

APPORTS ET LIMITES DES TESTS D'UTILISABILITE

Cécile VAN DE LEEMPUT¹, Vincent KOENIG²

Résumé

Ayant depuis peu mis en place un laboratoire d'utilisabilité au sein de notre unité de recherche, nous souhaitons présenter nos premières réflexions sur les apports et limites des tests d'utilisabilité, à partir des résultats de différents tests effectués. En effet, la définition théorique du concept ainsi que le déroulement de nos tests nous amènent à remettre en question une série de facteurs intervenant dans les tests d'utilisabilité. Partant de la définition normative du concept, nous situerons les différentes approches dans le contexte général des travaux sur l'utilisabilité. Dans un deuxième temps, nous passerons en revue plusieurs points critiques envisageant les «améliorations» potentielles. Finalement, nous concluons que l'utilisabilité ne peut s'approcher de manière expérimentale sans prendre en compte les apports du concept lors de la conception de la démarche expérimentale même.

Mots-clés : utilisabilité, U-Lab, évaluation, méthodologie, Internet

I. Introduction

Les recherches touchant au concept d'utilisabilité sont encore assez récentes. En effet, l'apparition du concept d'utilisabilité remonte aux recherches sur l'ergonomie des logiciels des années 80. Le concept-clef qui guidait ces recherches était celui d'interface homme-ordinateur. Cette dernière notion découle évidemment directement des considérations sur les systèmes homme(s)-machine(s). Ces dernières années, les publications spécifiques se sont même multipliées et ont rendu très connus leurs auteurs (cf. J. Nielsen par exemple). De plus, il semble que le développement de la majorité des produits récents ne se passe plus d'une prise en compte de facteurs d'utilisabilité. Dès lors, la raison d'être d'une recherche supplémentaire peut être mise en question. Or, si ces dernières années ont vu se populariser le concept d'utilisabilité, il va de pair avec un certain détournement de l'objectif initial. En effet, la notion en question s'est insérée majoritairement dans l'approche économique concurrentielle des stratégies de production actuelles aux dépens de son rapprochement intrinsèque de l'être humain, utilisateur d'un produit. Un deuxième courant de travaux a favorisé une approche se focalisant fortement sur

¹ Laboratoire de Psychologie Industrielle et Commerciale, Université Libre de Bruxelles, Av. F.-D. Roosevelt, 50 (CP 122), 1050 Bruxelles, Belgique, cecvdl@ulb.ac.be

² Laboratoire de Psychologie Industrielle et Commerciale, Université Libre de Bruxelles, Av. F.-D. Roosevelt, 50 (CP 122), 1050 Bruxelles, Belgique, vkoenig@ulb.ac.be

les aspects techniques des produits en proposant des standards à respecter dans la conception d'un produit à haut degré d'utilisabilité. Une troisième approche est constituée par le courant des recherches sur la gestion de l'information. La question de l'usage de l'information y est plus marquée (Le Coadic, 1998). Nous allons revenir sur ces différentes approches plus loin dans ce texte. Remarquons finalement que nous nous intéressons, dans ce contexte, exclusivement aux interfaces des systèmes informatisés (sites Internet, réseaux intranet p.ex.).

II. Utilisabilité : concept et pratiques

Par définition (norme ISO 9241-11), il faut entendre par utilisabilité, la capacité d'un produit à être utilisé par ses utilisateurs, cette évaluation tenant compte du produit, du contexte d'utilisation et des objectifs. Cette évaluation s'inscrit dans la démarche classique de l'ergonomie reprenant les aspects systémiques des situations étudiées. Le plus souvent, l'évaluation se réalise au sein d'un laboratoire appelé U-Lab (usability laboratory) qui constitue une installation spécifique à l'administration, à l'observation et à l'enregistrement des tests d'utilisabilité. La figure 1 présente la configuration spatiale de la salle d'expérimentation.

