios valiosos de los comentarios irrelevantes.
- Los usuarios introvertidos tienden a hablar cuando han pensado lo que van a decir. Es el caso del usuario que está callado durante la mayor parte de la tarea y emite una valoración –más o menos lapidaria– al final. Su peligro es que no dejan registrar esos pensamientos intermedios que le llevan a su conclusión. El papel del facilitador es animar a este usuario a decir lo que piensa, pero de manera muy cuidadosa, ya que necesita su tiempo y verse forzado a hablar puede incomodarlo.

Recordemos que antes de iniciar la ejecución de las tareas y el pensamiento manifiesto tenemos un tiempo de recepción del usuario que el facilitador tiene que aprovechar para calibrar su tiempo de respuesta o para adecuarse al perfil (¿es intravertido, extrovertido?).

5) Estar atento a la validez de los datos

Por muy insistentemente que hayamos comunicado al usuario que no le estamos juzgando, algunos usuarios se sienten siempre juzgados o simplemente observados. Esta sensación puede ser imposible de cambiar por lo que sólo se puede minimizar.

Si el facilitador percibe que el usuario no se está comportando de manera natural, hay que registrarlo y tenerlo en cuenta en el momento de hacer el análisis de los datos del test.

Si los datos de comportamiento del usuario no tienen fiabilidad, son útiles otras fuentes de datos como el seguimiento ocular⁶, el seguimiento de clics o cualquier tipo de registro informático⁷.

⁽⁶⁾En inglés, *eye-tracking*.

⁽⁷⁾En inglés, *log*.

6) Dejar que los usuarios hablen de sus propias experiencias

En ocasiones, especialmente si lo que quieren manifestar es negativo, los usuarios se sienten incómodos expresando su propio punto de vista y lo ponen en boca de otras personas:

“Para mí está bien, pero la gente en general lo encontrará difícil”.

“Para mi madre sería difícil”.

“La gente mayor tendrá problemas con esto”.

Son opiniones de usuarios que no aportan información al test, ya que los datos que buscamos son los que se recogen durante la prueba. No nos interesa que el usuario se comporte como si fuera un crítico o un profesional con opiniones subjetivas.

El facilitador debe extraer de estas afirmaciones de los usuarios lo que se refiere a ellos mismos y reformularlo como pregunta.

Ejemplo

USUARIO: “Para mí está bien, pero la gente en general lo encontrará difícil”.

FACILITADOR: “¿Quieres decir que para ti está bien?”.

USUARIO: “Sí, porque...”.

Ved también

Podéis ver el método del seguimiento ocular en el subapartado 3.3 de este módulo didáctico.

Ejemplo

USUARIO: “Para mi madre sería difícil”.

FACILITADOR: “De acuerdo. Para ti ¿te resulta difícil?”.

USUARIO: “Sí/No, porque...”.

7) Detectar la autocensura de los usuarios

Otra forma que encuentran los usuarios de “maquillar” sus opiniones negativas es rebajarlas.

Si el usuario encuentra difícil el producto o servicio, frecuentemente lo formula acompañándolo de una disculpa: “Es difícil, pero una vez que te acostumbras no está tan mal...”.

Una situación típica de la entrevista final de un test es que el usuario comente “Al principio era difícil, pero creo que la gente que tiene que usarlo lo entenderá”.

Lo que nos sugiere este usuario es que no considera que el suyo sea el perfil para usar el producto o servicio, porque ha tenido dificultades. El facilitador tiene que responder investigando cuáles son estas dificultades: “¿Qué te ha parecido más difícil al principio?”.

8) No nos conformamos sólo con los problemas, pero no buscamos soluciones

El resultado de nuestro test no puede limitarse a los problemas, necesitamos el porqué de los mismos.

Los usuarios no vieron el enlace, los usuarios no entendieron la etiqueta, los usuarios no utilizaron el botón... Estas conclusiones no son útiles a ningún equipo de desarrollo que busque solucionarlos, porque no concreta qué cabría modificar para que el problema se solucione. En definitiva, falta indagar sobre los porqués. El facilitador tiene que insistir al usuario para que reformule estos porqués y para que las acciones de éste vayan acompañadas de una explicación.

También es habitual que los usuarios acompañen el problema con la solución que consideran más sencilla desde un punto de vista muy personal: “No he visto este enlace, debería estar en negrita o en otro color”. Hay que escuchar la opinión de los usuarios y respetarla, pero al usuario le falta visión global del producto, no es arquitecto de información ni diseñador. El facilitador puede aprovechar que el usuario propone esta solución para extraer más información sobre el porqué del problema.

Ejemplo

USUARIO: “No he visto este enlace, debería estar en negrita o en otro color”.

FACILITADOR: “No has visto el enlace, ¿por qué crees que ha pasado?”

USUARIO: “Porque está escondido, porque no parece un enlace, porque esperaba un botón, etc.”.

Si en este diálogo el facilitador consigue que el usuario explore suficientemente el problema, puede definirse con él una solución.

FACILITADOR: “Por tanto, no has visto el enlace porque no parece un enlace, ¿qué aspecto debería tener?”

9) Objetivar lo subjetivo

Durante un test los usuarios hacen muchas cosas a las que no es fácil extraer un significado: sonrían, fruncen el ceño, se quedan mirando algo, suspiran... Nuestro equipo de observadores puede tener a un responsable de recoger todo este comportamiento no verbal, pero el facilitador puede ayudar a que esta información sea correctamente interpretada.

Una buena estrategia pasa porque el facilitador ponga de manifiesto objetivamente estos comportamientos; se trata simplemente de “traducir” sus gestos a palabras, pero poniendo cuidado en no calificar.

Ejemplo

El usuario sonrío.

FACILITADOR: “Has sonreído”.

USUARIO: “Sí, esto me ha recordado a...”.

En lo que se refiere al comportamiento no verbal del usuario, es importante mantener una distancia casi de antropólogo. Si el usuario sonrío y el facilitador le pregunta directamente “¿Por qué has sonreído?” es difícil que no se sienta cuestionado y puede que le resulte difícil responder.

1.4. Clasificación de los tests con usuarios

No existe una clasificación formal o estandarizada de los tests de tareas con usuarios; no obstante, podemos agruparlos según varias dimensiones, como las características físicas del entorno de test o el prototipo que se evalúa.

1.4.1. Presencial / remoto

Existen diferentes maneras de efectuar los tests con usuarios. En un principio, las pruebas de usabilidad se llevaban a cabo en laboratorios especializados que disponían de los medios necesarios para efectuarlos. Por ejemplo, un laboratorio con una sala ambientada como una oficina o un hogar donde el usuario lleva a cabo el test. Un espacio donde observadores podían visualizar al usua-

rio y las interacciones que éste hacía con la aplicación y un equipo técnico que se encargaba de registrar todo el proceso. Esta opción continúa siendo interesante para aquellos estudios en los que el entorno y el contexto de uso juegan un papel principal.

Como ya hemos indicado, también se pueden llevar a cabo pruebas con la ayuda de un ordenador que disponga de un software de grabación que permita ver el recorrido del usuario por la interfaz, y una cámara para grabar sus gestos y comportamientos. En cualquier caso, se requiere la presencia física del usuario y de los evaluadores en un mismo espacio; esto supone varias limitaciones:

- Temporales: porque sólo se puede hacer un test cada la vez.
- Económicos: porque involucrar a usuarios es costoso, ya que cuesta tiempo y dinero.
- Tamaño de la muestra: porque por las limitaciones temporales y económicas, el número de usuarios que se involucran en los tests no suele llegar a los 15 participantes.

Hace ya algún tiempo que existe otro modo de llevar a cabo un test con usuarios sin necesidad de tenerlos en la misma habitación con los expertos. Este tipo de test con usuarios se conoce como **test remoto**. El participante se encuentra en un lugar físico distante del experto.

En el caso de tener usuarios repartidos por distintas zonas geográficas o cuando el contexto de uso no puede ser reproducido en un laboratorio, un test de usabilidad remoto es la mejor solución, ya que los usuarios pueden participar en la prueba desde cualquier lugar.

Ventajas del test remoto frente a la opción de la presencialidad:

- El usuario realiza las tareas en su entorno habitual.
- Permite la participación de varios observadores.
- Evita desplazamientos y se puede evaluar con usuarios de todo el mundo.
- Simultaneidad en la ejecución del test.
- Reducción de costes.
- Mayor muestra de usuarios.

Inconvenientes del test remoto:

- No podemos ver la expresión facial ni los movimientos físicos del usuario, a no ser que este tenga una cámara.
- Problemas con la conexión a Internet o con el tipo de software (navegadores, plugins, controladores) que utiliza el usuario.

- Sensibilidad de la información entregada o del acceso a lugares restringidos para usuarios. por ejemplo, en el caso de un banco.

Podemos distinguir dos tipos de test remoto: moderado y no moderado.

Test remoto moderado

El test de usabilidad se lleva a cabo en línea mediante un software que permite la conexión remota de varios ordenadores. Mediante este software el usuario puede acceder al ordenador del moderador y visualizar el producto que se quiere evaluar. El procedimiento es el mismo que el de un test con usuarios no remoto, con la diferencia de que el usuario no está en el mismo sitio físico que el moderador. Se utilizan tecnologías de apoyo como las videoconferencias, en las que el usuario y los moderadores están conectados en línea. La pantalla del usuario y la conversación entre el usuario y la persona que modera el estudio se pueden grabar en vídeo como si fuera un test de usabilidad presencial. Además, la conversación puede ser seguida por otras personas.

¿Qué software se necesita? Para efectuar un test de usabilidad remoto moderado se necesitan los siguientes elementos:

- Ordenador
- Cámara (webcam)
- Aplicación específica para compartir pantallas, es decir, podremos acceder a un ordenador y éste al nuestro mediante este software (por ejemplo: TeamViewer).
- Teléfono o aplicación específica para hablar con el usuario (por ejemplo: Skype).
- Aplicación específica para grabar la conversación y la interacción del usuario con la interfaz (por ejemplo: Morae o Camtasia).

Llevar a cabo un test de usabilidad remoto moderado es prácticamente lo mismo que hacer un test de usabilidad estándar o convencional. La principal diferencia, como ya hemos comentado, es que los usuarios no están presencialmente en el momento del test. Por eso es imprescindible conseguir que se sientan cómodos y explicarles, antes de empezar, el procedimiento que se va a seguir.

El test de usabilidad remoto requiere cierta preparación por parte del usuario, si se compara con un test de usabilidad convencional. Los participantes deben disponer principalmente de software específico en su ordenador y por este

motivo se recomienda contactar con ellos con tiempo suficiente antes del test, mediante correo electrónico o por teléfono, y detallar el software necesario, y cómo instalarlo y configurarlo.

En general, al finalizar los tests de usabilidad se agradece al usuario su tiempo y dedicación mediante un obsequio o gratificación. En el caso del test de usabilidad remoto, podemos hacer llegar la gratificación vía correo postal o utilizar maneras “remotas” de compensación para usar en lugares de comercio electrónico.

Test remoto no moderado

Esta tipología de test permite sustituir al moderador por un software. Los especialistas y analistas diseñan el software para que éste cumpla la misma función, aunque siempre conviene recordar que hay que valorar el tipo de prueba que llevamos a cabo dadas las diferencias evidentes que puede haber entre la presencia humana y una máquina.

