30

Tests utilisateurs

(user tests)

Le test utilisateur est la méthode phare pour évaluer l'expérience utilisateur d'un système dans un processus itératif. Il consiste à mettre en situation l'utilisateur afin d'observer ses comportements, ses réactions et sa performance dans la réalisation de tâches prédéfinies. Les données recueillies donnent de précieuses informations sur les forces et les faiblesses du système évalué, ainsi que sur l'expérience vécue par les utilisateurs cibles. Méthode itérative par excellence, le test utilisateur s'utilise dès les premières maquettes du système et jusqu'au produit final. Il se combine dans une même session avec d'autres méthodes, comme les échelles d'utilisabilité ou d'UX, pour une multiplication des approches et une diversité des données recueillies. À vos tests, prêts, partez !

Quoi Évaluer l'UX d'un système en observant la manière dont les utilisateurs accomplissent leurs tâches.

Qui Des utilisateurs testent un système en réalisant un ensemble de tâches. Au moins un expérimentateur observe et analyse la performance des utilisateurs.

Où En situation naturelle pour l'utilisateur dans son environnement quotidien, ou en situation contrôlée dans un lieu dédié.

Quand Tout au long du processus de conception, pour évaluer itérativement concepts, maquettes et produits finaux.

Comment L'utilisateur est invité à réaliser un ensemble de tâches prédéfinies par l'équipe de conception. Ses interactions avec le système sont observées et analysées.

PLANIFICATION
●●●●○ Difficile
Durée : 1 à 2 jours

PASSATION
●●●○○ Moyen
Durée : 30 à 90 min./ utilisateur

ANALYSE DES RÉSULTATS
●●●●○ Difficile
Durée : 1 à 3 jours

EXPERTISE REQUISE
●●●●○ Avancé

Fiches liées : 6. Observation – 4. Entretien – 23. Échelles d'utilisabilité – 24. Échelles UX

La méthode du test utilisateur consiste principalement à évaluer l'UX d'un système en observant la façon dont les utilisateurs cibles l'utilisent pour réaliser des tâches. Elle est aussi une combinaison flexible de méthodes quantitatives et qualitatives permettant d'évaluer la performance des participants, leurs comportements ou encore leurs réactions. L'analyse des résultats offre de précieuses informations pour l'amélioration itérative du système.

Fondements théoriques

L'une des premières applications du test utilisateur en IHM est probablement celle de Chapanis (1953) pour la conception des tableaux de bord des avions américains durant la seconde guerre mondiale. La multiplication des cadrans dans le cockpit entraînait chez les pilotes de nombreuses erreurs de lecture, d'interprétation et par conséquent de pilotage. L'équipe de Chapanis a alors conçu et testé différents types d'afficheurs. Par exemple, pour les informations numériques, cinq types d'afficheurs ont été testés auprès de plusieurs pilotes. Le pourcentage d'erreurs commises pour la lecture de chaque cadran a été recueilli, permettant ainsi de recommander l'utilisation d'un afficheur adapté pour les nouveaux cockpits.

La première interface graphique Xerox Star (fiche 17. Iconographie) a également mobilisé des tests utilisateurs pour sa conception. Dans un article rétrospectif, Johnson *et al.* (1989) expliquent que la conception de la partie logicielle a reposé sur une forte articulation entre prototypages et tests utilisateurs dans une démarche itérative, conduisant ainsi à un système qui répond le mieux possible aux attentes des utilisateurs.

La méthode du test utilisateurs, plutôt centrée à ses débuts (efficacité, efficience et satisfaction des utilisateurs, Maguire, 2001), a progressivement évolué vers des modes de passation qui évaluent l'expérience utilisateur dans sa globalité (voir l'introduction de cet ouvrage). Ainsi, la passation de tests en situation réelle (in situ) entre peu à peu dans les pratiques des professionnels et des chercheurs, et prend tout son sens pour l'évaluation des technologies mobiles (Kjeldskov, 2004).

Sur le principe, la méthode du test utilisateur a toutefois peu évolué. Elle consiste toujours à observer les actions de l'utilisateur dans la réalisation de tâches prédéfinies, en limitant toute interférence sur la situation observée dans le respect d'une démarche ethnographique (fiche 6. Observation).

Pourquoi utiliser cette méthode ?

Tester pour une amélioration continue

La méthode du test utilisateur s'inscrit dans un processus itératif : elle s'applique tour à tour aux premiers concepts du produit, aux maquettes basse fidélité et aux différentes versions du prototype (fiche 18. Maquettage), jusqu'au système final avant et après son déploiement sur le marché. Chaque itération permet d'améliorer l'expérience utilisateur afin d'atteindre le

niveau de qualité attendue par l'équipe de conception. C'est un excellent outil de diagnostic et d'aide à la formulation de recommandations. Cette approche itérative favorise une amélioration continue du système, tout au long du cycle de développement, en révélant une grande partie des problèmes d'utilisation.

Mettre en situation

Le test utilisateur permet de confronter les utilisateurs à un système en simulant des tâches qu'ils accompliraient en situation réelle. C'est la performance de ces usages qui est évaluée. Les tests recueillent de nombreuses informations sur la façon dont les utilisateurs interagissent avec le système. Les comportements de ces derniers sont souvent imprévisibles : c'est en les observant, de façon non interférente, qu'il est possible d'identifier les difficultés d'interaction et d'analyser si le système offre une expérience positive ou non.

Combiner plusieurs méthodes

Les tests utilisateurs fournissent un cadre à de nombreuses techniques. En effet, ils sont rarement appliqués sans être complétés par d'autres méthodes UX, a minima durant l'étape de débriefing. Généralement, un entretien est mené à l'issue de la passation afin de recueillir les impressions de l'utilisateur vis-à-vis du système (fiche 4. Entretien). De même, des questionnaires sont souvent proposés après la réalisation des tâches (fiches 23. Échelles d'utilisabilité et 24. Échelles UX) pour recueillir des données complémentaires.

