Mètodes d'avaluació amb usuaris

Amaia Calvo-Fernández Rodríguez
Sergio Ortega Santamaría
Alicia Valls Saez

PID_00176614

UOC
Universitat Oberta de Catalunya
www.uoc.edu

Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-Compartir igual (BY-SA) v.3.0 Espanya de Creative Commons. Podeu modificar l'obra, reproduirla, distribuir-la o comunicar-la públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), i sempre que l'obra derivada quedi subjecta a la mateixa llicència que el material original. La llicència completa es pot consultar a http://creativecommons.org/licenses/by-sa/3.0/es/legalcode.ca

Índex

Introducció 5

1. L'avaluació de la usabilitat amb usuaris 9
1.1. Evolució del concepte d'*usuari* i de les interfícies 9
1.2. Primera aproximació als tests amb usuaris 13
1.2.1. Treball amb models mentals 14
1.2.2. Què és un test amb usuaris 17
1.2.3. Quan s'ha de dur a terme un test amb usuaris 18
1.2.4. Persones necessàries en un test amb usuaris 21
1.2.5. Laboratoris i eines per a l'avaluació de la usabilitat 25
1.3. Procés del test 31
1.3.1. Definició i objectius del test 32
1.3.2. Preparació del test 33
1.3.3. Captació dels participants 41
1.3.4. Execució del test: exemple pràctic 42
1.3.5. Consells per a ser un bon facilitador de test 45
1.4. Classificació dels tests amb usuaris 50
1.4.1. Presencial/remot 50
1.4.2. Formal / de guerrilla 54
1.4.3. Tests amb prototips d'alta/baixa fidelitat 55

2. Variants del test amb usuaris 56
2.1. Test de codescobriment 56
2.2. Seguiment ocular 57
2.3. Llapis i paper 58

3. Altres mètodes d'avaluació amb usuaris 61
3.1. Indagació contextual 62
3.2. Entrevistes 64
3.3. Enquestes 65
3.4. Qüestionaris 65
3.4.1. *Software usability measurement inventory* (SUMI) 66
3.4.2. *Measuring the usability of multi-media systems* (MUMMS) 66
3.4.3. *Website analysis and measurement inventory* (WAMMI) ... 67
3.4.4. *System usability scale* (SUS) 67
3.4.5. Altres qüestionaris estandarditzats 70
3.5. Dinàmiques de grup (*focus groups*) 71
3.6. Classificació de targetes (*card sorting*) 72
3.7. Registres de seguiment d'usuari 72
3.8. Sessions guiades (*journaled sessions*) 73

3.9. Registre informàtic de seguiment (*logging*) 73

4. Anàlisi de resultats..... 76
4.1. Organització de la informació 76
4.1.1. Organització sobre un criteri preestablert 77
4.1.2. Organització sense un criteri preestablert 77
4.2. Tractament de les dades: mètriques 78
4.2.1. Eficiència i efectivitat 79
4.2.2. Tractament de les dades numèriques 80
4.2.3. Mètriques en un test amb usuaris: exemple d'aplicació 84
4.3. Interpretació de les dades 84
4.3.1. Interpretació de les dades 85

Resum..... 89

Bibliografia..... 91

Introducció

El 2008, la revista *Fortune* entrevistava Steve Jobs, CEO d'Apple, i li preguntava sobre les claus de l'èxit de la companyia. Jobs il·lustrava una de les seves nombroses respostes amb una citació de Henry Ford, dissenyador i fabricant de cotxes i fundador de la Ford Motor Company.

Lectura complementària

Entrevista a Steve Jobs, CEO d'Apple:

- "Steve Jobs speaks out" [article en línia]. *Fortune* (2008, 7 de març).

> "Si hagués preguntat als meus clients què volien, m'haurien dit que un cavall més ràpid."
>
> Henry Ford

Aquesta afirmació constitueix una de les claus de les pàgines següents. Apple no sol recórrer a la investigació de mercats per a aconseguir el desenvolupament de productes, però sí que fa enquestes als seus usuaris i contínuament n'obté informació per a saber què és el que els agrada o què és el que poden necessitar en cada moment.

Tanmateix, ni Apple ni moltes altres empreses que treballen amb productes i serveis interactius no necessiten, en el sentit més literal de la frase, conèixer directament l'opinió dels consumidors. Hem de reconèixer que les persones que utilitzem els productes esmentats no som bones articulant i verbalitzant una resposta precisa i clara sobre les nostres necessitats.

En definitiva, no són els usuaris els qui han d'oferir solucions concretes –demanar cavalls veloços o màquines potents–, però sí que han de ser la font d'informació més fiable que ens permeti valorar-ne la participació en els processos de construcció.

Per a descobrir el millor disseny i optimitzar progressivament les nostres interfícies web, necessitem veure com els usuaris treballen, com executen les tasques que hem dissenyat, i tot això seguint tres "normes" ineludibles, que Nielsen (2001) destaca seguint la línia que hem marcat fins ara:

Lectura complementària

J. Nielsen (2001, agost). "First rule of usability? Don't lisen to users" [article en línia]. *Alertbox* (núm. 5).

1) Observar què fa i com es comporta realment la gent.

2) No creure allò que diuen que fan.

3) No creure allò que la gent prediu del futur i del que poden arribar a fer.

Insistirem en aquesta filosofia de treball en els pròxims apartats, però vegem què passa de la banda dels dissenyadors i desenvolupadors.

Quan un equip de desenvolupament crea un producte o servei interactiu, pensa que aquest serà útil i fàcil d'usar. Els seus membres creuen que han col·locat el botó en el lloc més visible de la pàgina, que han dissenyat un formulari clar i agradable, i que han ordenat els continguts d'un lloc de la manera més lògica possible.

Totes les idees preconcebudes per part d'aquest equip desapareixen en el moment en què un usuari es troba per primera vegada amb aquest producte o servei. Potser no troba el botó perquè pensava que havia de ser en un lloc determinat de la interfície, no aconsegueix ni entendre el formulari ni acabar-lo o no troba conceptes bàsics dins de l'arbre de continguts, que han estat pensats i analitzats estratègicament.

És llavors quan ens adonem que avaluar els nostres productes o serveis amb usuaris reals pot ser una bona manera de saber si anem pel bon camí, és a dir, si el producte o el servei que desenvolupem és eficient, i fàcil i agradable d'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 ayudará 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

L'avaluació de la usabilitat no sempre requereix la participació dels usuaris. Alguns mètodes d'avaluació faciliten la localització de problemes d'usabilitat i l'assoliment de millores en el disseny a partir del compliment de principis, la consulta a experts o l'anàlisi de mètriques, per assenyalar alguns exemples. De vegades, valorar l'ús d'aquestes metodologies abans d'implicar l'usuari ens pot ajudar a reduir una quantitat rellevant de problemes d'usabilitat que amb una avaluació amb usuaris podrien representar un cost elevat.

Per això, l'objectiu principal d'una avaluació amb usuaris, també coneguda com a *test d'usabilitat amb usuaris*, és verificar la presència de problemes més importants d'usabilitat en el nostre producte o servei interactiu (els que generen més impacte en l'usuari) per a així poder oferir de manera eficaç solucions als problemes trobats.

Hi ha bones raons per a dur a terme avaluacions amb usuaris durant el que s'anomena *cicle de vida complet del producte* o *servei*. Podem escollir diferents moments per a efectuar la prova:

- Abans de començar-lo a crear, per a detectar de manera precoç errors de disseny.

- Durant la construcció, per a no desviar-nos dels objectius o tasques.

- Després del llançament, per a comprendre què haurem de millorar en una versió següent o per a treballar sobre un redisseny.

Ara bé, quan parlem de disseny centrat en l'usuari (DCU) parlem d'un enfocament d'investigació i un desenvolupament de principis guiats pel coneixement constant dels usuaris i els seus models mentals a partir dels seus processos d'interacció. Això implica que si aconseguim obtenir una bona informació en etapes primerenques, les probabilitats d'aconseguir un producte i la seva funcionalitat adequada per a usuaris concrets seran més altes.

No hem d'oblidar que la cerca d'una millora progressiva del producte ens convida a conèixer com més aviat millor els errors de disseny.

1. L'avaluació de la usabilitat amb usuaris

La història del mètode d'avaluació amb usuaris està íntimament lligada al desenvolupament de la interacció persona-ordinador (IPO[1]), que al seu torn no s'entendria sense analitzar l'aparició de la tecnologia com a element de gran consum (en especial, l'ordinador personal a partir de 1980) i la creació de les interfícies gràfiques d'usuari (GUI[2]).

(1) *IPO* és la sigla d'*interacció persona-ordinador*.

(2) *GUI* és la sigla de l'expressió anglesa *graphical user interface*, en català *interfície gràfica d'usuari*.

Però també és important atendre altres aspectes crucials que van sorgir al final de la dècada dels setanta i al principi dels vuitanta, i que també ens permetran entendre el començament dels tests d'usabilitat. A saber:

- Evolució de la psicologia cognitiva i una valoració més gran dels processos cognitius humans que ens han permès conèixer com les persones senten, perceben, actuen, emmagatzemen i recuperen informació, etc.
- Proliferació d'estudis, investigacions i publicacions relacionades amb el processament humà de la informació.
- Augment de models sobre sensació, percepció, memòria, etc., i en conseqüència, augment d'experiments parcials i acadèmics centrats en l'execució de tasques.
- Desenvolupament de les interfícies gràfiques, que representen una millora en la comunicació amb els usuaris a partir d'avenços i millores en la interacció, en la visualització o en l'organització.

1.1. Evolució del concepte d'*usuari* i de les interfícies

Els propòsits principals dels primers tests d'usabilitat eren l'obtenció de validacions estadístiques o l'avaluació al final del procés per a demostrar que tot era correcte.

Però, atesos els diferents aspectes esmentats més amunt, es va començar a observar que aquest tipus d'actuació provocava massa frustracions en la creació de productes i serveis dirigits a un públic determinat. Ja no es tractava de saber i confirmar que tot funcionava correctament o de validar un nou model. El benefici principal era descobrir i esbrinar on eren els errors, com es produïen o què era el que no funcionava bé.

De tot això es desprèn que, gràcies als avenços en camps com la comunicació, les ciències de la informació, la psicologia cognitiva i la interacció persona-ordinador, ara sabem més coses sobre com els éssers humans processen la informació.

La consideració d'aquests avenços ens ha portat a incloure conceptes com avaluació precoç, continuada i iterativa en els enfocaments donats a les metodologies, com també a abandonar-ne l'exclusivitat científica o acadèmica per a convertir-se en una valuosa, àgil i eficient manera d'obtenir la informació adequada en la construcció dels productes interactius.

D'aquests avenços assolits en les últimes dècades, es pot esmentar la importància del progrés en les interfícies d'usuari. La seva evolució ha facilitat l'augment de la productivitat que s'ha promès durant anys gràcies a l'ús de la tecnologia.

Hem après que la interfície d'usuari és un dels elements crítics en l'acceptació i apropiació de tecnologies, productes i serveis per part de l'usuari. El seu disseny ha requerit plantejaments acurats perquè el model creat per l'usuari fos correcte i, en part, coincident amb el del que tenia el creador.

El disseny d'una interfície ha de prendre en consideració com les persones adquireixen, interioritzen, processen o exterioritzen informació a partir de les sensacions, les percepcions i les interaccions per aconseguir més productivitat, precisió i satisfacció.

En el seu llibre *En el principio... fue la línea de comandos*, Neal Stephenson ens parla sobre l'aparició de les GUI, no solament per entendre l'impacte de les interfícies en la nostra vida, sinó també sobre les contradiccions que s'han creat amb la seva aparició.

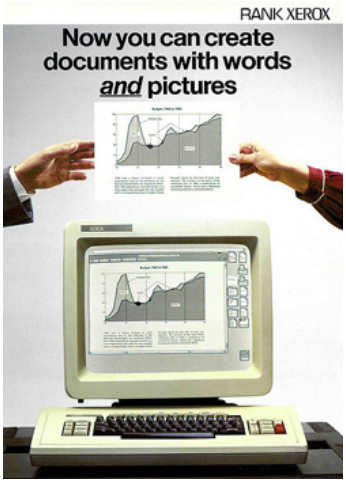
"Si el vídeo se hubiera inventado hace cien años, tendría una ruedecita para la sintonización y una palanca para avanzar y rebobinar, y una gran asa de hierro forjado para cargar o expulsar los casetes. Llevaría un gran reloj analógico delante, y habría que ajustar la hora moviendo las manillas en la 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 l'àmbit dels ordinadors personals, el salt inicial cap a aquesta interfície "que los clientes utilicen realmente" es va donar amb l'aparició de conceptes com ara *Finestra*, *Icona*, *Menú*, *Punter* (*Window*, *Icon*, *Menu*, *Pointing Device*, *WIMP*), aplicat per l'equip de Xerox Palo Alto Research Center (PARC) en el model Xerox Alto, que era experimental. El seguiria el Xerox Star, que es va comercialitzar el 1981.



Font: *Wikipedia* (CC-BY-SA 3.0)

L'intens treball en les proves d'usabilitat que els investigadors del PARC van dur a terme per al desenvolupament de l'Star i el desig d'ajudar l'usuari a comprendre i acceptar nous plantejaments, desenvolupaments, metàfores, etc., en les interfícies gràfiques d'usuari va conduir a una forma d'interacció persona-ordinador més enriquidora i beneficiosa.

Els membres de l'equip eren conscients d'aquests canvis i així ho feien saber en els informes i presentacions públiques.

> "L'enregistrament en vídeo ha estat una eina molt important. Per començar, la càmera ens permetia veure i escoltar el subjecte sense haver-nos d'estar a la mateixa habitació. En segon lloc, era un registre de totes les activitats, per la qual cosa no necessitàvem prendre notes perfectes durant l'experiment. En tercer lloc, els dissenyadors es convencien millor amb les cintes que amb els nostres freds nombres que la gent tenia problemes amb el sistema."
>
> "Human factors testing in the design of Xerox's 8010 Star office workstation". Conferència CHI de 1983.

Com dèiem al començament, en els anys vuitanta l'IPO vivia un període de grans avenços i la introducció de la psicologia cognitiva, especialment a partir de les aportacions de Card, Moran i Newell (1983), i de Norman i Draper (1986), ha facilitat la cerca d'un objectiu molt clar: conèixer i representar com interactua l'ésser humà amb l'ordinador.

Lectures complementàries

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.

D'aquesta manera, la conclusió anterior dels investigadors del PARC, òbvia des d'una perspectiva actual, resumeix perfectament aquest desig per conèixer i avaluar involucrant l'usuari real en els processos de disseny. Les virtuts dels tests amb usuaris començaven a aflorar.

Apple es va basar en la interfície del Xerox Star per crear la interfície d'Apple Lisa, i també va dur a terme tests amb usuaris, reclutant voluntaris entre la seva família i amics per a provar el nou sistema. Recollint novament una altra de les idees de Steve Jobs en l'entrevista a *Fortune*, la idea no és intentar enganyar la gent, no es tracta de convèncer-los de la necessitat d'una cosa que ells no volen. Hem d'esbrinar què és el que nosaltres volem i tenir el control adequat per a poder pensar si una gran quantitat de persones també ho voldran.

Lectures complementàries

"Steve Jobs speaks out" [article en línia]. *Fortune* (2008, 7 de març).

Potser la dificultat d'aquest procés és obtenir el significat amb relació a d'altres. Dit d'una altra manera, les nostres percepcions i significats dels productes que construïm no poden ser incompatibles amb la possibilitat de convertir-se en percepcions i significats compartits.

Durant les proves fetes per l'equip del Xerox PARC, els membres van recollir resultats que confirmaven aquesta idea de la percepció i el significat individual davant la percepció i el significat dels altres.

> "Quan el programari requeria confirmació de l'usuari, mostrava una petita finestra anomenada «quadre de diàleg», que contenia una pregunta i presentava dos botons, per a la confirmació positiva o negativa. Els botons tenien les etiquetes «Do It» i «Cancel». Els dissenyadors van observar durant els tests que els usuaris tendien a encallar-se en aquest quadre de diàleg i acabaven clicant sobre «Cancel», quan s'esperava que ho fessin sobre «Do It».
>
> Finalment, el facilitador del test va interrompre la sessió amb un usuari que semblava particularment molest per la finestra per preguntar-li quin era el problema. L'usuari va respondre «I'm not a dolt, why is the software calling me a dolt?» (No sóc un idiota, per què el programari em diu idiota?; «dolt» és la paraula anglesa per a «idiota»).
>
> Va resultar que l'usuari no percebia l'espai entre la «o» i la «I» de «Do It», i la tipografia del sistema feia que la «I» semblés una lletra ela. Per tant, els usuaris arribaven al quadre de diàleg i veien que havien d'escollir entre «Idiota» o «Cancel·lar».
>
> L'equip va decidir substituir el «Do It» per l'«OK», més directe, que coneixem bé, però que al principi havien descartat perquè els semblava excessivament col·loquial."
>
> Andy Hertzfeld, "Do It", *folklore.org* (1982, juny)

En qualsevol cas, tant Xerox Star com Apple Lisa van ser fracassos comercials, també provocats per l'alt preu dels equips i cert desencant del públic respecte a les prometedores interfícies gràfiques.

L'IPO també es va transformar en qüestionar les teories cognitives i en incorporar altres disciplines com l'antropologia per a tenir en compte també l'entorn i les situacions d'ús dels sistemes (Suchman, 1987; Löwgren, 1995) i valorar l'interès creixent que tenia la usabilitat.

En la dècada següent, serà Nielsen (1994) qui introduirà els *discount usability methods* a partir de l'enginyeria de la usabilitat (IU) com una manera més barata i qualitativa –però igualment operativa– d'efectuar proves d'usabilitat, especialment el test amb usuaris.

El desenvolupament i l'evolució d'Internet han consolidat una nova trajectòria en què la competitivitat, la productivitat i el compliment d'objectius han afavorit els esforços professionals i han millorat la realització de proves ràpides però contínues durant el procés de desenvolupament del producte o servei.

Lectures complementàries

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*. Linköpping: Linköpping University, Department of Computer and Information Science. IDA Technical Report.

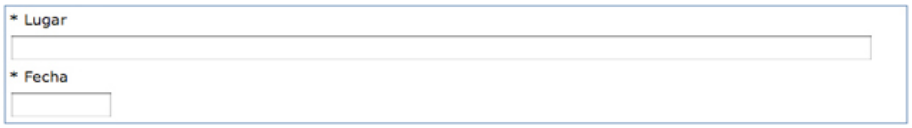
D. Norman; S. Draper (1986). *User centered system design: new perspectives on human-computer interaction*. Hillsdale: Erlbaum.

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

M. Ribera Turró (2005, novembre-desembre). "Evolución y tendencias en la interacción persona-ordenador". *El profesional de la información* (vol. 15, núm. 6, pàg. 414-422).

1.2. Primera aproximació als tests amb usuaris

Observeu els camps següents d'un formulari recollit d'una pàgina a Internet:



Sense tenir una contextualització clara, tan sols observant la posició, la forma i les mides, què escriuríeu en cada quadre i com ho faríeu?

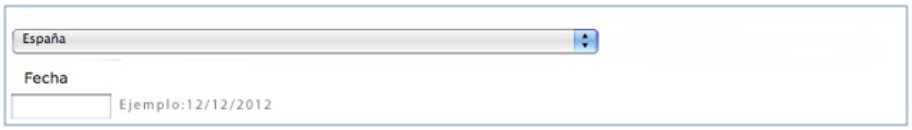
Podem deduir que el camp *Lugar* ens permet introduir una gran quantitat de caràcters, donant per fet que alguns llocs poden estar formats per paraules extenses. Per exemple: Vilafranca del Penedès (Barcelona).

El camp *Fecha*, per la seva banda, no aclareix com s'ha d'escriure la data i sembla que no és important per a recollir la informació. Això sí, potser deduirem que la podem escriure amb nombres perquè no cabria en el quadre de cap altra manera. Llavors, podríem escriure una data amb aquest format (12/12/12) o amb aquest altre (12-12-12), o fins i tot amb aquest últim (12 12 2012).

Un detall més. En tots dos camps veiem un asterisc al costat del nom. No s'hi indica res però podem fer dues suposicions: una seria que el camp és obligatori i no es pot deixar buit; una altra seria que hi ha una anotació en la part inferior de la pàgina que aclareix algun aspecte d'aquests camps marcats.

Amb aquest exemple notòriament confús, podem comprovar la quantitat de suposicions que es poden arribar a fer quan algú ens sol·licita unes simples dades i la manera d'introduir-les no és clara. Fins i tot, si demanéssim més opinions sobre la manera d'introduir les dades, trobaríem més variants possibles.

La solució per a aquesta tasca sembla senzilla. N'hi ha prou de no concedir tanta llibertat i marge d'error a l'usuari i evitar l'ambigüitat al màxim.



Si col·loquem els elements on s'espera que siguin, ens assegurarem que aquests seran percebuts. Però a més, si evitem la confusió i generem interaccions que concordin amb les expectatives, crearem experiències que l'usuari trobarà satisfactòries. Això significa estar en sintonia amb els models mentals dels nostres usuaris.

Ara bé, en un cas tan senzill com el que hem explicat podem intuir un disseny equilibrat i adequat que vagi d'acord amb les accions que farem en moltes altres pàgines. Però, què passaria si la resposta no fos tan senzilla?

Penseu en alguna cosa més complexa com, per exemple, un procés de compra, una configuració d'un compte o, per anar a una cosa més propera per a tots, penseu en les opcions de configuració d'una política de privacitat en una xarxa social. Com la percebran els usuaris? Quins conceptes es poden introduir? Quines interaccions seran necessàries?

1.2.1. Treball amb models mentals

Construïm models mentals a partir de la nostra experiència i els emmagatzemem en la nostra ment. Aquesta afirmació que sembla tan evident ens recorda, tal com vèiem en el subapartat anterior, que en tot el que fem participen les nostres idees, perspectives, pressuposicions o estratègies. És a dir, hi participen les nostres idees generals, que donen forma als nostres pensaments i a tots els nostres actes.

> Les estratègies que la gent empra per a resoldre un problema es fonamenten en el que cada individu ja coneix o creu que coneix.

Exemple de model mental

Lindgaard (1994) proposa un exemple que aclareix més el tema dels models mentals.

Imaginem que fem un pa i, quan el traiem del forn, no ha crescut com esperàvem. Què hem fet malament?

El forn era massa calent quan hem introduït el pastís, la qual cosa ha matat les cèl·lules de llevat, i per això el pa no ha pujat.

No s'han barrejat bé els ingredients i el llevat no ha pogut fer la seva funció.

La quantitat de llevat era insuficient per a la quantitat de massa.

No s'ha deixat fermentar prou la massa abans de posar-la al forn.

Algú ha obert la porta del forn abans d'hora i n'ha interromput la cocció.

Depenent de quina d'aquestes explicacions pensem que és la correcta, adoptarem una estratègia o una altra la pròxima vegada que fem un pa: farem millor la barreja, variarem les quantitats dels ingredients, controlarem millor la temperatura del forn. I l'elecció de l'explicació dependrà de la nostra experiència anterior en l'elaboració de pans; si som forners experts, podrem analitzar el color, el nivell de cocció de la massa o la textura, per a arribar a una conclusió.

Si mai no havíem fet un pa, tindrem menys eines per a analitzar les "proves del delicte" i arribar a una conclusió.

El que l'exemple vol il·lustrar és que, a l'hora de fer una elecció, el raonament difereix enormement entre unes persones i altres, entre usuaris novells i experts d'un sistema (forners o pocatraces).

Si ens traslladem al terreny de les interfícies, passa que en una avaluació, davant d'un problema o elecció, els nostres usuaris adopten una estratègia relacionada amb la manera en què entenen el problema que se'ls presenta: alguns pensen que el forn és massa calent, d'altres afegeixen més llevat, d'altres remenen amb més vigor. És a dir, actuen segons el seu model mental de "maneres de fer un pa perfecte".

