

CHAPITRE 9

L'ÉVALUATION ERGONOMIQUE DES DOCUMENTS ÉLECTRONIQUES

J.M. CHRISTIAN BASTIEN et ANDRÉ TRICOT

1. INTRODUCTION

Les documents électroniques dont il est question dans le présent chapitre englobent les sites Web, leurs moteurs de recherche et les sites Web de recherche d'information, mais aussi les bases de données documentaires qui permettent de chercher des documents et, ce qui est de plus en plus le cas, d'y avoir accès directement sous forme de document html ou PDF. Par conséquent, l'évaluation ergonomique des documents électroniques concerne d'une part l'évaluation des

interfaces de recherche et de présentation des résultats qui en sont issus, de consultation de ces documents et d'autre part l'évaluation des documents eux-mêmes, c'est-à-dire la façon dont ces documents sont présentés, la facilité avec laquelle on peut les lire, *etc.* On imagine donc la difficulté de la tâche d'évaluation ergonomique des documents électroniques compte tenu de tous les aspects devant être abordés.

L'objectif de ce chapitre est de présenter une brève revue des méthodes et des connaissances pouvant être utilisées pour l'évaluation ergonomique des documents électroniques. Il ne s'agit pas de présenter dans le détail les méthodes évoquées, mais de présenter les résultats des recherches récentes dans le domaine des méthodes et procédures d'évaluation ergonomique des logiciels interactifs pouvant être appliquées aux documents électroniques ou des recherches réalisées dans le domaine des documents électroniques pouvant être utilisées pour l'évaluation. Compte tenu des limites imparties, le lecteur intéressé trouvera de nombreuses références à des articles et ouvrages dans lesquels il pourra trouver de plus amples informations sur les méthodes et études présentées.

La première partie de ce chapitre propose une définition de la qualité ergonomique des documents électroniques et de ses enjeux.

Deux méthodes d'évaluation des documents électroniques sont présentées dans les deux parties suivantes. Ces méthodes ont fait l'objet de nombreuses publications et sont bien documentées : il s'agit des tests utilisateurs et de l'inspection ergonomique. Ces deux méthodes appartiennent respectivement à deux grandes catégories : les méthodes requérant la participation directe des utilisateurs et les méthodes s'appliquant aux caractéristiques de l'interface.

2. QU'EST-CE QUE LA QUALITÉ ERGONOMIQUE DES DOCUMENTS ÉLECTRONIQUES ?

Un document est un artefact dont l'utilisateur attend qu'il réponde à son besoin d'information. La qualité d'un document réside donc dans sa capacité à remplir ce besoin. La qualité du document et les dimensions qui la composent ont pu être définies de différentes façons depuis une quinzaine d'années. Il nous semble que ces différentes façons se recoupent entre elles, bien qu'en apparence elles s'opposent sur la façon d'intégrer ou pas différentes dimensions de l'évaluation.

De façon schématique, l'évaluation ergonomique s'intéresse à la situation suivante (figure 10.1). Les différentes dimensions évoquées dans cette figure sont détaillées plus bas.

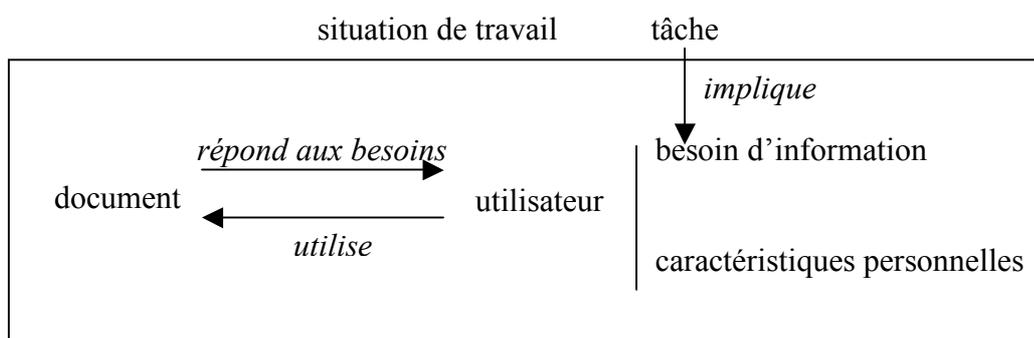


Figure 10.1. – *Les dimensions de l'évaluation ergonomique d'un document*

Une première série de critères (dimensions) concerne strictement la réponse au besoin d'information : ces critères sont donc relatifs au besoin d'information et à la tâche à réaliser. On regroupe souvent ces critères sous la catégorie « utilité ».