Convaincre votre manager ou client

1. Cette méthode consiste à observer les utilisateurs dans leurs interactions avec le système. Les données recueillies sont riches d'enseignements et conduisent à formuler des recommandations pour améliorer l'UX du système.
2. Plusieurs modes de passation sont possibles, adaptés à toutes les exigences ou contraintes d'un projet : en situation contrôlée, à distance ou en mode guérilla, il existe une forme de test qui vous conviendra.
3. Invitez votre client à un test utilisateur ou montrez-lui un enregistrement audiovisuel. Bien souvent, la façon dont les utilisateurs accomplissent leurs tâches ne correspond pas à ce que les commanditaires imaginent et cela éveille alors leur intérêt pour la méthode.

Avantages	Limites
• Révèle une grande partie des problèmes d'utilisation du système. • Recueille de nombreuses informations sur la performance, les attitudes, les comportements et les réactions de l'utilisateur vis-à-vis d'un système. • Les résultats des tests permettent d'étayer facilement un rapport de recommandations : chaque élément d'amélioration du système est justifié par les difficultés d'interaction rencontrées par les utilisateurs. • Méthode flexible qui soutient une approche méthodologique mixte (recueil de données qualitatives et quantitatives).	• Coûteux en temps, en particulier pour les passations individuelles qui requièrent de 30 à 90 minutes d'investissement. • Les scénarios d'usage peuvent avoir un côté « artificiel », trop éloigné des tâches qu'accompliraient naturellement les utilisateurs. • Le test utilisateur ne permet pas d'explorer toutes les fonctionnalités d'un système. L'équipe de conception doit faire des choix de scénarios principaux qu'elle souhaite observer.

Mise en pratique

Format

Les différents types de tests utilisateurs

Les tests utilisateurs peuvent être conduits de différentes manières selon les objectifs et les contraintes du projet de conception. On distingue quatre grandes catégories de tests : en situation contrôlée, en contexte naturel, à distance et en mode guérilla.

Les tests en situation contrôlée

Menés en laboratoire d'utilisabilité *(usability lab)* ou dans tout autre contexte contrôlé, ces tests cherchent à isoler le participant de ce qui pourrait le distraire de sa tâche afin de minimiser l'impact de variables externes sur les résultats observés.

Dans les laboratoires d'utilisabilité, le participant est placé dans une salle de test séparée d'une salle d'observation par une glace sans tain (figure 30-1). Un équipement audiovisuel enregistre ses interactions avec le système sous plusieurs angles. Dans la salle d'observation, un expert au moins observe le déroulement du test sans perturber l'activité de l'utilisateur.

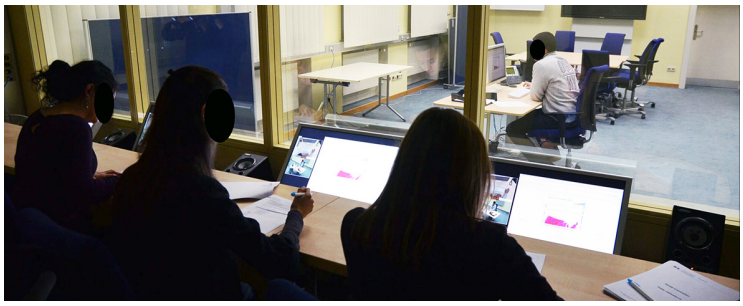

Figure 30–1
Exemple de laboratoire d'utilisabilité
(Groupe de recherche HCI, Institut COSA, université du Luxembourg)

Les tests utilisateurs « in sitro »

Si les laboratoires d'utilisabilité ont été très appréciés dans les années 1990 pour la rigueur méthodologique qu'ils confèrent aux tests utilisateurs, ils sont aujourd'hui moins plébiscités en raison de leur coût élevé de fonctionnement et du caractère assez artificiel de la situation d'interaction. De plus en plus de laboratoires (appelés dorénavant UserLabs, plates-formes d'observation ou encore laboratoires d'usage) cherchent alors à reproduire un contexte d'usage réaliste, en recréant pour les besoins d'une étude un environnement domestique, professionnel ou une situation particulière. L'objectif est de plonger l'utilisateur dans un contexte familier, car c'est en s'approchant d'une situation naturelle que l'expérience utilisateur sera en effet évaluée de la manière la plus valide.
Pour décrire ces environnements contrôlés mais proches du monde réel, Kjeldskov et Skov (2007) utilisent le terme de tests « in sitro », contraction des termes in situ, relatif aux études menées sur le terrain, et in vitro, qui se réfère aux tests en laboratoire. Les auteurs ont par exemple recréé un contexte hospitalier en laboratoire (de manière basique et à moindre coût), afin de tester l'utilisabilité d'un système portable destiné au personnel soignant.

Les laboratoires d'utilisabilité ne sont pas le seul moyen de contrôler l'environnement de l'utilisateur. Les tests peuvent aussi être organisés dans une salle de réunion ou un bureau en entreprise. Dans ce cas, ils s'appuient souvent sur des outils qui assistent l'enregistrement des données, en capturant en temps réel :

- l'écran de l'ordinateur sur lequel est installé l'outil ;
- les actions de l'utilisateur sur le clavier ;
- les expressions faciales de l'utilisateur à l'aide de la webcam ;
- les verbalisations de l'utilisateur lorsqu'on lui demande de penser à voix haute (voir encadré « Le protocole Think aloud » plus loin dans cette fiche).

Quelques outils de tests utilisateurs

Plusieurs outils de tests utilisateurs existent dans le commerce. Les plus connus et les plus recommandés sont payants.

- L'outil Morae ® (https://www.techsmith.com/morae.html) est l'un des plus utilisés par les professionnels de l'UX. En plus des enregistrements audiovisuels, Morae permet d'organiser les sessions de passation en présentant tour à tour les tâches que l'utilisateur doit exécuter.
- La société Noldus (www.noldus.com) propose quant à elle un ensemble de solutions logicielles pour l'encodage et l'analyse des comportements utilisateurs (The Observer XT ® et FaceReader ®), ainsi que pour la gestion des enregistrements vidéo (Media recorder ® et Viso ®).
- Ovo Logger ® (www.ovostudios.com/ovologger.asp) et Silverback ® (http://silverbackapp.com) proposent également un accompagnement dans la passation de tests, de l'organisation des tâches aux analyses statistiques.