Per tant, els models mentals no són "patrimoni" únic de les etapes de recollida de dades sobre el nostre usuari o del disseny i conceptualització del producte. Arriben a nosaltres durant el test amb usuaris i hi hem de parar atenció en el moment de l'anàlisi.

Un model mental és la representació de la realitat que cada persona es construeix per entendre fenòmens específics. Don Norman els va definir així: "En la interacció amb l'entorn, amb els altres, amb els artefactes tecnològics, la gent es forma models mentals interns d'ells mateixos i de les coses amb què interactuen. Aquests models proporcionen capacitat de preveure i explicar la interacció" (Getner i Stevens, 1983).

Des de la perspectiva de l'avaluació, que és la que abordem en aquestes pàgines, s'exposen algunes característiques importants dels models mentals que convé destacar abans de donar pas a les explicacions sobre els tests amb usuaris.

a) Els models mentals són incomplets

Imaginem que volem avaluar un nou cercador. Què passaria si el model mental de tots els nostres usuaris fos "això és com en Google"? Podem haver influït en part en la creació d'aquest model incomplet amb el disseny del nostre test?

Sovint les avaluacions també són parcials, per tant, cal jutjar críticament si el model que han adoptat els usuaris estava motivat per les característiques del nostre test:

- Si avaluem un prototip o no n'avaluem en profunditat totes les funcionalitats, és natural que donin lloc a un model mental incomplet.

- Si les tasques o escenaris no inclouen el conjunt del producte o servei, és natural que els usuaris no tinguin un model mental complet.

b) Els models mentals no tenen límits precisos: diferents dispositius i interaccions es confonen els uns amb les altres

Durant l'avaluació, els usuaris han utilitzat com a referència altres dispositius per a interactuar? Quan l'usuari ha expressat verbalment el que pensa, s'han registrat frases com "això funciona com això o allò" o "provaré a fer el que faria amb això o allò, per veure si funciona igual"?

c) Els models mentals tenen un component "acientífic" i "supersticiós"

Podem descobrir que els nostres usuaris han desenvolupat "rituals" d'interacció durant les proves d'avaluació. Definiríem els rituals com a patrons d'interacció que els usuaris adopten, que saben racionalment que són innecessaris, però que els permeten pensar que el sistema respon amb més eficàcia.

El comportament supersticiós es reforça en la interacció amb sistemes en els quals hi ha transacció econòmica. Per exemple, en unes proves amb màquines recreatives d'atzar (popularment, escurabutxaques), comprovem que els usuaris pensaven que la manera d'arrepenjar-se en la màquina o de prémer els botons influïa sobre el resultat.

Aquests comportaments són difícils de racionalitzar, però és necessari que, com a equip que dissenya o avalua la interacció amb el producte, en registrem l'existència i els tinguem en compte a l'hora d'analitzar els resultats d'una avaluació.

Altres "supersticions" d'usuari

Altres "supersticions" d'usuari que hem recollit en diferents projectes: "quan navego per una entitat bancària, tanco la resta de finestres", "quan un botó falla, el torno a prémer però durant més estona".

Don Norman apunta, "encertadíssimament"[3] (Getner i Stevens, 1983), que algunes d'aquestes supersticions s'han originat quan la persona ha trobat algun tipus de problema durant l'ús del producte o servei i creu que una seqüència particular d'accions el resoldrà o evitarà que es repeteixi. Per tant, també hem d'estar atents a l'origen de les supersticions dels nostres usuaris: emmascaren algun problema que hàgim de conèixer?

(3) Mai no s'enalteix prou Don Norman, us animem a fer-ho.

d) Els models mentals són estalviadors

Els usuaris prefereixen fer més interaccions que dedicar energia mental a pensar el millor camí per a resoldre el seu problema. Això és un indicador doble:

- D'una banda, del nivell de coneixement que tenen els nostres usuaris del producte o servei, ja que el comportament més adequat o l'ús de les dreceres denoten un coneixement més profund del sistema. Si ningú no utilitza dreceres, què pot significar?

- D'altra banda, que els models mentals tendeixen a simplificar el que és complex. Quins aspectes detectem que se simplifiquen?

Exemple

Mitjançant l'anàlisi de registres informàtics (*logs*) podem detectar, per exemple, que els usuaris d'una intranet fan una sèrie de passos innecessaris per arribar a una funcionalitat concreta. Això no és suficient per a afirmar que existeix un model mental determinat, però és una bona pista per a investigar mitjançant mètodes com ara entrevistes, dinàmiques de grup (*focus groups*) o tests de tasques.

1.2.2. Què és un test amb usuaris

El test amb usuaris és una prova d'usabilitat emmarcada en l'enfocament de disseny centrat en l'usuari (DCU[4]). Consisteix a avaluar un lloc web a partir de l'observació, expressió verbal i interacció dels usuaris mentre executen les tasques pròpies d'un producte o servei interactiu.

(4) *DCU* és la sigla de *disseny centrat en l'usuari.*

En un test d'usabilitat s'analitzen els punts forts i febles d'un producte o servei amb usuaris reals i representatius sota condicions reals d'ús. Es fixen uns objectius de test i es determina el reclutament d'un públic objectiu per a aquest test. Es creen les tasques i els escenaris que estructurin la prova.

Una persona experta –el facilitador o moderador– seria l'encarregada de moderar el test mentre observa el comportament de l'usuari; un o diversos observadors prendrien nota de les accions que aquest executa. Una vegada acabats els tests, s'analitzarien i interpretarien els resultats per a detectar els aspectes que s'han de corregir.

Tot i així, la naturalesa dels tests amb usuaris és iterativa. El procés pot començar una vegada i una altra per anar refinant els resultats; de fet, mai no s'esgotarien els errors possibles, sinó que la seva especificitat aniria en augment.

Els tests d'usabilitat permeten contrastar totes les idees de l'equip que ha desenvolupat el producte o servei amb la realitat de l'usuari que l'ha de consumir. Sovint dins d'un departament de desenvolupament es generen discussions sobre si un disseny és adequat o no, i moltes vegades el millor recurs per a posar fi a aquestes discussions és dur a terme un test; d'aquesta manera sabrem què és el que realment fa l'usuari. És una manera de corroborar que les decisions que es prenen tenen de veritat un impacte positiu sobre el producte o servei mitjançant una demostració vàlida amb fets tangibles.

En aquest sentit, els tests s'assemblen una mica als experiments que es duen a terme com a part del mètode científic: es formulen unes hipòtesis (pot emplenar l'usuari aquest formulari?), se seleccionen uns subjectes de test (els usuaris), es prenen mesures durant la prova, s'analitzen els resultats i s'intenta demostrar que la hipòtesi constitueix una explicació vàlida.

Al llarg d'aquest apartat explicarem què necessitem per a dur a terme aquest mètode d'avaluació.

1.2.3. Quan s'ha de dur a terme un test amb usuaris

Un test amb usuaris és una de les tècniques d'avaluació més costoses, per això es recomana utilitzar altres tècniques prèvies per a descobrir problemes bàsics d'ús. Per exemple, primer podem analitzar el producte o servei fent una anàlisi heurística basada en els principis generals d'usabilitat. D'aquesta manera trobarem els errors de disseny més bàsics i fàcilment perceptibles sense que la participació d'usuaris sigui necessària.

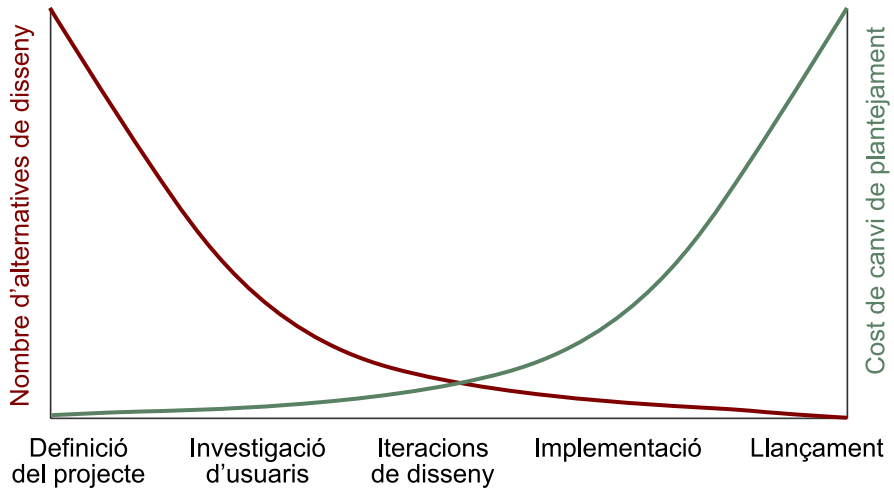
Abans de fer un test amb usuaris, ens hem d'assegurar que el servei o producte compleix els requisits bàsics d'usabilitat per a evitar malbaratar temps i diners (Hassan Montero i Martín Fernández, 2003).

Lectura complementària

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

És important, a més, que el test amb usuaris no es deixi per al final del procés de creació del producte o servei. Sovint, el test es fa massa tard per a arreglar les fallades que s'hi detecten. Com més ràpid detectem un problema, menys costosa en serà la reparació.

El gràfic següent reflecteix aquesta realitat: la línia descendent representa el nombre d'alternatives de disseny a mesura que avança el cicle de vida del projecte; la línia ascendent representa l'impacte econòmic de "canviar d'idea", és a dir, de corregir el plantejament del producte o servei. A mesura que el projecte avança en el temps i passa diferents fases –definició, investigació d'usuaris, iteracions de disseny, implementació i llançament–, el nombre d'alternatives de disseny es redueix i augmenta el cost de canviar de plantejament, ja que "es van tancant portes". Per aquesta raó el test amb usuaris, com qualsevol altre mètode d'avaluació, no s'hauria de deixar per al final (la fase d'implementació, per exemple), ja que, si ens hem equivocat en alguna de les decisions anteriors, serà molt costós tornar enrere.



Font: De la conferència de Jeffrey Veen a *Web Essentials 05*, "Beyond Usability: Designing the Complete User Experience"

Procurem no començar la casa per la teulada

Imaginem que volem construir un edifici de pisos. Si comencem per la teulada, no tindrem res que sustenti l'edifici. Sempre començarem pels fonaments, pel terra, basant-nos en els projectes que hagin desenvolupat els arquitectes responsables (és a dir, aplicant tècniques d'usabilitat, com ara una avaluació heurística).

Mitjançant avaluacions prèvies sabrem si l'edifici compleix les normatives legals de construcció. Per esbrinar si els materials i productes que s'empren per a la construcció són els adequats, haurem d'anar a comprovar-ne la qualitat pis per pis.

Encara que en el món de la construcció no sigui gaire comú, convidarem els futurs llogaters a passejar-se per cada pis perquè ens ajudin a esbrinar si l'edifici satisfà les seves necessitats i expectatives i inclou les característiques que aquests esperen. Si esperem que l'edifici estigui acabat perquè s'hi passegin, potser serà massa tard per a esmenar-ne els errors. Si, per contra, efectuem avaluacions a mesura que es construeixen les plantes, segurament evitarem repetir errors i ens assegurarem que els pisos s'adeqüen als projectes desenvolupats pels arquitectes (requisits bàsics d'usabilitat) i a les necessitats dels llogaters usuaris.

Així doncs, com més aviat es dugui a terme un test amb usuaris, millor serà el resultat i més baix serà el cost.

Un altre punt important del "quan" és que, per a fer un test amb usuaris, no cal tenir el producte completament acabat. El podem fer mitjançant prototips, d'alta o baixa fidelitat.

Vegeu també

Podeu veure els prototips d'alta o baixa fidelitat en el subapartat 1.4.3 d'aquest mòdul didàctic.

Sobre aquesta qüestió, Krug (2001) fa aquestes observacions importants:

- Provar un usuari és 100% millor que no provar-ne cap; provar sempre funciona. Fins i tot la pitjor prova amb l'usuari equivocat mostrarà coses que podem fer per millorar el producte o servei.

- Provar un usuari al principi del projecte és millor que provar-ne cinquanta gairebé al final; una prova al principi d'un projecte, quan encara tenim temps d'utilitzar i esmenar els errors que detectem, és gairebé sempre més valuosa que una sofisticada prova més tard. És important evitar construir malament des del principi.

- La importància de reclutar usuaris representatius està sobreestimada; és important fer els tests amb usuaris que utilitzaran el producte o servei, encara que és més important provar-ho aviat i sovint.

La vehemència de Steve Krug

Steve Krug és, abans que res, un comunicador de l'àrea de la usabilitat. En aquest sentit, dóna més importància a transmetre la passió per la seva feina que a ser metodològicament correcte en cada frase. Per què diem això? Perquè dir que "la importància de reclutar usuaris representatius està sobreestimada" és una frase provocadora. Com veurem més endavant, escollir bé els usuaris és un ingredient fonamental perquè els resultats d'una avaluació siguin fiables. Això no obstant, estem d'acord amb Krug en el fet que fer un test amb qualsevol usuari, per inadequat o poc comú que sigui el seu perfil per al producte o servei que avaluem (exemple: una persona gran i un telèfon mòbil Blackberry), és millor que no fer cap test amb cap usuari.

- El fet de provar no és aprovar o desaprovar alguna cosa. És informar de la seva decisió. Sovint a la gent li agrada pensar que pot utilitzar els tests per a comprovar si un sistema de navegació *a* és millor que un sistema *b*. Ningú no té els recursos per a establir el tipus d'experiment controlat que necessita. La prova ens pot proporcionar inversions incalculables que, al costat de l'experiència, decisió professional i sentit comú, faran que puguem escollir amb confiança entre *a* o *b*.

- Provar és un procés repetitiu; provar no és una cosa que només es faci una vegada. Hem de provar, solucionar i provar-ho de nou.

- Res no pot amb una reacció viva del públic; l'important és saber què volen els nostres usuaris. Hem de provar i provar amb ells fins a arribar a la decisió definitiva.

Quan comencem a dissenyar un producte o servei mai no és massa aviat per a efectuar un test d'usabilitat. Encara que únicament tinguem un prototip en paper, el podem ensenyar als usuaris per a saber què en pensen. A més, en no tenir disseny, no es distrauran amb detalls, sinó que se centraran més en el producte o servei en si mateix.

Lectura complementària

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

Una cosa que es dedueix d'aquests punts és que hem d'actuar com a científics i aprendre dels nostres experiments, siguin quins en siguin els resultats. Si aquests contradiuen els nostres models mentals, igualment aportaran informació valuosa si els parem l'atenció adequada.

A continuació presentarem els recursos humans i materials necessaris per a poder dur a terme una prova amb usuaris.

1.2.4. Persones necessàries en un test amb usuaris

En un test amb usuaris participen diferents persones que exerceixen diferents rols durant la prova: l'**usuari**, per a dur a terme la prova; un **facilitador**, que modera el test; i **observadors**, que recullen diferents aspectes del que passa durant el test.

L'usuari

Segons el diccionari de l'IEC, *usuari* és la persona que "usa alguna cosa".

> L'usuari és el consumidor potencial i habitual del producte o servei que avaluem.

La selecció dels usuaris adequats constitueix gran part de l'èxit final del test, ja que les observacions que en fem, com també les seves opinions, comentaris i reaccions, seran la matèria primera amb què s'elaboraran els resultats. Per aquesta raó, la mostra d'usuaris ha de ser la més representativa possible de qui esperem que finalment utilitzi el producte i servei que desenvolupem.

En què consisteix el rol d'usuari? Consisteix en el següent:

- Prendre part en el test, tant si és remot com presencial.
- Firmar els acords de confidencialitat i cessió vinculats al projecte.
- Fer les tasques del test, seguint les pautes marcades pel facilitador.
- Respondre als qüestionaris pretest i posttest.
- Verbalitzar els seus pensaments i reflexions sobre la tasca feta.

El facilitador o moderador

> El facilitador o moderador és la persona encarregada de conduir el test i acompanyar l'usuari durant tot el procés.

El rol de facilitador no és fàcil, ja que ha de conjugar tres funcions principals:

1) Salvaguardar la integritat física i psicològica dels participants. Pot semblar exagerat, però en el fons un test de tasques cerca que un usuari cometi errors davant d'un expert. Aquesta situació no és agradable per a ningú, per la qual cosa el facilitador ha de vetllar pels usuaris:

- Perquè se sentin còmodes. Un usuari incòmode no reacciona com ho faria en una situació real, la qual cosa invalida les dades.

- Estar atents a signes d'estrès o cansament, fins i tot decidir quan es fan les pauses perquè l'usuari descansi.

- Alleugerir els aclaparaments i les tensions de l'usuari davant de la tasca. El cert és que obtenim informació molt valuosa dels moments durs pels quals passa l'usuari. El facilitador té l'obligació d'alleugerir els neguits de l'usuari, transmetre-li serenitat en tot moment; transmetre-li la idea que no és un pocatraça pel fet que no sàpiga utilitzar alguna cosa, sinó que el producte necessita millorar i que la seva ajuda és imprescindible.

2) Maximitzar el flux d'informació cap als observadors. Malgrat que els observadors són a prop, sol ser el facilitador el qui rep més informació de l'usuari. El facilitador ha d'animar l'usuari a manifestar el que pensa i ocupar-se que aquesta informació arribi fins als observadors. El facilitador:

- És el responsable que es compleixi el protocol de pensament manifest[5], anomenat també *pensament en veu alta*.

- Ha de subratllar les accions rellevants de l'usuari, perquè els observadors les puguin registrar.

Exemple

En un test amb una pàgina web, l'usuari tria una cerca avançada en lloc de la cerca simple. El facilitador pot subratllar aquest fet simplement dient "has triat la cerca avançada". Això servirà perquè els observadors estiguin atents i perquè l'usuari probablement expliqui per què.

- Ha d'animar els usuaris a preguntar però no ha de donar cap resposta: les preguntes de l'usuari ajuden a determinar quines funcionalitats s'han de millorar.

Exemple

En un web de compra:

USUARI: Si toco aquest botó d'aquí vaig directe al carretó?

FACILITADOR: Ho provem?

- Ha de fomentar que l'usuari expressi el que pensa, formulant preguntes obertes: "és el que n'esperaves?", "què és el que intentes fer?".

(5)En anglès, *thinking aloud*.

Vegeu també

Podeu veure el concepte de *pensament manifest* en el subapartat 1.3.4 d'aquest mòdul didàctic.

- Ha de para atenció al llenguatge no verbal. Les expressions del rostre i la postura dels usuaris diuen molt del que passa per la seva ment. De vegades, hi ha un observador dedicat a registrar les reaccions emocionals de l'usuari que recull aquest llenguatge no verbal, però és convenient ajudar-lo perquè verbalitzi el que passa.

Exemple

Un esbufec de l'usuari sovint expressa disgust. No es tracta de preguntar "què t'ha fet esbufegar?", ja que probablement l'usuari ni tan sols n'és conscient. La pregunta oberta és una bona opció: "què ha passat?".

- Ha d'evitar que l'usuari es desviï del tema. Hi ha usuaris que s'esplaien en les pròpies opinions sobre el que s'avalua; cal recordar que no ens interessen les opinions de l'usuari, ens interessa el que aquest aconsegueix fer durant el test.

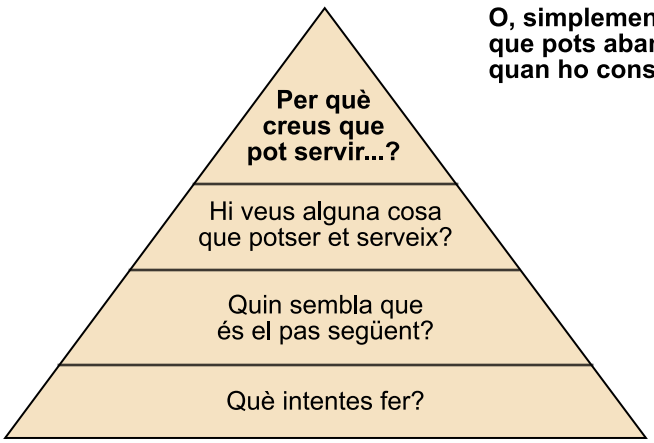
Exemple

Els usuaris tenen tendència a generalitzar des del que ells opinen respecte del món: "Crec que la gent tindrà problemes amb això". El facilitador ha de reaccionar: "Tu tens problemes?".

- Ha de decidir quan s'acaba una tasca. Si tenim una tasca en què l'usuari s'encalla, el facilitador li pot suggerir passar a la següent. Sovint tots els usuaris d'un test s'encallen en un mateix punt. Si l'equip ja té prou informació per a demostrar que hi ha un problema, el criteri del facilitador pot ser que no cal continuar.

Exemple

Els usuaris més voluntariosos insisteixen a reprendre tasques que han deixat inacabades si durant el test han entès com fer-les; també respon al criteri del facilitador el fet de permetre-ho. En tot cas, amb vista a l'anàlisi és molt important recollir que la tasca es va poder acabar en un segon intent.



O, simplement, et recordem que pots abandonar la tasca quan ho consideris convenient.

Lectura recomanada

En relació amb el llenguatge no verbal, és interessant revisar el projecte següent:

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

3) Mantenir el grau d'integritat màxim en la informació. El facilitador ha de tenir sempre en ment que un test amb usuaris és una prova orientada a obtenir dades fiables. Per tant, ha de vetllar perquè aquestes dades es contaminin el menys possible; el facilitador pot decidir que un usuari no compleix el perfil i que no ha de fer un test, o que una tasca ha quedat invalidada per alguna raó.

Reflexió

Es talla la connexió a Internet a mig fer una tasca amb un web i es perden les dades d'un formulari: què hem de fer? Permetre que l'usuari torni a començar?, demanar que passi a la tasca següent i desestimar aquesta?

La piràmide de la imatge mostra, de la base al cim, un seguit de preguntes que el facilitador pot fer a un usuari "encallat". Es tracta de no arribar a preguntar "per què creus que pot servir...?", ja que és una intervenció directa del facilitador en el test. Tot i així, queda a criteri de cada facilitador decidir si hi intervé o no.

Els observadors

Tan important com la tasca del facilitador o de l'usuari és la tasca de l'equip d'observació. Els observadors són persones encarregades de registrar dades de test per a la seva anàlisi posterior; el nombre i el paper dels observadors dependran dels objectius que ens hem fixat en el test, ja que hauran de recollir unes dades o unes altres; per exemple, pot interessar mesurar el nombre de clics que fan els usuaris, el nombre de vegades que usen el cercador o el nombre de cops que consulten el manual d'ajuda.

Una distribució bàsica d'observadors podria ser:

a) Observador encarregat dels temps i del resultat de les tasques:

- Aconsegueix l'usuari dur a terme la tasca?
- Quant de temps triga?

b) Observador encarregat de les manifestacions verbals dels usuaris.

c) Observador especialitzat A: per exemple, reaccions emocionals dels usuaris.

d) Observador especialitzat B: per exemple, en una prova de mòbils, quines tecles prem per tasca.

És important planificar la tasca dels observadors i estandarditzar-la abans del test i decidir amb antelació les dades que es recullen i en quin format. Resulta útil fer una reunió prèvia amb tot l'equip implicat en el test i plantejar preguntes com les següents:

- Quan considerarem que l'usuari ha fet amb èxit la tasca?
- Quan ho considerarem fracàs?
- Què passa si abandona la tasca? És un fracàs en la tasca?
- En quin format recollirem els temps dels usuaris?

Reflexió

En una ocasió, en acabar un test, ens vam trobar que els observadors encarregats de recollir els temps havien fet la seva tasca de manera no estandarditzada: un posava a zero el seu cronòmetre en cada tasca, un altre anotava els minuts d'enregistrament del vídeo i un altre les hores en què ocorria cada registre. Totes les dades eren vàlides, però va caldre fer feina extra d'homogeneïtzació.