Mientras el usuario está en su casa, el software va solicitando a los usuarios una serie de tareas y preguntas que éstos deben contestar a medida que el sistema va capturando datos sobre la interacción del usuario con el producto. Este sistema permite la entrada de varios usuarios simultáneamente sin la presencia de moderadores, por ello se pueden evaluar muestras de cientos de usuarios, y obtener datos estadísticamente representativos.

Elegir el tipo de test es importante porque variará según las necesidades que tengamos en cada momento. El test remoto no moderado, al no necesitar de un observador o facilitador que acompañe al usuario en todo el proceso de test, ofrecerá unos resultados más cuantitativos; en cambio, el test remoto moderado, muy parecido en cuanto a metodología al test de usabilidad convencional, nos permitirá observar el comportamiento del usuario, escuchar sus comentarios y visualizar la interacción de éste con la interfaz.

Cuando nos encontremos ante la decisión de elegir uno u otro, debemos analizar los objetivos que queremos alcanzar con el test. Si lo que necesitamos es averiguar cuántas veces los usuarios hacen clic en un sitio o la interacción de éstos con la interfaz sin que sea imprescindible su observación, quizá deberemos optar por un test remoto no moderado, el cual nos aportará una gran cantidad de información que deberemos analizar posteriormente. Si, por el contrario, lo que necesitamos es conocer la opinión de los usuarios, sus reacciones ante un sitio u observar su comportamiento, deberemos efectuar un test remoto moderado.

Software recomendado

A continuación se muestran algunos ejemplos de software de los que se puede utilizar para llevar a cabo un test remoto.

Enlaces de interés

Usability tools (en línea)
Remote Research (en línea)

Software necesario para un test remoto moderado

Producto	Descripción
Livellook	Servicio para compartir pantallas en remoto. El usuario puede visualizar la pantalla del facilitador y viceversa. No graba audio ni vídeo.
OpenVULab	Creado por la Universidad de York. Herramienta de código abierto que como UserVue permite grabar vídeo y audio en los ordenadores de los usuarios mientras se desarrolla el test.
Skype	Creado por Mac OS X. Herramienta gratis donde se puede compartir pantalla, vídeo y chat. Requiere tener el programa instalado en el ordenador.
TeamViewer	Servicio para compartir pantallas en remoto. El usuario puede visualizar la pantalla del facilitador y viceversa. No graba audio ni vídeo. No hace falta descargar el software para su utilización.

Software necesario para un test remoto no moderado

Producto	Descripción
ClicTale	Software que graba la interacción con un producto o servicio utilizando javascript. graba la interacción del usuario, así como un mapa donde se puede visualizar el recorrido del usuario: dónde han hecho clic, dónde ha desplazado la barra de desplazamiento (<i>scroll</i>), formularios...
UserZoom	Potente herramienta que permite controlar y dirigir múltiples proyectos de DCU, y que es capaz de proporcionar gran cantidad de información cuantitativa, generar las tareas para el test de usabilidad y controlarlas, <i>card sortings</i> , entrevistas o reclutamiento de usuarios desde un panel de control general.
KeyNote	Herramienta parecida a la anterior, en la que los usuarios contestan encuestas y completan tareas mediante listas desplegables (<i>pop-up</i>) sin necesidad de descargar el programa. Mediante un panel de control podemos visualizar los resultados exhaustivos del test. Generalmente se obtienen datos cuantitativos.
m-pathy	Herramienta que graba los movimientos y clics del ratón sin necesidad de instalar ningún tipo de software en el ordenador del usuario.
Chalkmark	Creado por OptimalWorkshop. El usuario puede resolver las tareas mediante imágenes estáticas. El sistema genera un mapa donde se puede visualizar el recorrido del usuario, similar a ClickTale. Es apropiado para realizar tests no demasiado complicados.
LabsMedia	Creado por LabsMedia. Proyecto de código libre similar a VULabs y ClickTale.
UserFocus	Software que permite obtener gran cantidad de información cuantitativa, generar las tareas para el test de usabilidad y controlarlas. No requiere descarga o instalación.
SMT	Creado por Luis Leiva. Herramienta de código libre que permite grabar el recorrido de un usuario por un producto o servicio.

1.4.2. Formal / de guerrilla

Como ya hemos indicado, una prueba de usabilidad se puede llevar a cabo en distintos lugares. Según las características de estos entornos, el test se puede considerar “formal” o de “guerrilla”:

a) **Tests formales:** son aquellas pruebas que se realizan en un laboratorio, adecuado y construido para llevar a cabo este tipo de técnicas. Generalmente disponen de tecnologías avanzadas para la grabación de la prueba además de va-

rias habitaciones dispuestas con el objetivo de crear un entorno real para el usuario, así como permitir a los observadores la visualización de las tareas que desarrolla el usuario mediante espejos unidireccionales. En definitiva, se trata de entornos contruidos con el objetivo exclusivo de efectuar este tipo de pruebas de evaluación.

b) Tests informales o “de guerrilla”: otra manera más económica y fácil de llevar a cabo un test con usuarios son los llamados tests de guerrilla. El procedimiento de test es exactamente el mismo al del anterior, lo que varía son las condiciones externas. Para este tipo de test no necesitamos unas instalaciones como en el test formal, únicamente es necesario disponer de un ordenador con el software adecuado, una webcam y, ¡como no!, a los usuarios que van a efectuar el test. Con este equipamiento, podemos realizar un test con usuarios con resultados y utilidades muy similares al anterior.

1.4.3. Tests con prototipos de alta/baja fidelidad

Como ya hemos comentado, se pueden efectuar diferentes tipos de tests con usuarios según la fase en la que se encuentre el proyecto; no hace falta esperar a tener una implementación final para llevar a cabo un test de usabilidad.

Existen diferentes tipos de pruebas dependiendo de la fase o del estado de detalle en el que se encuentre el producto o servicio que se analice:

a) Tests con un prototipo de alta fidelidad: se llaman *prototipos de alta fidelidad* aquellos que están muy cercanos al producto o servicio final, tanto en diseño o interactividad como en funcionalidades. Su ventaja es que representan prácticamente aquello con lo que se encontrará el usuario, pero esto los hace más costosos de producir; son más propios de etapas finales del proyecto.

b) Tests con un prototipo de baja fidelidad: se llaman *prototipos de baja fidelidad* aquellos que están lejos del diseño y las funcionalidades del producto o servicio final; un prototipo de baja fidelidad de una página web de venta de billetes de avión puede estar dibujado sobre una página de papel. Su ventaja es que son muy rápidos de crear y destruir, por lo que en etapas iniciales del proyecto resultan muy útiles para descartar ideas o para identificar requisitos.

Dependiendo del momento en el que se encuentre nuestro proyecto, de lo que pretendamos evaluar o del presupuesto que tengamos, utilizaremos un tipo de test u otro.

Siempre obtendremos mayor detalle de información al utilizar un prototipo de alta fidelidad que un prototipo de baja fidelidad, pero en determinadas fases de desarrollo no nos interesa tanto el detalle como evaluar las líneas generales del producto o servicio.

Lectura recomendada

Sobre las fases para la realización de un test informal, podéis consultar la obra siguiente:

Yusef Hassan Montero, Francisco J. Martín Fernández (2003). “Método de test con usuarios”, *No solo usabilidad* (9 de diciembre).

Lo importante es saber seleccionar el tipo de evaluación según las necesidades del proyecto. Es mejor ir haciendo pruebas constantes y en diferentes grados de calidad y fidelidad durante el proyecto, que esperar a tener un producto final para el test.

2. Variantes del test con usuarios

Hasta ahora hemos presentado el test con usuarios en su versión más convencional: un usuario, un facilitador, un laboratorio, observadores... Sin embargo, hay condicionantes que pueden hacer necesaria una variación sobre los elementos clásicos:

- Por necesidades específicas del producto o servicio: por ejemplo, que la forma de consumo del producto que estamos evaluando no sea individual, sino colectiva. No es lo mismo evaluar un cajero automático (interacción individual), que una máquina recreativa (interacción colectiva). El formato de nuestro test debe variar en consecuencia.
- Porque aparecen nuevas tecnologías que ayudan a aportar más datos sobre la interacción de los usuarios con el producto o servicio. La tecnología nos ofrece avances que pueden aplicarse a la evaluación con usuarios y variar el formato de nuestro test. La innovación más reciente ha sido la incorporación del seguimiento ocular de los usuarios, es decir, tecnología que permite saber dónde miran en una pantalla. El futuro nos ofrecerá cada día más y más herramientas que pueden hacer variar el clásico test.
- Por el estadio de desarrollo del producto interactivo; cuando nos encontramos ante una fase muy temprana de desarrollo, puede ser difícil realizar un test con un prototipo que funcione. Esto no tiene que llevar a la renuncia de una evaluación con usuarios: simplemente, hay que cambiar el formato de la evaluación.

A continuación repasamos algunas de las principales variantes del test de usuarios, que responden a estos condicionantes. Dependiendo del producto interactivo y de los objetivos de evaluación de la usabilidad, el test con usuarios puede diseñarse de manera que combine tipologías y variantes del test estándar o convencional y, de este modo, ajustar los objetivos de la evaluación y obtener mejores resultados.

2.1. Test de codescubrimiento

El **test de codescubrimiento** es en esencia un test con usuarios, pero ejecutado por dos usuarios simultáneamente que colaboran en la resolución de las tareas.

Este tipo de variante puede reflejar mejor determinados ambientes de uso de un sistema; por ejemplo, una aplicación de ofimática es más común que se use en una oficina en la que otros usuarios de la misma aplicación pueden ayudar cuando hay un problema.

Por otra parte, siempre será más natural para dos usuarios intercambiar impresiones entre ellos sobre el producto, que ir hablando en voz alta en solitario con el facilitador del test, lo cual ayuda a vencer uno de los riesgos de cualquier test con usuarios: que el usuario no se sienta cómodo y no actúe con naturalidad.

En este tipo de test, la labor del facilitador se complica, en tanto que tiene que ocuparse de que los dos usuarios hablen por igual y tomen decisiones de manera consensuada; nada peor en un test de codescubrimiento que los usuarios se enfaden entre sí por la resolución de las tareas.

2.2. Seguimiento ocular

El *eye-tracking* hace referencia a aquellas tecnologías que permiten registrar la exploración visual de un usuario ante la información que se le presenta. Podría ser traducido como *seguimiento ocular*, ya que la monitorización que se lleva a cabo es la del movimiento de los ojos, concretamente, la identificación y medición de las fijaciones. En este sentido, el *eye-tracking* es una herramienta de recopilación de datos cuantitativos, más que una metodología de test propiamente dicha.

Identificación y medición de fijaciones

Los saltos o movimientos rápidos que hacemos con los ojos se llaman *sacadas*. Entre una sacada y otra se produce una fijación, periodo de relativa quietud del ojo que posibilita ver nítidamente la zona enfocada (Hassan y Herrero-Solana, 2007).

Como herramienta para la evaluación de interfaces, se pueden utilizar dos alternativas posibles: cámaras⁸ colocadas en la cabeza de los participantes, o cámaras colocadas frente al usuario en el monitor donde va a dirigir su mirada.

⁽⁸⁾En inglés, *eye-trackers*.