Les tests en contexte naturel

Si l'utilisabilité se prête bien à une évaluation en situation contrôlée, l'UX présente de nouveaux défis à relever pour l'évaluation des systèmes. Parmi ces défis, l'importance du contexte d'interaction physique et social pour l'expérience vécue remet en question l'artificialité des tests en situation contrôlée. Les tests en contexte naturel évaluent l'UX sur le terrain. Les

participants sont sollicités chez eux, sur leur lieu de travail, dans la rue… ou tout autre endroit dans lequel s'effectue d'ordinaire leurs interactions avec le système. Le principe du test ne change pas : l'utilisateur est invité à réaliser plusieurs tâches avec un système. Ses interactions sont observées puis analysées.

Un courant en IHM, baptisé *the turn to the wild* (Crabtree *et al.*, 2013), défend d'ailleurs l'idée que puisque les technologies s'inscrivent pleinement dans la vie quotidienne des utilisateurs, c'est précisément dans ce contexte qu'elles doivent être évaluées.

Les tests à distance

À l'inverse des laboratoires d'utilisabilité, les tests utilisateurs à distance sont devenus un moyen de passation prisé par les experts de l'UX (Bolt, 2010). Il s'agit de situations de tests où l'observateur et l'utilisateur sont géographiquement distants (Bastien, 2010). Ils présentent l'avantage de pouvoir mobiliser rapidement un grand nombre d'utilisateurs cibles, sans se soucier de leur localisation. Il existe deux types de tests à distance :

- les tests à distance **synchrones** sont conduits en temps réel par l'évaluateur, qui gère et observe la passation à l'aide d'un système spécifique ou par vidéoconférence ;
- les tests à distance **asynchrones** sont conduits par des systèmes automatiques, qui gèrent les passations individuelles sans intervention directe de l'évaluateur. Ils offrent la possibilité de faire passer simultanément un grand nombre de participants.

Les tests en mode guérilla

Les tests utilisateurs en mode guérilla prennent une place de plus en plus importante dans la pratique des professionnels (MacDonald & Atwood, 2013). Ils s'affranchissent du formalisme des tests utilisateurs classiques pour réduire les contraintes logistiques et les coûts de passation. Bien souvent, il s'agit de solliciter spontanément des utilisateurs cibles, dans un lieu public, pour l'évaluation d'un système. Si leur caractère low cost est attirant, il faut néanmoins connaître leurs limites (fiche 2. Recrutement des utilisateurs, section « Le cas des méthodes guérillas ») et toujours leur préférer des tests plus conventionnels, et plus valides.

Un test particulier en mode guérilla : le Trunk test

Défini par Steve Krug (2005), ce mode guérilla a été conçu pour l'évaluation des sites web. Krug décrit le contexte : « *Imaginez que vous soyez enfermé dans le coffre d'une voiture les yeux bandés et emmené sur une page web au plus profond de l'arborescence d'un site ! Si la page est bien conçue, vous devriez pouvoir répondre à ces questions sans hésitation : de quel site s'agit-il ? Quelle est cette page ? Quelles sont les principales sections de ce site ? Que puis-je faire depuis cette page ? Des éléments de la page m'indiquent-ils où je suis sur le site ? Comment puis-je effectuer une recherche ?* »
L'objectif du *Trunk test* est de permettre aux utilisateurs de se repérer dans un site web depuis n'importe quelle page, avec des éléments qui sautent aux yeux. Le test comprend trois étapes en mode guérilla :

- choisissez n'importe quelle page d'un site et imprimez-la ;
- tenez la feuille à bout de bras en plissant les yeux, de sorte que vous ne puissiez pas étudier attentivement les détails de la page ;

- essayez de trouver le plus rapidement possible les éléments suivants et entourez-les : 1. L'identité du site, 2. Le nom de la page, 3. Les sections et sous-sections, 4. Les éléments de navigation locale, 5. Les éléments qui vous indiquent où vous vous trouvez dans le site, 6. Une fonction de recherche.

Si vous ne pouvez pas identifier tous ces éléments, proposez des recommandations afin de mieux les faire apparaître.

Les supports à tester

Les tests utilisateurs s'appliquent à différents supports, en fonction de l'état d'avancement du projet (tableau 30-1). Rappelons qu'il est conseillé de tester le système le plus tôt possible au cours du processus de conception, afin de disposer d'une latitude d'intervention aussi importante que possible.

Tableau 30–1 Utilisation de différents supports en fonction des objectifs visés par le test utilisateur

Objectif du test	Description	Support
Exploratoire	Tester le système dès la génération des premières solutions, afin d'identifier les problèmes de fond (structure de la page, enchaînement des actions…).	Maquettes
Évaluation	Tester le système dans une version suffisamment fonctionnelle pour simuler l'exécution de tâches spécifiques auprès des utilisateurs finaux.	Prototype fonctionnel
Validation	Tester le système dans sa version finale, avant la mise sur le marché, afin de vérifier qu'il répond bien aux exigences de qualité attendues.	Prototype final
Comparaison	Tester plusieurs alternatives d'un même système, ou plusieurs systèmes comparables entre eux, afin d'identifier les forces et faiblesses de chaque produit.	Maquettes, prototypes, produits finaux
Amélioration continue	Évaluer la conformité du système avec les besoins du marché dans un processus d'amélioration continue.	Produits sur le marché

Planification

Avant toute chose, vous devrez sélectionner le type de test le mieux adapté à vos besoins (voir précédemment la section « Format »). Rappelez-vous également qu'un test est un « écosystème » combinant plusieurs méthodes. Planifier des sessions de tests implique de réfléchir à la construction cohérente d'un plan de test (figure 30-2) et d'un protocole, qui décriront en détail le déroulement des passations.

