

Enquête sur les pratiques de tests d'utilisabilité

Naouel Moha, Li Qing, Ashraf Gaffar, Ahmed Seffah

Human-Centered Software Engineering Group
Université Concordia
1455 de Maisonneuve Blvd West
H3G 1M8, Montréal, Canada
{n_moha, q_lix03, gaffar, seffah}@encs.concordia.ca

RESUME

Cet article présente les résultats d'une récente enquête réalisée sur les tests d'utilisabilité impliquant approximativement une centaine de praticiens de l'utilisabilité. Cette enquête avait pour but d'identifier les pratiques des professionnels de l'utilisabilité dans la conduite de leurs tests. L'analyse des résultats de l'enquête a permis de présenter un état de l'art dans ce domaine et d'en relever les forces et faiblesses. Comme faiblesse, nous avons relevé entre autre le manque d'outils spécifiques pour la collecte et l'analyse des données de tests ; comme force, la progressive valorisation de ce domaine d'expertise au niveau des entreprises.

MOTS CLES : Praticiens de l'utilisabilité, tests d'utilisabilité, méthodes, processus, enquête en ligne, questionnaire.

ABSTRACT

This paper reports the results of a recent survey on usability testing involving nearly a hundred usability practitioners. The purpose of this survey was to identify the practices in usability testing. The analysis of those results allows to present a state of art in this domain and to identify the strengths and the weaknesses. As weakness, we have noted the lack of specific tools for collecting and analyzing test data; as strength, the promotion of this expert domain inside companies.

CATEGORIES AND SUBJECT DESCRIPTORS: H.5.2 [User Interfaces]: Evaluation/methodology|usability testing; D.2.5 [Testing and Debugging]: Testing tools.

GENERAL TERMS: Human Factors, Measurement, Verification.

KEYWORDS : Usability practitioners, usability testing,

methods, process, web survey, questionnaire.

INTRODUCTION

Cet article décrit les résultats d'une récente enquête de recherche destinée à interroger les professionnels de l'utilisabilité sur leurs pratiques dans la conduite des tests d'utilisabilité. Ces pratiques incluent les méthodes, les activités, les outils et les infrastructures de tests d'utilisabilité. Cet article s'intéresse également au positionnement de l'utilisabilité au sein des entreprises.

Le terme '*tests d'utilisabilité*' est utilisé pour faire référence au processus qui emploie des participants qui sont représentatifs d'une population cible pour évaluer le degré auquel un logiciel est conforme à des critères d'utilisabilité spécifiques [1]. La norme ISO 9241-11 définit l'utilisabilité de la manière suivante : « un système est utilisable lorsqu'il permet à l'utilisateur de réaliser sa tâche avec efficacité, efficacité et satisfaction dans le contexte d'utilisation spécifié ».

Peu de travaux ont été réalisés sur la modélisation et la formalisation des tests d'utilisabilité impliquant des utilisateurs. Nous pensons que la qualité du processus des tests d'utilisabilité a un grand impact sur la pertinence des données collectées et des problèmes identifiés à la fin des tests d'utilisabilité. Gellner présente une critique pertinente sur la faible modélisation des processus de tests d'utilisabilité existants [1, 2]. La plupart des processus sont des synthèses des expériences de travail de conduite des tests d'utilisabilité [3].

Un des buts de notre recherche est de définir un processus pour conduire des tests d'utilisabilité qui décrit l'ensemble des activités de tests, ainsi que les méthodes et outils pour réaliser ces activités. Mais avant d'être capables de présenter un tel processus, nous avons conduit cette enquête afin d'identifier les meilleures pratiques ainsi que les forces et faiblesses dans la conduite des tests d'utilisabilité. 86 professionnels de l'utilisabilité y ont participé. Les résultats de cette enquête vont nous permettre de corriger, redéfinir et valider le processus de conduite des tests d'utilisabilité que nous avons défini au sein de notre équipe.

Dans la première section de cet article, nous passons en revue les travaux similaires réalisés dans ce domaine. La section suivante présente les motivations qui ont conduit à une telle étude. Cette section est suivie de la description de la méthodologie suivie pour planifier et conduire cette enquête. Enfin, la dernière section présente l'analyse et les résultats de l'enquête.