D'un point de vue pratique, les tests d'utilisabilité se déroulent généralement en trois phases selon un schéma expérimental classique : un local de test, un sujet, un expérimentateur, des consignes de test. La première phase correspond à l'accueil du sujet. Le sujet est accueilli, mis en confiance en lui expliquant le déroulement de la procédure ainsi que les précautions déontologiques. Ensuite, le test en soi se caractérise par une confrontation du sujet à une tâche généralement prédéfinie (en fonction de l'approche et/ou du but poursuivis) à effectuer sur l'interface en question. L'observation et l'enregistrement se font en parallèle. La troisième phase correspond à la phase analytique qui a lieu après un bref débriefing effectué avec le sujet. Cette dernière phase consiste en l'échange entre les observateurs et le sujet et peut se compléter par une analyse détaillée de l'enregistrement. C'est l'expérimentateur qui prend en charge l'accueil, donne les consignes, mène le débriefing et assure la modération du groupe de discussion. Remarquons que ce n'est généralement pas l'expérimentateur seul qui observe l'expérience en cours. Habituellement, la salle d'observation accueille les concepteurs et autres personnes intéressées à la conception du produit ce qui conduit évidemment à rassembler un groupe de personnes posant des regards différents sur les réactions de l'utilisateur face au produit.

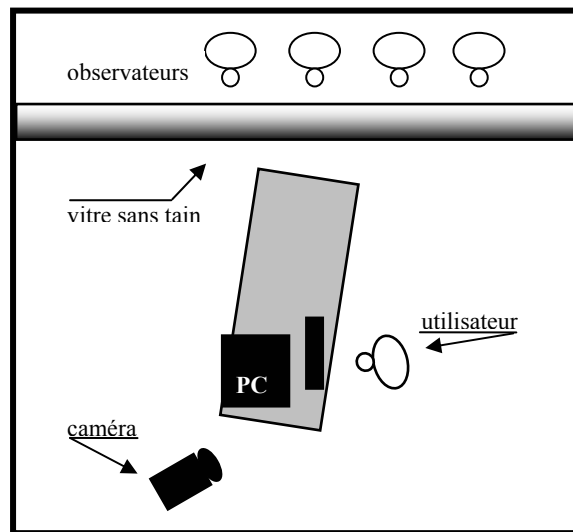


Figure 1 : configuration d'une salle d'expérimentation

Si le concept d'utilisabilité est devenu de plus en plus important ces dernières années, c'est principalement dû au fait que l'application de certains principes de base de ce concept apporte des avantages considérables. Différents auteurs (i.e. Shneiderman, 1998, Scapin, 1986, Sperandio, 1996) ont défini des critères spécifiques à prendre en compte lors de la conception de tels produits : compatibilité, homogénéité, contrôle explicite, flexibilité, guidage par feedback, et autres. Plus récemment, le concept d'utilisabilité s'est développé grâce aux travaux d'auteurs comme Jordan, Nielsen, Bevan ou Shneiderman.

Jordan (1998) donne une bonne vue d'ensemble des différents avantages qu'apporte une démarche intégrant des tests d'utilisabilité. Les entreprises concurrentielles sont arrivées aujourd'hui à des niveaux de qualité et d'innovation très similaires si bien qu'elles ne peuvent recourir que difficilement aux méthodes connues pour promouvoir leurs produits. L'aspect « design » devient alors de plus en plus important et permet aux entreprises de marquer leurs différences dans des segments de marché fort homogènes. Cette approche, à court ou à long terme, ne reflète cependant rien d'autre qu'un souci d'économie de temps et d'argent, associé à la pression constante de maximisation du profit. L'utilisabilité apparaît dès lors comme un concept à valeur économique prédominante. Toutefois, au sein même de cette valence économique, certaines entreprises attachent quelque importance à l'utilisateur final.

En effet, les produits sont choisis, achetés et utilisés par des consommateurs ou utilisateurs en raison de leur efficacité ou de leur efficience réelle ou perçue. L'utilisateur recherchera un produit facile à utiliser, fiable et offrant un certain confort d'utilisation. Dans ce cas, il se peut que les priorités de l'acheteur (l'entreprise) ne correspondent pas à celles de l'utilisateur (le travailleur) et cela aux dépens de ce dernier. De plus, l'utilisateur a trop souvent eu l'habitude de se voir frustré face à un produit peu utilisable, ne convenant pas à ses attentes spécifiques.