En un equip de test també arribem a definir el "fals èxit" i el "fals fracàs" com a dades que es poden recollir, ja que de vegades els usuaris resolen adequadament les tasques sense saber-ho, o erròniament sense adonar-se'n.

Exemple

- Com hem de recollir les dades?
- Apuntem l'hora del test al costat de cada expressió, declaració o manifestació verbal de l'usuari?

En qualsevol cas, els observadors no són sempre imprescindibles. La participació dels observadors dependrà dels objectius i del nivell de complexitat del test i dels recursos disponibles per a dur-lo a terme.

1.2.5. Laboratoris i eines per a l'avaluació de la usabilitat

Podem distingir dues classes de **laboratoris d'usabilitat** segons el tipus d'instal·lacions:

1) Laboratoris fixos o estables, amb instal·lacions pensades i adaptades a l'avaluació.

2) Laboratoris portàtils, pensats per a desplaçar-se.

D'altra banda, els **components del mobiliari, maquinari i programari** més habituals en un laboratori d'usabilitat solen ser:

1) Elements de sala, mobiliari i ambient: taules i cadires, mirall d'observació unidireccional, llums regulables i insonorització.

2) Elements de suport per a la recollida de dades: sistema d'enregistrament en vídeo, micròfons, monitors, equipament per a l'enregistrament en vídeo, ordinador i programari per al suport a la recollida de dades.

Laboratoris d'usabilitat

1) Laboratoris d'usabilitat estables

Aquest tipus de laboratoris d'usabilitat són, com el seu nom indica, espais dissenyats i adaptats especialment per al desenvolupament de proves d'avaluació d'usabilitat. Aquesta adaptació consisteix fonamentalment a aconseguir reproduir de manera realista un o diversos entorns d'ús i dotar l'equip d'avaluadors de les eines necessàries per a dur a terme la seva feina.

Reproduir de manera realista significa adaptar l'espai al context d'ús del producte o servei interactiu, tal com hi interactuaria l'usuari.

És per això que sovint l'espai del laboratori es divideix en dues sales:

a) Sala d'observació. S'hi situa l'equip d'observadors, l'equip d'enregistrament i, de vegades, un client que pugui assister al desenvolupament del test.

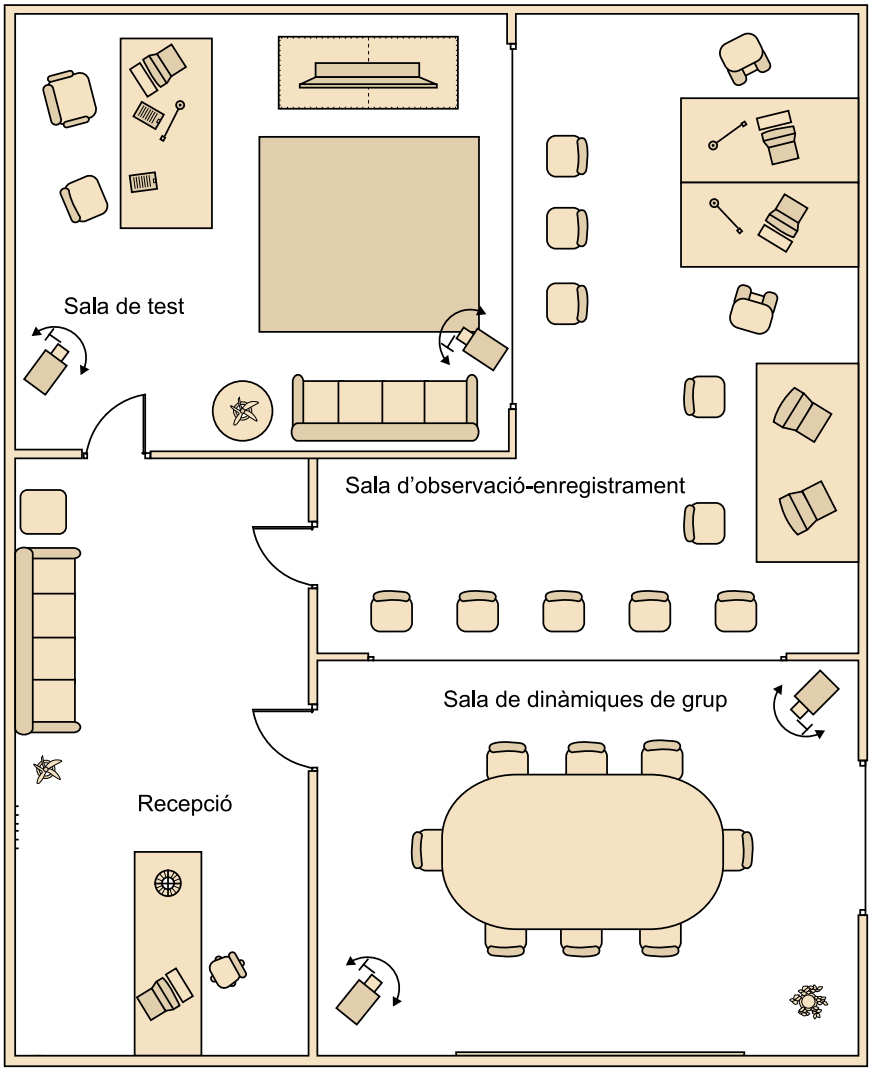
b) Sala de test, on té lloc la prova. Sovint està ambientada com un entorn domèstic –per exemple, una sala d'estar– o un entorn d'oficina.

Entre totes dues habitacions hi ha instal·lat un vidre d'observació unidireccional o de visió unilateral que permet veure els usuaris des de la sala d'observació mentre efectuen la prova, encara que no des de l'altre costat. Es completa amb càmeres d'enregistrament controlades remotament, que serveixen per a gravar la prova o tenir plans concrets de l'usuari.

Càmeres de Gesell

En l'àmbit científic, aquestes sales en què es desenvolupen els tests d'usabilitat es denominen *càmeres d'observació* o *càmeres de Gesell*. El psicòleg i pediatre Arnold Gesell va ser pioner en l'ús d'aquestes sales en les seves investigacions en el camp de l'educació.

En alguns laboratoris d'usabilitat hi ha sales addicionals on es pot fer un altre tipus de proves d'usabilitat que requereixen més infraestructura.



Els beneficis d'un laboratori d'usabilitat són els següents:

- Disposar d'una habitació separada ens permet disposar de més gent, a part del facilitador, que visualitzi el test i, per tant, més ulls centrats en els moviments, els comportaments i les interaccions de l'usuari sense que aquest en sigui conscient.

- En aquest tipus de laboratoris fixos o estables el facilitador no s'ha de preocupar per temes tècnics, com ara l'enregistrament d'àudio i vídeo, o de col·locar una càmera web enfocada a l'usuari. Aquests laboratoris ja disposen d'un equip tècnic muntat i permeten tenir tècnics especialistes que s'encarreguin dels detalls d'enregistrament.

Encara que els resultats de fer un test en un laboratori fix o portàtil poden ser molt semblants, de vegades és clau donar a entendre a l'organització que la usabilitat té un paper primordial en el desenvolupament i disseny, i que per això es construeixen laboratoris específics per al seu estudi. De vegades hi ha clients que poden donar importància a aquest tipus d'aspectes.

- En disposar d'observadors que prenen notes sobre com l'usuari es relaciona amb el producte o servei avaluat i n'observen el comportament, el facilitador té més llibertat per a centrar-se en altres aspectes, la qual cosa potser no seria possible en un laboratori portàtil, en trobar-se sol amb l'usuari, sense ningú que pugui prendre notes.

2) Laboratoris d'usabilitat portàtils

Convertir una oficina o una habitació en un laboratori és molt senzill. Únicament necessitem un ordinador, una càmera web i el programari disponible per a enregistrar la interacció, l'àudio i el vídeo del test d'usabilitat. D'aquesta manera podem fer proves on vulguem.

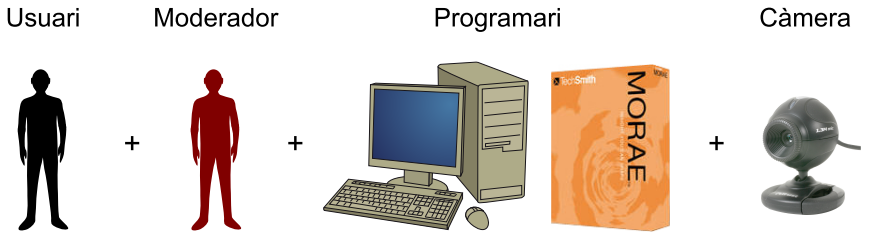
Els beneficis d'un laboratori portàtil són els següents:

- Més autonomia. És més fàcil organitzar un test d'usabilitat portàtil, ja que no depenem d'una sala específica per a fer el test. El podem dur a terme a qualsevol habitació de la nostra oficina, a casa de l'usuari, etc.

- Estalvi de costos. És més car fer un test d'usabilitat en un laboratori fix que muntar un laboratori portàtil.

- Comoditat per a l'usuari. De vegades pot semblar més fred i artificial fer el test en un laboratori d'usabilitat fix.

- Adaptació a l'entorn i context d'ús. De vegades, és important fer el test en l'entorn real d'ús del producte o servei, el qual no sempre es pot reproduir en una sala de laboratori fix.



Mobiliari, eines de maquinari i programari

1) Elements de la sala, mobiliari i ambient

Els elements de sala, mobiliari i ambient són els que ens permeten crear l'espai que dóna acollida als usuaris en cada prova.

a) Mobiliari. En l'elecció de mobiliari, les premisses són:

- Polivalència. És poc probable que disposem d'una sala per a desenvolupar cada mètode. El més habitual és que les sales de tests siguin polivalents i la seva estructura s'adapti al tipus de mètode que apliquem. Per això, l'elecció del mobiliari ha de permetre que treballem amb un únic usuari o amb un grup d'usuaris. En aquest sentit, són útils les taules modulars amb rodes o les cadires apilables.

- Comoditat. Qualsevol de les tipologies de test pot arribar a durar una hora, per la qual cosa el mobiliari ha de ser còmode i ergonòmic.

- Immersió. Si les instal·lacions ho permeten, podem disposar d'una sala amb un entorn immersiu que reprodueixi una casa o una oficina. Això implica la inclusió d'elements decoratius: quadres, plantes, catifes, etc. L'avantatge d'un espai immersiu és que ajuda els usuaris a sentir-se més "a casa" i afavoreix que durant les proves es comportin amb naturalitat.

b) Miralls d'observació unidireccional. Hem parlat de la separació entre dos ambients per un vidre de visió unilateral. Permeten a un observador mirar sense ser vist, des d'una sala a les fosques, el que ocorre a la sala adjacent il·luminada. Són un luxe que no es poden permetre gaires laboratoris, ja que requereixen una adaptació *ex professo* de l'estructura de les sales. Aquests vidres mirall no solament permeten que els observadors d'un test siguin fora de la sala, sinó que el possible client o part de l'equip tècnic que desenvolupen el producte també puguin assistir a la sessió d'avaluació en directe, sense haver de mirar les cintes. No oblidem que aquest tipus d'enregistrament s'ha d'acollir a unes normes legals i a unes polítiques d'ús i privacitat que s'han de respectar.

c) Llums regulables. Són molt recomanables si hem instal·lat els vidres d'observació unidireccional, ja que aquests funcionen il·luminant més una sala que l'altra; els llums regulables ens permeten no haver de treballar en la foscor absoluta. També són un element valuós per a reproduir diferents ambients: il·luminació total per a oficines, llums més tènues per a reproduir una casa, etc.

d) Insonorització. Si tenim la sort de disposar de més d'una sala, és recomanable insonoritzar-les. Això facilita la presència d'un equip d'observadors i de convidats externs en cada prova que als seus llocs corresponents podran fer comentaris o tenir converses que no afectin el desenvolupament de la prova.

2) Elements de suport per a la recollida de dades

Els elements de suport per a la recollida de dades serveixen per a facilitar la recollida de dades durant el test per a la seva anàlisi posterior, i augmentar així la riquesa de les dades i la documentació.

a) Sistema d'enregistrament en vídeo. Les càmeres de vídeo permeten enregistrar les sessions i filmar en detall aspectes com la cara dels usuaris. Podem arribar a utilitzar dues càmeres, una dins de la sala on hi ha l'usuari i una altra a la sala d'observació. La primera podrà enregistrar la pantalla, la postura de l'usuari o els moviments del ratolí. La segona càmera pot prendre plans detallats de la cara, les mans o de qualsevol anotació que l'usuari hagi de fer sobre un quadern situat a la taula. Les característiques tècniques del sistema d'enregistrament en vídeo dependran del pressupost disponible, de l'ús que se'ls vulgui donar i del nivell de competència del nostre equip humà. Un sistema analògic o digital d'enregistrament necessita un tècnic que entengui com es barregen les imatges de diferents càmeres (utilitzant un mesclador, per exemple) i sigui capaç de fer un muntatge o una postproducció. Si fos així, no hem d'oblidar donar les instruccions necessàries a aquesta persona sobre allò que volem recollir.

b) Micròfons. Es poden instal·lar micròfons omnidireccionals a les sales, tal com s'utilitzen en les dinàmiques de grup[6]. Tanmateix, per al test amb usuaris, és recomanable utilitzar micròfons "de corbata", que capten perfectament les paraules de l'usuari. També és útil que el facilitador disposi d'un sistema de comunicació amb la resta de l'equip d'observació (normalment, un auricular) per a rebre instruccions o saber si hi ha algun problema tècnic.

(6)En anglès, *focus group*.

c) Monitors. Disposar de diversos monitors connectats al sistema de vídeo permetrà als observadors veure amb detall el que passa en el test, tant en la pantalla de l'usuari com en les seves reaccions.

d) Ordinador. Un o més ordinadors són necessaris perquè l'usuari dugui a terme el test, per a recollir les dades, per a enregistrar la sessió digitalment, etc.

e) Programari de suport per a la recollida de dades. Aquest tipus de programari possibilita que el que passa durant el test es reculli de manera ordenada i vinculada als enregistraments de vídeo. Això té diversos avantatges:

- Homogeneïtza la recollida de dades entre diferents observadors.
- Permet que les dades del test estiguin disponibles de manera immediata, sense que s'hagin de "passar a net" les notes.
- Possibilita fer un seguiment dels temps d'enregistrament d'àudio i vídeo, i una extracció posterior de clips de vídeo o seqüències.
- Ofereix una automatització de les mètriques i estadístiques.
- Permet crear "marques" vinculades a esdeveniments del test. Per exemple, una marca que representi cada vegada que a l'usuari se li escapa una grolleria o un improperi; aquesta marca es configura de manera que es registra vinculada a un temps i un moment de l'enregistrament.

Són moltes les solucions gratuïtes o comercials que s'ajusten a aquestes i altres característiques. Algunes fins i tot faciliten l'enregistrament a partir d'una càmera web de manera presencial o remota, amb la qual cosa podríem estalviar costos.

Taula

Algunes solucions de suport per a la recollida de dades

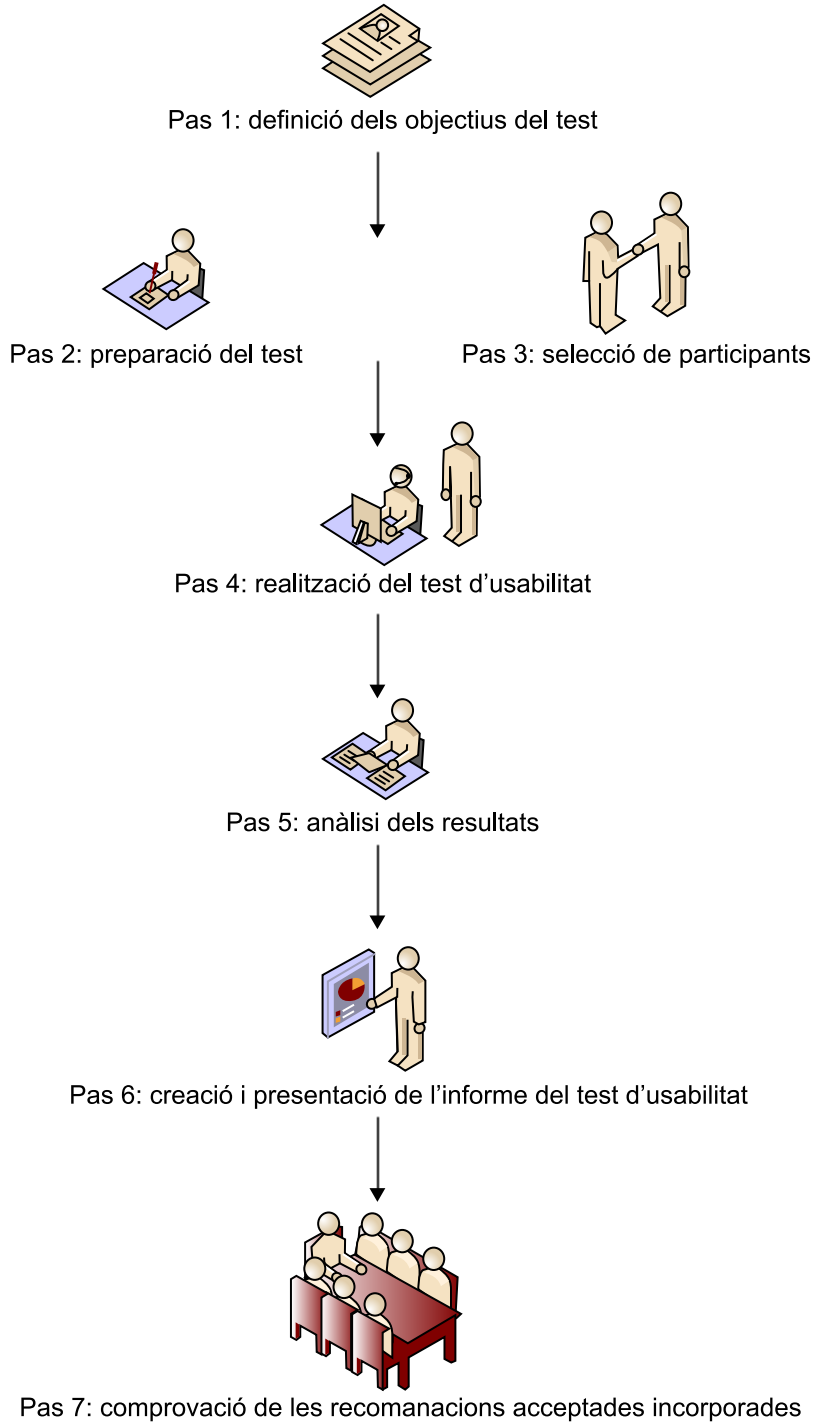
Nom	Web	Tipus
Silverback 2.0. (Clearleft)	silverbackapp.com	Programari de prova o *shareware* (30 dies)
Morae (Techsmith)	techsmith.com/morae.asp	Programari de prova (30 dies)
ObServant (User Works)	observant.com.au/	Comercial
Usability Activity Log (Bit Debris Solutions)	bitdebris.com	Comercial
The Observer XT (Noldus)	noldus.com	Comercial
Userfly	userfly.com/	Versió *free* limitada i versió *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/	Programari gratuït o *freeware*

Això no obstant, l'absència d'aquest tipus de programari no és un obstacle per a dur a terme un test amb usuaris. Es pot substituir per un simple full de càlcul, un editor de textos i una mica d'organització en la recollida de dades.

1.3. Procés del test

El primer que hem de fer és definir els objectius del test, és a dir, tenir clar el que volem aconseguir amb la prova d'usabilitat. Una vegada definits els objectius, podem començar a preparar el test, planificar tot el procediment del test (les tasques dels usuaris, els escenaris ficticis i els qüestionaris) segons els objectius especificats.

Exemple d'etapes de procés d'un test



Adaptat de: Abhay Rautela, ConeTrees.com (CC BY NC ND)

Quan tinguem clar què volem avaluar i com serà el procés del test, podrem començar a seleccionar els participants, sempre tenint en compte els objectius definits i les característiques del producte o servei que avaluem.

Quan tinguem els participants seleccionats, podrem començar el test. Una vegada efectuat, haurem de recollir totes les dades recopilades durant el test i analitzar els resultats.

1.3.1. Definició i objectius del test

Per a dur a terme un test amb usuaris és necessari definir prèviament els objectius que es volen aconseguir amb la finalitat d'aclarir i definir allò que volem avaluar del producte o servei.

> L'objectiu general d'una prova o test d'usabilitat és conèixer amb un nivell de detall elevat, la facilitat d'ús que presenta un producte o servei i el nivell d'eficàcia, eficiència i satisfacció que produeix en l'usuari.

A partir d'aquí, es poden proposar objectius específics. Els objectius específics es traduiran en tasques que els usuaris participants en el test duran a terme. Els objectius del test han de ser específics i raonables perquè les tasques que se'n desprenguin siguin properes als usuaris a qui s'adreça, factibles i amb una durada adequada. En resum, en primer lloc s'han de definir clarament els objectius del test i, una vegada definits, s'han de definir les tasques que hauran de fer els usuaris.

Reflexió

Fixeu-vos que bona part dels exemples que es proposen a continuació corresponen a l'avaluació de la usabilitat de llocs web, ja que és l'entorn més evolucionat i propici per a efectuar-la. Això no obstant, altres entorns, tecnologies, dispositius, productes i serveis també estan subjectes a avaluacions d'usabilitat.

A continuació passem a explicar àmbits d'actuació que es poden considerar a l'hora de marcar els objectius per al test i definir posteriorment les tasques necessàries.

a) Eficàcia i eficiència en la navegació

En aquest punt ens interessa saber si els usuaris troben allò que necessiten fàcilment. Es tracta de descobrir si els usuaris troben un patró de navegació que encaixi amb el seu model mental, és a dir, si les opcions de navegació són prou clares per a trobar el que busquen, si les categories són oportunes, si el llenguatge que s'utilitza és proper a l'usuari, etc.

Ens hem de preguntar el següent

Els usuaris troben el que necessiten de manera intuïtiva, fàcil i ràpida?

b) Contingut útil i pràctic

És necessari conèixer la informació que l'usuari necessita realment i organitzar-la de manera clara perquè la trobi fàcilment. Un usuari no necessita llegir tot el contingut d'un lloc per a trobar la informació, l'hem d'ajudar a trobar-la de la manera més útil i còmoda possible.

Ens hem de preguntar el següent

Quina informació és la que vol l'usuari?

Hem organitzat la informació de manera clara perquè els usuaris la trobin ràpidament?

c) Claredat en la presentació

El disseny gràfic, la font, la mida de la navegació i també el contingut són elements que poden ajudar l'usuari o, al contrari, el poden distreure i crear una barrera entre aquest i la informació.

Ens hem de preguntar el següent

El disseny, els colors, la font, etc., ajuden l'usuari o, al contrari, el distreuen i fan que se senti incòmode?

d) Proporcions de les tasques

En un test amb usuaris es creen una sèrie de tasques per a comprovar el grau d'eficiència del producte. Aconseguir que els usuaris executin les tasques de manera satisfactòria ens aportarà informació vàlida per al desenvolupament del producte o servei i ens ajudarà a saber com s'han sentit els usuaris en fer-les.

Ens hem de preguntar el següent

Els usuaris han estat capaços de fer les tasques plantejades? Si és així, com s'han sentit?

És important tenir aquests objectius al cap durant tot el procés per a no desviar el focus del projecte. Tenint clars els objectius del test, els resultats seran molt més precisos.

1.3.2. Preparació del test

En aquest subapartat definirem i organitzarem tots els elements que s'utilitzaran durant el test. Aquests elements formaran el guió de la nostra planificació.