Une deuxième série de critères concerne le document lui-même, ses qualités intrinsèques, qui le rendent plus ou moins utilisable. On regroupe souvent ces critères sous la catégorie « utilisabilité ».

Une troisième série de critères concerne l'adéquation du document aux caractéristiques personnelles de l'utilisateur, comme sa profession, son âge, sa familiarité avec le contenu du document ou son expertise dans l'utilisation des moteurs de recherche. On regroupe parfois ces critères sous la catégorie « accessibilité », insistant avec ce terme sur les cas où les caractéristiques personnelles relèvent des handicaps sensoriels ou physiques.

Une quatrième série de critères concernent l'adéquation du document avec la situation de travail (en termes de temps, d'espace, d'organisation, etc.). Différents mots désignent cette catégorie, notamment acceptabilité et compatibilité, la notion d'acceptabilité intégrant les dimensions de motivation de l'utilisateur.

Nous allons présenter maintenant quelques approches de l'évaluation ergonomique des documents, chacune ayant une façon particulière de prendre en compte les quatre dimensions que nous

venons de présenter. À titre d'exemple, nous aborderons les documents normatifs et notamment la norme ISO 9241-110 sur le dialogue (D'autres parties de la norme 9241, que nous n'aborderons pas ici sont aussi pertinentes pour les documents électroniques, c'est le cas par exemple la partie 11 sur l'utilisabilité, la partie 13 sur le guidage ou encore la partie 17 sur les formulaires) et les Critères Ergonomiques.

2.1. LA NORME ISO 9241-110 « PRINCIPES DE DIALOGUE »

La norme ISO 9241-110 (2006) intitulée « Principes de dialogue » n'est pas spécifique aux documents, mais aux dialogues humain - machine. Elle est aisément et fréquemment transposée au domaine des documents électroniques. Cette norme intègre toutes les dimensions de l'évaluation ergonomique sauf une, la compatibilité avec la situation de travail, qui n'est abordée qu'à travers la tâche.

La norme ISO 9241 retient sept principes : (1) Adaptation à la tâche : le dialogue permet de réaliser la tâche de façon efficace et efficiente (2) Caractère auto-descriptif : le dialogue est immédiatement compréhensible grâce au retour d'information du système ou expliqué à la demande de l'utilisateur (3) Contrôle utilisateur : l'utilisateur peut initier et contrôler la direction et le rythme de l'interaction (4) Conformité aux attentes utilisateurs : le dialogue est cohérent et correspond aux caractéristiques des utilisateurs (5) Tolérance aux erreurs le résultat prévu peut être obtenu avec ou sans actions correctrices (6) Aptitude à l'individualisation : l'interface peut être modifiée pour s'adapter aux besoins de la tâche ou aux préférences ou compétences des utilisateurs (7) Facilité d'apprentissage : le dialogue soutient et guide l'utilisateur dans l'apprentissage du système.

Elle indique également sept principes pour la présentation des informations : (1) Clarté : le contenu s'affiche rapidement et avec précision (2) Discriminabilité : les informations peuvent être distinguées avec précision, (3) Concision : seules les informations nécessaires à la tâche sont affichées (4) Cohérence : les mêmes informations sont présentées de manière identique sur toute l'application (5) Détectabilité : les informations sont codées de façon adéquate au bon endroit (6) Lisibilité : l'information est facile à lire (7)

Compréhensibilité : la signification des termes est clairement compréhensible.

2.2. LES CRITÈRES ERGONOMIQUES DE SCAPIN ET BASTIEN (1997)

Les critères ergonomiques de Scapin et Bastien (1997) constituent eux aussi une approche intégrée de l'évaluation ergonomique, mais en prenant en compte l'évaluation de certains aspects de l'utilité, plus particulièrement la « compatibilité » avec les tâches.