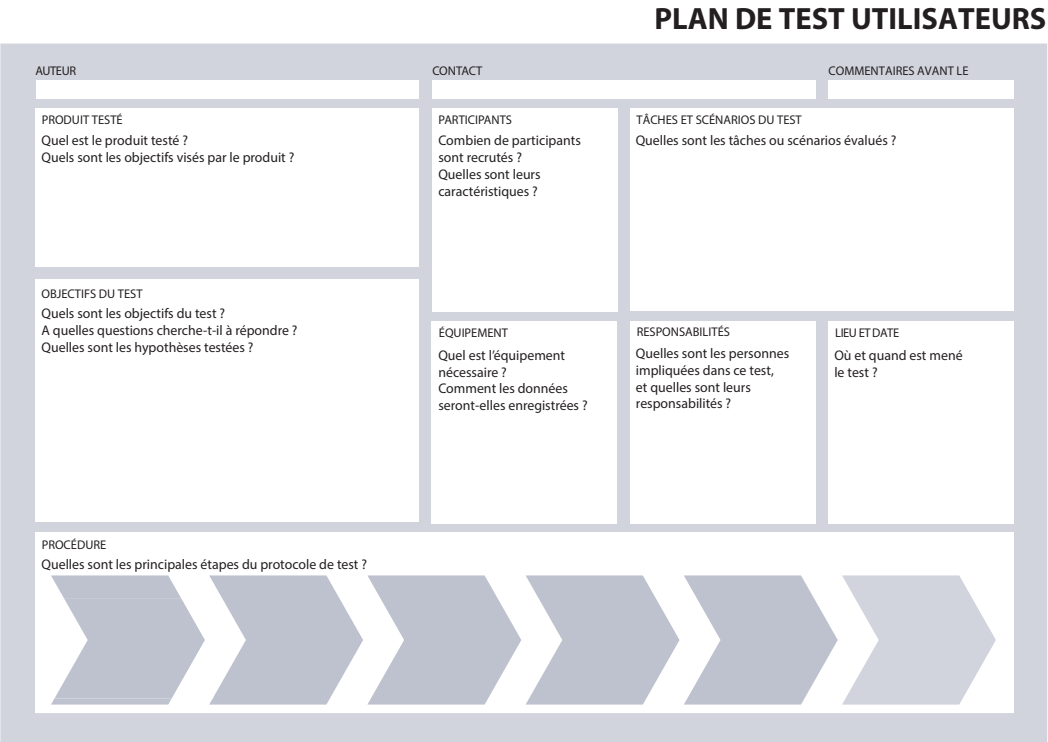

Figure 30–2 Le plan de test utilisateurs est un outil de planification et de communication.

Définir les scénarios d'usage

Les tests utilisateurs reposent sur la réalisation de tâches prédéfinies par l'évaluateur. Ces tâches, qui prennent la forme de scénarios d'usage, simulent des actions représentatives des buts utilisateurs et doivent donc être élaborées avec soin. Plusieurs types de scénarios sont possibles :

- dans le cas le plus fréquent, les scénarios d'usage représentent les buts principaux des utilisateurs (procéder à un achat sur un site e-commerce, naviguer vers un lieu grâce à un GPS…) ;
- il est également pertinent dans certains cas de tester les éléments du système qui peuvent poser le plus de problèmes. Les scénarios vont par exemple se baser sur les résultats d'une évaluation experte (fiches 26. Évaluation experte et 27. Inspection cognitive) en reprenant les points les plus critiques ;
- enfin, pour augmenter le réalisme du test, il est possible de demander à des utilisateurs de définir eux-mêmes les scénarios d'usage qui leur semblent les plus pertinents. Cela répond bien aux exigences d'une démarche UX, qui cherche à observer l'utilisateur dans des conditions les plus naturelles possibles.

Selon la difficulté des tâches et le nombre de fonctions que vous souhaitez tester, rédigez de 5 à 8 scénarios. Cela représente un bon compromis entre ce que souhaite explorer l'équipe de conception et la durée raisonnable d'un test. Les scénarios seront présentés aux utilisateurs par ordre de difficulté croissante. Débutez par un scénario facile que tous les utilisateurs réussiront.

Au niveau de leur contenu, les scénarios :

- sont rédigés sous la forme d'objectifs à atteindre, chaque scénario comprenant un seul objectif. Ne décrivez pas les scénarios sous la forme d'une succession d'actions (comme « cliquez ici », « ouvrez telle fenêtre ») ;
- sont rédigés sous la forme de phrases courtes et facilement compréhensibles ;
- s'adressent directement à l'utilisateur par le pronom « vous » ;
- indiquent les éléments additionnels nécessaires à la réalisation de la tâche (par exemple des identifiants si l'utilisateur doit consulter l'espace personnel d'un site, les informations à remplir dans certains champs…).

Une fois définis, il est indispensable de pré-tester vos scénarios pour valider leur cohérence avant la passation effective du test, et pour établir le temps nécessaire pour une passation (fiche 2. Recrutement des utilisateurs, section « Pré-tester le déroulement des sessions »).

Établir une grille d'observation

Pour faciliter l'observation des actions réalisées par l'utilisateur, il est conseillé d'établir une grille qui sera remplie au fur et à mesure de la passation. Dans cette grille, vous pourrez notamment renseigner :

- les activités et tâches réalisées ;
- le temps qu'il a fallu à l'utilisateur pour réaliser chaque tâche ;
- la succession des différents écrans parcourus par l'utilisateur ;
- les réussites et éléments qui ont généré de la satisfaction ;
- les échecs, erreurs, pannes ou problèmes rencontrés ;
- les commentaires des utilisateurs (voir encadré « Le protocole Think aloud », page 472) ;
- des éléments d'observation libre rédigés par l'observateur (par exemple la description de l'état émotionnel de l'utilisateur).

Le défi est d'établir une grille qui couvre un maximum de détails prédéfinis, sans pour autant surcharger l'observateur.

Recruter les utilisateurs

La fiche 2 (Recrutement des utilisateurs) vous fournira toutes les informations nécessaires pour la phase de recrutement à proprement parler. Mais combien d'utilisateurs faut-il recruter pour obtenir des résultats valides à un test utilisateur ? C'est la question la plus fréquente à cette étape de la planification et malheureusement il n'y a pas de réponse toute faite. Cela dépend des objectifs et des caractéristiques de votre projet.

En 1992, Virzi (1992) affirme que 5 utilisateurs cibles suffisent pour détecter 80 % des problèmes d'utilisabilité d'un système. Nielsen (2000) recommande le même chiffre, car il constitue selon lui un bon compromis entre le coût engagé pour la réalisation des tests et le nombre de problèmes détectés.