TRAVAUX SIMILAIRES

Plusieurs enquêtes ont été réalisées sur la conception centrée utilisateur ou UCD (*User-Centered Design*) [4-6]. La conception centrée utilisateur est définie comme étant une approche de conception itérative et multidisciplinaire basée sur une participation active des utilisateurs pour une meilleure compréhension de leurs besoins et de leurs tâches [5]. L'enquête réalisée par Vredenburg et al. portait essentiellement sur l'étude des pratiques UCD en terme de budget, d'impact organisationnel, de mesures de succès, et de méthodes et techniques les plus largement utilisées [5]. Cette enquête a montré que bien que les praticiens en UCD sont convaincus de leur impact positif au niveau des entreprises, ils ont des difficultés à estimer l'importance de leur contribution à travers des mesures claires et précises. L'enquête réalisée par Venturi et Troost investiguait l'intégration et l'adoption de l'UCD au sein des entreprises en termes de management, d'infrastructure et de communication [6]. Les tests d'utilisabilité constituent une technique spécifique utilisée dans la conception centrée utilisateur pour mesurer l'utilisabilité d'un produit. Plusieurs autres enquêtes se sont focalisées sur le test logiciel [7, 8]. Cependant, nous n'avons trouvé aucune enquête ou étude réalisée sur les tests d'utilisabilité qui permette d'avoir une idée précise sur les pratiques actuelles dans ce domaine.

MOTIVATIONS

Un certain nombre de raisons nous a poussés à conduire cette enquête. Premièrement, l'étude des résultats de cette enquête constitue un état de l'art des pratiques actuelles dans la conduite des tests d'utilisabilité. Les questions de cette enquête ont été conçues afin d'identifier les meilleures pratiques. Ces pratiques seront ensuite diffusées à l'ensemble des praticiens de l'utilisabilité. Deuxièmement, l'analyse de ces résultats permettra de valider un processus plus formel de conduite de tests d'utilisabilité basé sur l'expérience des professionnels du domaine. Enfin, ces résultats permettront de fournir des indications pour de futures directions de recherche et ainsi améliorer les pratiques de tests.

METHODOLOGIE

Le Questionnaire

Cette enquête a débuté en octobre 2004 et s'est déroulée sur une période de quatre mois. Elle se présentait sous la forme d'un questionnaire, subdivisé en 4 sections incluant un total de 12 questions [9]. Les deux premières sections étaient destinées à collecter des informations re-

latives aux répondants et aux entreprises dans lesquelles ils travaillent. La troisième section était consacrée aux méthodes et aux activités de tests. La dernière section présentait des questions relatives à l'infrastructure et aux outils utilisés durant les tests d'utilisabilité. A la fin du questionnaire, les répondants ont été invités à faire des commentaires libres et à donner des éléments de réponses à la question : *D'après vous, qu'est-il nécessaire d'améliorer dans les tests d'utilisabilité en termes d'activités, de méthodes et d'outils ?*

D'après les pré-tests que nous avons réalisés au préalable sur quatre sujets, le questionnaire pouvait être complété entre 10 et 15 minutes. La courte durée pour compléter ce questionnaire a été un des arguments présentés pour susciter le plus de réponses possibles. De plus, nous avons assuré aux répondants que l'information collectée à travers ce questionnaire sera strictement confidentielle et ne sera utilisée qu'à des fins de recherche. Aucune information personnelle (sauf l'adresse électronique et le nom du répondant) ni aucune information permettant d'identifier l'entreprise n'a été demandée.

Le questionnaire se présentait sous la forme d'un formulaire à remplir sur Internet et incluait un système de validation afin d'éviter des soumissions partielles. Toutes les données ont été collectées dans une base de données MySQL pour une analyse automatisée via l'outil d'analyse statistique SPSS [10].

La Population Cible

La population cible des répondants de cette étude était principalement constituée de professionnels de l'utilisabilité et de chercheurs qui ont une expérience dans la conduite des tests d'utilisabilité. Le questionnaire a été envoyé par courrier électronique à des listes de diffusion et des newsgroups liés à l'utilisabilité : ACM-SIGCHI, UPA Montréal, UPA Toronto, AFIHM, ErgoIHM, et Interactif parmi d'autres. 86 praticiens ont complété l'enquête sur une durée de trois mois. Les répondants provenaient principalement du Canada (24.4%), des Etats-Unis (32.6%) et de France (34.9%).

57 répondants se sont identifiés comme étant des spécialistes de l'utilisabilité et des interfaces utilisateurs. 25 comme étant des spécialistes du facteur humain et 14 comme des concepteurs d'interface web (cf. Tableau 1).