Qüestionari pretest

Un primer qüestionari ajuda a recopilar especialment dades sociodemogràfiques que ens ajudaran en el procés d'anàlisi dels resultats. Les consideracions sociodemogràfiques sempre són presents en el qüestionari pretest, com el nom, l'edat, el sexe, l'empresa, el càrrec, etc. La resta de preguntes del qüestionari dependran de l'aplicació que avaluem.

Segons l'aplicació

Si preparem un test d'usabilitat d'una pàgina de compres en línia, farem preguntes sobre la freqüència de connexió a la pàgina, el nombre de compres que s'efectuen per mes, etc.

Usabilitat d'una botiga de roba en línia

Imaginem que efectuem un test d'usabilitat d'una botiga de roba en línia. Si no tenim informació suficient sobre el grup de participants, podem formular algunes preguntes senzilles relacionades amb les seves dades personals i situació laboral. A continuació podrem treballar sobre qüestions específiques per al test d'usabilitat.

Alguns exemples:

- Freqüència de connexió a Internet:
  - Quantes hores a la setmana et connectes a Internet? 1-2 hores/setmana;
  - Des d'on et connectes habitualment?
- Compres en línia:
  - Has comprat alguna vegada en línia?
  - Em pots especificar els webs on has comprat?
  - Cada quant acostumes a comprar?
  - Amb quina freqüència accedeixes a Internet?
  - Quan va ser l'última vegada que hi vas comprar alguna cosa?
  - Has efectuat alguna compra amb el telèfon mòbil?

Abans de l'inici de la prova se sol deixar que cada usuari faci una petita exploració (durant uns quants minuts) i es comenci a familiaritzar amb el producte o servei. Una vegada passats aquests minuts li podem fer preguntes sobre aquesta primera exploració.

Escenaris i tasques

L'**escenari** és la descripció d'un personatge en una situació i context d'ús amb unes tasques i activitats assignades. Col·loquem els usuaris en situació i els donem les dades necessàries perquè assumeixin el paper i actuïn en conseqüència.

Exemple d'escenari

Imagineu que sou clients de la companyia X i un amic us comenta que des del web de la companyia podeu tramitar moltes de les accions que acostumeu a fer per telèfon.

No oblidem que el context fa referència a les condicions en què s'utilitzarà un producte interactiu. D'aquesta manera hem d'oferir a l'usuari la informació suficient per a aconseguir que se situï i s'hi adeqüi. Per exemple, si efectuem un test amb usuaris per a una companyia telefònica amb usuaris que no pertanyen a la companyia, hem de demanar-los que actuïn com si ho fossin, i aportar-los-en la informació necessària.

Per la seva banda, les **tasques** són el punt principal en el guió d'un test amb usuaris. Són les accions que es demana a l'usuari que dugui a terme amb el producte o servei al llarg del test.

Exemple de tasca

Accediu a la pàgina web de la companyia telefònica X i intenteu obtenir dades sobre el vostre consum durant el mes de gener.

Com s'elaboren els escenaris?

Per a elaborar els escenaris és imprescindible conèixer i tenir ben detectats els objectius que es volen assolir amb el test d'usabilitat, com també les funcionalitats o característiques que es volen analitzar del producte o servei.

Com ja hem comentat, els tests d'usabilitat són proves que ens ajuden a avaluar la qualitat d'un producte o servei i la mesura en què és usable per al públic a qui es dirigeix. L'usuari que fa el test pot ser una persona que conegui el producte o servei o no però, sigui com sigui, l'hem de posar en situació a fi que els resultats del test siguin el més fiables possible. Per això, en demanar-li de fer una sèrie d'accions que potser normalment no fa, les hem de dotar d'un sentit complet.

En general, els escenaris s'utilitzen per a introduir les tasques i crear un ambient imaginari que doni sentit a l'acció.

Per exemple, en introduir una tasca a l'usuari, hi ha una gran diferència entre les situacions següents:

- Sense escenari: "Accedeix a la pàgina principal d'un web de compres en línia i troba-hi un llibre que vulguis comprar" (tasca).

- Amb escenari: "Imagina que el dimarts de la setmana que ve és l'aniversari de la teva mare i saps que li faria il·lusió tenir l'última novel·la d'un gran escriptor. A més, saps que a la teva mare li agraden els llibres de butxaca perquè acostuma a llegir-los a la platja o a l'autobús. Podries accedir al lloc web de compres en línia X i intentar trobar el llibre que vol?" (tasca).

Com veiem, en l'exemple sense escenari, l'usuari no es troba en cap context, per la qual cosa durà a terme la tasca sense cap tipus de motivació i sense tenir una informació suficient que li permeti donar sentit a aquesta tasca. Al contrari, en el segon exemple, que incorpora un escenari, l'usuari es veu implicat en un context i ja disposa d'una motivació suficient per a fer l'acció.

Quan tinguem les característiques que vulguem avaluar i les tasques que en conseqüència vulguem que els usuaris duguin a terme, podrem dotar la situació d'un context, d'un escenari.

Com s'elaboren les tasques?

A l'hora d'escollir què s'avaluarà, el mateix client per a qui treballem ens pot proporcionar pistes que ens ajudin a generar les tasques i marcar els aspectes que poden ser crítics en el seu servei o producte.

Altres vegades, és important tenir l'opinió d'un usuari expert o fer reunions d'equip prèvies per a saber quin tipus de tasques són més habituals i quin tipus de tasques són crítiques.

Les tasques d'un test d'usabilitat s'han de basar en les característiques següents:

- **Raonables.** La tasca ha de ser raonable, és a dir, ha de reflectir una cosa que per a l'usuari sigui normal de fer, mai una cosa que no es faria mai. Altrament, l'usuari no es podria sentir prou integrat en el paper i no duria a terme la tasca correctament.

  Exemples

  Raonable: "Comprar dues entrades per a un concert".

  No raonable: "Comprar dues entrades per a sis concerts".

- **Orientades a la tasca.** Quan un usuari és sol a casa seva o a la feina utilitzant un producte o servei, ho fa per algun motiu. Per a l'usuari, aquest producte o servei és una eina amb la qual pot interactuar, ja que a canvi li proporciona alguna cosa. En posar l'usuari en un context irreal, en un laboratori d'usabilitat, aliè a la seva vida quotidiana, hem d'aconseguir que senti que l'acció que fa té un objectiu. La motivació de l'usuari quan fa una tasca és essencial. Per tant, hem de proporcionar un escenari tan realista com sigui possible.

  Exemples

  Orientada a tasca: "Comprar un CD nou per a un familiar proper perquè t'has adonat que el que es va comprar l'any passat està espatllat i no se sent bé".

  No orientada a tasca: "Comprar un nou CD".

- **Específiques.** Per a ser consistent i centrar-nos en la tasca, aquesta ha de tenir un objectiu final específic. D'aquesta manera l'usuari entendrà millor l'acció que ha de fer i tindrà una motivació específica. Hem d'anar amb compte de no utilitzar termes o noms inclosos en el producte o servei analitzat, ja que podem generar pistes a l'usuari, sense voler-ho, i alterar els resultats del test. Per exemple, si l'acció que l'usuari ha de fer és accedir a la pestanya *Contractes*, hem d'intentar no dir la paraula *contractes*, sinó formular la tasca de manera que no inclogui aquesta paraula.

Exemples

Específica: "Comprar l'últim llibre de Pérez-Reverte".

No específica: "Comprar un novel·la d'aventures".

- **Factibles.** No s'ha de proposar una tasca impossible a l'usuari, ja que no ajudaria a millorar el disseny i li podria provocar frustració. Un cas lleugerament exagerat seria demanar-li que busqui el preu d'un pis en un poble de Madrid fins i tot sabent que no en trobarà cap perquè la base de dades no reflecteix aquesta població. Avaluem el disseny, no l'usuari.

- **Seqüències de tasques realistes.** Les tasques han de reflectir l'ordre de treball natural amb la interfície. No ens hem d'inventar seqüències que no es corresponguin amb la realitat del producte o servei.

Exemples

Si pensem en un web de vehicles d'ocasió, la seqüència realista seria navegar per les categories que es presenten o fer una cerca per a després filtrar i seleccionar. Demanar a l'usuari que trobi un vehicle de baixa cilindrada amb tubs d'escapament cromats seria sortir de la seqüència –dels objectius– i no aconseguiríem analitzar la fluïdesa de l'usuari amb el producte o servei.

- **No especialitzades.** En proposar tasques hem de provar una cosa que tothom pugui fer o conèixer, no una cosa sobre la qual algú sàpiga molt. Una tasca ha de ser una cosa que l'usuari conegui, però s'ha de procurar que no sigui un expert en aquest camp. Per això hem d'evitar escollir usuaris que s'adaptin amb rapidesa excessiva als esquemes i als models mentals que ofereix el producte o servei que s'avalua, utilitzant les entrevistes prèvies.

Exemples

Si fem un test sobre motors d'avions, en la fase de reclutament podem preguntar: "Quin és el teu nivell de coneixement sobre avions?".

- **Durada raonable.** Generalment, les tasques que s'han de fer no són gaire complexes, per tant, l'usuari no ha de trigar més de deu minuts a completar cadascuna. Hem de repartir bé el temps per tasca. Si el test dura noranta minuts, les tasques haurien de durar deu o dotze minuts. D'aquesta manera les tasques es poden dividir en subtasques i, fins i tot, podem tenir algunes opcions en reserva per si veiéssim que la prova va bé de temps i els usuaris es troben còmodes.

Per a fer les proves d'usabilitat és recomanable crear una taula en què es col·locaran totes les tasques. Així, els observadors poden anar escrivint al costat de cada tasca els comentaris que susciten les accions que l'usuari du a terme i altres notes.

Per a cada característica que vulguem avaluar, hem de generar una tasca, i amb totes les tasques obtindrem el guió de tasques per a la prova.

Exemple

Objectiu de la tasca: "Volem comprovar si els usuaris utilitzen el cercador i si la seva ubicació és l'adequada".

Formulació de la tasca: "Se t'ha trencat l'aspirador de casa i necessites comprar-ne un de nou. Com t'ho faries per obtenir el mateix model de la llista de tots els aspiradors disponibles?".

Potser una cosa que cal valorar sobre les tasques són les circumstàncies d'ús d'una interfície. Parlarem del temps de dedicació, de la motivació de l'usuari i d'altres aspectes crucials. Però no hem d'oblidar que la tasca està determinada per unes circumstàncies específiques que també hem de definir. Per exemple, comprar un bitllet amb tres mesos d'antelació no és el mateix que trobar-se en la necessitat de comprar un bitllet d'un dia per a l'altre. El comportament no serà el mateix i s'apropará més o menys a una situació real depenent del lloc, la pàgina o les tasques que avaluem.

Una vegada tinguem seleccionades les tasques, és recomanable practicar algunes proves amb algun company d'equip per a veure si el test és massa llarg o si les tasques són adequades i comprensibles.

Exemple de document per a un test d'avaluació

A continuació, es mostra un exemple del tipus de document que es pot utilitzar en una prova en què es recullen les tasques-escenaris, les mètriques i possibles comentaris.

Escenari	Tasca	Èxit/fracàs	Temps	Comentaris
Imagina que la setmana que ve és l'aniversari de la teva mare i saps que li faria il·lusió tenir la novel·la de Pérez Reverte de què tant parla.	"Podries accedir al web X i tractar de trobar l'última novel·la d'Arturo Pérez Reverte?"	Es tracta d'anotar si l'usuari ha estat capaç de fer la tasca encomanada.	[hh:mm]	Per a anotacions que puguin ser interessants respecte a la tasca; per exemple: troba una novel·la de Reverte, però no és l'última.
	"Pots comprar el llibre?"			
	"Pots comprovar en quina data es lliurarà el llibre?"			
Fa poc has llegit *Els pilars de la terra* i vols deixar un comentari per compartir la teva opinió amb altres lectors.	"Busca el llibre en el web. "			
	"Puntua el llibre."			
	"Contesta el comentari d'un altre usuari."			

Mètriques

Les **mètriques** són dades quantitatives i qualitatives que recollim durant el test. Alguns exemples de mètriques possibles són:

- La velocitat amb què l'usuari fa una tasca (temps).
- La quantitat d'errors que comet (nombre d'errors).
- La freqüència amb què rectifica l'error (nombre de rectificacions).

- La quantitat d'usuaris que completen la tasca satisfactòriament (nombre d'usuaris).

Els objectius numèrics no són la meta del disseny. Un dels grans errors és pensar que si hi ha una cosa que no es pot mesurar, llavors no importa o no compta.

No hi ha un conjunt universal de mètriques que serveixin per a totes les proves, sinó que aquestes depenen dels objectius concrets que s'hagin definit i del tipus de producte o servei que s'avalua.

A diferents dispositius, diferents mètriques. Per exemple, un web o un dispositiu mòbil poden compartir la mètrica de temps per tasca, però mentre que en un web pot interessar l'observació de l'ús del botó de tornada (*back*), en un dispositiu mòbil aquesta opció no pot existir com a tal; en canvi, podem observar i registrar si l'usuari empra la marcació numèrica o per veu.

Apuntem aquí una divisió simple de les mètriques i alguns exemples habituals:

Guies d'exemples

Alguns exemples de guia per a tests amb usuaris (en anglès) es poden trobar en línia.

a) Quantitatives

- Èxit: si l'usuari és capaç de fer la tasca.
- Fracàs: si l'usuari no és capaç de fer la tasca.
- Temps: temps de cada usuari per tasca.

La definició d'*èxit* i *fracàs* és més difusa del que sembla. Si un usuari completa tots els passos d'una tasca excepte l'últim, podem parlar de fracàs?

Alguns professionals donen percentatges d'èxit a fracàs a cada usuari i a cada tasca. Tot i així, el grau de refinament respecte de qualsevol mètrica que vulguem recollir és una decisió que l'equip avaluador ha de preveure per a cada test.

Refinant una mica més l'èxit i el fracàs, podem intentar recollir mètriques dels casos en els quals els usuaris creuen que han completat correctament o incorrectament la tasca. Es poden recollir mètriques com "fals èxit" i "fals fracàs" a partir del pensament manifest de l'usuari, en què verbalitza si ha acabat la tasca o no.

b) Qualitatives

- **Observacions.** Anotacions sobre les dificultats, els comportaments inusuals o les situacions en les quals la causa de l'error no és òbvia.

- **Comportament.** Accions efectuades per l'usuari que permeten entendre l'execució de la tasca. Per exemple, en un test amb un web, la ruta que segueix un usuari per a resoldre-la: Home\opció 1\opció 2...

- **Literals.** Opinions subjectives sobre l'experiència i la interfície expressades pels participants, o comentaris que acompanyen l'acció i formen part del protocol de pensament manifest.

- **Anàlisi de l'experiència emocional.** Anotacions sobre les reaccions emocionals dels usuaris, com ara exclamacions de sorpresa, disgust, expressions de frustració, d'alegria, etc. Aquest tipus d'observacions emocionals són profundament subjectives i convé que l'observador que les reculli tingui ben tipificats els trets emocionals que ha de registrar abans de concloure que s'ha produït una emoció o una altra.

Lectura recomanada

Llegiu l'article següent:
Eva de Lera; Muriel Garreta-Domingo (2007). "Ten emotion heuristics: guidelines for assessing the user's affective dimension easily and cost-effectively".

Qüestionari posttest

Al final de la sessió de test hem de deixar un quant temps (com a mínim deu minuts) perquè l'usuari expressi la seva opinió sobre el producte o servei. Per això és necessari disposar d'un qüestionari que ens ajudarà a guiar aquestes preguntes i conèixer així el grau de satisfacció de l'usuari. Preguntes del tipus "Què és el que més t'ha agradat?", "Què és el que menys t'ha agradat?", "Què t'ha semblat confús?", etc., ajuden a conèixer l'opinió global de l'usuari. És a dir, aquest últim tipus de preguntes se centra especialment a avaluar la usabilitat subjectiva, la percepció que l'usuari ha tingut del lloc web i de la tasca.

El qüestionari posttest es pot fer en forma de preguntes tancades o preguntes obertes. En general, en tots dos casos, al final sempre es deixa un temps perquè l'usuari faci observacions o comentaris oberts.

Exemples de qüestionaris posttest

Preguntes tancades:

- En general, creus que és fàcil utilitzar el lloc web?

Poc 0○ 1○ 2○ 3○ 4○ 5○ 6○ 7○ Molt

- La teva impressió general sobre l'ús del lloc web és:

Pèssima 0○ 1○ 2○ 3○ 4○ 5○ 6○ 7○ Excel·lent

- Has sabut en quina pàgina estaves en tot moment?

Mai 0○ 1○ 2○ 3○ 4○ 5○ 6○ 7○ Sempre

- Consideres que el disseny gràfic t'ha ajudat a trobar el que buscaves?

Mai 0○ 1○ 2○ 3○ 4○ 5○ 6○ 7○ Sempre

- El lloc web t'ha permès complir les teves tasques?

Poc 0○ 1○ 2○ 3○ 4○ 5○ 6○ 7○ Molt

Preguntes obertes:

- Què és el que més t'ha agradat del lloc web?
- Què és el que menys t'ha agradat del lloc web?
- Què te n'ha semblat el disseny gràfic?
- Creus que tornaries a fer servir aquesta pàgina en altres ocasions?

Carta d'autorització

La **carta d'autorització** és un document en matèria de protecció de dades i privacitat de l'usuari en què es requereix el seu consentiment previ per a capturar i emmagatzemar informació, i que aclareix la finalitat del seu tractament i l'ús que s'hi donarà.

Cada usuari ha de firmar aquesta carta, que autoritza el facilitador a prendre notes, enregistrar àudio i vídeo, etc., i confirma que la informació és confidencial per a l'empresa encarregada de dur a terme el test.

Exemple de carta d'autorització

Jo, _____, amb DNI número _____, autoritzo la captació de la imatge, veu i testimoni a _____, amb domicili a _____ i NIF/CIF _____, els quals es podran utilitzar amb l'única i exclusiva finalitat d'anàlisi de l'estudi de què he estat informat.

En cas de revocació del consentiment s'ha d'avisar per escrit a l'adreça postal següent:

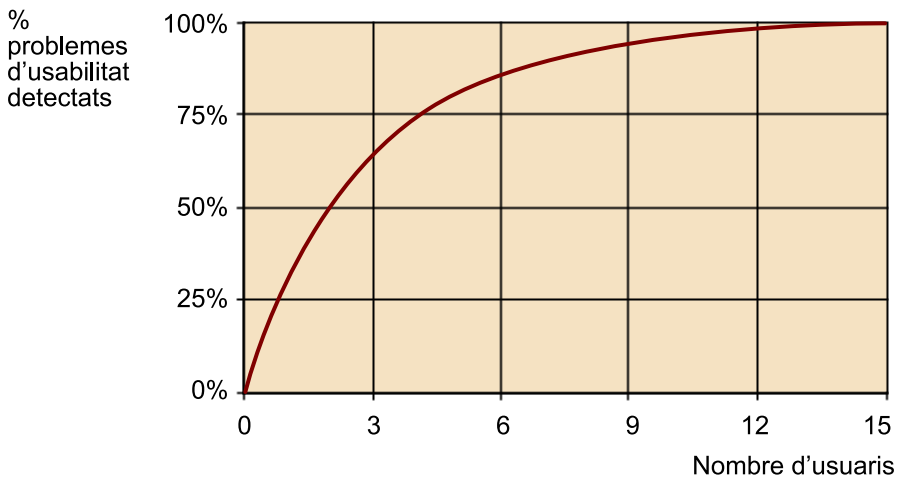
XXXXXX

També accepto que tota la informació intercanviada és propietat exclusiva de XXXX.

Firmat:

1.3.3. Captació dels participants

S'ha conjecturat molt sobre el nombre adequat d'usuaris en fer un test d'usabilitat, però els resultats obtinguts en diferents proves conclouen que un nombre elevat no es correspon amb millors resultats. Nielsen (2000) considera que un nombre òptim de participants en un test amb usuaris ha de ser almenys de cinc. Fins i tot recomana que en lloc de treballar amb quinze usuaris en un sol test, és millor dur a terme tres proves en diferents moments del procés amb cinc usuaris.



No es tracta de donar solució al 100% dels problemes o de buscar totes les debilitats del sistema. El propòsit és millorar el disseny i de manera iterativa anar donant respostes a la millora de la usabilitat.

Per començar a reclutar usuaris per a un test amb perfils adequats, hem de tenir molt clar qui és el públic objectiu del lloc web.

Exemple de públic objectiu

Si desenvolupem una intranet per a una empresa, haurem de preguntar al nostre client per usuaris que la facin servir, ja que aquests seran els usuaris potencials del nostre test.

Encara que nosaltres mateixos ens podem encarregar del reclutament, posant anuncis en pàgines concretes, hi ha empreses especialitzades en aquest tipus de serveis que agilitaran el procés. Una vegada seleccionats els participants, ens hi hem de posar en contacte per acordar un dia, un lloc i una hora per fer el test.

Generalment, en els tests amb usuaris s'ofereix una gratificació pel temps dedicat. Aquestes gratificacions per col·laboració ajuden i motiven els usuaris, cosa que no hem de perdre de vista en cap moment.

1.3.4. Execució del test: exemple pràctic

Segons les explicacions anteriors, podríem dividir una prova d'usabilitat en sis passos o fases de desenvolupament: benvingura i explicació, agenda, qüestionari pretest, tasques, qüestionari posttest, i firma de l'autorització i lliurament de la gratificació.

Primer pas: benvinguda i explicació del test

Abans de començar amb el test d'usabilitat, el qüestionari i les tasques, hem de crear un ambient proper i confortable. L'usuari abordarà sol la tasca, encara que hi hagi gent al voltant, i se sentirà observat davant d'una persona desconeguda (facilitador) i una càmera de vídeo o una càmera web que l'enregistrarà. És per això que hem de trencar el gel perquè se senti còmode i intenti actuar amb tota normalitat. Altrament, els resultats podrien sortir esbiaixats.

Hem de començar explicant que l'objectiu de la prova és avaluar la qualitat d'ús del producte o servei, mai no s'avalua l'usuari. Hem de reiterar que si l'usuari comet algun error durant el test no serà culpa seva, sinó del disseny. És de summa importància que això li quedi molt clar.

També s'ha d'insistir en l'aplicació del protocol de pensament manifest durant el test. Aquesta tècnica és la que permet a l'equip d'avaluadors saber el que l'usuari pensa mentre executa les tasques (usabilitat percebuda), o almenys apropar-s'hi una mica; per tant, és útil per a detectar què li crida l'atenció, què el motiva a executar cada acció, què és el que no comprèn, etc.

S'ha d'explicar a l'usuari que no som al seu cap i, per tant, no sabem què pensa en cada moment. Escoltar les seves opinions sobre el nostre producte o servei durant tot el test implica dir en veu alta tot el que li passi pel cap. Al llarg del test el podem ajudar amb frases com "Què n'opines?", "Què hi veus ara?", "Què et sembla aquesta nova pantalla?", etc.

Una bona manera d'explicar el pensament manifest a l'usuari és mostrar-li un exemple pràctic del que esperem d'ell.

A més, ha de quedar molt clar que la funció del facilitador és la d'observar, en cap cas ajudar l'usuari. Per tant, l'usuari ha de saber que el facilitador no podrà donar suport a les seves decisions en la consecució de les tasques.

Finalment, és important deixar clar el temps que durarà el test i mantenir-ho. Hem de respectar els horaris dels participants.

Segon pas: agenda

Explicar a l'usuari els passos que seguirem durant el test per a situar-lo.

Exemple

"Primerament et faré un seguit de preguntes abans de començar amb el test per conèixer-te una mica millor. A continuació et demanaré que facis una sèrie de tasques.

T'aniré posant un conjunt d'escenaris ficticis per posar-te en situació i hauries de resoldre les tasques que et vagi proposant. Recorda que jo no et puc ajudar.