Les tests d'utilisabilité ont donc été privilégiés par rapport à une approche faisant appel à l'expertise de quelques spécialistes en utilisabilité. Il apparaît que l'approche systémique développée de longue date dans l'approche ergonomique, a également été intégrée dans ces tests.

Finalement, il importe de mentionner l'importance primordiale des recherches sur l'utilisabilité de systèmes à risques. A côté des critères économiques cités ci-dessus, la prise en compte de la fiabilité et de la maniabilité de systèmes complexes est également à souligner dans le développement de tests d'utilisabilité, que ce soit dans les secteurs des produits de consommation (tels que le secteur automobile, le secteur du multimédia) ou dans les secteurs professionnels (tels que les secteurs de l'aéronautique, les secteurs de production d'électricité, ...).

III. Deux approches de l'utilisabilité

Dans les paragraphes qui précèdent, nous avons situé les aspects les plus importants de l'utilisabilité dans son contexte général. Les approches que des « experts » en utilisabilité poursuivent diffèrent cependant souvent. Pour mieux comprendre ces différences, nous nous pencherons dans les paragraphes qui suivent sur une analyse plus approfondie des méthodologies poursuivies.

III.1. Les courants impliqués

Le concept d'utilisabilité, tel que nous l'abordons, s'articule autour de trois courants différents. La première place l'être humain au centre de la problématique ce qui qualifie cette approche de fondamentalement psychologique. Deuxièmement, mettant en avant des considérations d'ordre technique dans un contexte d'interface homme-ordinateur, il est indispensable de prendre en compte les données techniques informatiques. Finalement, la dimension ergonomique est également prédominante, le concept d'utilisabilité présentant le double objectif d'améliorer l'efficacité et l'efficience d'une part et la satisfaction de l'utilisateur d'autre part.

III.2. Approcher l'utilisabilité : deux manières différentes

A côté de ces trois courants, il apparaît qu'autour du concept d'utilisabilité se sont développées deux approches que nous appellerons respectivement approche d'intervention et approche de recherche. La première, typique des fins économiques, vise surtout à améliorer les produits, à les rendre plus performants, ceci dans un but quasi strictement concurrentiel. Les publications de Nielsen (1993) s'insèrent bien dans cette approche. L'approche de recherche par contre, poursuit un but de compréhension et d'identification des critères d'utilisabilité. Cette deuxième approche met l'individu au centre des tests et le définit comme « standard » par rapport auquel doit être conçu un produit à haut degré d'utilisabilité, alors que la première approche met le produit au centre des tests.

Si nous opposons ces deux approches dans un premier temps en fonction des finalités poursuivies, il importe cependant de relever que d'autres différences en découlent. Ainsi, les méthodes et les outils mis en œuvre ne sont pas nécessairement identiques et l'utilisation qui en est faite peut varier elle aussi. Pour les deux approches retenues, nous avons identifié les méthodes et les outils, les buts, les

Apports et limites des tests d'utilisabilité

résultats et les priorités suivis et nous les avons regroupés sous forme d'un tableau synthétique (tableau 1).

	Approche d'intervention	Approche de recherche
Méthodes / outils	Style souvent exploratoire, peu de soucis méthodologiques → tests exploratoires	Suivi strict de scénarios, rigueur méthodologique → tests scénarisés
	Enregistrement Interviews et discussions	Enregistrement Interviews et discussions
Buts	Favoriser la vente du produit Faire face à la concurrence Rechercher la rentabilité	Identifier les défauts Evaluer les stratégies individuelles Définir les critères d'évaluation
Résultats	Evaluation globale (unidimensionnelle)	Evaluation par critères (plurifactorielle)
Priorité	Evaluation du produit Amélioration du produit	Etude scientifique des critères d'évaluation Etude du comportement humain