Les répondants avaient la possibilité de sélectionner plusieurs positions professionnelles parmi la liste qui leur a été proposée et d'éventuellement spécifier d'autres catégories de positions. Ceux qui appartiennent à la catégorie 'Autre' sont essentiellement des chercheurs du domaine. Il s'agit donc d'un échantillon pertinent composé principalement de professionnels de l'utilisabilité ayant une expérience en interaction homme machine et en ergonomie.

Catégorie	Fréquence
Consultant	36
Spécialiste en facteur humain	25
Marketing	4
Programmeur/Développeur	9
Testeur de logiciel	4
Ingénieur système	1
Support technique	2
Spécialiste utilisabilité et interface utilisateur	57
Concepteur Web	14
Autre	20

Tableau 1 : Catégories professionnelles des répondants.

La Marge d'erreur

Est-ce que ces 86 répondants constituent un échantillon représentatif ? Pour répondre à cette question, nous avons calculé la marge d'erreur. La marge d'erreur représente la mesure de précision des résultats d'une enquête et est égale à 1 divisé par la racine carrée du nombre de personnes de l'échantillon. Ainsi, un échantillon de 86 personnes nous donne une marge d'erreur de plus ou moins 10.8% ce qui signifie qu'il y a une forte probabilité (95% de chance) que la plupart des personnes de la population cible dans son ensemble donneront des réponses similaires (avec une précision de 10.8%) aux réponses obtenues de l'échantillon.

Validation des Données

Avant d'entreprendre des analyses détaillées, nous avons parcouru les réponses afin de détecter des incohérences. Le système de validation en JavaScript que nous avons défini permettait d'éviter d'omettre des questions ou de fournir des réponses biaisées comme une adresse électronique erronée ou une chaîne de caractères au lieu d'un chiffre. De plus, le questionnaire était constitué de questions à choix multiples et de listes de sélection pour éviter d'avoir des données incohérentes. Aucune donnée manquante ou incorrecte n'a été détectée.

Partitionnement des Données

Après avoir validé les données, nous avons partitionné les réponses en différents sous-groupes basés sur l'expertise des répondants et des catégories d'entreprises auxquelles ils appartiennent. La subdivision des sous-groupes des répondants est détaillée au niveau de la section ci-dessous dédiée au profil des répondants. Le partitionnement aide à interpréter plus facilement et plus fidèlement les données.

RESULTATS ET ANALYSE DE L'ENQUETE

Les résultats de l'enquête sont organisés comme suit. Dans les deux premières parties, nous présentons des informations relatives au profil des répondants et aux entreprises dans lesquelles ils travaillent. Puis, une partie est dédiée aux méthodes, activités et matériels de tests d'utilisabilité. Les parties suivantes sont quant à elles

consacrées aux infrastructures et aux outils de tests. La dernière partie donne les résultats de l'analyse qualitative des commentaires récoltés.

Le Profil des Répondants

Comme l'indique l'analyse de la population cible, les répondants sont principalement des professionnels de l'utilisabilité ayant une expérience significative dans le domaine.

Les répondants ont été répartis en trois différentes catégories selon deux critères : le nombre de tests qu'ils ont conduit durant les trois dernières années et le niveau de leur expertise d'après leur autoévaluation. Ces trois catégories sont les suivantes : les novices, les intermédiaires et les experts (cf. Figure 1). Nous avons dénombré 8 experts, 66 intermédiaires et 12 novices.

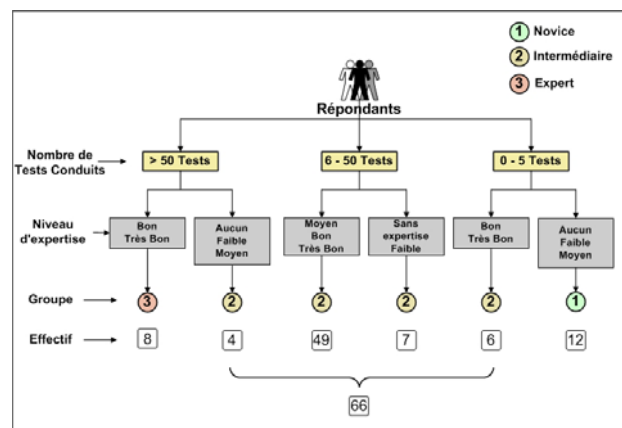


Figure 1 : Partitionnement des répondants.

Ce partitionnement a permis d'interpréter les réponses de la majorité des questions de l'enquête en se basant sur l'expertise des répondants. Ainsi, les réponses fournies par des experts avaient un poids de valeur 3 plus important que les réponses fournies par des intermédiaires ou des novices ; de même que les réponses fournies par les intermédiaires avaient un poids de valeur 2 par rapport aux novices dont le poids de valeur était seulement de 1.