Si no saps com resoldre la tasca, m'ho comentes i passem a la següent. No et preocupis si no la fas bé o no l'acabes, ja que, com t'he comentat, no t'avaluo a tu, sinó l'aplicació.

En acabar les tasques et faré unes preguntes relacionades amb el test per tenir-ne una opinió global.

Aquest procés acostuma a durar una hora, així que, quan vulguis, comencem."

Tercer pas: qüestionari pretest

El qüestionari pretest, el pot emplenar l'usuari o el pot preguntar el facilitador.

Recomanem que sigui el facilitador qui formuli les preguntes i intenti conèixer el màxim possible l'usuari, d'aquesta manera també pot aprofitar aquest moment per a apropar-se a l'usuari i trencar el gel, si encara no se sent còmode.

Quart pas: tasques

Indiquem a l'usuari que començarem, fent una sèrie de tasques que haurà de dur a terme.

Primerament explicarem a l'usuari el producte o servei que analitzarem i deixarem que hi accedeixi i dediqui uns minuts a observar-lo, tocar-lo, etc., en definitiva, a familiaritzar-s'hi. Al cap d'uns minuts li preguntarem què és el que veu, de què creu que es tracta, quin creu que n'és l'objectiu i la seva opinió sobre el primer cop d'ull.

En aquest sentit es pot aprofitar aquest primer contacte de l'usuari amb la pàgina per a efectuar una breu tasca prèvia denominada *test dels cinc segons*. Consisteix en un exercici senzill que pot aportar informació sobre la pàgina del lloc web que inicialment mostrem a l'usuari. Durant cinc segons podrà veure aquesta pàgina i, passada aquesta breu estona, l'haurà de minimitzar, ocultar, o haurà d'apagar la pantalla. Després li demanarem que ens expliqui tot el que recorda, la primera impressió i quin tipus de contingut creu que conté el lloc web.

Lectura complementària

Sobre el test dels cinc segons, podeu consultar l'obra següent:

C. Perfetti (2005). "5-Second Tests: Measuring Your Site's Content Pages" [article en línia]. *User interface engineering*.

Tota la informació que ens ofereixi l'usuari ens permetrà conèixer aspectes relacionats amb el disseny de la pàgina i amb la jerarquia visual del contingut. Si és capaç de donar una explicació coherent del propòsit d'aquesta, destacar el que en resulta més rellevant, i això es correspon amb els objectius de la pàgina, tindrem una primera dada sobre la comunicació del nostre lloc i sobre l'ordenació visual i la claredat dels continguts.

Es tracta d'una tècnica ràpida i fàcil d'aplicar que podem dur a terme abans o després de situar els usuaris en un context determinat.

Un cop acabada aquesta primera part, analitzarem la facilitat d'ús del producte o servei mitjançant tasques i escenaris concrets.

Durant el transcurs de les tasques podem interrompre l'usuari, si creiem que és necessari. De la mateixa manera, si percebem que l'usuari no sap acabar una tasca, li direm amablement que no hi ha cap problema i passarem a la següent. Hem de recordar que un usuari mai no ha de pensar que fa malament una tasca. És el producte o el servei el que genera l'error, o les persones que l'han dissenyat.

Si veiem que no s'expressa, li hem de recordar que és important que ens expliqui les seves impressions per a saber el que pensa.

Cinquè pas: qüestionari posttest

Una vegada els usuaris han completat les tasques, podem passar al qüestionari posttest per a avaluar l'opinió que tenen del producte o servei i perquè expliquin quines han estat les seves percepcions i sensacions durant la prova. Podem aprofitar per a sol·licitar-los qualsevol altre comentari o observació que vulguin afegir.

Sisè pas: firma de l'autorització i lliurament de la gratificació

Un cop acabat el test, els lliurarem la carta d'autorització i, en els casos que s'escaigui, es lliurarà la gratificació o un petit obsequi o regal que tingui relació amb la pàgina, amb l'empresa, o que simplement permeti recordar a l'usuari la seva participació en la prova.

Exemple

"Finalment, et demano que llegeixis i, si hi estàs d'acord, signis aquest paper d'acord amb les dades obtingudes en aquesta prova, així com l'àudio i el vídeo. És la nostra política de privacitat, que permet fer-te saber que no podem fer ús de la informació per a cap altra cosa que no sigui cobrir els objectius i els propòsits del projecte en qüestió.

També et lliurem un obsequi com a agraïment per haver-nos dedicat el teu temps.

Moltes gràcies"

1.3.5. Consells per a ser un bon facilitador de test

Com que el rol de facilitador és fonamental en un test amb usuaris, a continuació presentem una sèrie de consells pràctics per a dur-lo a terme.

1) Ser conscient dels prejudicis i projeccions

Un facilitador té dos mantres: "No jutgem l'usuari" i "No hi ha respostes bones o dolentes". Creure en aquestes dues premisses constitueix la meitat de la peça clau de la seva tasca; actuar en conseqüència és l'altra meitat.

El comportament del facilitador, implícit o explícit, influeix sobre l'usuari.

La dedicació professional plena a la usabilitat implica, com en moltes altres professions, un tracte directe amb clients i usuaris. És natural que hi hagi gent amb qui no hi hagi afinitat, fins i tot envers la qual es pugui sentir cert rebuig per la seva actitud. Això, que és un fet, no s'ha d'evidenciar en cap moment. Tots els usuaris són iguals davant del facilitador.

És tan difícil controlar el grat com el desgrat; aquesta és la raó per la qual també es desaconsella que un facilitador moderi un test amb familiars o amics. Entre la gent que es coneix hi ha tot tipus de condicionants emocionals que contaminen el test.

De la mateixa manera, el sentiment no pot ser mutu. No cal convertir-se en el millor amic de l'usuari, però tampoc no s'ha de deixar que ens rebutgi, ja que l'èxit del test rau en el fet que confiï en nosaltres com a facilitadors i que s'esforci amb la nostra sola presència en l'execució de les tasques.

Al cap i a la fi, un test és una avaluació. Quan mirem "amb ull crític" el que fa una altra persona, des del punt de vista psicològic es produeix una projecció de les nostres percepcions: veiem en l'altre aspectes que nosaltres podríem evitar o que faríem d'una altra manera.

Què es pot fer amb això? Ser conscient d'aquesta realitat és l'única manera de prendre mesures per a evitar-ho i comportar-nos de la manera més natural i neutra possible.

En termes pràctics, una manera de frenar els prejudicis del facilitador és evitar respostes que impliquen valoració: comentaris com "Excel·lent" o "Molt bé", com a reacció al protocol de pensament manifest, es poden substituir per altres de més neutres com "Ja ho entenc" o "D'acord". Fins i tot les preguntes que incideixin en les respostes de l'usuari poden ajudar que continuï amb els seus comentaris: "Llavors, dieu que el menú de navegació no us funciona?".

2) Ser transparent i guanyar-se la confiança de l'usuari

Imagineu que en ple test amb usuaris es produeix un error tècnic en l'enregistrament. Pànic a la sala d'observació. Al facilitador, que està en comunicació directa amb l'equip d'observadors, li diuen que hi ha un problema tècnic amb la cinta i que el test no es pot continuar enregistrant. El facilitador es remou inquiet des del seu lloc al costat de l'usuari, "Què faig?, Aturo la prova?, Continuo?". L'usuari nota que el facilitador està preocupat i es pregunta: "Tan malament ho faig?", "Tant l'avorreixo?".

L'usuari no és aliè al comportament del facilitador, per la qual cosa convé que aquest sigui sincer i transparent. Per exemple, quan passa una cosa poc habitual, com és la interrupció d'un enregistrament o un tall en la connexió a Internet, per posar-ne dos exemples, es pot explicar amb naturalitat a l'usuari i indicar-li que farem una breu pausa per solucionar el problema.

Fins i tot es pot capitalitzar un comportament no previst. Si la connexió a Internet falla, podem recopilar més informació i preguntar a l'usuari què faria en un cas així i com afectaria la seva interacció amb el producte o servei. No cal fingir que hem provocat la fallada en la connexió a propòsit.

L'axioma de la transparència també s'aplica a les reaccions del facilitador davant del que fa l'usuari. Encara que les paraules del facilitador siguin neutres, si el seu comportament no verbal indica sorpresa, desgrat o avorriment, l'usuari se sentirà insegur.

Esbufegar, remoure's a la cadira si s'està assegut, moure excessivament les mans, tamborinar amb els dits, etc. La llista de comportaments involuntaris i no verbals que provoquen una impressió negativa sobre l'usuari és llarga.

3) L'últim test, com el primer

Molt relacionat amb el que hem exposat en el punt anterior, el facilitador no ha de deixar que el cansament o la monotonia afectin el seu comportament durant qualsevol de les sessions de test. De vegades, la tasca de facilitació esdevé repetitiva, ja que en un bloc de diverses proves, els usuaris cauen recurrentment en els mateixos problemes de disseny del producte o servei. Aquest fet no ha de desvirtuar l'últim test amb l'últim usuari. Si un error es repeteix una vegada més, també és una dada valuosa que devem a aquest últim usuari.

4) Adaptar-se a l'estil de comunicació de l'usuari

Sovint després d'un test es cau en la temptació de "comentar la jugada" de l'usuari, especialment la seva capacitat de dur a terme el protocol de pensament manifest. L'usuari és catalogat com a massa xerraire o massa callat, massa tímid, massa exagerat, etc.

En comptes de jutjar, es pot recordar que és responsabilitat del facilitador que l'usuari faci una bona feina. Cal adaptar-se a les característiques de l'usuari, no esperar que ell s'adapti al nostre protocol:

- Els usuaris extravertits tendeixen a pensar i parlar alhora. És el cas de l'usuari que enfila en una mateixa frase "no-ho-trobo-és-aquí-ai-això-no-era-on-sóc". El perill d'aquests és que durant el test aboquen un excés d'informació no sempre rellevant. Per tant, el paper del facilitador és centrar aquest usuari i ser capaç de separar els comentaris valuosos dels comentaris irrellevants.

- Els usuaris introvertits tendeixen a parlar quan han pensat el que diran. És el cas de l'usuari que està callat durant la major part de la tasca i emet una valoració –més o menys lapidària– al final. El perill d'aquests és que no deixen registrar aquests pensaments intermedis que el porten a la seva conclusió. El paper del facilitador és animar aquest usuari a dir el que pensa, però de manera molt delicada, ja que necessita el seu temps i veure's forçat a parlar el pot incomodar.

Recordem que abans d'iniciar l'execució de les tasques i el pensament manifest tenim un temps de recepció de l'usuari que el facilitador ha d'aprofitar per a calibrar el seu temps de resposta o per a adequar-se al perfil (és introvertit, extravertit?).

5) Estar atent a la validesa de les dades

Per molt insistentment que hàgim comunicat a l'usuari que no el jutgem a ell, alguns usuaris se senten sempre jutjats o simplement observats. Aquesta sensació pot ser impossible de canviar, per la qual cosa només es pot minimitzar.

Si el facilitador percep que l'usuari no es comporta de manera natural, cal registrar-ho i tenir-ho en compte en el moment de fer l'anàlisi de les dades del test.

Si les dades de comportament de l'usuari no són fiables, són útils altres fonts de dades com ara el seguiment ocular[7], el seguiment de clics o qualsevol tipus de registre informàtic de seguiment[8].

(7)En anglès, *eye-tracking*.

(8)En anglès, *log*.

Vegeu també

Podeu veure el mètode del seguiment ocular en el subapartat 3.3 d'aquest mòdul didàctic.

6) Deixar que els usuaris parlin de les seves pròpies experiències

De vegades, especialment si el que volen manifestar és negatiu, els usuaris se senten incòmodes quan expressen el punt de vista propi i el posen en boca d'altres persones:

"Per a mi està bé, però la gent en general ho trobarà difícil."

"Per a la meva mare seria difícil."

"La gent gran tindrà problemes amb això."

Són opinions d'usuaris que no aporten informació al test, ja que les dades que busquem són les que es recullen durant la prova. No ens interessa que l'usuari es comporti com si fos un crític o un professional amb opinions subjectives.

El facilitador ha d'extreure d'aquestes afirmacions dels usuaris el que es refereix a ells mateixos i reformular-ho com a pregunta.

Exemple

USUARI: "Per a la meva mare seria difícil."
FACILITADOR: "D'acord. Per a tu és difícil?"
USUARI: "Sí/no, perquè..."

Exemple

USUARI: "Per a mi està bé, però la gent en general ho trobarà difícil."
FACILITADOR: "Vols dir que per a tu està bé?"
USUARI: "Sí, perquè..."

7) Detectar l'autocensura dels usuaris

Una altra manera que troben els usuaris de "maquillar" les seves opinions negatives és rebaixar-les.

Si l'usuari troba difícil el producte o servei, sovint ho formula acompanyant el comentari d'una disculpa: "És difícil, però una vegada que t'hi acostumes no està tan malament...".

Una situació típica de l'entrevista final d'un test és que l'usuari comenti "Al principi era difícil, però crec que la gent que l'ha de fer servir ho entendrà".

El que ens suggereix aquest usuari és que no considera que el seu sigui el perfil per a usar el producte o servei, perquè ha tingut dificultats. El facilitador ha de respondre investigant quines són aquestes dificultats: "Què t'ha semblat més difícil al principi?".

8) No ens conformem només amb els problemes, però no busquem solucions

El resultat del nostre test no es pot limitar als problemes, necessitem el perquè. Els usuaris no han vist l'enllaç, els usuaris no han entès l'etiqueta, els usuaris no han utilitzat el botó, etc. Aquestes conclusions no són útils per a cap equip de desenvolupament que busqui solucionar els problemes, perquè no concreta què caldria modificar per a solucionar-los. En definitiva, falta indagar sobre els perquès. El facilitador ha d'insistir a l'usuari perquè reformuli aquests perquès i perquè les accions d'aquest vagin acompanyades d'una explicació.

També és habitual que els usuaris acompanyin el problema amb la solució que consideren més senzilla des d'un punt de vista molt personal: "No he vist aquest enllaç, hauria d'estar en negreta o en un altre color". Cal escoltar l'opinió dels usuaris i respectar-la, però a l'usuari li falta visió global del producte, no és arquitecte d'informació ni dissenyador. El facilitador pot aprofitar que l'usuari proposa aquesta solució per a extreure més informació sobre el perquè del problema.

Exemple

USUARI: "No he vist aquest enllaç, hauria d'estar en negreta o en un altre color."
FACILITADOR: "No has vist l'enllaç, per què creus que ha passat?"
USUARI: "Perquè està amagat, perquè no sembla un enllaç, perquè esperava un botó, etc."

Si en aquest diàleg el facilitador aconsegueix que l'usuari explori prou el problema, es pot definir amb ell una solució.

FACILITADOR: "Per tant, no has vist l'enllaç perquè no sembla un enllaç, quin aspecte hauria de tenir?"

9) Objectivar allò que és subjectiu

Durant un test, els usuaris fan moltes coses de les quals no és fàcil extreure un significat: somriuen, arrufen les celles, es queden mirant alguna cosa, sospiren, etc. El nostre equip d'observadors pot tenir un responsable de recollir tot aquest comportament no verbal, però el facilitador pot ajudar que aquesta informació s'interpreti correctament.

Una bona estratègia passa perquè el facilitador posi de manifest objectivament aquests comportaments; simplement es tracta de "traduir" els seus gestos a paraules, però parant atenció a no qualificar.

Exemple

L'usuari somriu.

FACILITADOR: "Has somrigut."
USUARI: "Sí, això m'ha recordat..."

Referent al comportament no verbal de l'usuari, és important mantenir una distància gairebé d'antropòleg. Si l'usuari somriu i el facilitador li pregunta directament per què ha somrigut, és difícil que no se senti qüestionat i pot ser que li resulti difícil respondre.

1.4. Classificació dels tests amb usuaris

No hi ha una classificació formal o estandarditzada dels tests de tasques amb usuaris; això no obstant, els podem agrupar segons diverses dimensions, com ara les característiques físiques de l'entorn de test o el prototip que s'avalua.

1.4.1. Presencial/remot

Hi ha hagut diferents maneres d'efectuar els tests amb usuaris. Al principi, les proves d'usabilitat es duien a terme en laboratoris especialitzats que disposaven dels mitjans necessaris per a efectuar-los. Per exemple, un laboratori amb una sala ambientada com una oficina o una casa on l'usuari du a terme el test. Un espai on els observadors podien visualitzar l'usuari i les interaccions

que d'aquest amb l'aplicació i un equip tècnic que s'encarregava d'enregistrar tot el procés. Aquesta opció continua essent interessant per als estudis en què l'entorn i el context d'ús tenen un paper principal.

Com ja hem indicat, també es poden dur a terme proves amb l'ajuda d'un ordinador que disposi d'un programari d'enregistrament que permeti veure el recorregut de l'usuari per la interfície, i una càmera per a enregistrar-ne els gestos i els comportaments. En qualsevol cas, es requereix la presència física de l'usuari i dels avaluadors en un mateix espai; això comporta diverses limitacions:

- Temporals: perquè només es pot fer un test cada vegada.

- Econòmiques: perquè utilitzar usuaris és costós, ja que costa temps i diners.

- Grandària de la mostra: perquè per les limitacions temporals i econòmiques, el nombre d'usuaris que participen en els tests no sol arribar a quinze.

Ja fa un quant temps que hi ha una altra manera de dur a terme un test amb usuaris sense necessitat de tenir-los a la mateixa habitació amb els experts. Aquest tipus de test amb usuaris es coneix com a **test remot**. El participant es troba en un lloc físic distant de l'expert.

En cas de tenir usuaris repartits per diferents zones geogràfiques o quan el context d'ús no es pot reproduir en un laboratori, un test d'usabilitat remot és la millor solució, ja que els usuaris poden participar en la prova des de qualsevol lloc.

Avantatges del test remot enfront de l'opció de la presencialitat:

- L'usuari fa les tasques en el seu entorn habitual.
- Permet la participació de diversos observadors.
- Evita desplaçaments i pot avaluar amb usuaris de tot el món.
- Simultaneïtat en l'execució del test.
- Reducció de costos.
- Mostra més gran d'usuaris.

Inconvenients del test remot:

- No podem veure l'expressió facial ni els moviments físics de l'usuari, tret que aquest tingui una càmera.
- Problemes amb la connexió a Internet o amb el tipus de programari (navegadors, connectors[9], controladors) que utilitza l'usuari.
- Sensibilitat de la informació lliurada o de l'accés a llocs restringits per a usuaris, per exemple, en el cas d'un banc.

(9) En anglès, *plug-in*.

Hi ha dos tipus de test remot: moderat i no moderat.

Test remot moderat

El test d'usabilitat es du a terme en línia mitjançant un programari que permet la connexió remota de diversos ordinadors. Amb aquest programari l'usuari pot accedir a l'ordinador del moderador i visualitzar el producte que es vol avaluar. El procediment és el mateix que el d'un test amb usuaris no remot, amb la diferència que l'usuari no és al mateix lloc físic que el moderador. S'utilitzen tecnologies de suport com les videoconferències, en les quals l'usuari i els moderadors estan connectats en línia. La pantalla de l'usuari i la conversa entre l'usuari i la persona que modera l'estudi es poden gravar en vídeo com si fos un test d'usabilitat presencial. A més, la conversa pot ser seguida per altres persones.

Quin programari es necessita? Efectuar un test d'usabilitat remot moderat necessita els elements següents:

- Ordinador
- Càmera (*webcam*)
- Aplicació específica per a compartir pantalles, és a dir, podrem accedir a un ordinador i aquest al nostre mitjançant aquest programari (per exemple: TeamViewer).
- Telèfon o programa específic per a parlar amb l'usuari (per exemple: Skype).
- Programa específic per a enregistrar la conversa i la interacció de l'usuari amb la interfície (per exemple: Morae o Camtasia).

Dur a terme un test d'usabilitat remot moderat és pràcticament el mateix que fer un test d'usabilitat. La principal diferència, com ja hem comentat, és que els usuaris no estan presencialment en el moment del test. Per això és imprescindible aconseguir que se sentin còmodes i explicar-los, abans de començar, el procediment que se seguirà.

El test d'usabilitat remot requereix certa preparació per part de l'usuari, si es compara amb un test d'usabilitat convencional. Els participants han de disposar principalment de programari específic en el seu ordinador i per aquest motiu es recomana contactar-hi amb temps suficient abans del test, per correu electrònic o per telèfon, i detallar el programari necessari, i com instal·lar-lo i configurar-lo.

En general, en finalitzar els tests d'usabilitat s'agraeix a l'usuari el seu temps i dedicació mitjançant un obsequi o gratificació. En el cas del test d'usabilitat remot, podem fer arribar la gratificació per correu postal o utilitzar maneres "remotes" de compensació per a usar en llocs de comerç electrònic.

Test remot no moderat

Aquesta tipologia de test permet substituir el moderador per un programari. Els especialistes i els analistes dissenyen el programari perquè aquest compleixi la mateixa funció, encara que sempre convé recordar que cal valorar el tipus de prova que duem a terme ateses les diferències evidents que hi pot haver entre la presència humana i una màquina.

Mentre l'usuari és a casa seva, el programari proposa als usuaris una sèrie de tasques i fa preguntes que han de contestar a mesura que el sistema captura dades sobre la interacció de l'usuari amb el producte. Aquest sistema permet l'entrada de diversos usuaris simultàniament sense la presència de moderadors, per això es poden avaluar mostres de centenars d'usuaris i obtenir dades estadísticament representatives.

Escollir el tipus de test és important perquè variarà segons les necessitats que tinguem en cada moment. El test remot no moderat, en no necessitar un observador o facilitador que acompanyi l'usuari en tot el procés de test, oferirà uns resultats més quantitatius; en canvi, el test remot moderat, molt semblant quant a metodologia al test d'usabilitat convencional, ens permetrà observar el comportament de l'usuari, escoltar els seus comentaris i visualitzar la interacció d'aquest amb la interfície.

Quan hem de triar entre l'un o l'altre, hem d'analitzar els objectius que volem assolir amb el test. Si el que necessitem és esbrinar quantes vegades els usuaris fan clic en un lloc o la seva interacció amb la interfície sense que en sigui imprescindible l'observació, potser haurem d'optar per un test remot no moderat, que ens aportarà una gran quantitat d'informació que haurem d'analitzar posteriorment. Si, al contrari, el que necessitem és conèixer l'opinió dels usuaris, les seves reaccions davant d'un lloc o observar-ne el comportament, haurem d'efectuar un test remot moderat.

Programari necessari

A continuació es mostra algun programari dels que es poden utilitzar per a dur a terme un test remot.

Enllaços d'interès

Usability tools (en línia)
Remote Research (en línia)

Programari necessari per a un test remot moderat

Producte	Descripció
UserVue	Programari de TechSmith. Permet visualitzar la pantalla dels participants en temps real. Poden observar la sessió sense que l'usuari els vegi tants observadors com es vulgui. La crida i els moviments que l'usuari hagi fet queden enregistrats al final de la sessió.
Livelook	Servei per a compartir pantalles en remot. L'usuari pot visualitzar la pantalla del facilitador i viceversa. No enregistra àudio ni vídeo.
OpenVULab	Creat per la Universitat de York. Eina de codi obert que com UserVue permet enregistrar vídeo i àudio en els ordinadors dels usuaris mentre dura el test.

Producte	Descripció
Skype	Creat per Mac OS X. Eina gratuïta en què es pot compartir pantalla, vídeo i xat. Requereix tenir el programa instal·lat en l'ordinador.
TeamViewer	Servei per a compartir pantalles en remot. L'usuari pot visualitzar la pantalla del facilitador i viceversa. No enregistra àudio ni vídeo. No cal descarregar el programari per a utilitzar-lo.