Ceci n'est toutefois pas une règle qui s'applique à tous les projets. Sauro et Lewis (2012) soulignent que 5 utilisateurs peuvent rencontrer un unique problème, ou au contraire rencontrer

chacun des problèmes différents. Ces deux cas de figure n'auront pas le même impact sur l'évaluation de votre système. Certains auteurs (Rubin & Chisnell, 2008) recommandent de tester auprès de 8 utilisateurs (ou 30 dans le cadre d'une démarche scientifique afin de pouvoir traiter statistiquement les résultats).

Calculez la taille idéale de votre échantillon d'utilisateurs

La taille optimale d'un échantillon d'utilisateurs dépend étroitement de la complexité du système et du nombre de problèmes qu'il est susceptible de comporter. Plus il est probable que les utilisateurs rencontrent de nombreux problèmes avec votre système, plus vous aurez besoin de participants pour votre test. Une équation de probabilité sert à calculer la taille idéale de votre échantillon (Sauro & Lewis, 2012) :

$$P(x \geq 1) = 1 - (1 - p)^n$$

p désigne la probabilité qu'un problème soit rencontré par l'utilisateur, c'est-à-dire la fréquence d'apparition d'un problème, *n* est le nombre d'utilisateurs qui passent le test et $P(x \geq 1)$ est la probabilité qu'un problème soit rencontré au moins une fois parmi les *n* participants.
Le plus difficile à cette étape est d'estimer le nombre de problèmes potentiels de votre système (*p*). Vous pouvez pour cela pré-tester votre système auprès de 5 utilisateurs et calculer la fréquence avec laquelle chaque problème a été rencontré (avec 5 utilisateurs, un problème peut avoir été rencontré de 1 à 5 fois). Calculez ensuite la moyenne de ces fréquences pour obtenir la valeur *p*. Le tableau 30-2 prend l'exemple de 6 problèmes et de 5 utilisateurs. Chaque case cochée correspond à un problème rencontré par un utilisateur.

Tableau 30–2 Exemple de calcul du nombre de problèmes potentiels d'un système (*P*)

		Problème rencontré						
		1	**2**	**3**	**4**	**5**	**6**	
Utilisateur	**1**	x			x			**Moyenne**
	2					x	x	
	3		x	x				
	4							
	5		x			x		
Fréquence		**0,2**	**0,4**	**0,2**	**0,2**	**0,4**	**0,2**	***P* = 0,27**

Un outil en ligne (www.measuringu.com/problem_discovery.php) permet de calculer votre taille d'échantillon idéale en fonction de *p* (champ *Problem occurrence*) et de $p(x \geq 1)$ (champ *Discover [85 %] of all problems*, sachant que 85 % est généralement un bon compromis). Avec notre exemple où $p = 0{,}27$, le nombre d'utilisateurs requis est de 6,03 (soit 6 utilisateurs).

Alors faut-il s'en tenir à 5 (ou 8) utilisateurs ? Cet échantillon sera suffisant si votre système a déjà fait l'objet de tests au cours de précédentes itérations, ou si vous souhaitez dégager les problèmes les plus bloquants. En revanche, cela ne suffira pas pour couvrir toutes les questions relatives à l'UX de votre système ou pour évaluer un système complexe.

Passation

Durée d'un test utilisateur

Un test utilisateur ne devrait pas durer plus de 90 minutes, accueil des participants et étape de débriefing inclus. Le tableau 30-3 suggère le déroulement typique d'une session.

Tableau 30–3 Exemple de déroulement d'un test utilisateur

Étape de la passation	Durée	Durée cumulée
Accueil du candidat et instructions	10 min.	10 min.
Réalisation des scénarios d'usage	30 min.	40 min.
Passation de questionnaires	15 min.	55 min.
Entretien de débriefing	15 min.	1 h 10
Remerciement et dédommagement	5 min.	1 h 15

Le test utilisateur peut être associé à de nombreuses autres méthodes. Avant la passation, le test des 5 secondes (fiche 29) est utile pour recueillir la première impression de l'utilisateur sur le système. On aura aussi recours à des outils de mesure psychophysiologiques ou à des systèmes de suivi des mouvements oculaires par *eyetracking* si cela fait sens pour le projet (voir encadré « Mesure des mouvements oculaires par eyetracking » dans la section suivante).

Avant la passation

Votre passation débute par l'accueil du participant. Mettez-le à l'aise et expliquez-lui le déroulement de la séance avant de lui faire signer le formulaire de consentement (fiche 3. Déontologie et éthique). Quand il se sent prêt à démarrer le test, expliquez-lui la consigne et remettez-lui les scénarios de tâches sous forme écrite. En disposant chaque scénario sur une page d'un petit fascicule, vous suivrez plus facilement l'état d'avancement de l'utilisateur au cours du test en vous référant à la page qu'il consulte. Rappelez-lui enfin que c'est le système qui est évalué et non sa performance. Assurez-vous que le participant a bien compris les consignes et répondez à toutes ses questions avant de démarrer.

Pendant la passation

Pendant la passation du test, n'interrompez pas le participant. Observez ses interactions et remplissez vos grilles d'observation. Si vous êtes dans la même pièce que lui, abstenez-vous de tout commentaire ou de toute autre réaction qui pourrait le désorienter. Si le participant rencontre un problème bloquant sur l'un des scénarios, vous pouvez lui donner pour instruction de passer au scénario suivant.

Le protocole Think aloud – Penser à voix haute

Le protocole *Think aloud* (Jaspers *et al.*, 2004) consiste à demander aux participants de penser à voix haute durant leurs interactions avec le système. Les utilisateurs sont invités à exprimer verbalement leur raisonnement au cours de la réalisation de leurs tâches, leurs opinions sur celles-ci, leurs réactions… en somme, tout ce qui leur passe par la tête ! C'est un protocole très utile durant un test, car il donne souvent du sens aux actions des utilisateurs et facilite l'interprétation des observations. Il faut toutefois être conscient que ce type de verbalisation n'est pas naturel pour les participants et peut parfois biaiser un peu le ressenti de l'interaction.
Du point de vue de l'évaluateur, le principal inconvénient du Think aloud est que les participants oublient rapidement de s'exprimer à voix haute, surtout lorsqu'ils sont isolés comme dans les laboratoires d'utilisabilité. Puisque vous ne devez pas intervenir durant la passation pour le leur rappeler, l'une des bonnes pratiques consiste à noter la consigne du Think aloud avant chaque scénario.