Le Profil des Entreprises

Le tableau suivant donne la proportion en fréquence et en pourcentage des différentes catégories d'entreprises auxquelles les répondants appartiennent.

Il apparaît que 24.4% des répondants proviennent d'entreprises qui comptent plus de 5000 employés. 12.8% des répondants viennent d'entreprises qui comptent entre 1001 et 5000 employés. Néanmoins, 27.9% des répondants sont soit des praticiens indépendants (11.6%) soit appartiennent à une entreprise de moins de 10 employés (16.3%).

Nombre d'employés	Fréquence	Pourcentage
Consultant indépendant	10	11.6
2-10 employés	14	16.3
11-20 employés	5	5.8
21-50 employés	4	4.7
51-100 employés	5	5.8
101-250 employés	2	2.3
251-500 employés	6	7.0
501-1000 employés	5	5.8
1001-5000 employés	11	12.8
Plus de 5000 employés	21	24.4
Je ne sais pas	3	3.5
Total	86	100.0

Tableau 2 : Catégorie des entreprises.

33.7% de l'ensemble de ces entreprises n'ont qu'entre 3 et 5 ans d'expérience dans la conduite des tests d'utilisabilité (cf. Figure 2). Très peu d'entreprises ont plus de 15 années d'expérience.

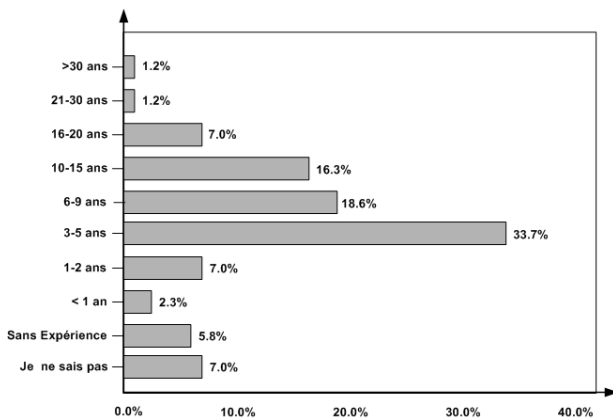


Figure 2 : Nombre d'années d'expérience dans les tests d'utilisabilité des entreprises.

Le tableau ci-dessous donne le nombre d'années d'expérience en utilisabilité par rapport à la taille de l'entreprise.

Taille de l'entreprise	Expérience
Grande (+ de 1000 employés)	~ 6 ans
Moyenne (100 < employés < 1000)	3 à 5 ans
Petite (< 100 employés)	3 à 5 ans

Tableau 3 : Années d'expérience en utilisabilité par rapport à la taille de l'entreprise.

Cette analyse permet de montrer combien la conduite des tests d'utilisabilité est une pratique encore jeune au sein des entreprises que ce soit dans les grandes ou les petites entreprises et ceci explique en partie son manque

de formalité et de maturité, et surtout son manque de valorisation au sein des entreprises.

Nous nous sommes également intéressés au nombre de spécialistes de l'utilisabilité qui travaillent dans ces entreprises. Dans les grandes entreprises de plus de 1000 employés, un praticien en utilisabilité représente moins de 1% des employés de l'entreprise soit 10 personnes. Le tableau ci-dessous présente la proportion des professionnels de l'utilisabilité par rapport à l'ensemble des employés de l'entreprise.

Taille de l'entreprise	Proportion
Grande (+ de 1000 employés)	< 1% ~ 10 personnes
Moyenne (100 < employés < 1000)	< 5% ~ 10 personnes
Petite (< 100 employés)	~ 10% ~ 10 personnes

Tableau 4 : Proportion des professionnels de l'utilisabilité par rapport à la taille de l'entreprise

Ces résultats montrent que peu importe la taille de l'entreprise, les équipes de tests d'utilisabilité sont généralement composées d'un petit groupe d'environ 10 personnes. A noter que l'enquête conduit par Venturi et Troost a également montré que les praticiens en UCD sont employés par des grandes entreprises et représentent moins de un pour cent de l'ensemble des employés [6]. Il serait pertinent dans le cadre d'une autre enquête d'évaluer le nombre de professionnels de l'utilisabilité par rapport au nombre de développeurs et d'identifier le secteur de l'entreprise et les budgets alloués aux tests d'utilisabilité afin de mieux interpréter ces résultats.