No se trata de una tecnología innovadora, puesto que han existido dispositivos que permitían este tipo de análisis desde los años cincuenta, pero se trataba de aparatos muy intrusivos que incomodaban a los usuarios, ya que eran propios de la investigación científica más ortodoxa.

Un ejemplo de nueva tecnología

Eye Tracker Tobii Glasses es un ejemplo de la última tecnología de seguimiento ocular.

En la última década, algunas empresas han desarrollado hardware y software de seguimiento ocular capaz de registrar el comportamiento natural del usuario en entornos reales sin necesidad de incomodar al usuario con tecnología abrumadora. Aun así, el coste, la escasez de información, la interpretación de resultados o los inconvenientes del propio método son razones suficientes para entender por qué no se utiliza tanto como ayuda a la evaluación en el ámbito de la experiencia de uso.

Los sistemas de seguimiento ocular que se utilizan en las evaluaciones de usabilidad se componen, básicamente, de los elementos siguientes:

- Una pantalla adaptada, capaz de detectar mediante infrarrojos el movimiento ocular del usuario mientras éste desarrolla las tareas. Esta detección requiere que antes de comenzar la prueba se calibre el aparato para adaptarlo a las particularidades físicas de cada participante.
- Un software de interpretación de los registros de la pantalla. En esencia, se recoge dónde miraba el usuario (coordenadas de la pantalla), durante cuánto tiempo, de dónde venía y dónde miró después. Son necesarias herramientas de software que transformen todos estos datos en información útil para el evaluador.

La presentación de los resultados de un seguimiento ocular puede variar según el software que se utilice para interpretar los datos, pero consiste, a grandes rasgos, en:

- Mapas de recorrido, que representan el “camino” de los ojos del usuario por la pantalla.
- Mapas de “calor”⁹, que representan visualmente el tiempo que cada usuario mira hacia determinadas zonas de la pantalla como un mapa de calor: cuanto más rojo, más tiempo.
- Datos por áreas, que muestran valores numéricos de los tiempos de fijación del usuario en un área previamente definida, por ejemplo, el buscador de una página.

⁽⁹⁾En inglés, *hotspots*.

Si bien la riqueza de los datos que proporciona el seguimiento ocular es grande, esta herramienta presenta dificultades a la hora de interpretar qué significan realmente todos estos datos. Por ejemplo, si el usuario ha mirado un elemento de la pantalla, ¿significa realmente que lo ha visto? Lo que no mira, ¿quiere decir que no le interesa o que no lo entiende? Es por esto por lo que a menudo se realizan pruebas con seguimiento ocular que incorporan el método del pensamiento manifiesto¹⁰.

⁽¹⁰⁾En inglés, *think-aloud*.

2.3. Lápiz y papel

¿Qué es?

La técnica de lápiz y papel es una variante del test de usabilidad donde usuarios representativos de un producto o servicio llevan a cabo tareas a través de una versión en papel de la interfaz del producto o servicio controlada por un moderador (Snyder, 2003).

Lectura complementaria

Podéis consultar más acerca de la técnica del lápiz y papel en la obra siguiente:

C. Snyder (2003). *Paper prototyping: the fast and easy way to design and refine user interfaces*. San Francisco: Morgan Kaufmann.

En este caso, al no disponer de una maqueta funcional, el moderador es quien debe ir alternando los papeles, como si fueran las diferentes vistas de interacción con el producto, cuando el usuario lo indique.

Son muchas las veces que durante el desarrollo de un test con usuarios nos encontramos con participantes que opinan sobre el diseño de una aplicación, iconos o tipografías. No es algo desdeñable, pero puede desviar el centro de discusión inicial. Al hacer la prueba con lápiz y papel nos evitamos este tipo de debates y únicamente nos centramos en la esencia del producto; su objetivo, funcionalidad e interacción.

¿Cómo llevarlo a cabo?

Una vez definidas las funcionalidades y características del producto o servicio con el que trabajamos y que queremos evaluar, lo único que necesitamos es dibujar los prototipos en papel que vamos a utilizar durante el test.

El método del lápiz y papel

Para entender mejor cómo funciona el método del lápiz y papel, proponemos ver el siguiente vídeo:

- Usability Testing with a Paper Prototype

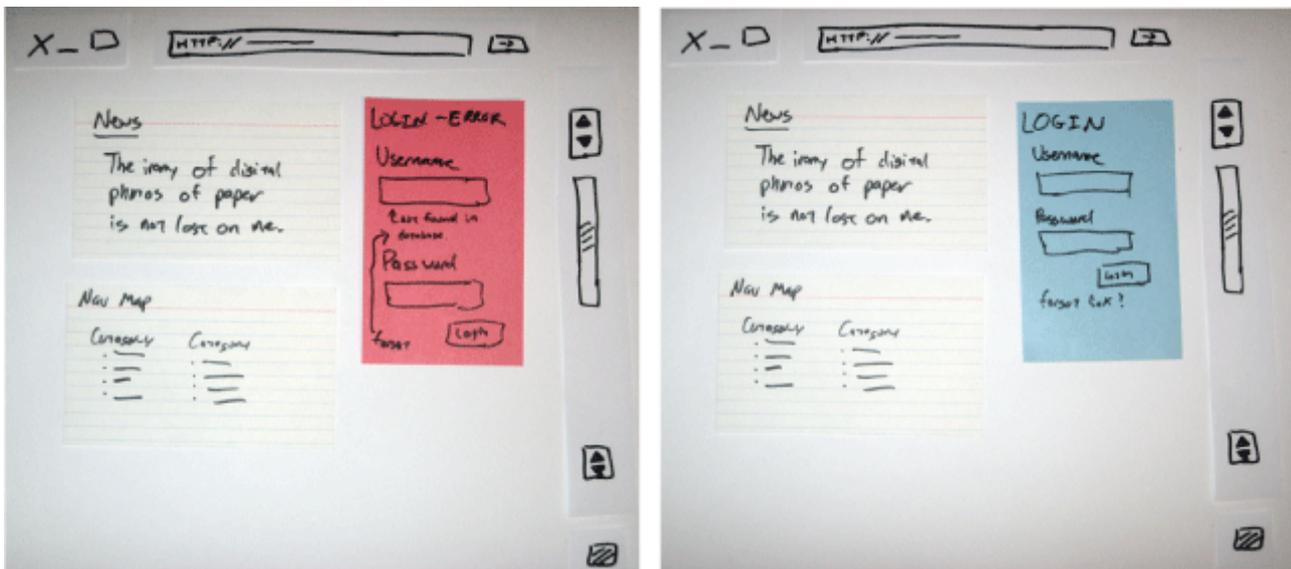


Ilustración de alistapart.com

Estos prototipos pueden ser dibujados a mano o con herramientas digitales básicas tales como Word o PowerPoint.

¿Quién participa?

- **Facilitador:** es la persona encargada de dirigir y moderar la sesión.
- **Usuarios:** son las personas susceptibles de utilizar la aplicación y quienes comprobarán, mediante el test con lápiz y papel, el correcto funcionamiento del producto o servicio.
- **Observadores:** puede ocurrir que otras personas del departamento o los mismos clientes quieran observar la sesión y tomar notas, aunque no siem-

pre. En este caso, se recomienda que el usuario no sea consciente de esta observación y que, por tanto, los observadores se ubiquen en otra sala.

Ventajas de lápiz y papel

Los tests de lápiz y papel son relativamente fáciles de llevar a cabo; además, no necesitan ser maquetados o programados, por tanto, son más económicos que otros tipos de tests.

Con este tipo de test es fácil detectar problemas de uso desde las fases iniciales del proyecto. No necesitamos software adicional para su realización. Con lápiz y papel podemos tanto crear como corregir los prototipos.

Inconvenientes de los prototipos de papel

Al ser un prototipo sencillo no se puede utilizar para evaluar diseño.

Con prototipos hechos en lápiz y papel solo se puede simular la respuesta del sistema y acercarnos a una situación real, pero nunca proyectarla tal y como es.

En los prototipos de lápiz y papel a veces no se tienen en cuenta criterios como el tamaño y forma del producto, por lo que puede llevar a diseños que después no se puedan aplicar.

En este tipo de test no es importante valorar esta parte, ya que sobre papel acostumbramos a dibujar interfaces irreales con tamaños a veces exagerados y con combinaciones de elementos difíciles. No tenemos que imaginarnos la interfaz en su sentido más literal. Podemos apoyarnos en medidas, dimensiones y tamaños cercanos a la realidad de una pantalla.

3. Otros métodos de evaluación con usuarios

Existen otros métodos de evaluación que implican a los usuarios potenciales de un producto interactivo.

Los métodos de evaluación en términos generales pueden clasificarse de maneras muy diversas, dependiendo de si atendemos a unos elementos u otros, como: quién lleva a cabo la metodología, el coste, el tiempo, el tipo de datos que se obtienen...

Utilizaremos aquí la siguiente clasificación, que tiene en cuenta el tipo de objetivo que se busca con la aplicación de cada método y el evaluador que la aplica:

a) Métodos de revisión¹¹: agrupa los métodos que sirven para revisar las características del producto o servicio que impactan sobre la experiencia de uso o sobre la accesibilidad; en general, el trabajo de revisión se lleva a cabo por expertos en diseño centrado en el usuario (DCU). La evaluación heurística sería un tipo de método de inspección.

⁽¹¹⁾ Aparecen a menudo en la bibliografía sobre el tema como *inspection methods* o *métodos de inspección*.

b) Métodos de recopilación¹²: son los métodos que buscan recopilar datos cualitativos o cuantitativos sobre la experiencia de uso del producto o servicio; en estos métodos lo más común es implicar al usuario final y están orientados a conocer quién es este usuario y qué necesita de nuestro producto o servicio.

⁽¹²⁾ También aparecen a menudo en la bibliografía sobre el tema como *inquiry methods*.

c) Métodos de test: métodos de análisis que buscan corroborar con el usuario que el producto o servicio satisface sus necesidades y expectativas. Frecuentemente, estos métodos definen tareas para que este usuario las realice y se analiza hasta qué punto le resulta satisfactorio llevarlas a cabo. Serían los métodos que hemos tratado en los apartados anteriores.

Ya que este módulo está dedicado a los métodos de evaluación con usuarios, nos centraremos en describir brevemente los de recopilación y test, que son los que implican al usuario final para poder aplicarlos.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Nombre: Tipología: Descripción										Momento de aplicación: <input type="checkbox"/> Presencial / <input type="checkbox"/> Remoto <input type="checkbox"/> Cuantitativo / <input type="checkbox"/> Cualitativo Coste: <input type="checkbox"/> Bajo <input type="checkbox"/> Medio <input type="checkbox"/> Alto Tiempo: <input type="checkbox"/> Corto <input type="checkbox"/> Medio <input type="checkbox"/> Largo Hardware/Software: <input type="checkbox"/> Necesario				
Versión imprimible														

3.1. Indagación contextual

Beyer y Holtzblatt (1998) describieron un conjunto de técnicas centradas en el usuario cuyo propósito era ayudar a los equipos de diseño y desarrollo a tomar decisiones acordes con las necesidades de los primeros.

La indagación contextual¹³ fue una de ellas y se ha incorporado como un método de entrevista de campo, basado en unos principios básicos que la diferencian de otros métodos tradicionales de entrevistas.