Programari necessari per a un test remot no moderat

Producte	Descripció
ClicTale	Programari que enregistra la interacció amb un producte o servei utilitzant JavaScript. Enregistra la interacció de l'usuari, com també un mapa en què es pot visualitzar el recorregut de l'usuari: on han fet clic, on ha desplaçat la barra de desplaçament (*scroll*), formularis, etc.
UserZoom	Eina potent que permet controlar i dirigir múltiples projectes de DCU, i que és capaç de proporcionar gran quantitat d'informació quantitativa, generar les tasques per al test d'usabilitat i controlar-les (*card sortings*, entrevistes o reclutament d'usuaris) des d'un tauler de control general.
KeyNote	Eina semblant a l'anterior, en la qual els usuaris contesten enquestes i completen tasques mitjançant llistes desplegables (*pop-up*) sense necessitat de descarregar el programa. Mitjançant un tauler de control podem visualitzar els resultats exhaustius del test. Generalment s'obtenen dades quantitatives.
m-pathy	Eina que enregistra els moviments i clics del ratolí sense necessitat d'instal·lar cap tipus de programari en l'ordinador de l'usuari.
Chalkmark	Creat per OptimalWorkshop. L'usuari pot resoldre les tasques mitjançant imatges estàtiques. El sistema genera un mapa en què es pot visualitzar el recorregut de l'usuari, similar a ClickTale. És apropiat per a fer tests no gaire complicats.
LabsMedia	Creat per LabsMedia. Projecte de codi lliure similar a VULabs i ClickTale.
UserFocus	Programari que permet obtenir una gran quantitat d'informació quantitativa, generar les tasques per al test d'usabilitat i controlar-les. No requereix descàrrega o instal·lació.
SMT	Creat per Luis Leiva. Eina de codi lliure que permet enregistrar el recorregut d'un usuari per un producte o servei.

1.4.2. Formal / de guerrilla

Com ja hem indicat, una prova d'usabilitat es pot dur a terme en diferents llocs. Segons les característiques d'aquests entorns, el test es pot considerar formal o de guerrilla:

a) Tests formals. Són les proves que es fan en un laboratori, adequat i construït per a dur a terme aquest tipus de tècniques. Generalment disposen de tecnologies avançades per a l'enregistrament de la prova a més de diverses habitacions disposades amb l'objectiu de crear un entorn real per a l'usuari, com també de permetre als observadors veure les tasques que l'usuari desenvolupa mitjançant miralls unidireccionals. En definitiva, es tracta d'entorns construïts amb l'objectiu exclusiu d'efectuar aquest tipus de proves d'avaluació.

b) Tests informals o de guerrilla. Una altra manera més econòmica i fàcil de dur a terme un test amb usuaris són els anomenats *tests de guerrilla*. El procediment de test és exactament el mateix que l'anterior, el que varia són les condicions externes. Per a aquest tipus de test no necessitem unes instal·lacions com en el test formal, únicament és necessari disposar d'un ordinador amb el programari adequat, una càmera (*webcam*) i, per descomptat!, els usuaris que faran el test. Amb aquest equipament, podem fer un test amb usuaris amb resultats i utilitats molt similars a l'anterior.

Lectura recomanada

Sobre les fases per a la realització d'un test informal, podeu consultar l'obra següent:

Yusef Hassan Montero; Francisco J. Martín Fernández (2003, 9 de desembre). "Método de test con usuarios" [article en línia]. *No Solo Usabilidad*.

1.4.3. Tests amb prototips d'alta/baixa fidelitat

Com ja hem comentat, es poden efectuar diferents tipus de tests amb usuaris segons la fase en què es trobi el projecte; no cal esperar a tenir una implementació final per a dur a terme un test d'usabilitat.

Hi ha diferents tipus de proves depenent de la fase o de l'estat de detall en el qual es trobi el producte o servei que s'analitzi:

a) Tests amb un prototip d'alta fidelitat. S'anomenen *prototips d'alta fidelitat* els que són molt propers al producte o servei final, tant en disseny o interactivitat com en funcionalitats. L'avantatge d'aquests tests és que representen pràcticament allò amb què es trobarà l'usuari, però això els fa més costosos de produir; són més propis de les etapes finals del projecte.

b) Tests amb un prototip de baixa fidelitat. S'anomenen *prototips de baixa fidelitat* els que estan lluny del disseny i les funcionalitats del producte o servei final; un prototip de baixa fidelitat d'una pàgina web de venda de bitllets d'avió pot estar dibuixat en una pàgina de paper. L'avantatge d'aquests tests és que es creen i es destrueixen molt de pressa, per la qual cosa en etapes inicials del projecte resulten molt útils per a descartar idees o per a identificar requisits.

Depenent del moment en què es trobi el nostre projecte, del que vulguem avaluar o del pressupost que tinguem, utilitzarem un tipus de test o un altre.

Sempre obtindrem una informació més detallada si utilitzem un prototip d'alta fidelitat que no pas si n'utilitzem un de baixa fidelitat, però en determinades fases de desenvolupament no ens interessa tant el detall com avaluar les línies generals del producte o servei.

L'important és saber seleccionar el tipus d'avaluació segons les necessitats del projecte. És millor fer proves constants i en diferents graus de qualitat i fidelitat durant el projecte, que no esperar a tenir un producte acabat per al test.

2. Variants del test amb usuaris

Fins ara hem presentat el test amb usuaris en la versió més convencional: un usuari, un facilitador, un laboratori, observadors, etc. Tanmateix, hi ha condicionants que poden fer necessària una variació sobre els elements establerts:

- per necessitats específiques del producte o servei: per exemple, que la forma de consum del producte que estem avaluant no sigui individual, sinó col·lectiva. No és el mateix avaluar un caixer automàtic (interacció individual), que una màquina recreativa (interacció col·lectiva). El format del nostre testha de variar en conseqüència.

- perquè apareixen noves tecnologies que ajuden a aportar més dades sobre la interacció dels usuaris amb el producte o servei. La tecnologia ens ofereix avenços que es poden aplicar a l'avaluació amb usuaris i variar el format del nostre test. La innovació més recent ha estat la incorporació del seguiment ocular dels usuaris, és a dir, tecnologia que permet saber on miren en una pantalla. El futur ens oferirà cada dia més i més eines, que poden fer variar el clàssic test.

- per l'estadi de desenvolupament del producte interactiu: quan ens trobem davant d'una fase molt primerenca de desenvolupament, pot ser difícil fer un test amb un prototip que funcioni. Això no ha de portar a renunciar a avaluar amb usuaris: simplement, cal canviar el format de l'avaluació.

A continuació repassem algunes de les variants principals del test d'usuaris, que responen a aquests condicionants. Depenent del producte interactiu i dels objectius de l'avaluació d'usabilitat, el test amb usuaris es pot dissenyar de manera que combini tipologies i variants del test estàndard o convencional i, d'aquesta manera, ajustar els objectius de l'avaluació i obtenir resultats millors.

2.1. Test de codescobriment

En essència, el **test de codescobriment** és un test amb usuaris, però executat per dos usuaris simultàniament que col·laboren en la resolució de les tasques.

Aquest tipus de variant pot reflectir millor determinats ambients d'ús d'un sistema. Per exemple, és més freqüent que s'utilitzi una aplicació d'ofimàtica en una oficina on, quan hi ha un problema, altres usuaris de la mateixa aplicació poden ajudar.

D'altra banda, per a dos usuaris, sempre serà més natural intercanviar impressions sobre el producte entre ells que anar parlant en veu alta i en solitari amb el facilitador del test, la qual cosa ajuda a evitar un dels riscos de qualsevol test amb usuaris: que l'usuari no se senti còmode i no actuï amb naturalitat.

En aquest tipus de test, la tasca del facilitador es complica, en tant que s'ha d'ocupar que els dos usuaris parlin per igual i prenguin decisions de manera consensuada; no hi ha res pitjor en un test de codescobriment que els usuaris s'enfadin entre ells per la resolució de les tasques.

2.2. Seguiment ocular

L'***eye-tracking*** fa referència a les tecnologies que permeten registrar l'exploració visual d'un usuari davant de la informació que se li presenta. Podria ser traduït com a *seguiment ocular*, ja que el monitoratge que es du a terme és el del moviment dels ulls, concretament la identificació i el mesurament de les fixacions. En aquest sentit, l'*eye-tracking*, és una eina de recopilació de dades quantitatives, més que una metodologia de test pròpiament dita.

Identificació i mesura de fixacions

Els salts o moviments ràpids que fem amb els ulls es diuen *moviments sacàdics*. Entre un moviment sacàdic i un altre es produeix una fixació, període de relativa quietud de l'ull que permet veure nítidament la zona enfocada (Hassan i Herrero-Solana, 2007).

Com a eina per a l'avaluació d'interfícies, es poden utilitzar dues alternatives: càmeres[10] col·locades al cap dels participants, o càmeres col·locades davant l'usuari al monitor on dirigirà la mirada.

(10)En anglès, *eye-trackers*.

No es tracta d'una tecnologia innovadora, ja que han existit dispositius que permetien aquest tipus d'anàlisi des dels anys cinquanta, però es tractava d'aparells molt intrusius que incomodaven els usuaris, ja que eren propis de la investigació científica més ortodoxa.

En l'última dècada, algunes empreses han desenvolupat maquinari i programari de seguiment ocular capaç d'enregistrar el comportament natural de l'usuari en entorns reals sense necessitat d'incomodar-lo amb tecnologia aclaparadora. Tot i així, el cost, l'escassetat d'informació, la interpretació de resultats o els inconvenients del mètode mateix són raons suficients per a entendre per què no s'utilitza tant com una ajuda per a l'avaluació en l'àmbit de l'experiència d'ús.

Un exemple de nova tecnologia

Eye Tracker Tobii Glasses és un exemple de l'última tecnologia de seguiment ocular.

Els sistemes de seguiment ocular que s'utilitzen en les avaluacions d'usabilitat consten bàsicament dels elements següents:

- Una pantalla adaptada, capaç de detectar mitjançant infrarojos el moviment ocular de l'usuari mentre aquest du a terme les tasques. Aquesta detecció requereix que abans de començar la prova es calibri l'aparell per a adaptar-lo a les particularitats físiques de cada participant.

- Un programari d'interpretació dels registres de la pantalla. En essència, es recull on mirava l'usuari (coordenades de la pantalla), durant quant de temps, d'on venia i on ha mirat després. Són necessàries eines de programari que transformin totes aquestes dades en informació útil per a l'avaluador.

La presentació dels resultats d'un seguiment ocular pot variar segons el programari que s'utilitzi per a interpretar les dades, però a grans trets consisteix en el següent:

- Mapes de recorregut, que representen el camí dels ulls de l'usuari per la pantalla.

- Mapes de calor[11], que representen visualment el temps que cada usuari mira cap a determinades zones de la pantalla com un mapa de calor: com més vermell, més temps.

(11)En anglès, *hotspots*.

- Dades per àrees, que mostren valors numèrics dels temps de fixació de l'usuari en una àrea definida prèviament, per exemple, el cercador d'una pàgina.

Si bé la riquesa de les dades que proporciona el seguiment ocular és gran, aquesta eina presenta dificultats a l'hora d'interpretar què signifiquen realment totes aquestes dades. Per exemple, si l'usuari ha mirat un element de la pantalla, significa realment que l'ha vist?, allò que no mira, vol dir que no li interessa o que no ho entén? És per això que sovint es fan proves amb seguiment ocular que incorporen el mètode del pensament manifest[12].

(12)En anglès, *think-aloud*.

2.3. Llapis i paper

Què és?

La tècnica de llapis i paper és una variant del test d'usabilitat en què usuaris representatius d'un producte o servei duen a terme tasques mitjançant una versió en paper de la interfície del producte o servei controlada per un moderador (Snyder, 2003).

En aquest cas, en no disposar d'una maqueta funcional, el moderador és qui ha d'anar alternant els papers, com si fossin les diferents vistes d'interacció amb el producte, quan l'usuari ho indiqui.

Lectura complementària

Podeu consultar més coses sobre la tècnica del llapis i paper en l'obra següent:

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

Són moltes les vegades que durant el desenvolupament d'un test amb usuaris ens trobem amb participants que opinen sobre el disseny d'una aplicació, icones o tipografies. No és res gens menyspreable, però pot desviar el centre de discussió inicial. En fer la prova amb llapis i paper ens evitem aquest tipus de debats i únicament ens centrem en l'essència del producte, el seu objectiu, la funcionalitat i la interacció.

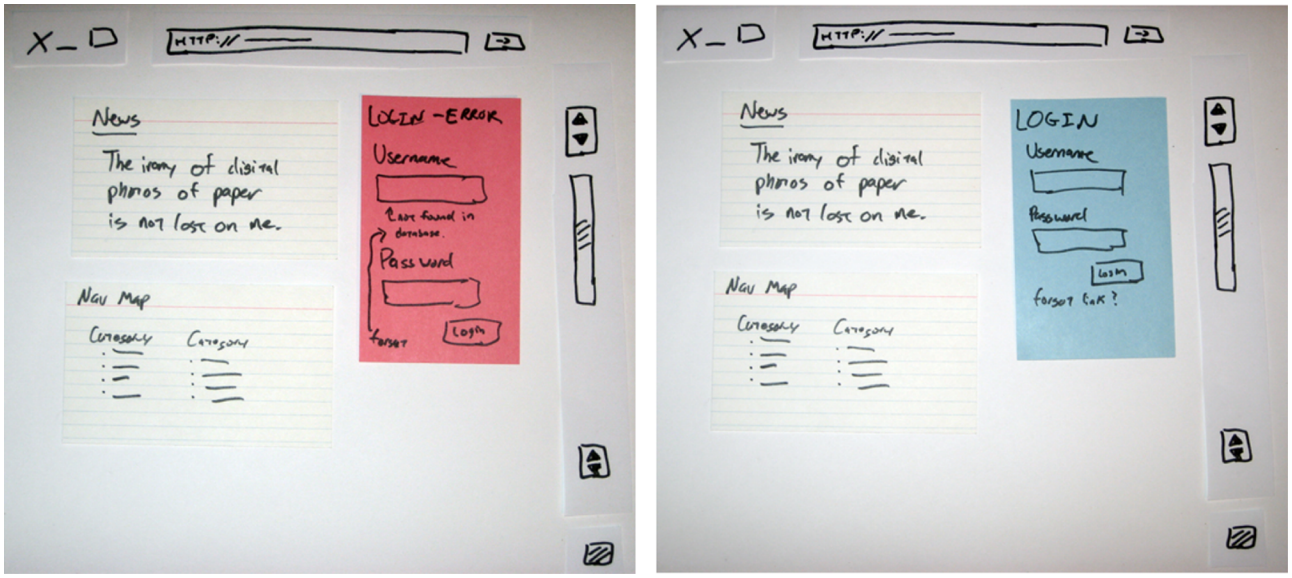
El mètode del llapis i paper

Per entendre millor com funciona el mètode del llapis i paper, proposem veure el vídeo següent:

- *Usability Testing with a Paper Prototype*

Com es du a terme?

Una vegada definides les funcionalitats i les característiques del producte o servei amb què treballem i que volem avaluar, l'únic que necessitem és dibuixar els prototips en paper que utilitzarem durant el test.



Il·lustració d'alistapart.com

Aquests prototips es poden dibuixar a mà o amb eines digitals bàsiques com ara el Word o el PowerPoint.

Qui hi participa?

- **Facilitador.** És la persona encarregada de dirigir i moderar la sessió.

- **Usuaris.** Són les persones susceptibles d'utilitzar l'aplicació i que comprovaran, mitjançant el test amb llapis i paper, el funcionament correcte del producte o servei.

- **Observadors.** Encara que no sempre serà això, és possible que altres persones del departament o els clients mateixos vulguin observar la sessió i prendre notes. En aquest cas, es recomana que l'usuari no sigui conscient d'aquesta observació i que, per tant, els observadors s'ubiquin en una altra sala.

Avantatges del llapis i paper

Els tests de llapis i paper són relativament fàcils de dur a terme; a més, no cal que estiguin maquetats o programats, per tant, són més econòmics que altres tipus de tests.

Amb aquest tipus de test és fàcil detectar problemes d'ús des de les fases inicials del projecte. No necessitem programari addicional per a fer el test. Amb llapis i paper podem tant crear com corregir els prototips.

Inconvenients dels prototips de paper

En ser un prototip senzill, no es pot utilitzar per a avaluar disseny.

Amb prototips fets amb llapis i paper només es pot simular la resposta del sistema i apropar-se a una situació real, però mai projectar-la tal com és.

En els prototips de llapis i paper de vegades no es tenen en compte criteris com la mida i la forma del producte, per la qual cosa poden portar a dissenys que després no es puguin aplicar. En aquest tipus de test no és important valorar aquesta part, ja que sobre paper acostumem a dibuixar interfícies irreals amb mides de vegades exagerades i amb combinacions d'elements difícils. No ens hem d'imaginar la interfície en el sentit més literal. Podem buscar suport en mesures, dimensions i mides properes a la realitat d'una pantalla.

3. Altres mètodes d'avaluació amb usuaris

Hi ha altres mètodes d'avaluació en què participen els usuaris potencials d'un producte interactiu.

Els mètodes d'avaluació, en termes generals, es poden classificar de maneres molt diverses, depenent de si es tenen en compte uns elements o uns altres com, per exemple: qui du a terme la metodologia, el cost, el temps, el tipus de dades que s'obtenen, etc.

Aquí utilitzarem la classificació següent, que té en compte el tipus d'objectiu que es busca amb l'aplicació de cada mètode i l'avaluador que l'aplica:

a) Mètodes de revisió[13]. Agrupen els mètodes que serveixen per a revisar les característiques del producte o servei que impacten en l'experiència d'ús o en l'accessibilitat. En general, el treball de revisió el duen a terme experts en disseny centrat en l'usuari. L'avaluació heurística seria un tipus de mètode d'inspecció.

(13)Apareixen sovint en la bibliografia sobre el tema com a *inspection methods* o *mètodes d'inspecció*.

b) Mètodes de recopilació[14]. Són els mètodes que busquen recopilar dades qualitatives o quantitatives sobre l'experiència d'ús del producte o servei. En aquests mètodes el més comú és fer participar l'usuari final i estan orientades a saber qui és aquest usuari i què necessita del nostre producte o servei.

(14)També apareixen sovint en la bibliografia sobre el tema com a *inquiry methods*.

c) Mètodes de test. Mètodes d'anàlisi que busquen corroborar amb l'usuari que el producte o servei satisfà les necessitats i expectatives. Freqüentment, aquests mètodes defineixen tasques perquè aquest usuari les faci i s'analitza fins a quin punt això li resulta satisfactori. Serien les metodologies de què hem parlat en els apartats anteriors.

Atès que aquest mòdul està dedicat als mètodes d'avaluació amb usuaris, ens centrarem a descriure breument els de recopilació i test, que són els que fan participar l'usuari final per a poder aplicar-les.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Nom:
Tipologia:

Descripció

Moment d'aplicació:

☐ **Presencial** / ☐ **Remot**
☐ **Quantitatiu** / ☐ **Qualitatiu**
Cost:
☐ Baix ☐ Mitjà ☐ Alt
Temps:
☐ Curt ☐ Mitjà ☐ Llarg
Maquinari/programari:
☐ Necessari

Versió imprimible

3.1. Indagació contextual

Beyer i Holtzblatt (1998) van descriure un conjunt de tècniques centrades en l'usuari el propòsit de les quals era ajudar els equips de disseny i desenvolupament a prendre decisions que s'adiguin amb les necessitats del primer.

La indagació contextual[15] va ser una d'aquestes tècniques i s'ha incorporat com un mètode d'entrevista de camp, basat en uns principis bàsics que la diferencien d'altres mètodes tradicionals d'entrevistes.

(15) En anglès, *contextual inquiry*.

La **indagació contextual** és un estudi de camp en què no s'analitza l'usuari sinó que s'aprèn d'ell. Es tracta de visitar l'usuari en el seu lloc de treball, a casa seva o en l'espai físic on efectuï les seves activitats rutinàries, per a analitzar, observar i aprendre els seus hàbits, activitats, característiques i factors d'entorn.

Recollir els detalls i les motivacions implícites de les persones quan són davant els nostres productes o serveis interactius permet traspassar als dissenyadors part de les necessitats reals dels usuaris i aconseguir un coneixement compartit molt valuós en la seva feina.

La persona que du a terme la investigació de camp ha de recopilar informació sobre l'usuari mitjançant l'observació i l'entrevista, mentre l'usuari fa la seva feina, deixant que aquest se senti còmode per a arribar a entendre la situació de la millor manera possible.

En la investigació de camp és molt important crear un bon clima perquè l'usuari se senti còmode amb l'entrevistador, mantenir una relació de confiança entre tots dos, fins i tot deixar que l'usuari lideri la situació i l'entrevistador aprengui d'aquest, ja que, tal com hem dit, es tracta d'aprendre de l'usuari.

Per a dur a terme l'entrevista de camp és essencial tenir prèviament un objectiu clar i definit. És recomanable no prendre notes davant de l'usuari perquè aquest no se senti avaluat. Podem portar amb nosaltres una màquina enregistradora i prendre els apunts necessaris en finalitzar l'entrevista.

Els principis bàsics de l'entrevista de camp són els següents: context, treball col·laboratiu[16], interpretació i objectiu.

(16)En anglès, *partnership*.

1) Context

L'observació dels usuaris en el seu entorn permet obtenir informació precisa sobre les seves necessitats. És necessari concretar descripcions i dades sobre els usuaris i el seu entorn per a evitar que aquestes siguin abstractes.

L'entrevista s'efectua al lloc de treball on l'usuari utilitza el producte que s'avalua.

Exemple

Si, per exemple, avaluem un campus virtual, anirem a casa de l'usuari:

USUARI: "Generalment m'arriben els exercicis per correu electrònic."
ENTREVISTADOR: "En tens cap? El podem mirar junts?"

2) Treball col·laboratiu

L'objectiu del treball col·laboratiu és que l'usuari i l'entrevistador creïn un vincle d'unió i col·laborin per entendre el treball que s'ha de dur a terme.

Per a això, és necessari canviar l'estructura típica d'entrevista i que els usuaris percebin que realment interessa el treball que fan amb el producte avaluat.

Consells:

- No hem de fer únicament una entrevista convencional, per tant, hem d'evitar el model pregunta-resposta.
- L'entrevistador tampoc no hi acudeix per respondre preguntes, sinó per aprendre de l'usuari.
- L'entrevistador ha d'evitar sentir-se un convidat. Al contrari, s'ha de sentir un company.

3) Interpretació

En la investigació de camp, és fonamental saber distingir entre les dades, les interpretacions de les dades i les interpretacions que comporten aquestes dades. Per a assegurar-nos de distingir correctament aquests conceptes, hem de compartir les interpretacions que elaborem durant l'entrevista amb els usuaris.

4) Objectiu

És imprescindible tenir un objectiu de la investigació clar, detectat i ben definit per a evitar utilitzar únicament un qüestionari. Ens ajudarà a gestionar el temps de l'entrevista.

D'altra banda, tenir un objectiu clau ens ajudarà a entendre'ns millor amb l'usuari i a saber dirigir l'entrevista adequadament.

Fases de l'entrevista

a) Entrevista convencional (15 minuts)

Introducció i objectiu de l'entrevista per a anar coneixent l'usuari i aquest, al seu torn, l'entrevistador.

b) Transició (30 segons)

Expliquem les normes de la investigació contextual i donem les indicacions a l'usuari perquè faci la seva feina mentre l'entrevistador l'observa. Abans que comenci amb les tasques que li corresponen, és important aclarir que podem interrompre l'usuari quan ho considerem necessari.

c) Entrevista contextual (1-2 hores)

L'usuari comença amb les seves tasques mentre l'entrevistador observa i interpreta.

d) Contrast (15 minuts)

L'entrevistador proposa un breu resum del que ha entès amb l'objectiu de confirmar la seva percepció i contrastar-la amb l'usuari.