Tableau 1 : tableau synthétique des approches d'intervention et de recherche

IV. Analyse critique des approches

Après cette synthèse critique du concept tel qu'il est abordé dans la littérature, nous développerons, dans la suite de cette communication, les apports de nos propres recherches en U-Lab. Ceux-ci nous ont permis d'être largement confrontés à la problématique présentée. Au fil des premières séances de test déjà, il apparaissait que le concept ne pouvait pas s'approcher d'une manière aussi facile que bon nombre de publications le font croire. Les propos qui précèdent témoignent de cette démarche de remise en question ainsi que de la réflexion sur l'approche à adopter. Après quelques mois de fonctionnement, nous pouvons dresser les constats suivants à partir de recherches effectuées soit dans l'approche d'intervention soit dans l'approche de recherche. Toutes les recherches ont concerné des tests d'utilisabilité de sites Internet (sites commerciaux, sites de loisirs, sites expérimentaux, sites éducatifs).

IV.1. L'approche de la recherche renvoie au problème de la « mesure » et des critères

La question des méthodes que nous avons évoquée plus haut renvoie évidemment aussi à la problématique des critères de l'utilisabilité. Si l'approche d'intervention se restreint généralement à l'observation pure et l'enregistrement vidéo (tests exploratoires), l'approche de recherche, par contre, aspire à des connaissances plus approfondies grâce à des paradigmes expérimentaux plus sophistiqués (tests scénarisés). Dans le premier cas, l'enregistrement ne constitue généralement qu'une trace du test tandis que l'approche de recherche en fait un usage de support à l'analyse. C'est dans ce dernier cas que le problème de la définition des critères et de la mesure prend toute son importance.

Ainsi les critères ergonomiques tels que Scapin et Bastien les définissent (guidage, charge de travail, contrôle explicite, adaptabilité, gestion des erreurs, etc.), sont des concepts souvent difficiles à opérationnaliser en tant que paramètres dans un paradigme expérimental qui vise à les contrôler, voire les mesurer. Ils servent donc de repères à une approche exploratoire mais s'avèrent incomplets face aux objectifs de recherche. En effet, ces critères peuvent se voir évalués au sens binaire d'une prise en compte ou non ; la détermination des degrés respectifs auxquels ces facteurs seraient respectés s'avère difficile. La mesure reste donc un problème à résoudre.

IV.2. L'interface, l'utilisateur et leurs relations

Le concepteur a souvent tendance à rejeter les défauts d'application sur l'utilisateur et à déclarer son produit technique comme excellent et abouti. Le chercheur qui s'intéresse à l'utilisabilité a lui trop souvent tendance à déclarer que toute défaillance est à mettre sur le compte du produit. Cette opposition montre bien que le même phénomène peut s'approcher de différentes manières. Nous ne croyons cependant pas que la « réalité » se trouve approchée par une de ces positions extrêmes. L'utilisateur ne pourra jamais agir en faisant abstraction de facteurs largement individuels et variables comme les émotions ou la motivation par exemple. Dès lors il devient flagrant combien la recherche d'une certaine homogénéité interindividuelle peut souffrir du phénomène d'hétérogénéité intra individuelle.

IV.3. L'approche d'intervention favorise le résultat, moins les méthodes

Dans l'approche d'intervention, le test d'utilisabilité gagne sa raison d'être par les résultats qu'il produit. Dans cette approche, Nielsen (1993) propose que cinq à sept sujets seraient largement suffisants pour « évaluer » l'utilisabilité d'un site Internet. Aussi propose-t-il de multiplier le nombre de tests et non le nombre de sujets, une démarche soulevant des questions de fiabilité des résultats et suggérant donc une approche différente en recherche. Il est vrai que les soucis méthodologiques poursuivis dans une approche de recherche sont incompatibles avec ce genre de propos et que les soucis méthodologiques priment largement sur le résultat même du test. Nous n'avons pu récolter, lors de nos tests, aucun indice soutenant l'idée qu'un nombre très restreint de sujets suffirait à évaluer l'utilisabilité d'un site Internet.