⁽¹³⁾En inglés, *contextual inquiry*.

La **indagación contextual** es un estudio de campo donde no se analiza al usuario sino que se aprende de él. Se trata de visitar al usuario en su lugar de trabajo, en su casa o en el espacio físico donde efectúe sus actividades rutinarias, para analizar, observar y aprender sus hábitos, actividades, características y factores de entorno.

Recoger los detalles y las motivaciones implícitas de las personas cuando se enfrentan a nuestros productos o servicios interactivos permite traspasar a los diseñadores parte de las necesidades reales de los usuarios y lograr un conocimiento compartido muy valioso en su trabajo.

La persona que lleva a cabo la investigación de campo debe recopilar información sobre el usuario mediante la observación y la entrevista, al tiempo que el usuario realiza su trabajo, dejando que este se sienta cómodo para poder llegar a entender la situación del mejor modo posible.

En la investigación de campo es muy importante crear un buen clima para que el usuario se sienta confortable con el entrevistador, mantener una relación de confianza entre ambos, incluso dejar que el usuario lidere la situación y el entrevistador aprenda de éste, ya que, tal y como hemos dicho, se trata de aprender del usuario.

Para llevar a cabo la entrevista de campo, previamente, es esencial tener un objetivo claro y definido. Es recomendable no tomar notas delante del usuario para que éste no se sienta evaluado. Podemos llevar con nosotros una grabadora y al finalizar la entrevista tomar los apuntes necesarios.

Los principios básicos de la entrevista de campo son: contexto, trabajo colaborativo¹⁴, interpretación y objetivo.

⁽¹⁴⁾En inglés, *partnership*.

1) Contexto

La observación de los usuarios en su entorno permite obtener información precisa sobre sus necesidades. Es necesario concretar descripciones y datos sobre los usuarios y su entorno para evitar que éstas sean abstractas.

La entrevista se efectúa en el lugar de trabajo donde el usuario utiliza el producto que se evalúa.

Ejemplo

Si, por ejemplo, evaluamos un campus virtual, iremos a casa del usuario:

USUARIO: Generalmente me llegan los ejercicios por correo electrónico.

ENTREVISTADOR: ¿Tienes alguno? ¿Lo podemos mirar juntos?

2) Trabajo colaborativo

El objetivo del trabajo colaborativo es que el usuario y el entrevistador creen un vínculo de unión y colaboren para entender el trabajo que se tiene que llevar a cabo.

Para ello, es necesario cambiar la estructura típica de entrevista y que los usuarios perciban que realmente interesa el trabajo que realizan con el producto evaluado.

Consejos:

- No vamos a hacer únicamente una entrevista convencional, por tanto, debemos evitar el modelo pregunta-respuesta.
- El entrevistador tampoco acude para responder preguntas, sino para aprender del usuario.
- El entrevistador debe evitar sentirse un invitado. Por el contrario, debe sentirse un compañero.

3) Interpretación

En la investigación de campo, es fundamental saber distinguir entre los datos, las interpretaciones de los datos y las interpretaciones que conllevan estos datos. Para asegurarnos de distinguir correctamente estos conceptos, debemos compartir las interpretaciones que vamos elaborando durante la entrevista con los usuarios.

4) Objetivo

Es imprescindible tener un objetivo de la investigación claro, detectado y bien definido para evitar utilizar únicamente un cuestionario. Nos ayudará a gestionar el tiempo de la entrevista.

Por otra parte, tener un objetivo clave nos ayudará a entendernos mejor con el usuario y a saber dirigir adecuadamente la entrevista.

Fases de la entrevista

a) Entrevista convencional (15 minutos)

Introducción y objetivo de la entrevista para ir conociendo al usuario y éste a su vez al entrevistador.

b) Transición (30 segundos)

Explicaremos las normas de la investigación contextual y le daremos las indicaciones al usuario para que haga su trabajo mientras el entrevistador le observa. Antes de que empiece con sus tareas, es importante aclarar que podremos interrumpir al usuario cuando creamos necesario.

c) Entrevista contextual (1-2 horas)

El usuario empieza con sus tareas mientras el entrevistador observa e interpreta.

d) Contraste (15 minutos)

El entrevistador propone un breve resumen de lo que ha entendido con el objetivo de confirmar su percepción y contrastarla con el usuario.

3.2. Entrevistas

Las **entrevistas** son una técnica mediante la cual los expertos formulan cuestiones previamente analizadas para investigar con usuarios reales, con el objetivo de conocer sus reacciones. Al ser una técnica flexible, permite alcanzar una gran cantidad de conocimiento sobre el usuario y su comportamiento.

Este tipo de técnica conlleva trabajar con usuarios reales *in situ* de manera presencial, además de contar con un moderador, que es la persona experta que se encarga de llevar a cabo la entrevista y que a la vez fomenta el diálogo y crea empatía con los usuarios.

Las entrevistas pueden ser estructuradas o abiertas. En las primeras el entrevistador sigue el guión preestablecido, por lo tanto, el proceso es más rígido. La entrevista abierta es más flexible y deja espacio a la improvisación.

Durante la entrevista se recomienda grabar el audio y no tomar notas, ya que de este modo el usuario no se distrae y el entrevistador se puede centrar en la entrevista. Es aconsejable comenzar con las preguntas de menos complejidad y permanecer neutral hacia el usuario, sin tratar de justificar ninguna decisión de diseño o de concepto. En el caso de las preguntas abiertas, éstas deben ayudar a responder al usuario con frases concretas (evitar los “sí” y “no”).

3.3. Encuestas

Las **encuestas** son una técnica donde se formulan una serie de cuestiones relacionadas con el producto analizado con el fin de obtener información de interés para su estudio.

La diferencia entre las encuestas y las entrevistas es que estas últimas son más enriquecedoras e interactivas.

3.4. Cuestionarios

Los **cuestionarios** son una técnica exploratoria basada en la recogida de información mediante las respuestas que proporcionan los usuarios. Esta información suele ser una lista de preguntas escritas que el usuario debe contestar, por tanto, deben ser lo más claras y objetivas posibles.

Se trata de descubrir cómo los usuarios utilizan un producto interactivo y sus preferencias. Al no ser una técnica donde se evalúa la interfaz *in situ*, más bien se recogen las opiniones de los usuarios sobre ese producto, es muy importante tener claras las preguntas que se van a formular, y requiere un mayor esfuerzo por parte del usuario al tener que rellenar y responder claramente las preguntas.

Si bien es posible crear un cuestionario desde cero con preguntas adaptadas al producto o servicio que evaluamos, existen algunos cuestionarios estandarizados; estos cuestionarios ya tienen preestablecidas las preguntas y las métricas que se emplean para su análisis, y lo que es más importante, debido a su utilización a lo largo del tiempo, se ha visto su utilidad e interés. Por esta razón, algunos de los cuestionarios estandarizados son de pago.

A continuación, se muestran algunos ejemplos de cuestionarios estandarizados. Cada cuestionario estandarizado evalúa distintos aspectos de la utilidad, el uso y la usabilidad de productos interactivos, especialmente de productos de software.

El SUMI, el MUMMS y el WAMMI fueron creados por el Human Factors Research Group, del University College de Cork. Se ocupan de evaluar, respectivamente, software, productos multimedia y páginas web.

El SUS fue creado por la Digital Equipment Corporation y se centra fundamentalmente en la evaluación de la usabilidad percibida (aspectos subjetivos de la usabilidad), pero sobre cualquier sistema; es, por tanto, más genérico.

3.4.1. Software usability measurement inventory (SUMI)

El SUMI¹⁵ es un cuestionario utilizado para la evaluación de la calidad de un conjunto software desde el punto de vista del usuario final. Este cuestionario puede ser utilizado para evaluar nuevos productos, efectuar comparaciones con versiones previas y establecer objetivos para desarrollos futuros. Consiste en 50 puntos a los que el usuario ha de responder: “De acuerdo”, “No lo sé”, “En desacuerdo”.

Algunos ejemplos de preguntas

“1. Este software responde demasiado despacio a las entradas que se le proporcionan.”

“3. Las instrucciones y advertencias son de ayuda.”

“13. El modo en el que se presenta la información del sistema es clara y comprensible.”

“22. No me gustaría tener que utilizar este software todos los días.”

El número de usuarios necesarios para obtener resultados interesantes oscila entre 12 y 30.

3.4.2. Measuring the usability of multi-media systems (MUMMS)

El cuestionario MUMMS¹⁶ tiene el mismo objetivo que el SUMI, pero pretende aplicarse al software multimedia; se caracteriza por la consideración de cinco subescalas:

- La medida en que el producto capta las respuestas emocionales del usuario.
- El grado de control con el que el usuario siente que él, y no el producto, va procediendo paso a paso.
- El grado de eficiencia con el que el usuario siente que puede conseguir los objetivos de su interacción con el producto.

⁽¹⁵⁾SUMI: Software usability measurement inventory

Ejemplo de cuestionario SUMI

Un ejemplo de cuestionario SUMI en inglés en línea:

- Software usability measurement inventory

⁽¹⁶⁾MUMMS: Measuring the usability of multi-media systems

- El nivel de ayuda y asistencia que el producto parece prestar al usuario.
- La facilidad de aprendizaje con la que un producto puede empezar y aprender nuevas características del producto.

3.4.3. Website analysis and measurement inventory (WAMMI)

WAMMI¹⁷, en cambio, se centra en la evaluación de sitios web. Se basa en el cuestionario que cumplimentan los usuarios, que proporciona una medida de la utilidad y facilidad de uso que encontraron en el sitio en cuestión. Puede ser utilizado de tres modos:

⁽¹⁷⁾WAMMI: Website analysis and measurement inventory

- Predicción: antes del lanzamiento de un nuevo sitio, un test con WAMMI puede dar una orientación sobre la reacción de los visitantes.
- Monitorización: en este caso, interesa conocer las razones de ciertos comportamientos, por qué algunos clientes repiten, por qué otros no vuelven nunca, etc.
- Referencia: interesa conocer la opinión de la audiencia en relación a otros sitios web, fundamentalmente por cuestiones de diseño.

3.4.4. System usability scale (SUS)

El cuestionario SUS¹⁸ fue desarrollado en 1986 por la Digital Equipment Corporation como parte de la aplicación de ingeniería de usabilidad a los sistemas de oficina. Recordemos que en la década de los ochenta era una novedad el uso de ordenadores en las oficinas; se hablaba incluso de la “oficina electrónica”, en contraste con la tradicional.

⁽¹⁸⁾SUS: System usability scale

Su propósito era proporcionar un test fácil de completar (número mínimo de cuestiones), fácil de puntuar y que permitiera establecer comparaciones cruzadas entre productos; es decir, que fuera posible comparar unas soluciones con otras y optar por la mejor.

Historia: cómo se creó la escala

La escala SUS se construyó sobre un conjunto original de 50 preguntas. Se eligieron dos sistemas, uno muy sencillo de usar y otro casi imposible, incluso para los técnicos; 20 personas, con perfiles desde administrativo hasta programador utilizaron los dos sistemas y respondieron a las 50 preguntas con una puntuación de 1 a 5 entre “completamente de acuerdo” y “completamente en desacuerdo”.