Les Méthodes, Activités et Matériels de Tests d'Utilisabilité

Un certain nombre de questions étaient dédiées aux activités, aux méthodes et matériels de tests d'utilisabilité. Ceux-ci ont été sélectionnés afin de pouvoir valider et raffiner le processus de conduite de tests d'utilisabilité que nous avons défini au sein de notre équipe de recherche et qui intègrent ces méthodes, activités et matériels. Ces questions consistaient à évaluer l'importance et la pertinence sur une échelle constituée des 5 valeurs suivantes : Très faible, Faible, Moyen, Bon et Très bon.

La figure 3 présente les résultats de l'évaluation de l'importance de l'activité 'Sélectionner des séquences vidéo' en fonction de l'expertise des répondants. Comme le montre la figure, la majorité des experts ont répondu qu'il s'agissait d'une activité peu importante alors que les novices ont répondu que cette activité avait un niveau d'importance moyen. Les avis sont partagés selon le niveau d'expertise des répondants. Il apparaît donc que cette activité n'est pas très pertinente dans la conduite des tests d'utilisabilité.

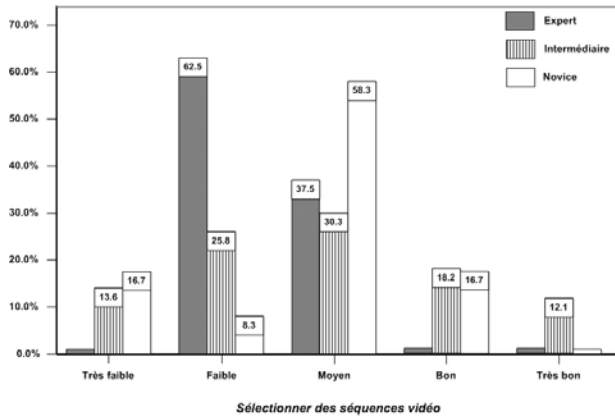


Figure 3 : Activité de tests : Sélectionner des séquences vidéo.

Une analyse similaire a été réalisée pour l'ensemble des activités, méthodes et matériels de tests. Les trois tableaux ci-dessous résument les résultats de l'analyse de l'évaluation de l'importance des méthodes, des activités et des matériels de tests dans la conduite de tests d'utilisabilité.

Activité	Niveau
Définir un profil utilisateur	Très bon
Définir les profils des équipes de tests et des testeurs	Moyen
Conduire un pré-test vous-même	Moyen
Conduire un test pilote	Très bon
Collecter les commentaires informels du participant après le test	Moyen
Revoir les questionnaires de pré-test et de post-test avec le participant	Moyen
Ajouter les commentaires des testeurs et des observateurs	Bon
Identifier les patterns et les conclusions des résultats de tests	Très bon
Préparer un rapport préliminaire	Bon
Sélectionner des séquences vidéo	Faible
Produire ou modifier des guidelines de conception	Bon

Tableau 5 : Niveau d'importance des activités de tests d'utilisabilité.

Méthode	Niveau
L'enquête contextuelle Étude de l'activité de vrais utilisateurs directement sur leur lieu de travail et dans les conditions réelles.	Bon
Les incidents critiques Les utilisateurs identifient des incidents spécifiques qu'ils ont rencontrés personnellement.	Bon
L'étude ethnographique Étude sur le comportement et les pratiques d'un groupe d'individus, par exemple, à partir d'observations sur le terrain ou à partir de questionnaires.	Bon
Les entretiens Interrogation des utilisateurs aux moyens de questions posées verbalement sur leur expérience et leurs préférences avec le produit.	Très bon

Le prototypage Évaluation de l'utilisabilité du produit final à partir d'un prototype de celui-ci.	Très bon
Les questionnaires Listes écrites de questions distribuées aux utilisateurs.	Moyen
Les scénarios de tests Scénarios spécifiant comment les participants doivent exécuter certaines tâches dans un contexte précis.	Très bon
Le prototypage papier (Storyboard) Prototype de basse fidélité constitué d'une série d'esquisses d'écrans sur papier.	Moyen
Penser à voix haute (Thinking aloud) Encouragement des participants à dire haut ce qu'ils pensent.	Bon
L'évaluation cognitive Procédure spécifique pour simuler les processus cognitifs des utilisateurs quand ils sont en interaction avec une interface.	Bon
Le magicien d'OZ Technologie non implémentée pour être évalué en utilisant un humain pour simuler la réponse du système.	Bon
Le focus group Collecte d'information lors de réunions informelles des utilisateurs pour discuter sur un sujet spécifique.	Faible

Tableau 6 : Niveau d'importance des méthodes de tests d'utilisabilité.