Mesure des mouvements oculaires par eyetracking

La diminution progressive des coûts des systèmes d'eyetracking encourage de plus en plus de professionnels de l'UX à s'équiper de ces outils qui soutiennent les tests utilisateurs. L'eyetracking consiste à suivre et enregistrer, en continu et en temps réel, la position du regard de l'utilisateur sur l'écran (Nielsen & Pernice, 2010). Les systèmes d'eyetracking sont constitués d'une part d'un équipement de capture des mouvements oculaires (lunettes ou système déporté au-dessus ou en dessous de l'écran) et, d'autre part, d'une application qui enregistre les points de fixation et les saccades des regards de l'utilisateur.
Ces systèmes présentent plusieurs avantages, dont la possibilité d'identifier quels éléments de l'interface seront les plus vus, ce qui permet d'évaluer la qualité de l'organisation des informations à l'écran. Ils renseignent également sur les parcours des regards des utilisateurs, afin d'identifier les éléments qui sont perçus en premier. Ils apportent enfin des informations complémentaires aux données enregistrées lors des tests utilisateurs, afin de mieux comprendre les problèmes rencontrés par les participants et leurs stratégies d'exploration du système évalué.
En revanche, cette méthode nécessite un lourd traitement de données, nombreuses et complexes. De plus, elle requiert malgré tout d'interroger les utilisateurs pour interpréter les résultats obtenus et comprendre pourquoi une personne s'est attardée sur tel ou tel élément de l'interface. Pour finir, ces outils, bien que de moins en moins intrusifs, nécessitent encore à ce jour un équipement conséquent peu adapté à des tests en contexte naturel.
Le plus souvent, deux types de représentations des résultats sont proposés par les systèmes d'*eyetracking* : une carte de chaleur *(heatmap)* (figure 30-3) et les parcours oculaires. Ces représentations mettent en avant les zones où le regard de l'utilisateur s'est le plus attardé, ainsi que le séquencement des mouvements oculaires.

Figure 30–3
Carte de chaleur de la page d'accueil du site Service-Public.fr (Drusch & Bastien, 2014)

Après la passation : le débriefing

Au moment du débriefing, plusieurs méthodes complémentaires au test utilisateur sont mobilisées. L'entretien (fiche 4) est la plus courante et reste incontournable. Vous pouvez également mobiliser d'autres méthodes de verbalisation, comme la complétion de phrases (fiche 21). Le protocole *Think aloud rétrospectif* (voir l'encadré « Le protocole Think aloud »), où les utilisateurs sont invités à exprimer après la passation ce qu'ils ont pensé durant le test, permet également de recueillir des participants davantage d'explications et des suggestions d'amélioration du système. Il présente aussi l'avantage de ne pas induire de biais durant la passation, les verbalisations pendant l'interaction étant souvent décrites par les participants comme non naturelles voire perturbantes. Bien souvent, des échelles d'utilisabilité (fiche 23) ou UX (fiche 24) sont soumises au participant afin de recueillir des données subjectives quantitatives.

L'entretien de débriefing sert aussi à vérifier que le participant se sent à l'aise et a bien vécu la situation de test (fiche 3. Déontologie et éthique).

Analyse et interprétation des résultats

Selon les métriques définies dans les objectifs du test, les résultats obtenus peuvent inclure :

- la proportion de succès et d'échecs pour chaque scénario ;
- le nombre d'erreurs commises pour chaque scénario ;
- le nombre et le type des erreurs répétées (lorsqu'un utilisateur commet plus d'une fois la même erreur) ;
- le nombre d'actions effectuées par rapport au nombre d'actions minimal requis pour réaliser la tâche ;
- le temps moyen mis pour la réalisation de chaque scénario ;
- le contenu des verbalisations des utilisateurs ;
- le contenu des entretiens au cours du débriefing ;
- les scores aux questionnaires d'utilisabilité ou d'UX, si vous en avez fait passer.

Les résultats pour chacune de ces observations seront repris dans un tableau récapitulatif (tableau 30-4).

Tableau 30–4 Exemple de synthèse des résultats d'un test utilisateur

		Taux de réussite	Nombre d'erreurs	Temps moyen	Verbalisations
Scénario A	**Action 1**	100 %	4	22 sec.	Je ne comprends pas pourquoi il faut d'abord choisir la date.
	Action 2	100 %	0	43 sec.	J'adore ce système de notation.
	Action 3	85 %	8	87 sec.	Ah ! C'est ici qu'il faut cliquer ?
Total		**95 %**	**12**	**152 sec.**	

Reportez-vous aux fiches 4, 23 et 24 pour le traitement des résultats relatif à chacune de ces méthodes.

La méthode RITE

La méthode RITE *(Rapid Iterative Testing and Evaluation)* est une technique de tests utilisateurs rapide et itératif, qui implique toute l'équipe de conception (Dumas & Fox, 2009). Elle repose sur l'identification des erreurs commises par un seul utilisateur, lors de son interaction avec un système. Le principe est qu'au lieu de réaliser des tests standardisés sur une même version du système, celui-ci va être corrigé et amélioré après chaque passation.
Concrètement, l'équipe de conception identifie pour un utilisateur ce qui semble être bloquant ou source d'erreur. Le problème est, si possible, immédiatement corrigé à l'issue de la passation. Un autre utilisateur passe ensuite le test sur le système amélioré, ses problèmes sont notés et corrigés immédiatement, et ainsi de suite jusqu'à ce qu'une forte proportion d'erreurs ait été corrigée.
La méthode repose ainsi sur la capacité de l'équipe de conception à formuler rapidement des recommandations pour l'amélioration du système. Les solutions implémentées seront testées par l'utilisateur suivant, et ce jusqu'à ce que 5 utilisateurs à la suite ne rencontrent aucun problème !