Se seleccionaron las preguntas que recibieron respuestas más extremas y con correlación más estrecha (es decir, que funcionaban juntas: si una tenía una respuesta de 5, la otra tenía una de 4 o 5). De esta manera, se redujo de 50 preguntas a las 10 actuales:

	Completamente de acuerdo				Completamente en desacuerdo
1. Creo que me gustará usar con frecuencia este sistema	1	2	3	4	5
2. Encontré el sistema innecesariamente complejo	1	2	3	4	5
3. Pensé que era fácil utilizar el sistema	1	2	3	4	5
4. Creo que necesitaría del apoyo de un experto para recorrer el sistema	1	2	3	4	5
5. Encontré las diversas posibilidades del sistema bastante bien integradas	1	2	3	4	5
6. Pensé que había demasiada inconsistencia en el sistema	1	2	3	4	5
7. Imagino que la mayoría de las personas aprenderían muy rápidamente a utilizar el sistema.	1	2	3	4	5
8. Encontré el sistema muy incómodo de usar	1	2	3	4	5
9. Me sentí muy confiado en el manejo del sistema	1	2	3	4	5
10. Necesito aprender muchas cosas antes de manejarme en el sistema	1	2	3	4	5

La escala SUS es de dominio público y de uso libre, siempre que se cite adecuadamente la fuente: © Digital Equipment Corporation, 1986.

Cómo utilizar la escala SUS

La escala SUS se utiliza en general inmediatamente después de que el usuario haya probado el sistema. Se solicitará a los usuarios el registro inmediato de su respuesta a cada punto, sin dejar que piense largamente en los mismos. No deben dejarse ninguna pregunta por contestar.

Si el usuario no se siente capaz de responder a alguna cuestión en particular, habrá de señalar el valor central de la escala (3).

Puntuación de la escala SUS e interpretación de resultados

La escala SUS es una escala de estilo Likert y su resultado final es un valor único que representa una medida compuesta de la usabilidad del sistema global sometido a estudio. Es decir, que al acabar de procesar los datos, tendremos un único valor que puede ser, digamos, 60 o 35, como la nota final de un examen. Esto es lo que permite que podamos comparar fácilmente sistemas evaluados mediante la escala SUS.

Hay que señalar que las puntuaciones independientes no son significativas por sí mismas. Es decir, que si a la primera pregunta hemos tenido una masiva respuesta de “completamente de acuerdo”, no es un dato significativo mientras esté aislado del resto.

Para calcular la puntuación del SUS, hay que sumar primero las contribuciones de cada punto. La contribución de cada punto valdrá entre 0 y 4:

- Para los puntos 1, 3, 5, 7 y 9, la contribución será la posición de la escala menos 1.
- Para los puntos 2, 4, 6, 8 y 10, la contribución será 5 menos la posición en la escala.

Por último, se multiplica la suma de los resultados por 2,5 para obtener el valor global del SUS. El resultado estará entre 0 y 100.

Ejemplo de puntuación de la escala SUS

Veamos un ejemplo para el cálculo de la puntuación:

1. *Creo que me gustará visitar con frecuencia este sistema*

1 2 3 4 5
En completo desacuerdo Completamente de acuerdo

Recordemos: para este punto, el cálculo es la posición en la escala menos 1. La puntuación será $5 - 1 = 4$.

2. *Encontré el sistema innecesariamente complejo*

1 2 3 4 5
En completo desacuerdo Completamente de acuerdo

Recordemos: para este punto, el cálculo es 5 menos la posición en la escala. La puntuación será $5 - 4 = 1$.

3. *Pensé que era fácil utilizar el sistema*

1 2 3 4 5
En completo desacuerdo Completamente de acuerdo

La puntuación será $2 - 1 = 1$.

Escala Likert

“La escala de tipo Likert es una escala psicométrica comúnmente utilizada en cuestionarios, y es la escala de uso más amplio en encuestas para la investigación. Cuando respondemos a un elemento de un cuestionario elaborado con la técnica de Likert, lo hacemos especificando el nivel de acuerdo o desacuerdo con una declaración (elemento, ítem o reactivo). La escala se llama así por Rensis Likert, que publicó un informe describiendo su uso.”

Wikipedia

4. *Creo que necesitaría del apoyo de un experto para recorrer el sistema.*

1 2 3 4 5
 En completo desacuerdo Completamente de acuerdo

La puntuación dará $5 - 1 = 4$.

5. *Encontré las diversas posibilidades del sistema bastante bien integradas.*

1 2 3 4 5
 En completo desacuerdo Completamente de acuerdo

La puntuación será $2 - 1 = 1$.

6. *Pensé que había demasiada inconsistencia en el sistema.*

1 2 3 4 5
 En completo desacuerdo Completamente de acuerdo

La puntuación será $5 - 3 = 2$.

7. *Imagino que la mayoría de las personas aprenderían muy rápidamente a utilizar el sistema.*

1 2 3 4 5
 En completo desacuerdo Completamente de acuerdo

La puntuación será $2 - 1 = 1$.

8. *Encontré el sistema muy grande al recorrerlo.*

1 2 3 4 5
 En completo desacuerdo Completamente de acuerdo

La puntuación será $5 - 4 = 1$.

9. *Me sentí muy confiado en el manejo del sistema.*

1 2 3 4 5
 En completo desacuerdo Completamente de acuerdo

La puntuación será $5 - 1 = 4$.

10. *Necesito aprender muchas cosas antes de manejarme en el sistema.*

1 2 3 4 5
 En completo desacuerdo Completamente de acuerdo

La puntuación será $5 - 2 = 3$.

Los resultados totales: $4 + 1 + 1 + 4 + 1 + 2 + 1 + 1 + 4 + 3 = 22$.

Multiplicamos por 2,5. La puntuación SUS: $22 \cdot 2,5 = 55$. Por tanto, la puntuación en la escala SUS que recibe este sistema es de 55.

3.4.5. Otros cuestionarios estandarizados

- ASQ (after-scenario questionnaire)
- CSUQ (computer system usability questionnaire)
- CUSI (computer user satisfaction inventory)
- Ergonorm questionnaire
- IsoMetrics
- ISONORM 9241/10
- PSSUQ (post-study system usability questionnaire)
- PUEU (perceived usefulness and ease of use)
- PUTQ (purdue usability testing questionnaire)
- QUIS (questionnaire for user interaction satisfaction)
- USE (usefulness, satisfaction, and ease of use)

3.5. Dinámicas de grupo (*focus groups*)

La metodología de las dinámicas de grupo¹⁹ no nace de la usabilidad, sino que es una herramienta común en ámbitos como la sociología o el marketing.

⁽¹⁹⁾En inglés, *focus group*.

Se trata de reunir un grupo reducido de personas, 6 a 12 habitualmente, con el objetivo de conocer sus impresiones, opiniones, mejoras, reacciones... ante un producto o servicio o ante una situación; siempre hay un moderador que se encarga de dirigir la sesión y canalizar el debate hacia aspectos relevantes.

En usabilidad se aplican las dinámicas de grupo para discutir aspectos del producto o servicio que se analiza; habitualmente, el papel de moderador lo ejerce un experto en usabilidad y dirige la discusión hacia unos objetivos que se han marcado inicialmente por el equipo de diseño y desarrollo. Éste es un aspecto que no es fácil de controlar, ya que el moderador tiene que conseguir respuestas y al mismo tiempo dejar que fluya la conversación entre los participantes sin dirigirlos excesivamente.

Es importante que la metodología incluya más de una sesión con distintos grupos y que los participantes sean significativos en la medida en que podrían ser usuarios reales del servicio o usuarios potenciales.

Respecto al análisis de los datos generados, las sesiones de dinámica de grupos se tienen que plantear como una metodología creativa y no analítica, ya que generan datos cualitativos de múltiple interpretación. Es decir, la dinámica sirve, por ejemplo, para detectar los puntos más importantes que tenemos que mejorar en nuestra web antes de iniciar un proceso de DCU, pero no para evaluar si ésta es usable o no.

3.6. Clasificación de tarjetas (*card sorting*)

La **clasificación de tarjetas** (*card sorting*) es una técnica para explorar y observar cómo los usuarios ordenan y clasifican las categorías de información de un sitio web, producto o servicio.

Esta técnica es utilizada en la arquitectura de la información y ayuda a aumentar la probabilidad de que los usuarios finales del producto o servicio sean capaces de encontrar lo que buscan conforme a su modelo mental.

Destacamos dos tipologías de clasificación de tarjetas:

- 1) **Abierta:** los usuarios reciben categorías sin ningún tipo de grupo establecido, por tanto, pueden agrupar las categorías libremente como mejor crean.
- 2) **Cerrada:** los usuarios reciben categorías con grupos previamente definidos, con lo que el usuario debe ubicar cada categoría en el grupo que crea conveniente.

Es recomendable utilizar esta tipología de clasificación de tarjetas cuando añadimos nuevos contenidos a un sitio web en una estructura ya existente.

3.7. Registros de usuario

Los registros de usuario²⁰ forman parte de las técnicas de evaluación remota con usuarios. Los propios usuarios son quienes crean un registro de acciones, reacciones, elementos que les llaman la atención del producto que se evalúa.

Los usuarios efectúan una prueba de usabilidad con el producto siguiendo un guión preestablecido y registrando todas las acciones que han llevado a cabo para completar cada tarea; esta información incluye desde el enlace que han preferido utilizar para acceder a una información determinada, hasta los datos con los que han completado un formulario.

La diferencia con otras técnicas es que el usuario-participante realiza la tarea de registro de lo que sucede durante el test. En los tests remotos puede haber un sistema automático el que recoge lo que sucede.

Sin embargo, esta técnica presenta debilidades que hacen que los datos resultantes puedan tener una baja fiabilidad:

Enlaces de interés

Donna Spencer; Todd Warfel (2004) "Card sorting: a definitive guide". *Boxes and arrows* (7 de abril).

"Card sorting" (Information&Design)

Yusef Hassan Montero; Francisco J. Martín Fernández (2004). "Card Sorting: Técnica de categorización de contenidos". *No solo usabilidad* (23 de marzo).

⁽²⁰⁾En inglés, *self-reporting logs*.

- Excesiva dependencia de los usuarios de la prueba para recoger los datos de lo que sucede.
- Los participantes, consciente o inconscientemente, puede que no reflejen objetivamente cada paso.
- Se pierde toda la información sobre las expresiones del rostro del usuario y el lenguaje no verbal.

Una variante de esta metodología consiste en que el usuario registre, no de forma textual, sino mediante capturas de pantalla, lo que va sucediendo en cada momento. Se denomina *screen snapshots*.

3.8. Sesiones guiadas (*journaled sessions*)

Las sesiones guiadas²¹ fueron el germen de los actuales tests remotos de usabilidad. Por tanto, forma parte de las técnicas remotas de evaluación con usuarios.

⁽²¹⁾En inglés, *journaled sessions*.

Las **sesiones guiadas** consisten en hacer llegar a los usuarios del test un prototipo del producto que incluye un sistema que captura sus acciones. Por ejemplo, un CD con una versión beta de una web que incorpora algún sistema de registro. También es necesario proporcionar a los usuarios un guión de las tareas que tienen que realizar.

Los usuarios realizan un test convencional con el producto y sus acciones quedan registradas. Después graban un nuevo CD y se lo hacen llegar al equipo evaluador.