3.2. Entrevistes

Les **entrevistes** són una tècnica mitjançant la qual els experts formulen qüestions analitzades prèviament per investigar amb usuaris reals, amb l'objectiu de conèixer-ne les reaccions. En ser una tècnica flexible, permet adquirir una gran quantitat de coneixements sobre l'usuari i el seu comportament.

Aquest tipus de tècnica comporta treballar amb usuaris reals *in situ*, de manera presencial, a més de disposar d'un moderador, que és la persona experta que s'encarrega de dur a terme l'entrevista i que alhora fomenta el diàleg i crea empatia amb els usuaris.

Les entrevistes poden ser estructurades o obertes. En les primeres, l'entrevistador segueix el guió preestablert, per tant, el procés és més rígid. L'entrevista oberta és més flexible i deixa espai a la improvisació.

Durant l'entrevista es recomana enregistrar l'àudio i no prendre notes, ja que d'aquesta manera l'usuari no es distreu i l'entrevistador es pot centrar en l'entrevista. És aconsellable començar amb les preguntes menys complexes i restar neutral envers l'usuari, sense tractar de justificar cap decisió de disseny o de concepte. En el cas de les preguntes obertes, aquestes han d'ajudar l'usuari a respondre amb frases concretes (evitar els *sí* i *no*).

3.3. Enquestes

Les **enquestes** són en una tècnica en què es formulen una sèrie de qüestions relacionades amb el producte analitzat a fi d'obtenir informació d'interès per al seu estudi.

La diferència entre les enquestes i les entrevistes és que aquestes últimes són més enriquidores i interactives.

3.4. Qüestionaris

Els **qüestionaris** són una tècnica exploratòria basada en la recollida d'informació mitjançant les respostes que proporcionen els usuaris. Aquesta informació se sol obtenir d'una llista de preguntes escrites que l'usuari ha de contestar, per tant, han de ser tan clares i objectives com sigui possible.

Es tracta de descobrir com els usuaris utilitzen un producte interactiu i les seves preferències. En no ser una tècnica en què s'avalua la interfície *in situ*, sinó que més aviat es recullen les opinions dels usuaris sobre aquest producte, és molt important tenir clares les preguntes que es formularan, i requereix més esforç per part de l'usuari pel fet d'haver de respondre clarament les preguntes.

Si bé és possible crear un qüestionari des de zero amb preguntes adaptades al producte o servei que avaluem, hi ha alguns qüestionaris estandarditzats; aquests qüestionaris ja tenen preestablertes les preguntes i les mètriques que s'empren per a analitzar-los, i el que és més important, a causa de la utilització que se n'ha fet al llarg del temps, s'ha comprovat la utilitat i l'interès que tenen. Per això alguns dels qüestionaris estandarditzats són de pagament.

A continuació, es mostren alguns exemples de qüestionaris estandarditzats. Cada qüestionari estandarditzat avalua diferents aspectes de la utilitat, l'ús i la usabilitat de productes interactius, especialment de productes de programari.

El SUMI, el MUMMS i el WAMMI van ser creats per l'Human Factors Research Group de la University College de Cork. S'ocupen d'avaluar, respectivament, programari, productes multimèdia i pàgines web.

El SUS va ser creat per la Digital Equipment Corporation i se centra fonamentalment en l'avaluació de la usabilitat percebuda (aspectes subjectius de la usabilitat), però sobre qualsevol sistema; és, per tant, més genèric.

3.4.1. *Software usability measurement inventory* (SUMI)

El SUMI[17] és un qüestionari utilitzat per a avaluar la qualitat d'un conjunt de programari des del punt de vista de l'usuari final. Aquest qüestionari es pot utilitzar per a avaluar nous productes, efectuar comparacions amb versions prèvies i establir objectius per a desenvolupaments futurs. Consisteix en cinquanta punts a què l'usuari ha de respondre: "D'acord", "No ho sé" o "En desacord".

(17)SUMI: *software usability measurement inventory.*

Exemple de qüestionari SUMI

Un exemple de qüestionari SUMI en anglès en línia:

- Software usability measurement inventory

Alguns exemples de preguntes

"1. Aquest programari respon massa a poc a poc a les entrades que se li proporcionen."

"3. Les instruccions i els advertiments són d'ajuda."

"13. La manera de presentar la informació del sistema és clara i comprensible."

"22. No m'agradaria haver d'utilitzar aquest programari cada dia."

El nombre d'usuaris necessaris per a obtenir resultats interessants oscil·la entre 12 i 30.

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

El qüestionari MUMMS[18] té el mateix objectiu que el SUMI, però pretén aplicar-se al programari multimèdia; es caracteritza per la consideració de cinc subescales:

(18)MUMMS: *measuring the usability of multi-media systems.*

- La mesura en què el producte capta les respostes emocionals de l'usuari.
- El grau de control amb què l'usuari sent que ell, i no el producte, procedeix pas a pas.
- El grau d'eficiència amb què l'usuari sent que pot aconseguir els objectius de la seva interacció amb el producte.

- El nivell d'ajuda i assistència que el producte sembla proporcionar a l'usuari.

- La facilitat d'aprenentatge amb què un usuari pot començar i aprendre noves característiques del producte.

3.4.3. *Website analysis and measurement inventory* (WAMMI)

WAMMI[19], en canvi, se centra en l'avaluació dels llocs web. Es basa en el qüestionari que emplenen els usuaris, que proporciona una mesura de la utilitat i facilitat d'ús que han trobar en el lloc en qüestió. Es pot utilitzar de tres maneres:

(19)WAMMI: *website analysis and measurement inventory.*

- Predicció: abans del llançament d'un nou lloc, un test amb WAMMI pot donar una orientació sobre la reacció dels visitants.

- Monitoratge: en aquest cas, interessa conèixer les raons de certs comportaments, per què alguns clients repeteixen, per què d'altres no tornen mai, etc.

- Referència: interessa conèixer l'opinió de la concurrència amb relació a altres llocs web, fonamentalment per qüestions de disseny.

3.4.4. *System usability scale* (SUS)

El qüestionari SUS[20] va ser desenvolupat el 1986 per la Digital Equipment Corporation com a part de l'aplicació d'enginyeria d'usabilitat als sistemes d'oficina. Recordem que en la dècada dels vuitanta era una novetat l'ús d'ordinadors a les oficines; es parlava fins i tot de l'"oficina electrònica", en contrast amb la tradicional.

(20)SUS: *system usability scale.*

El seu propòsit era proporcionar un test fàcil de completar (nombre mínim de qüestions), fàcil de puntuar i que permetés establir comparacions creuades entre productes; és a dir, que fos possible comparar unes solucions amb d'altres i optar per la millor.

Història: com es va crear l'escala

L'escala SUS es va construir sobre un conjunt original de 50 preguntes. Es van triar dos sistemes, un de molt senzill d'usar i un altre de gairebé impossible, fins i tot per als tècnics; 20 persones, amb perfils des d'administratiu fins a programador, van utilitzar els dos sistemes i van respondre les 50 preguntes amb una puntuació d'1 a 5 entre "completament d'acord" i "completament en desacord".

Es van seleccionar les preguntes que van rebre respostes més extremes i amb correlació més estreta (és a dir, que funcionaven juntes: si una tenia una resposta de 5 l'altra en tenia una de 4 o 5). D'aquesta manera, es va reduir de 50 preguntes a les 10 actuals:

	Completament d'acord				**Completament en desacord**
1. Crec que m'agradarà usar sovint aquest sistema	1	2	3	4	5
2. Vaig trobar el sistema innecessàriament complex	1	2	3	4	5
3. Vaig pensar que el sistema era fàcil d'utilitzar	1	2	3	4	5
4. Crec que necessitaria el suport d'un expert per a recórrer el sistema	1	2	3	4	5
5. Vaig trobar les diverses possibilitats del sistema bastant ben integrades	1	2	3	4	5
6. Vaig pensar que hi havia massa inconsistència en el sistema	1	2	3	4	5
7. Imagino que la majoria de les persones aprendrien molt ràpidament a utilitzar el sistema	1	2	3	4	5
8. Vaig trobar el sistema molt incòmode d'usar	1	2	3	4	5
9. Em vaig sentir molt confiat en el maneig del sistema	1	2	3	4	5
10. Necessito aprendre moltes coses abans de manegar amb comoditat el sistema	1	2	3	4	5

L'escala SUS és de domini públic i d'ús lliure, sempre que se'n citi adequadament la font: © Digital Equipment Corporation, 1986.

Com utilitzar l'escala SUS

L'escala SUS s'utilitza generalment immediatament després que l'usuari hagi provat el sistema. Se sol·licitarà als usuaris el registre immediat de la seva resposta a cada punt, sense deixar que hi pensi llargament. No s'ha de deixar cap pregunta per contestar.

Si l'usuari no se sent capaç de respondre alguna qüestió en particular, haurà d'assenyalar el valor central de l'escala (3).

Puntuació de l'escala SUS i interpretació de resultats

L'escala SUS és una escala d'estil Likert el resultat final de la qual és un valor únic que representa una mesura composta de la usabilitat del sistema global sotmès a estudi. És a dir, que en acabar de processar les dades, tindrem un únic valor que pot ser, posem per cas, 60 o 35, com la nota final d'un examen. Això és el que permet que puguem comparar fàcilment sistemes avaluats mitjançant l'escala SUS.

Escala Likert

"L'escala de tipus Likert és una escala psicomètrica comunament utilitzada en qüestionaris, i és l'escala que es fa servir més àmpliament en enquestes per a la recerca. Quan responem un element d'un qüestionari elaborat amb la tècnica de Likert, ho fem especificant el nivell d'acord o desacord amb una declaració (element, ítem o reactiu). L'escala es diu així per Rensis Likert, que va publicar un informe que en descrivia l'ús."

Wikipedia

Cal assenyalar que les puntuacions independents no són significatives per si mateixes. És a dir, que si a la primera pregunta hem tingut una resposta massiva de "completament d'acord", no és una dada significativa mentre estigui aïllada de la resta.

Per a calcular la puntuació del SUS, cal sumar primer les contribucions de cada punt. La contribució de cada punt valdrà entre 0 i 4:

- Per als punts 1, 3, 5, 7 i 9, la contribució serà la posició de l'escala menys 1.
- Per als punts 2, 4, 6, 8 i 10, la contribució serà de 5 menys la posició en l'escala.

Finalment, es multiplica la suma dels resultats per 2,5 per a obtenir el valor global del SUS. El resultat estarà entre 0 i 100.

Exemple de puntuació de l'escala SUS

Vegem un exemple per al càlcul de la puntuació:

1. Crec que m'agradarà visitar sovint aquest sistema

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☒

Completament en desacord — Completament d'acord

Recordem-ho: per a aquest punt, el càlcul és la posició en l'escala menys 1. La puntuació serà 5 – 1 = 4.

2. Vaig trobar el sistema innecessàriament complex.

1 ☐ 2 ☐ 3 ☐ 4 ☒ 5 ☐

Completament en desacord — Completament d'acord

Recordem-ho: per a aquest punt, el càlcul són 5 menys la posició en l'escala. La puntuació serà 5 – 4 = 1.

3. Vaig pensar que era fàcil utilitzar el sistema.

1 ☐ 2 ☒ 3 ☐ 4 ☐ 5 ☐

Completament en desacord — Completament d'acord

La puntuació serà 2 – 1 = 1.

4. Crec que necessitaria el suport d'un expert per a recórrer el sistema.

1 [X] 2 [ ] 3 [ ] 4 [ ] 5 [ ]

Completament en desacord Completament d'acord

La puntuació donarà 5 – 1 = 4.

5. Vaig trobar les diverses possibilitats del sistema bastant ben integrades.

1 [ ] 2 [X] 3 [ ] 4 [ ] 5 [ ]

Completament en desacord Completament d'acord

La puntuació serà 2 – 1 = 1.

6. Vaig pensar que hi havia massa inconsistència en el sistema.

1 [ ] 2 [ ] 3 [X] 4 [ ] 5 [ ]

Completament en desacord Completament d'acord

La puntuació serà 5 – 3 = 2.

7. Imagino que la majoria de les persones aprendrien molt ràpidament a utilitzar el sistema.

1 [ ] 2 [X] 3 [ ] 4 [ ] 5 [ ]

Completament en desacord Completament d'acord

La puntuació serà 2 – 1 = 1.

8. Vaig trobar el sistema molt gran en recórrer-lo.

1 [ ] 2 [ ] 3 [ ] 4 [X] 5 [ ]

Completament en desacord Completament d'acord

La puntuació serà 5 – 4 = 1.

9. Em vaig sentir molt confiat en el maneig del sistema.

1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [X]

Completament en desacord Completament d'acord

La puntuació serà 5 – 1 = 4.

10. Necessito aprendre moltes coses abans de manegar amb comoditat el sistema.

1 [ ] 2 [X] 3 [ ] 4 [ ] 5 [ ]

Completament en desacord Completament d'acord

La puntuació serà 5 – 2 = 3.

Els resultats totals: 4 + 1 + 1 + 4 + 1 + 2 + 1 + 1 + 4 + 3 = 22.

Multipliquem per 2,5. La puntuació SUS: 22 · 2,5 = 55. Per tant, la puntuació en l'escala SUS que rep aquest sistema és de 55.

3.4.5. Altres qüestionaris estandarditzats

- ASQ (*after-scenario questionnaire*)

- CSUQ (*computer system usability questionnaire*)
- CUSI (*computer user satisfaction inventory*)
- *Ergonorm questionnaire*
- *IsoMetrics*
- ISONORM 9241/10
- PSSUQ (*poststudy 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. Dinàmiques de grup (*focus groups*)

La metodologia de les dinàmiques de grup[21] no neix de la usabilitat, sinó que és una eina comuna en àmbits com la sociologia o el màrqueting.

(21)En anglès, *focus group*.

> Es tracta de reunir un grup reduït de persones, habitualment de sis a dotze, amb l'objectiu de conèixer-ne les impressions, opinions, millores, reaccions, etc., davant d'un producte o servei o davant d'una situació. Sempre hi ha un moderador que s'encarrega de dirigir la sessió i canalitzar el debat cap a aspectes rellevants.

En usabilitat s'apliquen les dinàmiques de grup per a discutir aspectes del producte o servei que s'analitza. Habitualment, el paper de moderador l'exerceix un expert en usabilitat i dirigeix la discussió cap a uns objectius que l'equip de disseny i desenvolupament ha marcat inicialment. Aquest és un aspecte que no és fàcil de controlar, ja que el moderador ha d'obtenir respostes i alhora deixar que la conversa flueixi entre els participants sense dirigir-los excessivament.

És important que la metodologia inclogui més d'una sessió amb diferents grups i que els participants siguin significatius en la mesura en què podrien ser usuaris reals del servei o usuaris potencials.

Respecte a l'anàlisi de les dades generades, les sessions de dinàmica de grup s'han de plantejar com una metodologia creativa, i no analítica, ja que generen dades qualitatives que es poden interpretar de diverses maneres. És a dir, la dinàmica serveix, per exemple, per a detectar els punts més importants que hem de millorar en el nostre web abans d'iniciar un procés de DCU, però no per a avaluar si aquest és usable o no.

3.6. Classificació de targetes (*card sorting*)

La **classificació de targetes** (***card sorting***) és una tècnica per a explorar i observar com els usuaris ordenen i classifiquen les categories d'informació d'un lloc web, producte o servei.

Aquesta tècnica és emprada en l'arquitectura de la informació i ajuda a augmentar la probabilitat que els usuaris finals del producte o servei siguin capaços de trobar el que busquen conforme al seu model mental.

Destaquem dues tipologies de classificació de targetes:

1) Oberta: els usuaris reben categories sense cap tipus de grup establert, per tant, poden agrupar les categories lliurement com els sembli millor.

2) Tancada: els usuaris reben categories amb grups definits prèviament, amb la qual cosa han d'ubicar cada categoria en el grup que considerin convenient.

És recomanable utilitzar aquesta tipologia de classificació de targetes quan afegim nous continguts en un lloc web en una estructura que ja existeix.

Enllaços d'interès

Donna Spencer; Todd Warfel (2004) "Card sorting: a definitive guide" [article en línia]. *Boxes and arrows* (7 d'abril).

"Card sorting" [article en línia]. *Information&Design*

Yusef Hassan Montero; Francisco J. Martín Fernández (2004). "Card Sorting: Técnica de categorización de contenidos" [article en línia]. *No Solo Usabilidad* (23 de març).

3.7. Registres de seguiment d'usuari

Els registres de seguiment d'usuari[22] formen part de les tècniques d'avaluació remota amb usuaris. Els usuaris mateixos són els qui creen un registre de seguiment d'accions, reaccions, elements que els criden l'atenció del producte que s'avalua.

(22)En anglès, *self-reporting logs*.

Els usuaris efectuen una prova d'usabilitat amb el producte seguint un guió preestablert i enregistrant totes les accions que han dut a terme per completar cada tasca. Aquesta informació inclou des de l'enllaç que han preferit utilitzar per a accedir a una informació determinada, fins a les dades amb què han emplenat un formulari.

La diferència amb altres tècniques és que l'usuari participant fa la tasca d'enregistrament del que passa durant el test. En els tests remots pot haver-hi un sistema automàtic el que recull el que passa.

Tanmateix, aquesta tècnica presenta debilitats que fan que les dades resultants puguin tenir una fiabilitat baixa:

- Dependència excessiva dels usuaris de la prova per a recollir les dades sobre allò que passa.

- És possible que, conscientment o inconscientment, els participants no reflecteixin objectivament cada pas.

- Es perd tota la informació sobre les expressions del rostre de l'usuari i el llenguatge no verbal.

En una variant d'aquesta metodologia, l'usuari enregistra, no de manera textual sinó mitjançant captures de pantalla, allò que passa en cada moment. Es denomina *screen snapshots*.

3.8. Sessions guiades (*journaled sessions*)

Les sessions guiades[23] van ser el germen dels tests remots d'usabilitat actuals. Per tant, formen part de les tècniques remotes d'avaluació amb usuaris.

(23)En anglès, *journaled sessions*.

Les **sessions guiades** consisteixen a fer arribar als usuaris del test un prototip del producte que inclou un sistema que en captura les accions. Per exemple, un CD amb una versió beta d'un web que incorpora algun sistema d'enregistrament. També és necessari proporcionar als usuaris un guió de les tasques que han de dur a terme.

Els usuaris fan un test convencional amb el producte i les seves accions queden enregistrades. Després enregistren un nou CD i el fan arribar a l'equip avaluador.

Amb l'avenç tecnològic que permet a usuaris i avaluadors interactuar en línia amb el producte i aconseguir un registre minuciós de les accions, aquest mètode ha estat substituït pel test remot o l'anàlisi de registres informàtics. En aquest sentit, les sessions guiades presenten els mateixos inconvenients metodològics que els tests remots però amb una gestió més complexa.

En qualsevol cas, podria ser una opció interessant per a l'avaluació de sistemes no interactius en els quals no fos possible el desplaçament de l'usuari a un laboratori. L'enregistrament de les accions de l'usuari mitjançant programari hauria de ser substituït per un altre mètode d'enregistrament, per exemple, l'enregistrament en vídeo.

3.9. Registre informàtic de seguiment (*logging*)

Un *log* és un registre d'esdeveniments d'un sistema, entenent per *esdeveniment* les dades o la informació referides a quin usuari, què ha fet, on ho ha fet, quan ho ha fet i, finalment, per què el sistema ha reaccionat així. Aquestes dades s'emmagatzemen en fitxers de text formats per les peticions i els resultats.

A priori, el *log* és un registre utilitzat en seguretat i administració de sistemes informàtics, però que pot tenir un gran valor per a avaluar el comportament dels usuaris.

Un format habitual de *log* web és el *common logfile format* i l'*extended log file format*, i pot tenir aquest aspecte:

Lectura complementària

Adaptat de:
Jorge Sánchez Sánchez. *Análisis de accesos a un servidor web de contenidos dinámicos*. Projecte de final de carrera. Universitat de les Illes Balears.

```
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)"
```

Aquesta informació es pot interpretar de la manera següent:

Camp	Contingut	Comentaris
Adreça IP del client	81.37.129.172	
Adreça del servidor	www.sanostra.es	
Nom d'usuari		L'usuari no està autenticat en el servidor
Data i hora	[10/Aug/2004 08:25:24 -0100]	
Petició/Recurs	GET homeweb.nsf/fwHo-me?ReadForm&lang=02HTTP/1.1	Pàgina, sol·licitada pel mètode GET utilitzant el protocol HTTP/1.1
Codi de resposta (estatus)	200	OK (petició correcta)
Mida de la resposta	20266	Mida en bytes
Referer	http: //www.google.es/search?hl=es&q=SA +NOSTRA	Enllaç des del cercador Google
User agent	Opera/8.5 (Macintosh; PPC Mac OS X; U; en)	Navegador Opera (versió 8.5) sobre un ordinador PowerPC i sistema operatiu Mac OS X de Macintosh

L'anàlisi d'aquestes dades des d'una perspectiva d'avaluació d'usabilitat pot ajudar a entendre, per exemple:

- Quin és el comportament habitual dels usuaris, quines parts del nostre web o quines eines del nostre sistema utilitzen més. Aquesta informació és fonamental, per exemple, per a detectar quins elements del sistema són més populars i quins ho són menys, i investigar-ne el perquè.

- Quins problemes es troben en utilitzar el sistema, i quan arriben a una resposta KO, per què passa.

El servidor web no és l'única manera de generar *logs*, sinó que es poden instal·lar programes que recullin les activitats dels usuaris en els seus propis ordinadors.

La metodologia de *logging* té l'avantatge que pot recollir una quantitat ingent de dades a un cost molt baix i de manera remota. L'esforç real del *logging* és d'interpretació d'aquest volum de dades.

4. Anàlisi de resultats

Després de fer la prova d'usabilitat ens trobem amb una gran quantitat de dades de naturalesa diversa a les nostres mans; i ara, què? Arriba el moment de fer-ne un ús correcte sense prendre-les com a mesures úniques i absolutes que donaran resposta immediata a l'avaluació.

Els resultats aporten més informació al nostre treball, ens orienten perquè cada dia corregim el rumb i milloren les solucions que anem proposant. Però hem de pretendre resoldre de manera automàtica qüestions que us preocupen o arribar a conclusions precipitades sense tenir una visió àmplia des de diverses perspectives. Potser aquest és l'error més habitual de les proves d'usabilitat. Per això, una de les parts més importants de les proves consisteix a completar una anàlisi responsable de resultats.

A continuació es presenten les etapes per a analitzar resultats, que poden diferir depenent del projecte d'avaluació i dels mètodes utilitzats. En qualsevol cas, l'objectiu d'aquestes etapes és establir l'ordre que s'ha de seguir per a transformar les dades i la informació en resultats que després es puguin utilitzar en les presentacions als clients o en l'equip de treball d'un producte interactiu.

Les etapes per a analitzar els resultats són:

- Organització de la informació
- Tractament de les dades
- Anàlisi de les dades

4.1. Organització de la informació

La primera etapa en l'anàlisi de resultats és l'organització de la informació. Per a això cal recopilar i organitzar tota la informació que hem recollit durant l'avaluació:

- En un test amb usuaris, per a cada tasca, els temps dedicats, els resultats obtinguts, els comentaris dels usuaris, etc.

- En una dinàmica de grup, els comentaris dels participants.

- En un *card sorting*, les organitzacions i les ordenacions que han fet els participants, les targetes o les categories que s'han rebutjat, comentaris dels participants, fotos, etc.

La informació recollida es pot organitzar de diferents maneres depenent dels objectius finals.

4.1.1. Organització sobre un criteri preestablert

Es tracta d'organitzar les dades segons un criteri que decidim prèviament. Aquest criteri permetrà crear unitats d'informació més manejables per a ferne l'anàlisi.

Com es tria el criteri?