Exploitation des résultats

Les résultats des tests vont permettre de corriger les problèmes d'utilisabilité du système évalué. Il est souvent utile de donner un ordre de priorité pour distinguer les tâches les plus bloquantes de celles posant le moins de problème. Une matrice identifiant la fréquence de chaque problème aide à identifier plus clairement un ordre de priorité (tableau 30-5). Les fréquences les plus élevées indiquent les problèmes récurrents à traiter.

Tableau 30–5 Identification des problèmes les plus fréquents

		Problème rencontré						
		1	**2**	**3**	**4**	**5**	**6**	**Fréquence moyenne**
Utilisateur	**1**	x	x			x	x	
	2	x			x	x		
	3		x			x		
	…		x	x	x	x	x	
Fréquence		**0,4**	**0,8**	**0,2**	**0,4**	**1**	**0,6**	**0,56**

Les verbalisations, données d'observation et données obtenues par d'autres méthodes utilisées lors du test (questionnaires, entretiens, etc.) sont également analysées pour dégager des éléments complémentaires relatifs à l'expérience utilisateur.

Toutes les données obtenues lors des tests utilisateurs sont retranscrites sous forme d'un rapport (le site www.usability.gov propose des modèles à télécharger, en anglais). Celui-ci peut

prendre la forme de captures d'écrans annotées, incluant notamment les observations et recommandations principales, ainsi que le degré de sévérité des problèmes identifiés (figure 30-4). Des propositions d'amélioration seront définies par l'équipe de conception à l'aide de méthodes d'idéation (par exemple un brainstorming, fiche 9) et de génération (par exemple l'élaboration de nouvelles maquettes, fiche 18). Ces solutions seront évaluées lors de nouveaux tests utilisateurs, dans une démarche itérative.

Figure 30–4 Présentation visuelle d'un rapport de test utilisateur

La conception d'un tableau de bord (*scorecard*, tableau 30-6) aidera à suivre le nombre de recommandations implémentées dans une nouvelle version du système pour chaque itération, ainsi que différents indicateurs de performance (taux moyen de réussite, temps moyen de réalisation des tâches…).

Tableau 30–6 Exemple de tableau de bord pour le suivi des améliorations de votre système

Version du système	Taux moyen de réussite	Temps moyen	Recommandations			Remarques libres
			Émises	Retenues	Implémentées	
v1	54 %	37 min.	32	27	15	Difficulté d'intégration technique des actions glisser-déposer
v2	73 %	26 min.	19	18	12	Intégration de la fonctionnalité "achat" demandée par le client
v3	92 %	18 min.	8	8	6	Version finale retenue pour commercialisation

Votre système pourra être déployé sur le marché lorsque vous estimerez avoir atteint le degré de qualité attendu.

Le test A/B

Le test A/B, aussi appelé *split testing*, est une méthode d'évaluation qui consiste à comparer deux versions d'un même système afin de déterminer laquelle obtient les meilleurs résultats. Les résultats sont souvent mesurés en termes de taux de conversion, qui correspond au nombre d'utilisateurs réalisant une action particulière (comme acheter un produit sur un site e-commerce ou s'inscrire à une newsletter) par rapport au nombre total d'utilisateurs ayant utilisé le système.
Cette méthode est très utilisée en marketing des sites e-commerce, afin de mesurer l'influence des éléments d'une page sur le comportement des utilisateurs. La page qui obtient le meilleur taux de conversation est ainsi conservée. Puis d'autres alternatives de pages sont testées, en retenant à chaque fois la plus performante. Cette méthode est donc hautement itérative et peut être appliquée en continu sans limite dans le temps.
Concrètement, deux versions A et B d'une page sont accessibles sur Internet. Les visiteurs ne voient que l'une ou l'autre, avec une répartition automatique et aléatoire (50 % des utilisateurs accèdent à la version A et 50 % à la version B). Les taux de conversion sont calculés à l'aide de systèmes d'analyse des consultations des pages web, comme Google Analytics.

TRUCS ET ASTUCES

- Entre les passations, n'oubliez pas de remettre à zéro les éléments de votre système pour ne pas biaiser vos scénarios et votre recueil de données : suppression de l'historique de navigation, déconnexion d'un compte utilisateur, retour au menu principal, etc.
- Pour les systèmes complexes, confrontez lors de l'entretien de débriefing les utilisateurs aux enregistrements de leur passation. Cette auto-confrontation permet de recueillir des explications sur les actions réalisées, les erreurs commises et le ressenti de l'utilisateur
- Réalisez une évaluation experte (fiches 26 et 27) avant un test utilisateur. Cela vous aidera à corriger les problèmes les plus évidents et ainsi à mieux exploiter les bénéfices des tests utilisateurs.
- La technique de la co-découverte peut aider à recueillir des verbalisations plus naturelles et spontanées en situation contrôlée. Au lieu d'un seul utilisateur, ce sont deux utilisateurs (qui se connaissent) qui vont réaliser le test ensemble. Par leurs commentaires et discussions, vous en apprendrez beaucoup sur l'UX de votre système.

Exemple d'application

Dans le cadre d'un projet européen, l'appli OdoMap (https://play.google.com/store/apps/details?id=com.spacebel.odoMap&hl=fr) a été développée. Un test a été conduit auprès de 10 utilisateurs cibles sur un prototype fonctionnel.

Parmi les sept scénarios d'usage définis, les utilisateurs devaient enregistrer la position GPS de leur domicile comme position par défaut dans le système.

Les tests ont été conduits au domicile des participants. Les interactions avec le smartphone ont été enregistrées à l'aide d'un support disposant d'une micro-caméra et les réactions faciales de l'utilisateur ont été filmées de manière synchronisée (figure 30-5). Des taux d'échec et de réussite ont été calculés pour chaque scénario, permettant d'identifier les éventuelles difficultés (figure 30-6).