Con el avance tecnológico que permite a usuarios y evaluadores interactuar en línea con el producto y alcanzar un minucioso registro de las acciones, este método ha sido sustituido por el test remoto o el análisis de registros informáticos. En este sentido, las sesiones guiadas presentan los mismos inconvenientes metodológicos que los tests remotos pero con mayor complejidad en la gestión.

En cualquier caso, podría ser una opción interesante para la evaluación de sistemas no interactivos en los que no fuera posible el desplazamiento del usuario a un laboratorio. El registro de las acciones del usuario mediante software tendría que ser sustituido por otro método de registro, por ejemplo, la grabación en vídeo.

3.9. Registro informático (*logging*)

Un *log* es un registro de eventos de un sistema, entendiendo por *evento* los datos o la información referente a ¿qué usuario?, ¿qué ha hecho?, ¿dónde lo ha hecho?, ¿cuándo lo ha hecho? y finalmente, ¿por qué reaccionó así el sistema? Estos datos se almacenan en ficheros de texto formados por las peticiones y los resultados.

A priori, el *log* es un registro utilizado en seguridad y administración de sistemas informáticos, pero que puede tener un gran valor para la evaluación del comportamiento de los usuarios.

Un formato habitual de *log* web es el Common Logfile Format y Extended Log File Format, y puede tener este aspecto:

```
81.37.129.172 www.sanostra.es - [10/Aug/2004:08:25:24 -0100]
"GET /homeweb.nsf/fwHome?ReadForm&lang=02 HTTP/1.1" 200 20266
"http://www.google.es/search?hl=es&q=SA+NOSTRA"
"Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
```

Lectura complementaria

Adaptado de:

Jorge Sánchez Sánchez.
"Análisis de accesos a un servidor web de contenidos dinámicos". Proyecto final de carrera. Universidad de las Islas Baleares

Esta información se puede interpretar de la manera siguiente:

Campo	Contenido	Comentarios
Dirección IP del cliente	81.37.129.172	
Dirección del servidor	www.sanostra.es	
Nombre de usuario		El usuario no está autenticado en el servidor
Fecha y hora	[10/Aug/2004 08:25:24 -0100]	
Petición / Recurso	GET homeweb.nsf/fwHome?ReadForm&lang=02HTTP/1.1	Página, solicitada por el método GET utilizando el protocolo HTTP/1.1
Código de respuesta (status)	200	OK (petición correcta)
Tamaño de la respuesta	20266	Tamaño en bytes
Referer	http://www.google.es/search?hl=es&q=SA+NOSTRA	Enlace desde el buscador Google
User agent	Opera/8.5 (Macintosh; PPC Mac OS X; U; en)	Navegador Opera (versión 8.5) sobre un ordenador PowerPC y sistema operativo Mac OS X de Macintosh

El análisis de estos datos desde una perspectiva de evaluación de usabilidad puede ayudar a entender, por ejemplo:

- ¿cuál es el comportamiento habitual de los usuarios?; ¿qué partes de nuestra web o qué herramientas de nuestro sistema utilizan más? Esta infor-

mación es fundamental, por ejemplo, para detectar qué elementos del sistema son más populares o menos, e investigar el porqué.

- ¿qué problemas se encuentran en el uso del sistema?; ¿cuándo llegan a una respuesta KO?; ¿por qué sucede?

La única forma de generar *logs* no es mediante el servidor web, sino que se pueden instalar programas que recojan las actividades de los usuarios en sus propios ordenadores.

La metodología de *logging* tiene la ventaja de que puede recoger una cantidad ingente de datos a un coste muy bajo y de manera remota; el esfuerzo real del *logging* es el de interpretación de este volumen de datos.

4. Análisis de resultados

Después de hacer la prueba de usabilidad nos encontramos con una gran cantidad de datos de diversa naturaleza en nuestras manos, ¿y ahora qué? Llega el momento de hacer un uso correcto de los mismos sin tomarlos como mediciones únicas y absolutas que van a dar respuesta inmediata a la evaluación.

Los resultados aportan más información a nuestro trabajo, nos orientan para ir corrigiendo el rumbo todos los días y mejoran las soluciones que vamos proponiendo. Pero no hay que pretender resolver de forma automática cuestiones que preocupan o llegar a conclusiones precipitadas sin tener una visión amplia desde diferentes perspectivas. Quizás este sea el error más habitual de las pruebas de usabilidad.

De ahí que una de las partes más importantes de las pruebas sea completar un análisis responsable de resultados.

A continuación se presentan las etapas para analizar resultados, que pueden diferir dependiendo del proyecto de evaluación y de los métodos utilizados. En cualquier caso, el objetivo de estas etapas es establecer el orden que debe seguirse para transformar los datos y la información en resultados que posteriormente puedan ser utilizados en las presentaciones a clientes o al equipo de trabajo de un producto interactivo.

Las etapas para analizar los resultados son:

- Organización de la información
- Tratamiento de los datos
- Análisis de los datos

4.1. Organización de la información

La primera etapa en el análisis de resultados es la organización de la información. Para ello hay que recopilar y organizar toda la información que hemos recogido durante la evaluación:

- En un test con usuarios, para cada tarea, los tiempos dedicados, los resultados obtenidos, los comentarios de los usuarios, etc.
- En una dinámica de grupo, los comentarios de los participantes.

- En un *card sorting*, las organizaciones y ordenaciones que han hecho los participantes, las tarjetas o categorías que se han rechazado, comentarios de los participantes, fotos, etc.

La información recogida se puede organizar de distintas maneras dependiendo de los objetivos finales.

4.1.1. Organización sobre un criterio preestablecido

Se trata de organizar los datos según un criterio que decidamos previamente. Este criterio permitirá crear unidades de información más manejables para hacer el análisis.

Si en la fase de definición de objetivos de la evaluación hemos definido las características del producto o servicio que queremos evaluar, también podemos organizar los datos en función de estas características.

No hay una opción ideal o definitiva para organizar los resultados. Dependiendo del tiempo disponible para hacer el análisis, podemos optar incluso por varias organizaciones de datos. Esto tiene una ventaja, ya que es como mirar el mismo paisaje desde dos ventanas diferentes: la perspectiva cambia y ves más detalles. Lamentablemente, la situación más habitual es que tendremos un tiempo limitado para organizar nuestros datos.

Las perspectivas de organización que suelen ser más útiles son las que giran en torno a las tareas o también en torno a las características del producto o sistema. En cambio, no se recomienda una organización de la información que se centre en el usuario, por la sencilla razón de que no evaluamos al usuario, sino al producto o servicio.

Una vez establecido el criterio que vamos a seguir, podremos tener una organización de datos similar a este ejemplo, que se organiza siguiendo las características del sistema.

¿Cómo se decide el criterio?

Por ejemplo, ¿qué es más productivo, organizar los datos de una dinámica de grupo por “comentarios de participante” o por “temas tratados”?; o, en el caso de un test ¿por usuario o por tarea?

Depende...

La palabra *depende* es una palabra constante en el ejercicio del diseño centrado en el usuario. Corresponde al porcentaje de alternativas a las que cada profesional está sujeto por la definición, características, circunstancias, contexto, usuarios... de cada producto o servicio creado.

Características (criterio)	Descripción	Usuario	Tiempo	Comentarios
Navegación	El usuario no pudo encontrar la información que buscaba.	Usuario 3	0:03:35	El usuario probó diversas opciones para encontrar la información deseada y acabó desistiendo.
Navegación	El usuario atinó a la primera sobre la opción correcta.	Usuario 1	0:00:20	El usuario declaró que la opción tenía la misma etiqueta que en otro sistema conocido.

4.1.2. Organización sin un criterio preestablecido

Este tipo de organización se basa en agrupar los datos en función de lo sucedido durante la prueba.

En primer lugar, se tienen que poner en común los problemas y aspectos más importantes que se encontraron durante el test con usuarios. Después, se agrupan por similitud y se etiquetan utilizando categorías. Finalmente, se aplican estas categorías a los datos, y se estructura toda la información en una tabla similar a la del ejemplo anterior.

Con esta organización, quizás a simple vista no sea posible detectar con facilidad los problemas de usabilidad, por lo que también podemos tratar y analizar los datos mediante la aplicación de métricas antes de organizarlos de este modo.

4.2. Tratamiento de los datos: métricas

En este punto, es interesante recordar a Nielsen (2003), que define la usabilidad como un atributo de calidad que cuenta con diversas variables a partir de las cuales puede ser medida:

- Facilidad de aprendizaje²²
- Eficiencia de uso²³
- Facilidad para recordar²⁴
- Errores
- Satisfacción

La definición ISO 9241-11 se refiere también a criterios mensurables para definir la usabilidad: efectividad, eficiencia, satisfacción. De este modo, es lógico que en el análisis de los resultados de un proceso de evaluación de la usabilidad, se trabajen los indicadores mensurables o métricas.

A continuación presentamos una lista con métricas sencillas que podemos utilizar en un proceso de evaluación:

- Porcentaje de éxito o fracaso por tarea de los usuarios: literalmente, cuántos usuarios fueron capaces de realizar la tarea correctamente (efectividad).
- Tiempo por tarea.
- Tiempo por tarea con relación al éxito o fracaso (eficiencia).
- Número de errores cometido por los usuarios.

Lectura recomendada

J. Nielsen (2003). "Usability 101: Introduction to usability. Useit.com". *Alertbox*.

⁽²²⁾En inglés, *learnability*.

⁽²³⁾En inglés, *efficiency*.

⁽²⁴⁾En inglés, *memorability*.

ISO 9241-11 (1998)

Recordemos que La ISO 9241-11 (1998) define la usabilidad como:

"El grado por el cual un producto puede ser usado por unos usuarios específicos para alcanzar ciertas metas especificadas con efectividad, eficiencia y satisfacción en un contexto de uso especificado."

- Satisfacción subjetiva de los usuarios, que normalmente habremos recogido mediante algún cuestionario o entrevista final.

4.2.1. Eficiencia y efectividad

Eficiencia y efectividad son dos criterios que merecen un análisis independiente, pero veamos algunos aspectos importantes de estos conceptos. Primeramente debemos ser conscientes de que son métricas que pueden cambiar de componentes entre una prueba y otra ¿qué significa esto?

La eficiencia es un concepto que generalmente asociamos con ahorro: un coche eficiente es el que necesita menos combustible para cubrir una determinada distancia. En el diseño centrado en el usuario también tiene que ver con los recursos que el usuario tiene que “consumir” en su interacción con el sistema: tiempo, esfuerzo mental, dinero, intentos, trabajo, etc. Algunos ejemplos de datos que pueden componer la medida de eficiencia son:

- Tiempo en cada intento de resolver una tarea.
- Tiempo en las tareas de una tipología concreta (por ejemplo, de registro).
- Tiempo para realizar una tarea en relación con el tiempo de un experto.
- Número de etiquetas del sistema aprendidas.
- Número de clics, número de pulsaciones de teclado.
- Número de veces que se ha utilizado la función “atrás” del navegador (en pruebas relacionadas con web).
- Número de iconos que recuerda el usuario tras completar la prueba.
- Número de consultas al manual de uso, ayuda en línea, facilitador, etcétera.