Per exemple, què és més productiu, organitzar les dades d'una dinàmica de grup per comentaris de participant o per temes tractats? o, en el cas d'un test, per usuari o per tasca?

Si en la fase de definició d'objectius de l'avaluació hem definit les característiques del producte o servei que volem avaluar, també podem organitzar les dades en funció d'aquestes característiques.

No hi ha una opció ideal o definitiva per a organitzar els resultats. Depenent del temps disponible per a fer l'anàlisi, fins i tot podem optar per diverses organitzacions de dades. Això té un avantatge, ja que és com mirar el mateix paisatge des de dues finestres diferents: la perspectiva canvia i se'n veuen més detalls. Lamentablement, la situació més habitual és que tenim un temps limitat per a organitzar les nostres dades.

Depèn...

La paraula *depèn* és una paraula constant en l'exercici del disseny centrat en l'usuari. Correspon al percentatge d'alternatives a què cada professional està subjecte per la definició, característiques, circumstàncies, context, usuaris, etc., de cada producte o servei creat.

Les perspectives d'organització que solen ser més útils són les que giren a l'entorn de les tasques o també l'entorn de les característiques del producte o sistema. En canvi, no es recomana una organització de la informació que se centri en l'usuari, per la senzilla raó que no avaluem l'usuari, sinó el producte o servei.

Una vegada establert el criteri que seguirem, podem tenir una organització de dades similar a aquest exemple, que s'organitza seguint les característiques del sistema.

Característiques (criteri)	Descripció	Usuari	Temps	Comentaris
Navegació	L'usuari no ha pogut trobar la informació que buscava.	Usuari 3	0.03.35	L'usuari ha provat diverses opcions per trobar la informació que volia i ha acabat desistint.
Navegació	L'usuari ha encertat a la primera l'opció correcta.	Usuari 1	0.00.20	L'usuari ha declarat que l'opció tenia la mateixa etiqueta que en un altre sistema conegut.

4.1.2. Organització sense un criteri preestablert

Aquest tipus d'organització es basa a agrupar les dades en funció del que ha passat durant la prova.

En primer lloc, s'han de posar en comú els problemes i els aspectes més importants que s'han trobat durant el test amb usuaris. Després, s'agrupen per similitud i s'etiqueten per categories. Finalment, s'apliquen aquestes categories a les dades, i s'estructura tota la informació en una taula similar a la de l'exemple anterior.

És possible que amb aquesta organització, a simple vista, no es puguin detectar amb facilitat els problemes d'usabilitat, per la qual cosa també podem tractar i analitzar les dades aplicant mètriques abans d'organitzar-les d'aquesta manera.

4.2. Tractament de les dades: mètriques

En aquest punt, és interessant recordar Nielsen (2003), que defineix la usabilitat com un atribut de qualitat que té diverses variables a partir de les quals es pot mesurar:

Lectura recomanada

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

- Facilitat d'aprenentatge[24]
- Eficiència d'ús[25]
- Facilitat per a recordar[26]
- Errors
- Satisfacció

(24)En anglès, *learnability*.

(25)En anglès, *efficiency*.

(26)En anglès, *memorability*.

La definició ISO 9241-11 també es refereix a criteris mesurables per definir la usabilitat: efectivitat, eficiència, satisfacció. D'aquesta manera, és lògic que en l'anàlisi dels resultats d'un procés d'avaluació de la usabilitat es treballin els indicadors mesurables o mètriques.

ISO 9241-11 (1998)

Recordem que la ISO 9241-11 (1998) defineix la usabilitat així:

"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."

A continuació presentem una llista amb mètriques senzilles que podem utilitzar en un procés d'avaluació:

- Percentatge d'èxit o fracàs per tasca dels usuaris: literalment, quants usuaris han estat capaços de fer la tasca correctament (efectivitat).
- Temps per tasca.
- Temps per tasca amb relació a l'èxit o fracàs (eficiència).
- Nombre d'errors comesos pels usuaris.
- Satisfacció subjectiva dels usuaris, que normalment haurem recollit mitjançant algun qüestionari o entrevista final.

4.2.1. Eficiència i efectivitat

Eficiència i efectivitat són dos criteris que mereixen una anàlisi independent, però vegem alguns aspectes importants d'aquests conceptes. Primerament hem de ser conscients que són mètriques que poden canviar de components entre una prova i una altra. Què significa això?

L'eficiència és un concepte que generalment associem a l'estalvi: un cotxe eficient és el que necessita menys combustible per a cobrir una distància determinada. En el disseny centrat en l'usuari també té a veure amb els recursos que l'usuari ha de "consumir" durant la interacció amb el sistema: temps, esforç mental, diners, intents, feina, etc. Alguns exemples de dades que poden compondre la mesura d'eficiència són:

Lectura complementària

Exemples adaptats de:
David Travis (2003). "Discount Usability: Time to Push Back the Pendulum?" [article en línia]. *Userfocus* (2 de maig).

- Temps en cada intent per a resoldre una tasca.
- Temps en les tasques d'una tipologia concreta (per exemple, de registre).
- Temps per a fer una tasca en comparació amb el temps d'un expert.
- Nombre d'etiquetes del sistema apreses.
- Nombre de clics, nombre de pulsacions de teclat.
- Nombre de vegades que s'ha utilitzat la funció "Enrere" del navegador (en proves relacionades amb el web).
- Nombre d'icones que recorda l'usuari després de completar la prova.
- Nombre de consultes al manual d'ús, ajuda en línia, facilitador, etc.

L'efectivitat és la capacitat de l'usuari d'assolir els seus objectius, cosa que en la realitat d'un test d'usabilitat es reflecteix, per exemple, en el nombre de tasques acabades amb èxit. També podem trobar altres dades que componguin una mesura d'eficiència:

- Nombre de funcions del producte o servei utilitzades.
- Nombre de tasques completades en un primer intent.
- Nombre d'errors, errors més repetits.
- Percentatge d'errors per tasca.
- Percentatge d'usuaris que han completat les tasques d'una tipologia concreta (per exemple, un procés de compra).
- Percentatge de tasques resoltes amb ajuda (manuals, assistència en línia, facilitador, etc.).
- Nombre de tasques resoltes sense ajuda.

En el moment de dissenyar la prova haurem definit les dades que recollirem i que posteriorment seran els components de les mètriques d'eficiència i efectivitat.

Per a convertir les dades brutes en mètriques necessitarem cert tractament numèric o estadístic.

4.2.2. Tractament de les dades numèriques

Alguns dels mètodes de disseny centrat en l'usuari impliquen la recollida de resultats quantitatius com, per exemple, el test amb usuaris, enquestes i qüestionaris. Els resultats quantitatius que s'obtenen amb aquests mètodes són, per exemple, el temps per a fer una tasca o el nombre de persones que han respost afirmativament a una pregunta determinada. De vegades, la presentació de les conclusions dels resultats inclou afirmacions com "un 10% dels usuaris han fallat en la tasca".

A continuació tractarem d'aspectes bàsics d'anàlisi estadística orientats als resultats de l'avaluació de la usabilitat, tot presentant les mètriques principals que podem aplicar a les dades quantitatives.

Mitjana, mediana i moda

Tres mètriques molt útils per a ajudar-nos a donar sentit a les dades i a detectar patrons, són la mitjana, la mediana i la moda. Posem, per exemple, els temps per usuari en una tasca i el nombre d'errors comès per cada un.

Tasca 1		
Usuari	Temps	Errors
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
mitjana	0:01:47	5
mediana	0:01:37	3
moda	0:00:50	2

La **mitjana** és un dels conceptes més familiars. Correspon a la mitjana d'un conjunt de nombres. En l'exemple, la mitjana de temps de tots els usuaris és d'1 minut i 47 segons. El problema de la mitjana és que es veu molt afectada pels valors extrems (els més alts i els més baixos).

El nostre exemple conté usuaris que gairebé han fet el doble de temps que la mitjana i usuaris que han fet gairebé la meitat de temps que la mitjana. En el cas del nombre d'errors, la mitjana és de 5, però tenim dos usuaris que no han comès cap error en absolut durant la tasca. És evident que en l'elaboració de la nostra anàlisi no ens podem guiar únicament per la mitjana com a indicatiu.

La **mediana** és el valor que deixa el mateix nombre de resultats per davant i per darrere, és a dir, que hi ha tantes dades per sobre de la mediana com per sota. En el nostre exemple, la mediana de temps és d'1 minut i 37 segons, la qual cosa vol dir que la quantitat d'usuaris que superen aquest temps és exactament la mateixa que la quantitat d'usuaris que han fet un temps inferior.

En aquest cas, la diferència entre la mitjana i la mediana és de 10 segons, la qual cosa és significativa. Com més gran sigui aquesta diferència, més extrems seran els valors la presència dels quals indicarà.

Valors per sobre 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
Valors per sota 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 és el valor més repetit entre tots els que considerem. Curiosament, en el nostre exemple, els dos valors més repetits són un temps de 50 segons per a fer la tasca i 2 errors comesos.

Per a tractar les dades obtingudes en un test amb usuaris, convé conèixer els conceptes de *mitjana*, *mediana* i *moda*, i extreure'n conjuntament els valors per a poder tenir així una aproximació realista del que ha passat.

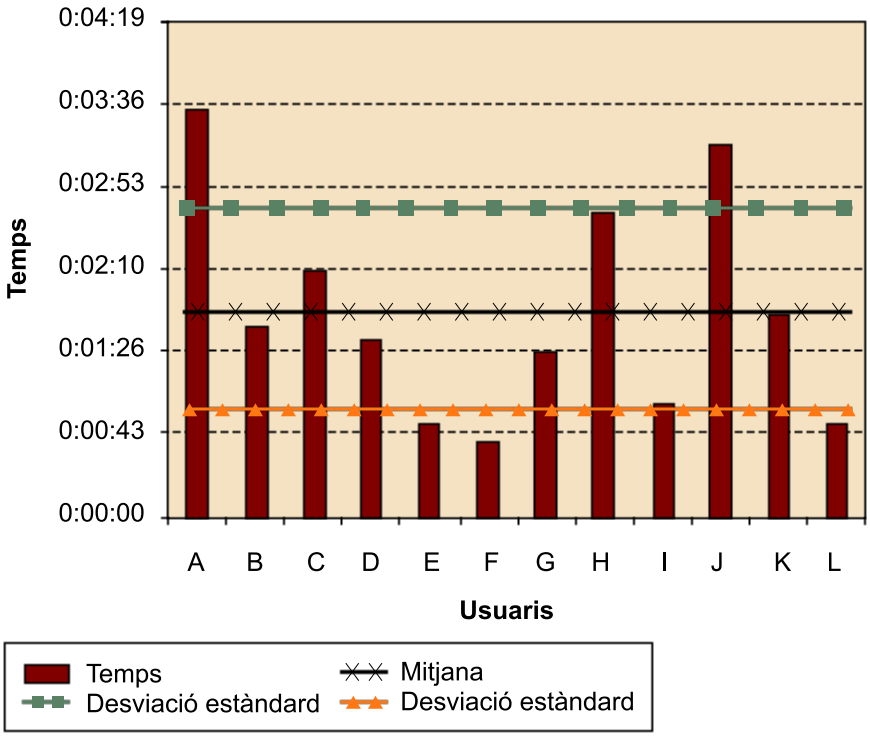
Desviació típica o estàndard

La desviació estàndard és una mesura del grau de dispersió de les dades respecte del valor mitjà. Dit d'una altra manera, la desviació estàndard és simplement la variació esperada respecte de la mitjana (*Wikipedia*).

En el cas de l'avaluació de l'experiència d'ús, les dades de les quals ens interessa saber la disparitat són el resultat de la interacció de l'usuari amb un sistema. Tornant a l'exemple anterior amb els temps i els errors:

Tasca 1		
Usuari	Temps	Errors
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
mitjana	0:01:47	5
mediana	0:01:37	3
moda	0:00:50	2
desviació estàndard	0:00:53	4,546061

La desviació estàndard del temps és de 53 segons. Això vol dir que hi ha un interval de gairebé 1 minut sencer de temps de dispersió entre les dades de temps. En el cas dels errors, la desviació és de més de 4. Aquestes dades ens confirmen la presència de valors extrems que haurem d'analitzar.



Conclusions

L'aplicació i la utilització de mètriques ens ajuda a tractar i interpretar les dades, però hi ha alguns perills:

a) Els nombres són "addictius": allargar indefinidament els càlculs sobre les nostres dades no sempre facilita la tasca d'anàlisi. Cal aplicar les tècniques justes –i mitjana, mediana i moda són bastant útils– perquè, del conjunt de dades inicial sense ordenar, puguem començar a treure'n alguna conclusió útil.

b) Les mètriques que hem presentat pertanyen a l'estadística descriptiva. Ens ajuden a entendre millor el significat de les dades que recollim i s'ajusta als objectius de l'avaluació de la usabilitat mitjançant un test amb usuaris. Això no obstant, si alguna vegada es presenta la necessitat de treure conclusions d'un procés d'avaluació que siguin extrapolables a conjunts més amplis de població, haurem d'utilitzar l'estadística inferencial i extremar el rigor en el disseny de la prova (la mostra, les hipòtesis de treball, les variables, etc.).

Estadística inferencial

La **inferència estadística** o **estadística inferencial** és una part de l'estadística que comprèn els mètodes i els procediments per a deduir propietats (fer inferències) d'una població, a partir d'una petita part de la població (mostra). La bondat d'aquestes deduccions es mesura en termes probabilístics, és a dir, tota inferència s'acompanya de la seva probabilitat d'encert (*Wikipedia*).

En conclusió, l'estadística posa al nostre abast eines que ens poden ajudar a entendre millor les dades d'un test amb usuaris en un procés de disseny centrat en l'usuari. Mitjana, mediana, moda i desviació típica poden ser un senyal que ens estigui indicant "mira aquí!, què ha passat?" i, d'aquesta manera, centri la nostra atenció en problemes d'usabilitat. Durant la fase de conclusions hem de completar les dades numèriques amb altres dades qualitatives per a obtenir així conclusions més significatives i completes.

4.2.3. Mètriques en un test amb usuaris: exemple d'aplicació

A continuació, a tall d'exemple, presentem una llista de dades i mètriques que podríem extreure de la informació recollida en un test amb usuaris:

- Mètriques sobre les dades de tasca.

- Mitjana de temps per a completar la tasca.

- Mitjana de temps per a completar la tasca respecte del temps estimat. En alguns tests, l'equip avaluador fa la tasca anticipadament, se'n registra el temps i se'n comparen els resultats.

- Mediana de temps per a completar la tasca, i nombre d'usuaris per sobre o per sota.

- Desviació estàndard del temps per a completar la tasca.

- Percentatge d'usuaris que han fet la tasca correctament.

- Percentatge d'usuaris que han fet la tasca correctament i en un temps determinat, que podria ser:
  - Temps estimat per l'equip avaluador.
  - Mediana del temps.
  - Desviació típica del temps.
  - Eficiència i efectivitat per usuari i tasca.

- Dades dels qüestionaris: percentatges de resposta a les diferents qüestions.

- Altres dades que podem haver recollit durant el test:
  - Reaccions emocionals dels usuaris durant la prova.
  - Nombre de comentaris positius i negatius.
  - Nombre d'usuaris que han afirmat que recomanaran el producte o servei durant la prova.

4.3. Interpretació de les dades

Una vegada s'han ordenat les dades i s'han obtingut les mètriques, arriba la fase d'interpretació de dades. Es tracta d'analitzar els perquès que hi ha al darrere dels fenòmens observats en cada procés d'avaluació.

Molts professionals de la usabilitat entenen aquesta fase com un art, com un procés que requereix habilitat i experiència. Certament, l'anàlisi vol buscar els patrons de comportament dels usuaris que ajudin a expressar les dades i per a això és necessari un ull entrenat. Tant és així que, a mesura que s'adquireix aquesta experiència, molts professionals de la usabilitat tenen menys necessitat de fer les anàlisis més formals i ja comencen a "veure" aquests patrons durant les sessions de test.

Lectura recomanada

El 2006, la Usability Professionals Association (UPA) va celebrar una sessió durant la seva conferència anual sobre l'anàlisi de resultats en avaluacions d'usabilitat:

Analyzing usability study results: Is it magic behind the curtain?

A continuació proposem una sèrie de recomanacions o direccions vers les quals un analista hauria de dirigir l'atenció. També recuperem el concepte de model mental d'usuari i l'impacte que pot tenir en el moment d'avaluar els resultats d'un test.

4.3.1. Interpretació de les dades

Amb les dades organitzades i les mètriques a la mà, el pas següent és enumerar els problemes d'usabilitat que evidencien les dades. És molt útil fer aquesta anàlisi tant en positiu com en negatiu:

- En negatiu: es tracta d'analitzar totes les dades relatives al fracàs en la resolució de tasques i argumentar el perquè.

- En positiu: es tracta d'analitzar totes les dades relatives a l'èxit en la resolució de tasques i argumentar el perquè.

Algunes recomanacions per a dur a terme l'anàlisi són:

a) Dir i fer

Les dades relacionades amb les tasques fetes pels usuaris són més fiables que les informacions llançades en veu alta durant el test d'usabilitat o després, en l'entrevista posterior o el qüestionari.

Aquesta recomanació té sentit en l'aplicació de mètodes com el test amb usuaris seguint el protocol del pensament manifest, que implica la realització de tasques acompanyades de verbalitzacions. No té sentit en altres mètodes d'avaluació que siguin eminentment verbals, com una dinàmica de grup o una entrevista.

D'alguna manera els comentaris de l'usuari sempre són un suport a l'explicació sobre el que ha passat en el test, però no són una conclusió per si mateixos. Podem prendre en consideració un comentari d'un usuari que expressi la dificultat de la tasca ("que difícil és això") si està acompanyat d'unes dades de tasca que indiquin una eficiència i efectivitat baixes. Si no, estarem donant importància a comentaris que poden provenir, per exemple, per la simple falta

d'atenció o nerviosisme per la prova. Fins i tot, com hem comentat en analitzar el paper del facilitador, l'usuari pot entendre que ha de dir el que el facilitador vol sentir, i exagerar en els comentaris positius o negatius.

És important no oblidar en cap moment que aquests tests es basen en la premissa de no analitzar l'usuari. Quin sentit tindria que baséssim les nostres conclusions únicament en els seus comentaris?

> Els comentaris dels usuaris són la sang a l'escena del crim, però no el cadàver. Podem imaginar els comentaris dels usuaris com les esquitxades de sang. La forma de cada gota ens dóna més pistes sobre com va ocórrer el crim, però no és el crim.

b) Busquem perquès

Analitzant les dades podem veure, per exemple, que l'efectivitat dels usuaris en una tasca determinada ha estat molt baixa, això és, molt pocs usuaris han estat capaços de dur-la a terme. Quina en pot ser la raó?, què havien de fer i què han fet?, per què no es donaven les condicions d'èxit?

En un test amb usuaris fet per a avaluar l'ús d'un forn microones, cap dels usuaris no va ser capaç d'interrompre el procés d'escalfar una tassa de llet sense obrir la porta del microones. Per tant, utilitzant un exemple d'ús quotidià d'un artefacte, tenim que un 0% dels participants va ser capaç de resoldre la tasca. La raó era que cap dels usuaris no premia el botó adequat, ja que és el mateix que s'empra per a posar en marxa la cocció.

Diversos usuaris comenten que no veuen el botó per a parar.

Quina en podria ser l'anàlisi? La dada és el 0% d'efectivitat, el comportament observat és que no premen el botó, el comportament previst és prémer el botó.

Per tant, podem argumentar que hi ha un error en la interfície. El mateix botó facilita dues funcions, una cosa complicada en l'ús de sistemes, i no hi ha cap element de disseny que ajudi l'usuari a entendre-ho. A més, queda reforçat amb els comentaris dels usuaris.

c) Buscar el que és diferent, buscar el que és igual

Què van fer tots els usuaris i què no va fer ningú i esperàvem que fessin?

Quin comportament era igual en totes les tasques (en el cas d'un test) i va canviar en una tasca en concret?

Quina tasca que havien estat capaços de fer al principi del test no van ser capaços de repetir al final?

Quines targetes va ordenar tothom de la mateixa manera, en el cas d'un *card sorting*, i quines targetes van ser disposades de manera diferent per tots els participants?

Quin recorregut visual va ser més repetit en un experiment amb *eye-tracking*, però quina àrea visual va quedar totalment obviada per tots els participants?

Totes aquestes consideracions i moltes més ens poden ajudar a detectar problemes en la interfície.

d) Buscar l'extrem

Les dades que indiquin comportaments extrems han d'atreure la nostra atenció. Si alguna cosa ha passat massa o massa poc, probablement ens revela que un element falla.

3) Reprendre les notes del test

Si durant els tests prenem notes d'aspectes que ens han cridat l'atenció, és el moment de veure si les dades ens confirmen alguna suposició o hipòtesi. Observacions com "tots els usuaris fan servir l'ajuda" es poden corroborar.

e) Pensem que ens equivoquem

És habitual que l'analista senti que ha descobert el secret del que ha passat en l'avaluació. Hi ha la temptació de fer encaixar les dades en una teoria implícita, és a dir, de forçar les dades de manera que encaixin en la nostra idea dels problemes de la interfície en lloc de detectar quins són aquests problemes per mitjà de les dades.

Si l'analista manté la humilitat suficient per a pensar que es pot equivocar, estarà molt més atent a les dades que poden refutar la seva teoria.

També és important intercanviar impressions amb la resta d'observadors i l'equip, que poden fer una revisió crítica de les nostres conclusions.

f) La navalla d'Ockham

"Keep it simple" (l'explicació més senzilla sol ser la correcta).

Evitem que el resultat de l'anàlisi siguin teories complicades sobre la interacció dels usuaris. El treball de l'analista ha de ser pragmàtic.

Sovint no cal tenir una visió completa del model mental dels usuaris per a poder apuntar molts errors d'una interfície.

La navalla de Ockham

La navalla d'Ockham (de vegades escrit Occam o Ockam), *principi d'economia* o *principi de parsimònia*, és un principi filosòfic atribuït a Guillem d'Ockham (1280-1349), segons el qual quan dues teories en igualtat de condicions tenen les mateixes conseqüències, s'ha de preferir la més simple a la més complexa (*Wikipedia*).

Resum

L'avaluació de la usabilitat amb usuaris és un aspecte clau en el procés de disseny centrat en l'usuari, ja que ens permet detectar els aspectes que funcionen i els que no, d'un producte o sistema interactiu. Tenint en compte aquest context, en aquest mòdul didàctic s'han vist els mètodes principals per a l'avaluació de la usabilitat amb usuaris.

El test d'usabilitat és l'instrument principal d'avaluació de la usabilitat amb usuaris. El test amb usuaris persegueix resoldre problemes en la interfície d'un producte interactiu a partir de l'obtenció d'informació directament dels usuaris que l'utilitzaran en el futur.

L'avaluació de la usabilitat no garanteix que un producte sigui completament usable en el futur, però permet fer una iteració més del procés de disseny centrat en l'usuari i eliminar així errors i redissenyar aspectes clau per a l'experiència d'ús.

Bibliografia

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, maig). "Apprenticing with the customer: A collaborative approach to requirements definition" [article en línia]. *Communications of the ACM*.

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" [article en línia]. *No Solo Usabilidad* (núm. 6).

Hassan Montero, Y.; Martín Fernández, F. J. (2003). "Método de test con usuarios" [article en línia]. *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". Conferència SIGCHI sobre factores humanos en sistemas de computación (pàg. 387-392). Gaithersburg (MD).

Löwgren, J. (1995). *Perspectives on usability*. Linköpping: Linköpping University, Department of Computer and Information Science. 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, agost).

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". *El profesional de la información* (vol. 15, núm. 6, pàg. 414-422, novembre-desembre).

Salant, P.; Dillman, D. A. (1994). *How to conduct your own survey*. Nova 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*. Nova York: Cambridge University Press.