Tout comme l'enquête menée par Venturi et Troost, notre analyse a montré que les entretiens et le prototypage sont des techniques très recommandées par les professionnels de l'utilisabilité. Les résultats de l'enquête réalisée par Vredenburg et al. sont en accord avec nos analyses : ils ont également trouvé que le focus group avait une importance faible par rapport aux autres méthodes UCD et que les enquêtes contextuelles ont un bon niveau d'importance. Par contre, leurs résultats ont montré que les enquêtes ont un niveau d'importance assez faible.

Matériel	Niveau
Questionnaire de sélection des participants aux tests	Bon
Questionnaire de pré-test	Bon
Questionnaire de post-test	Bon
Plan des sessions de test	Bon
Accord de confidentialité	Moyen
Accord d'enregistrement	Bon
Check list de l'environnement de test	Bon
Formulaire de collection de données	Bon
Rapport préliminaire de tests	Moyen
Rapport final de tests	Bon

Tableau 7 : Niveau d'importance des matériels de tests d'utilisabilité.

Ces résultats ont permis de relever quels sont les activités, méthodes et matériels de tests à considérer dans la conduite des tests d'utilisabilité selon les praticiens et experts du domaine.

Les Logiciels de Tests d'Utilisabilité

Une question de l'enquête consistait à demander aux répondants de spécifier les différents outils et logiciels

qu'ils utilisent lors de la conduite de leurs tests d'utilisabilité. Le tableau 8 présente la liste des outils et logiciels spécifiés par les répondants. Certains utilisent des logiciels de bureau tels que Excel ou Word, d'autres conçoivent leurs propres outils de tests (cf. outils en gras dans le tableau 8). Morae semble toutefois devenir de plus en plus populaire. Morae est un logiciel conçu par TechSmith qui permet d'enregistrer une session de tests à distance et d'annoter la vidéo de tests en temps réel pour des analyses futures [11]. L'analyse des outils montre qu'il n'y a pas d'outils ou de logiciels de tests prédominants pour la conduite des tests d'utilisabilité comme ceux qu'on pourrait trouver pour le développement de logiciels.

Outils	Fréquence
Bailey's Testing	1
Camera	1
Camlan	1
Custom logging database	1
Digital Video Camera + Excel	1
Excel (data analysis)	2
Free software	1
HTML Prototype and video	1
In house data logger	1
Kronos	1
Leximancer	1
Macshapa	1
Mobile wireless UT lab	1
Morae	8
MSN Encoder	1
Own ad hoc software tools	2
Own logger	1
Own logging software	1
RapidPlus	1
Self-developed tool	2
Test logger (very old DOS based)	1
Usability Datalogger (see TZ)	1
Video recording	2
Virage	1
WebEx for remote testing	1
Word	1
Mobile wired UT Lab	1
Notes by hand	1
Specific software patches	1
WebSurveyor	1

Tableau 8 : Outils et logiciels de tests d'utilisabilité.

Les Infrastructures de Tests

Différentes infrastructures permettent de conduire des tests d'utilisabilité : les plus connus sont les laboratoires fixes et mobiles d'utilisabilité. Les tests à distance via Internet sont une autre nouvelle forme d'infrastructure qui émerge dans ce domaine. Les tests à distance sont des tests d'utilisabilité où les testeurs en charge d'opérer

l'observation et l'analyse des participants sont séparés dans l'espace et/ou dans le temps des participants [12]. Dans certains cas, la conduite des tests d'utilisabilité ne nécessite pas de réelles infrastructures et sont conduits principalement sur le site ou lieu de travail du participant aux tests.

Au vu de l'analyse des résultats de l'enquête, il apparaît que les laboratoires d'utilisabilité fixes et les tests sur site restent prédominants à 68% dans la conduite des tests d'utilisabilité. Cependant, même si les laboratoires mobiles et les tests à distance sont encore peu répandus, ils deviennent de plus en plus exploités (Cf. Tableau 9). Par contre, les laboratoires d'utilisabilité de location ou dits externes sont utilisés de manière occasionnelle par 40% des répondants. A Montréal (Canada), le coût de location d'un laboratoire d'utilisabilité est assez excessif; il s'élève à environ 600€ par jour et atteint 1200€ par jour si des spécialistes sont mis à disposition.