La efectividad es la capacidad del usuario de alcanzar sus objetivos, lo que en la realidad de un test de usabilidad se refleja, por ejemplo, en el número de tareas realizadas con éxito. También podemos encontrar otros datos que compongan una medida de eficiencia:

- Número de funciones del producto o servicio utilizadas.
- Número de tareas completadas en un primer intento.
- Número de errores, errores más repetidos.

Lectura complementaria

Ejemplos adaptados de:
David Travis (2003). “Discount Usability: Time to Push Back the Pendulum?”. *Userfocus* (2 de mayo).

- Porcentaje de errores por tarea.
- Porcentaje de usuarios que completaron las tareas de una tipología concreta (por ejemplo, de proceso de compra).
- Porcentaje de tareas resueltas con ayuda (manuales, asistencia en línea, facilitador...).
- Número de tareas resueltas sin ayuda.

En el momento de diseñar la prueba habremos definido los datos que recogeremos y que posteriormente serán los componentes de las métricas de eficiencia y efectividad.

Para convertir los datos brutos en métricas requeriremos de cierto tratamiento numérico o estadístico.

4.2.2. Tratamiento de los datos numéricos

Algunos de los métodos de diseño centrado en el usuario implican la recogida de resultados cuantitativos, como por ejemplo el test con usuarios, encuestas y cuestionarios. Los resultados cuantitativos que se obtienen con estos métodos son, por ejemplo, el tiempo de ejecución para hacer una tarea o el número de personas que respondieron afirmativamente a una pregunta determinada. La presentación de las conclusiones de los resultados a veces incluye afirmaciones como “un 10% de los usuarios fallaron en la tarea”.

A continuación vamos a tratar aspectos básicos de análisis estadístico, orientados a los resultados de la evaluación de la usabilidad, presentando las principales métricas que podemos aplicar a los datos cuantitativos.

Media, mediana y moda

Tres métricas muy útiles para ayudarnos a dar sentido a los datos y a detectar patrones, son la media, la mediana y la moda. Pongamos por ejemplo los tiempos por usuario en una tarea y el número de errores cometido por cada uno.

Tarea 1		
Usuario	Tiempo	Errores
A	0:03:35	13
B	0:01:40	2
C	0:02:10	3
D	0:01:33	11
E	0:00:50	0
F	0:00:40	1
G	0:01:27	3
H	0:02:40	9
I	0:01:00	2
J	0:03:16	13
K	0:01:49	3
L	0:00:50	0
media	0:01:47	5
mediana	0:01:37	3
moda	0:00:50	2

La **media** es uno de los conceptos más familiares. Corresponde al promedio de un conjunto de números. En el ejemplo, la media de tiempo de todos los usuarios es 1 minuto y 47 segundos. El problema de la media es que se ve muy afectada por los valores extremos (los más altos y los más bajos).

Nuestro ejemplo contiene usuarios que hicieron casi el doble de tiempo que la media y usuarios que hicieron casi la mitad de tiempo que la media. En el caso del número de errores, la media es de 5, pero tenemos dos usuarios que no cometieron ningún error en absoluto durante la tarea. Es evidente que en la elaboración de nuestro análisis no podemos guiarnos únicamente por la media como indicativo.

La **mediana** es el valor que deja el mismo número de resultados antes y después que él mismo, es decir, que hay tantos datos por encima de la mediana como por debajo. En el ejemplo, la mediana de tiempo es de 1 minuto y 37 segundos, lo cual quiere decir que hay exactamente la misma cantidad de usuarios que superan ese tiempo como la cantidad de usuarios que hicieron un tiempo inferior.

La diferencia entre la media y la mediana en este caso es de diez segundos, lo que es significativo. Cuanto mayor sea esta diferencia, indicará la presencia de valores muy extremos.

Valores por encima de la mediana	A	0:03:35	A	13
	B	0:01:40	C	3
	C	0:02:10	D	11
	H	0:02:40	H	9
	J	0:03:16	J	13
	K	0:01:49	K	4
Valor de la mediana		0:01:37		3
Valores por debajo de la mediana	D	0:01:33	B	2
	E	0:00:50	E	0
	F	0:00:40	F	1
	G	0:01:27	G	2
	I	0:01:00	I	2
	L	0:00:50	L	0

La moda es el valor más repetido entre todos los que consideramos. Curiosamente, en el ejemplo, los dos valores más repetidos son un tiempo de 50 segundos para realizar la tarea y 2 errores cometidos.

Para tratar los datos obtenidos en un test con usuarios, conviene conocer los conceptos de media, mediana y moda, extraer conjuntamente sus valores para, de este modo, poder tener una aproximación realista de lo sucedido.

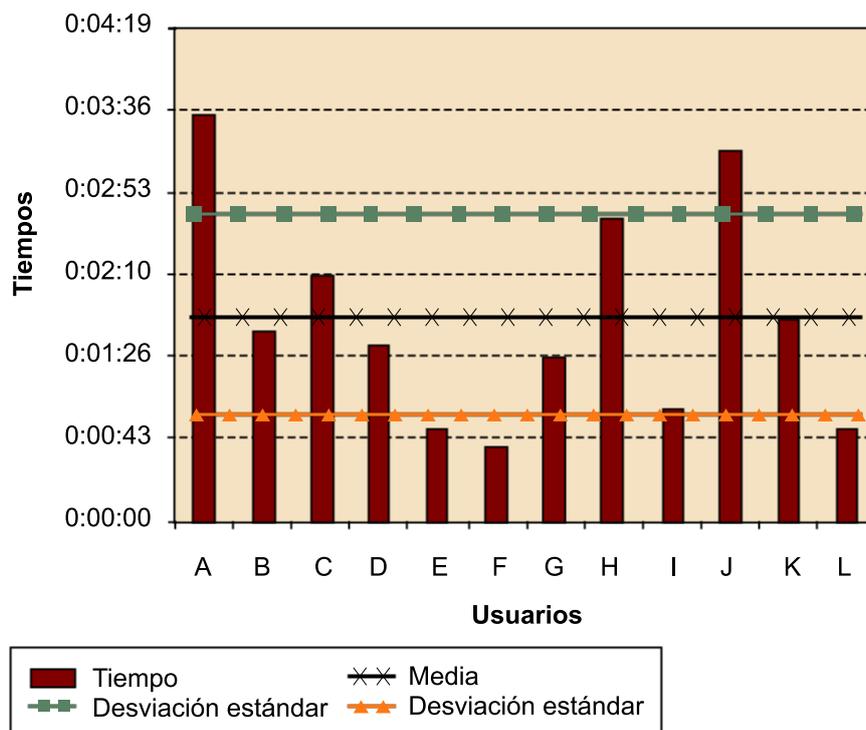
Desviación típica o estándar

La desviación estándar es una medida del grado de dispersión de los datos con respecto al valor promedio. Dicho de otra manera, la desviación estándar es simplemente el “promedio” o variación esperada con respecto a la media (Wikipedia).

En el caso de la evaluación de la experiencia de uso, aquellos datos de los que nos interesa saber la disparidad son el resultado de la interacción del usuario con un sistema. Volviendo al ejemplo anterior con los tiempos y los errores:

Tarea 1		
Usuario	Tiempo	Errores
A	0:03:35	13
B	0:01:40	2
C	0:02:10	3
D	0:01:33	11
E	0:00:50	0
F	0:00:40	1
G	0:01:27	2
H	0:02:40	9
I	0:01:00	2
J	0:03:16	13
K	0:01:49	4
L	0:00:50	0
media	0:01:47	5
mediana	0:01:37	3
moda	0:00:50	2
desviación estándar	0:00:53	4,546061

La desviación estándar del tiempo es de 53 segundos. Eso quiere decir que hay un intervalo de casi un minuto entero de tiempo de dispersión entre los datos de tiempo. En el caso de los errores, la desviación es de más de cuatro. Estos datos nos están confirmando la presencia de valores extremos que deberemos analizar.



Conclusiones

La aplicación y utilización de métricas nos ayuda a tratar e interpretar los datos, pero hay algunos peligros:

a) Los números son “adictivos”: alargar indefinidamente los cálculos sobre nuestros datos no siempre facilita la tarea de análisis. Hay que aplicar las técnicas justas –y media, mediana y moda son bastante útiles– como para que del conjunto de datos inicial sin ordenar podamos comenzar a sacar alguna conclusión útil.

b) Las métricas que hemos presentado pertenecen a la estadística descriptiva. Nos ayudan a entender mejor el significado de los datos que recogemos y se ajusta a los objetivos de la evaluación de la usabilidad mediante un test con usuarios. No obstante, si alguna vez se presenta la necesidad de extraer conclusiones de un proceso de evaluación que sean extrapolables a conjuntos más amplios de población, deberíamos utilizar la estadística inferencial y extremar el rigor en el diseño de la prueba (la muestra, las hipótesis de trabajo, las variables, etcétera).

En conclusión, la estadística pone a nuestro alcance herramientas que pueden ayudarnos a entender mejor los datos de un test con usuarios en un proceso de diseño centrado en el mismo. Media, mediana, moda y desviación típica pueden ser una señal que nos esté indicando “¡mira aquí!, ¿qué ha pasado?” y, de este modo, centrar nuestra atención en problemas de usabilidad. Durante la fase de conclusiones debemos completar los datos numéricos con otros datos cualitativos y así, obtener conclusiones más significativas y completas.

Estadística inferencial

La inferencia estadística o estadística inferencial es una parte de la estadística que comprende los métodos y procedimientos para deducir propiedades (hacer inferencias) de una población, a partir de una pequeña parte de la misma (muestra). La bondad de estas deducciones se mide en términos probabilísticos, es decir, toda inferencia se acompaña de su probabilidad de acierto. (Wikipedia)

4.2.3. Métricas en un test con usuarios: Ejemplo de aplicación

A continuación presentamos, a modo de ejemplo, una lista de datos y métricas que podríamos extraer de la información recogida en un test con usuarios:

- Métricas sobre los datos de tarea.
- Media de tiempo para completar la tarea.
- Media de tiempo para completar la tarea respecto del tiempo estimado. En algunos tests, el equipo evaluador realiza la tarea anticipadamente, se registra ese tiempo y se comparan los resultados con él.
- Mediana de tiempo para completar la tarea, y número de usuarios por encima o por debajo.
- Desviación estándar del tiempo para completar la tarea.

- Porcentaje de usuarios que realizaron la tarea correctamente.
- Porcentaje de usuarios que realizaron la tarea correctamente y en un tiempo determinado, que podría ser:
 - Tiempo estimado por el equipo evaluador
 - Mediana del tiempo
 - Desviación típica del tiempo
 - Eficiencia y efectividad por usuario y tarea
 - Datos de los cuestionarios: porcentajes de respuesta a las diferentes cuestiones
- Otros datos que podemos haber recogido durante el test:
 - Reacciones emocionales de los usuarios durante la prueba.
 - Número de comentarios positivos y negativos.
 - Número de usuarios que afirmaron que recomendarían el producto o servicio durante la prueba.

4.3. Interpretación de los datos

Una vez ordenados los datos y obtenidas las métricas, llega la fase de interpretación de datos. Se trata de analizar los “porqués” que hay detrás de los fenómenos observados en cada proceso de evaluación.