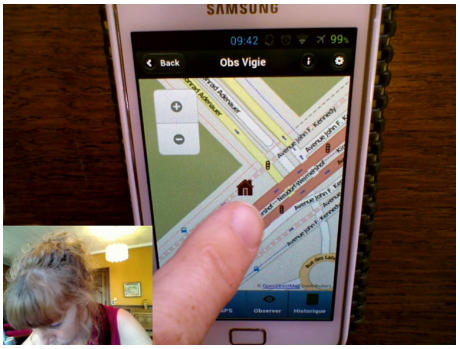

Figure 30–5 Exemple de test utilisateur sur smartphone

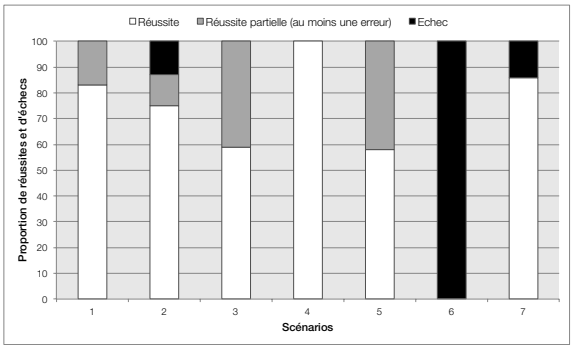

Figure 30–6 Taux de réussite et d'échec pour chaque scénario

Pour le scénario 6, huit problèmes ont été relevés. Par exemple, tous les utilisateurs appuyaient sur l'icône « maison » pour enregistrer la position GPS de leur domicile alors que l'icône servait en réalité à afficher la dernière position GPS enregistrée. Quatre recommandations ont été formulées pour améliorer la réalisation de ce scénario, puis testées auprès de 5 nouveaux participants.

Exercice pratique

L'application FreeMind (http://freemind.sourceforge.net) sert à créer des *mindmaps*. Son utilisation pose toutefois quelques problèmes, à cause de commandes peu claires et d'icônes ambigües.

- Identifiez cinq scénarios d'usage représentatifs de l'utilisation de ce système.
- Définissez un protocole de test et créez une grille d'observation. Rédigez également quelques questions à poser au participant lors de l'entretien de débriefing.
- Sélectionnez au minimum cinq participants. Testez l'application en demandant aux utilisateurs de réaliser les scénarios tout en pensant à voix haute.
- Analysez et synthétisez les données recueillies. Classez les tâches par ordre croissant de problèmes rencontrés, en les représentant dans un tableau ou sur un graphique. Sélectionnez la tâche qui a rencontré le plus de difficultés et dégagez de vos données les raisons des échecs éventuels.
- Rédigez un court rapport pour présenter le protocole de votre test utilisateur et ses résultats. Formulez des recommandations d'amélioration du système basées sur vos observations.

Questions sur la fiche

1. Qu'est-ce qu'un test utilisateur en mode guérilla ? Est-ce une pratique que vous recommanderiez ?
2. Quels sont les avantages et les limites des tests en laboratoire ?
3. En quoi consiste le protocole Think aloud ?
4. Pourquoi le test utilisateur s'inscrit-il bien dans une démarche itérative de conception ?

Bibliographie

Bastien, J.M.C. (2010). Usability testing: a review of some methodological and technical aspects of the method. *International Journal of Medical Informatics*, *79*(4), e18–23.

Bolt, N. (2010). *Remote Research: Real Users, Real Time, Real Research*. Rosenfeld Media.

Chapanis, A. (1953). Psychology and the Instrument Panel. *Scientific American*, *188*(4), 74–82.

Crabtree, A., Chamberlain, A., Grinter, R.E., Jones, M., Rodden, T., & Rogers, Y. (2013). Introduction to the special issue of "The Turn to The Wild." *ACM TOCHI*, *20*(3), 0–3.

Drusch, G., Bastien, J.M.C., & Paris, S. (2014). Analysing eye-tracking data: From scanpaths and heatmaps to the dynamic visualisation of areas of interest. *Proc. of AHFE 2014*, Krakow, Poland.

Dumas, J.S., & Fox, J.E. (2009). Usability testing : current practice and future directions. In A. Sears & J.A. Jacko (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (pp. 231–251). Taylor & Francis.

Jaspers, M.W.M., Steen, T., van den Bos, C., & Geenen, M. (2004). The think aloud method: a guide to user interface design. *International Journal of Medical Informatics*, *73*(11-12), 781–95.

Johnson, J., Roberts, T.L., Verplank, W., Smith, D.C., Irby, C., Beard, M., & Mackey, K. (1989). The Xerox Star: A retrospective. *Computer*, *22*(9), 11–26.

Kjeldskov, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies*, *60*(5-6), 599–620.

Kjeldskov, J., & Skov, M.B. (2007). Studying Usability In Sitro : Simulating Real World Phenomena in Controlled Environments. *International Journal of Human-Computer Interaction*, *22*(1-2), 7–36.

Krug, S. (2005). *Don't Make Me Think: A Common Sense Approach to Web Usability*. New Riders.

MacDonald, C.M., & Atwood, M. (2013). Changing Perspectives on Evaluation in HCI: Past, Present and Future. *Proc. of CHI 2013. Paris, France.* 1969–1978.

Maguire, M. (2001). Methods to support human-centred design. *International Journal of Human-Computer Studies*, *55*(4), 587–634.

Nielsen, J., & Pernice, K. (2010). *Eyetracking Web Usability*. New Riders.

Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing. How to plan, design and conduct effective tests*.. Wiley Publishing.

Sauro, J., & Lewis, J.R. (2012). *Quantifying the User Experience*. Elsevier.

Virzi, R.A. (1992). Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors*, *34*(4), 457–468.

Webographie

Guerilla testing : www.uxbooth.com/articles/the-art-of-guerrilla-usability-testing

Guerilla HCI : www.nngroup.com/articles/guerrilla-hci

Quelques exemples des choses à faire et à ne pas faire pour la définition des scénarios d'usage : www.nngroup.com/articles/task-scenarios-usability-testing

De nombreuses informations pratiques pour planifier et réaliser des tests utilisateurs : www.usability.gov/how-to-and-tools/methods/usability-evaluation/index.html

Nielsen, J. (2000). Why You Only Need to Test with 5 Users : www.nngroup.com/articles/why-you-only-need-to-test-with-5-users

Plusieurs modèles disponibles au format .doc pour vos rapports de tests utilisateurs : www.usability.gov/how-to-and-tools/resources/templates.html