	Jamais	Occasionnel	Souvent
Laboratoire mobile	49%	31%	20%
Tests à distance	60%	29%	10%
Moyenne	54.5%	30%	15%

Tableau 9 : Proportion des fréquences d'utilisation des infrastructures de tests.

Les consultants travaillent souvent sur site et utilisent de manière occasionnelle les laboratoires fixes d'utilisabilité. Néanmoins, ce sont les premiers avec les petites entreprises de moins de 100 employés à utiliser des infrastructures mobiles et à distance.

Nous pensons que les tests à distance via Internet gagnent en popularité et l'analyse des commentaires libres de cette enquête le témoigne. En effet, nombreux sont les professionnels de l'utilisabilité qui en ont vanté les mérites. Les tests d'utilisabilité à distance ont cet avantage sur les laboratoires traditionnels de permettre le test d'un nombre de participants plus élevé et plus divers dans leur propre environnement de travail, à budget plus réduit et en un temps relativement court [13].

ANALYSE QUALITATIVE

L'analyse qualitative porte essentiellement sur les commentaires récoltés à la fin du questionnaire. Cette analyse a permis de relever un certain nombre de points importants liés à la formalisation de la conduite des tests d'utilisabilité, au développement d'outils spécifiques et à la valorisation du domaine au sein des entreprises.

Les répondants ont souligné l'effort initié par la communauté des professionnels de l'utilisabilité pour une standardisation des procédures de tests et des rapports des résultats de tests. UPA qui est l'association des professionnels de l'utilisabilité est actuellement entrain de travailler sur la création d'une base de connaissance pour

les professionnels de l'utilisabilité qui regrouperait toutes les pratiques de tests. Le CIF (Common Industry Format) en est un exemple. Ce format industriel commun pour le rapport des résultats de tests d'utilisabilité est devenu un standard ANSI en décembre 2001 (ANSI/NCITS 354-2001) et est entrain de devenir un standard international. Le même phénomène se produit dans le plus large domaine de la conception centrée utilisateur. En effet, la plupart des entreprises ont développé des terminologies communes, des modèles et des outils pour l'échange des données entre les différentes professions de l'UCD [6]. Par contre, ce domaine souffre d'un manque de critères et de mesures claires et précises pour estimer l'importance et l'impact de leur contribution dans le monde industriel [5]. Tous ces efforts de standardisation et de formalisation sont un défi pour la communauté de l'utilisabilité afin d'assurer leur développement et leur acceptabilité au sein des entreprises.

Nombreux sont les répondants qui ont exprimé leur besoin d'avoir sur le marché des logiciels de récolte et d'analyse des données de tests mais également des logiciels d'annotation et d'indexation des vidéos. Ils ont besoin d'outils qui leur permettent d'indexer automatiquement les vidéos, de les annoter et d'extraire des séquences vidéo afin de traiter la considérable quantité d'information vidéo. L'outil Morae répond en partie à ce besoin mais il est actuellement le seul outil sur le marché à offrir de telles fonctionnalités. Ils ont également exprimé le besoin d'avoir des outils d'analyse de données qui soient non seulement faciles à utiliser et à configurer, mais également qui soient fiables et abordables en termes de prix.

Mais surtout et avant tout, les commentaires récoltés ont montré que les tests d'utilisabilité sont de plus en plus valorisés au sein des entreprises et les managers des entreprises sont de plus en plus convaincus de l'utilité de tels tests. Ainsi, ils sont plus disposés à investir des budgets pour l'achat des outils et infrastructures et à mobiliser du personnel pour conduire les tests. Ceci offre une belle perspective pour le développement et la formalisation des tests d'utilisabilité.

CONCLUSION

Cet article a présenté les résultats d'une enquête sur la conduite des tests d'utilisabilité impliquant 86 praticiens de l'utilisabilité. Cette enquête a permis de présenter l'état de l'art sur les pratiques actuelles dans la conduite des tests d'utilisabilité et de situer l'utilisabilité au sein des entreprises.

L'analyse de cette enquête a permis de relever les différents points suivants :

- Quelles étaient les activités, méthodes et matériels de tests à considérer dans la conduite des tests d'utilisabilité.

- Même si les laboratoires fixes d'utilisabilité et les tests sur sites restent prédominants comme infrastructures dans la conduite des tests, les tests à distance offrent une nouvelle plateforme de tests de plus en plus populaire.
- La nécessité de développer des outils pour la collecte et l'analyse des données de tests ainsi que pour l'annotation et l'indexation des vidéos.
- L'utilisabilité est un domaine encore jeune qui n'est pas encore bien implantée au niveau des entreprises mais ce domaine est de plus en plus valorisé et a ainsi de bonnes perspectives de développement. Une volonté de formaliser la conduite des tests d'utilisabilité se concrétise de plus en plus.