Muchos profesionales de la usabilidad entienden esta fase como un arte, como un proceso que requiere habilidad y experiencia. Ciertamente, el análisis pretende buscar los patrones de comportamiento de los usuarios que ayuden a expresar los datos y para ello es necesario un ojo entrenado. Tanto es así que, a medida que se adquiere esta experiencia, muchos profesionales de la usabilidad tienen menos necesidad de realizar los análisis más formales y comienzan a “ver” estos patrones ya durante las sesiones de test.

A continuación proponemos una serie de recomendaciones o direcciones hacia las que un analista debería dirigir su atención. También recuperamos el concepto de modelo mental de usuario y el impacto que puede tener en el momento de evaluar los resultados de un test.

4.3.1. Interpretación de los datos

Con los datos organizados y las métricas en la mano, el siguiente paso es enumerar los problemas de usabilidad que evidencian los datos. Es muy útil realizar este análisis tanto en positivo como en negativo.

- Negativo: se trata de analizar todos los datos relativos al fracaso en la resolución de tareas y argumentar el porqué.

Lectura recomendada

En el 2006, la Usability Professionals Association (UPA) celebró una sesión durante su conferencia anual sobre el análisis de resultados en evaluaciones de usabilidad:
Analyzing usability study results: Is it magic behind the curtain?

- Positivo: se trata de analizar todos los datos relativos al éxito en la resolución de tareas y argumentar el porqué.

Algunas recomendaciones para llevar a cabo el análisis son:

a) Decir y hacer

Los datos relacionados con las tareas realizadas por los usuarios tienen más fiabilidad que las informaciones lanzadas en voz alta durante el test de usabilidad, o después, en la entrevista posterior o en el cuestionario.

Esta recomendación tiene sentido en la aplicación de métodos como el test con usuarios siguiendo el protocolo de pensamiento manifiesto, que implica la realización de tareas acompañadas de verbalizaciones. No tiene sentido en otros métodos de evaluación que sean eminentemente verbales como una dinámica de grupo o una entrevista.

De alguna manera los comentarios del usuario son siempre un apoyo a la explicación sobre lo que ha pasado en el test, pero no son una conclusión en sí mismos. Podemos tomar en consideración un comentario de un usuario que exprese la dificultad de la tarea ("Qué difícil es esto") si está acompañado de unos datos de tarea que indican una baja eficiencia y efectividad. Si no, estaremos dando importancia a comentarios que pueden provenir, por ejemplo, de la simple falta de atención o nerviosismo de la prueba. Incluso, como comentamos cuando analizamos el papel del facilitador, el usuario puede entender que tiene que decir lo que el facilitador quiere oír, y exagerar en sus comentarios positivos o negativos.

Es importante no olvidar en ningún momento que estos tests se basan en la premisa de no analizar al usuario; ¿qué sentido tendría que basáramos nuestras conclusiones únicamente en sus comentarios?

Los comentarios de los usuarios son la sangre en la escena del crimen, pero no el cadáver. Podemos imaginar los comentarios de los usuarios como las salpicaduras de sangre. La forma de cada gota nos está dando más pistas sobre cómo ocurrió el crimen, pero no es el crimen.

b) Buscamos porqués

Analizando los datos podemos ver, por ejemplo, que la efectividad de los usuarios en una tarea determinada ha sido muy baja, esto es, muy pocos usuarios han sido capaces de llevarla a cabo. ¿Cuál puede ser la razón?, ¿qué tenían que hacer y qué han hecho?, ¿por qué no se daban las condiciones de éxito?

En un test con usuarios realizado para evaluar el uso de un horno microondas, ninguno de los usuarios fue capaz de interrumpir el proceso de calentar una taza de leche que no fuera abrir la puerta del microondas. Por tanto, utilizando un ejemplo cotidiano de uso de un artefacto, tenemos que un 0% de los participantes fue capaz de resolver la tarea. La razón era que ninguno de los usuarios pulsaba el botón adecuado, ya que es el mismo que se emplea para poner en marcha la cocción.

Varios usuarios comentan que “no ven el botón para parar”.

¿Cuál podría ser el análisis?: el dato es el 0% de efectividad, el comportamiento observado es “no pulsar el botón”, el comportamiento previsto es “pulsar el botón”.

Por tanto, podemos argumentar que hay un error de la interfaz. El mismo botón facilita dos funciones, algo complicado en el uso de sistemas, y no hay ningún elemento de diseño que ayude al usuario a entenderlo. Además, queda reforzado con los comentarios de los usuarios.

c) Buscar lo que es diferente, buscar lo que es igual

¿Qué hicieron todos los usuarios y qué no hizo ninguno y esperábamos que hicieran?

¿Qué comportamiento era igual en todas las tareas (en el caso de un test) y cambió en una tarea en concreto?

¿Qué tarea, que habían sido capaces de hacer al principio del test, no fueron capaces de repetir al final?

¿Qué tarjetas ordenó todo el mundo de la misma manera, en el caso de un card sorting, y qué tarjetas fueron dispuestas de manera diferente por todos los participantes?

¿Qué recorrido visual fue más repetido en un experimento con eyetracking, pero qué área visual quedó totalmente obviada por todos los participantes?

Todas estas consideraciones y muchas más pueden ayudarnos a detectar problemas en la interfaz.

d) Buscar lo extremo

Los datos que indiquen comportamientos extremos deben atraer nuestra atención. Si algo ha sucedido demasiado, o demasiado poco, probablemente nos revela que un elemento está fallando.

e) Retomar las notas del test

Si durante los tests tomamos notas de aspectos que nos llamaron la atención, es el momento de ver si los datos nos confirman alguna suposición o hipótesis. Observaciones como “todos los usuarios hacen uso de la ayuda” pueden ser corroboradas.

f) Pensemos que nos equivocamos

Es habitual que el analista sienta que ha descubierto el secreto de lo sucedido en la evaluación. Existe la tentación de hacer encajar los datos sobre una teoría implícita, es decir, forzar los datos de manera que encajen en nuestra idea de los problemas de la interfaz en lugar de detectar cuáles son esos problemas por medio de los datos.

Si el analista mantiene la suficiente humildad como para pensar que se puede estar equivocando, estará mucho más atento a los datos que pueden refutar su teoría.

También es importante intercambiar impresiones con el resto de observadores y el equipo, que pueden hacer una revisión crítica de nuestras conclusiones.

g) La navaja de Ockham

“Keep it simple.” (La explicación más sencilla suele ser la correcta.)

Evitemos que el resultado del análisis sean teorías complicadas sobre la interacción de los usuarios. El trabajo del analista debe ser pragmático.

En numerosas ocasiones no hace falta tener una visión completa del modelo mental de los usuarios para poder apuntar muchos errores de una interfaz.

La navaja de Ockham

La navaja de Ockham (a veces escrito Occam u Ockam), *principio de economía* o *principio de parsimonia*, es un principio filosófico atribuido a Guillermo de Ockham (1280-1349), según el cual cuando dos teorías en igualdad de condiciones tienen las mismas consecuencias, debe preferirse la teoría más simple a la más compleja. (Wikipedia)

Resumen

La evaluación de la usabilidad con usuarios es un aspecto clave en el proceso de diseño centrado en el usuario, ya que nos permite detectar aquellos aspectos que funcionan y aquellos que no, de un producto o sistema interactivo. Teniendo en cuenta este contexto, en este módulo didáctico se han tratado los principales métodos para la evaluación de la usabilidad con usuarios.

El test de usabilidad es el instrumento principal de evaluación de la usabilidad con usuarios. El test con usuarios persigue resolver problemas en la interfaz de un producto interactivo obteniendo información directamente de los usuarios que lo utilizarán en el futuro.

La evaluación de la usabilidad no garantiza que un producto sea completamente usable en el futuro, pero permite realizar una iteración más del proceso de diseño centrado en el usuario, permitiendo eliminar errores y rediseñar aspectos clave para la experiencia de uso.

Bibliografía

- Alreck, P. L.; Settle, R. B.** (1994). *The survey research handbook*. Chicago: Irwin Professional Publishing.
- Beyer, H.; Holtzblatt, K.** (1998). *Contextual design. defining customer-centered systems*. San Francisco: Morgan Kaufmann.
- Beyer, H.; Holtzblatt, K.** (1995). Apprenticing with the customer: A collaborative approach to requirements definition. *Communications of the ACM* (mayo).
- Card, S.; Moran, T.; Newell, A.** (1983). *The psychology of human-computer interaction*. Hillsdale: Lawrence Erlbaum Associates.
- Dumas, J. S.; Redish, J.** (1993). *A Practical Guide to Usability Testing Ablex*. Norwood.
- Foddy, W.** Constructing questions for interviews and questionnaires: Theory and practice in social research. Cambridge Univ Pr (Pap Txt).
- Hassan Montero, Y.; Herrero Solana, V.** (2007). "Eye-Tracking en Interacción Persona-Ordenador". *No Solo Usabilidad* (núm. 6).
- Hassan Montero, Y.; Martín Fernández, F. J.** (2003). "Método de test con usuarios". *No Solo Usabilidad* (núm. 2).
- Kirakowski, J.** "Questionnaires in usability engineering"
- Krug, S.** (2001). *No me hagas pensar*. Madrid: Prentice Hall.
- Lessler, J. L.** *Questionnaire design in the cognitive research laboratory*.
- Lewis, C.; Mack, R.** (1982) (ed.). "Learning to use a text processing system: Evidence from thinking aloud protocols". Conferencia SIGCHI sobre factores humanos en sistemas de computación (págs. 387-392). Gaithersburg (MD).
- Löwgren, J.** (1995). *Perspectives on usability*. Linköping: Department of Computer and Information Science, Linköping University. IDA Technical Report.
- Macaulay, L. A.** (1996). *Requirements engineering*. Berlín: Springer Verlag Series on Applied Computing.
- Nielsen, J.** (1994). *Usability engineering*. San Francisco: Morgan Kaufmann.
- Nielsen, J.** (2000). "Why you only need to test with 5 users". Useit.com Alertbox.
- Nielsen, J.** (2001). "First rule of usability? Don't listen to users". *Alertbox* (núm. 5, agosto).
- Nielsen, J.** (2003). "Usability 101: Introduction to usability. Useit.com". *Alertbox*.
- Norman, D.; Draper, S.** (1986). *User centered system design: new perspectives on human-computer interaction*. Hillsdale: Erlbaum.
- Oppenheim, A. N.** (1992). *Questionnaire design, interviewing and attitude measurement*. Pinter Pub Ltd.
- Preece, J.; Rogers, Y.; Sharp, H.; Benyon, D.; Holland, S.; Carey, T.** (1994). *Human-Computer Interaction*. Reading MA: Addison-Wesley.
- Ribera Turró, M.** (2005). "Evolución y tendencias en la interacción persona-ordenador". En: *El profesional de la información* (vol. 15, nov.-dic., núm. 6, págs. 414-422).
- Salant, P.; Dillman, D. A.** (1994). *How to conduct your own survey*. Nueva York, NY: John Wiley & Sons.
- Snyder, C.** (2003). *Paper prototyping: the fast and easy way to design and refine user interfaces*. San Francisco: Morgan Kaufmann.
- Suchman, L. A.** (1987). *Plans and situated actions: the problem of human-machine communication*. Nueva York: Cambridge University Press.