IV.4. Simulation versus travail réel

En différenciant les deux approches, nous avons constaté que les soucis méthodologiques en recherche pèsent sur le degré de réalisme des tests d'utilisabilité. Jordan (1998) avance que l'aspect artificiel de ce genre de tests peut agir sur les réactions de l'utilisateur. En effet, plus la situation de travail est simulée à un degré de réalisme élevé, plus il devient difficile de respecter la contrainte de réplicabilité (standardisation) d'une expérience scientifique. La situation de travail comporte en soi des zones de liberté de décision et d'action que le test ne peut que difficilement prévoir. Inversement, plus nous augmentons les exigences méthodologiques, plus le sujet se retrouve dans une situation parfois trop éloignée du travail réel que pour nous informer sur son comportement probable en situation réelle de travail. La solution à ce problème ne semble pas évidente et nous pensons que le scénario du test doit être suffisamment complexe pour pouvoir rendre compte de cette multitude de comportements possibles. De plus, des « zones de liberté » pourraient faire partie du scénario, tout en étant conçues de manière à ne pas influencer les tâches en question. Il restera cependant que nous ne pouvons, au meilleur de cas, que prêter beaucoup d'attention à cette problématique et bien élaborer le scénario ; le test restera cependant inévitablement une situation « artificielle », une situation de laboratoire.

IV.5. Composante évaluative de l'utilisateur

Comme nous l'avons expliqué plus haut (cf. déroulement des tests d'utilisabilité), ces expériences ne peuvent se passer d'explications sur le déroulement et sur les précautions déontologiques. Une très grande importance doit toujours être accordée à la conscience que devrait avoir le sujet du fait qu'il n'est pas l'objet intrinsèque de l'étude pour garantir son état de bien-être et une relative absence de stress. De plus, il importe évidemment de garantir le plus possible une reproduction d'une situation réelle de travail pour l'utilisateur.

En partant des réflexions sur la situation expérimentale, nous avons essayé d'instaurer de plus en plus clairement une « zone de liberté » pour l'utilisateur. Ainsi, dans nos derniers tests, nous avons inclus une introduction orale aussi bien que rédigée qui invitait le sujet explicitement à aborder les questions dans l'ordre préféré par l'utilisateur, de profiter du temps largement alloué et d'explorer les sites Internet en question comme bon lui semble. Le résultat en était surprenant : aucun des sujets ne prenait vraiment la liberté d'explorer le site. De plus, la tâche était remplie dans un minimum de temps possible ainsi que dans l'ordre des questions posées. L'ordre n'était violé que pour les seuls cas où le sujet se trouvait dans l'incapacité de répondre à une question et qu'il y revenait à un moment ultérieur. Cette démarche était alors complétée par l'insertion d'une question invitant le sujet à explorer le site pour trouver l'information demandée. Face à cette condition, les sujets ont « ignoré » la question et ainsi évité le comportement d'exploration. Cependant, cette dernière condition devra être testée avec un échantillon de personnes plus important.

Questionnés sur leurs comportements, la majorité des utilisateurs ont expliqué l'absence de ce comportement par le vécu subjectif d'un test servant à les évaluer à l'exemple d'un test scolaire (« Oui, mais dans ma tête, j'ai cru que c'était un test,

comme à l'école... »). Il semble donc qu'en dépit de toutes les précautions, les craintes du sujet ainsi que son estime de soi restent des facteurs extrêmement importants et dont l'influence ne semble que difficilement contrôlable. Les utilisateurs présentent donc une propension au vécu d'un « test » en dépit des consignes explicites contraires. Dans ce contexte, Jordan (2000) propose d'ailleurs de prendre en compte des traits de personnalité et de les intégrer à l'investigation.