D'autres directions de recherche ont pu être identifiées suite à cette enquête. Un premier axe de recherche est de concevoir et développer d'une part des outils d'aide à la collecte et l'analyse des données de tests et d'autre part des outils pour l'annotation et l'indexation des séquences vidéo. Le second axe de recherche est de proposer une plateforme pour conduire des tests d'utilisabilité incluant la spécification du processus, des logiciels, et des méthodes de tests. Les résultats de cette enquête vont être mis à profit pour valider et raffiner le processus de conduite de tests d'utilisabilité que nous avons défini au sein de notre équipe de recherche. Néanmoins, il est important de préciser que cette enquête, de part son nombre de questions, n'a pu traiter tous les problèmes liés aux tests d'utilisabilité et aux autres méthodes d'évaluation en général. Cependant, des enquêtes et études complémentaires sont envisagées pour étudier d'autres aspects qui n'ont pu être traités dans cette enquête tels que les motivations des professionnels pour choisir les tests utilisateurs (par exemple : temps, budget, exigence du client, etc.), les étapes du cycle de vie où ces évaluations sont réalisées et les secteurs d'entreprise où l'utilisabilité est pratiquée.

REMERCIEMENTS

Un grand merci à toutes les personnes qui ont accordé de leur temps pour compléter le questionnaire et partager leurs expériences dans le domaine. Tous nos remerciements vont également à nos collègues qui nous ont fait profiter de leurs expertises pour la conception du questionnaire et l'analyse des données. Nous remercions personnellement: Daniel Engelberg, Gabriel Michel, Armand Nganou, Bill Reimer et Mohamed Taleb.

BIBLIOGRAPHIE

1. Rubin, J. (eds.). *Handbook of usability testing: how to plan, design, and conduct effective tests*. Wiley, New York, 1994.
2. Nielsen, J. (eds.). *Usability Engineering*. Morgan Kaufmann, 1994.
3. Gellner, M. and Forbrig, P. Modeling the Usability Evaluation Process with the Perspective of Developing a Computer Aided Usability Evaluation (CAUE) System. In *Proceedings of INTERACT 2001 Workshop on "Usability throughout the entire systems development lifecycle"* (July 10-14, 2001, Tokyo, Japan), 2001.
4. Rosenbaum, S., Rohn, J.A., and Humburg, J. A Toolkit for Strategic Usability: Results from Workshops, Panels and Surveys. In *Proceedings of CHI 2000 : Conference on Human Factors in Computing Systems* (April 1-6, 2000, The Hague, Netherlands), 2000, pp. 337-344.
5. Vredenburg, K., Mao, J.Y., and al., e. *A survey of User Centred Design Practice*. CHI letters, Vol. 4, No. 1, 2002.
6. Venturi, G. and Troost, J. Survey on the UCD integration in the industry. In *Proceedings of NordiCHI04 conference proceedings* (October, 2004, Tampere, Finland), 2004.
7. Ng, S.P., Murnane, T., Reed, K., Grant, D., and Chen, T.Y. A Preliminary Survey on Software Testing Practices in Australia. In *Proceedings of Australian Software Engineering Conference (ASWEC'04)* (2004, Melbourne, Australia), 2004.
8. Torkar, R. and Mankefors, S. A Survey on Testing and Reuse. In *Proceedings of IEEE International Conference on Software-Science, Technology & Engineering* (November 04-05, 2003, Herzlia, Israel), 2003.
9. Survey on Usability Testing Process and Practices. Disponible à l'adresse <http://rana.cs.concordia.ca/survey/>
10. SPSS, logiciel d'analyse statistique. Disponible à l'adresse <http://www.spss.com/>
11. Morae, logiciel de conduite et d'analyse des tests d'utilisabilité pour les applications web et les logiciels. Disponible à l'adresse <http://www.techsmith.com/products/morae/default.asp>
12. Hartson, H.R., Castillo, J.C., Kelso, J., Kamler, J., and Neale, W.C. Remote Evaluation: The Network as an Extension of the Usability Laboratory. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '96)* 1996, pp. 228-235.
13. Moha, N., Li, Q., Seffah, A., and Michel, G. Towards a Platform for Usability Remote Tests via Internet. In *Proceedings of OZCHI2004* (November 22-24, 2004, University of Wollongong, Australia), 2004.