IV.6. L'utilisateur face à son observation

Les résultats précédents montrent à l'évidence que le vécu émotionnel et subjectif de l'utilisateur ne doit pas être sous-estimé. Nos tests ont montré que les sujets peuvent éprouver beaucoup de difficultés face à la situation d'observation pendant les tests d'utilisabilité. Les explications et consignes fournies au sujet étant déjà adaptées spécifiquement pour favoriser un climat le moins stressant possible, nous nous sommes mis à varier la disposition de l'installation à l'intérieur de la salle de test. Partant de la position de la vitre sans tain (immuable), nous avons essayé de positionner l'utilisateur de manière à ce qu'il soit le moins dérangé par la présence de ce dispositif d'observation. Pour ce faire, nous avons procédé à des questionnements spécifiques et nous avons confronté les réponses des sujets aux informations tirées du débriefing (satisfaction du sujet, malaise éventuel, stress éventuel, facilité relative à sortir de la situation d'observation, etc.). Finalement, nous avons trouvé qu'une position approchant celle de la vue de profil avait le plus de chances de faire oublier la présence de la vitre sans tain (cf. figure 1). La position « dos par rapport à la vitre » n'était pas acceptée par les sujets ; la position « face à la vitre » convenait mais rappelait trop aux sujets la présence des observateurs.

En fonction des réactions que nous avons pu observer, il semble que la position de la caméra (ou des caméras) est moins susceptible de déranger le sujet que la vitre sans tain. L'avantage en est que nous pouvons positionner le sujet prioritairement par rapport à la vitre sans tain et par rapport à l'interface à évaluer et que la caméra peut être disposée de manière à compléter au mieux l'observation directe à travers la vitre sans tain. Dans les limites du possible, nous évitons alors de mettre la caméra en position trop frontale par rapport à l'utilisateur. Cette configuration a l'avantage d'être jugée moins intrusive et de permettre en même temps une observation efficace.

V. Vers une approche de « méta-utilisabilité »

En dressant le bilan des constatations que nous venons de faire, il apparaît que bon nombre de facteurs interviennent avant l'expérience d'utilisabilité même. Une analyse en détail des remises en question et des améliorations partielles que nous proposons montre que ces réflexions concernent d'autant plus la méthodologie à mettre en place l'utilisabilité de la situation elle-même. Il devient alors clair que l'utilisabilité ne nous concerne pas seulement en tant qu'objet des tests en question mais que son importance dépasse le statut de variable dépendante. En effet, ces réflexions nous montrent que l'utilisabilité ne doit pas être considérée comme concept à part, délié des motivations et craintes de sujets. La dimension psychologique de la situation de test vient jouer un rôle extrêmement important dans

la systémique des tests d'utilisabilité. En effet, cette dimension agit sur les mesures tout aussi bien que sur la dynamique relationnelle de l'utilisateur (avec l'expérimentateur, avec les observateurs et avec le produit) et sur les aspects déontologiques. Il devient clair que l'utilisabilité constitue un concept complexe qui dépasse la problématique de la pure évaluation.

Ainsi, les critères présumés de l'utilisabilité doivent s'appliquer également à la situation d'observation même avant de prétendre tester ou mesurer un concept d'utilisabilité. En d'autres termes, les variables indépendantes sont concernées, elles aussi, par des apports du concept en question. Finalement, passant en revue nos réflexions, nous pouvons affirmer que les tests d'utilisabilité dans un contexte de recherche ne peuvent se passer d'une approche que nous qualifions de méta-utilisabilité.

Bibliographie

- Bevan, N. (1997). Quality of use : incorporating human factors into the software engineering lifecycle. *Proceedings of the third international symposium and forum on software engineering standards, ISEES'97 conférence*.
- Green, W. S., & Jordan, P. W. (1999). *Human factors in product design*. London, Taylor & Francis.
- Jordan, P. W. (1998). *An introduction to usability*. London, Taylor & Francis
- Jordan, P. W. (2000). *Designing pleasurable products*. London, Taylor & Francis
- Le Coadic, Y.. (1998). *Usages et usagers de l'information*. Paris, Edition Nathan Université.
- Nielsen J. (1993). *Usability engineering*. New York, Academic Press.
- Scapin, D. (1986). *Guide ergonomique de conception des interfaces homme-ordinateur*. Institut National de Recherche en Informatique et en Automatique (INRIA).
- Shneiderman, B. (1998). *Designing the user interface*. Reading, Massachusetts, Addison Wesley Longman, 3ème édition.
- Sperandio, J-Cl. (1996); *L'ergonomie face aux changements technologiques et organisationnels du travail humain*. Toulouse, Ed. Octarès.