1. Guidage
 - 1.1. Incitation
 - 1.2. Groupement / Distinction entre Items
 - 1.3. Feedback Immédiat
 - 1.4. Lisibilité

2. Charge de Travail
 - 2.1. Brièveté
 - 2.2. Densité Informationnelle

3. Contrôle Explicite
 - 3.1. Actions Explicites
 - 3.2. Contrôle Utilisateur

4. Adaptabilité
 - 4.1. Flexibilité
 - 4.2. Prise en compte de l'expérience de l'utilisateur

5. Gestion des Erreurs
 - 5.1. Protection contre les Erreurs
 - 5.2. Qualité des Messages d'Erreurs
 - 5.3. Correction des Erreurs

6. Homogénéité / Cohérence

7. Signifiante des Codes et Dénominations

8. Compatibilité

2.3. LE MODÈLE DE NIELSEN (1993)

Pour Nielsen, l'acceptabilité d'un système est la combinaison de l'acceptabilité sociale et de l'acceptabilité « pratique ». L'acceptabilité sociale fait référence au degré avec lequel le système obéit et/ou satisfait aux exigences de la société (éthique, légalité, etc.). L'utilisabilité et l'utilité « théorique » (c'est-à-dire le but que le système est censé permettre d'atteindre) sont deux dimensions de l'utilité « pratique » (c'est-à-dire le but que le système permet effectivement d'atteindre). L'utilité pratique est elle-même une sous-dimension de l'acceptabilité pratique. Les différentes dimensions sont présentées dans la figure 10.2. Aucune des dimensions que nous avons évoquées plus haut n'est ignorée.

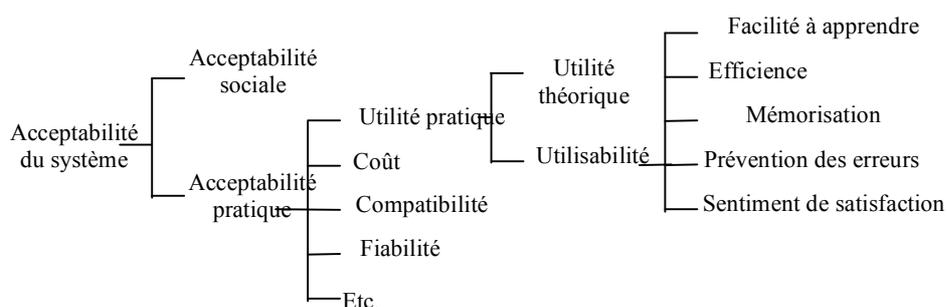


Figure 10.2. – *Le modèle de Nielsen (1993)*

Le schéma peut se lire de droite à gauche, chaque ensemble d'éléments de droite constituant une sorte de pré-condition pour l'élément placé à sa gauche. Dit autrement, l'acceptabilité dépend de l'utilité pratique qui dépend de l'utilisabilité ; mais l'utilisabilité ne dépend ni de l'acceptabilité ni de l'utilité. Si cette conception des relations entre les différentes dimensions de l'évaluation a valeur de référence aujourd'hui, quelques critiques peuvent être émises.

- Dans ce modèle, l'acceptabilité ne peut avoir d'effet sur l'utilisabilité ou l'utilité. Or, dans une étude à propos d'un site Web de formation pour les techniciens de maintenance véhicule auto, Amiel, Tricot et Mariné (2004) ont montré que l'utilisation et l'utilité du document dépendaient de sa compatibilité avec l'organisation du travail, en termes d'emplois du temps, de disponibilité d'un ordinateur, de

gestion des priorités entre tâches et de gestion de l'espace de travail. La possibilité d'accéder au document dépend donc de sa compatibilité avec la situation de travail, mais aussi avec les situations de handicaps physiques et sensoriels, avec la motivation des opérateurs pour accéder au document, etc.

- Dillon et Morris (1996) notent qu'avec cette conception, l'utilisabilité, qui est pourtant au centre des préoccupations de Nielsen et de la plupart des évaluateurs, (1) apparaît comme une condition nécessaire mais non suffisante (2) ne permet de rien dire sur l'usage.

2.4. LE MODÈLE DE DILLON ET MORRIS (1996)

Selon le modèle de Dillon et Morris présenté dans la figure 10.3, (lui-même inspiré du modèle de Davis, 1989), il faut introduire la notion d'attitude de l'utilisateur (perceptions, affects) pour pouvoir comprendre les relations entre utilité, utilisabilité et acceptabilité. En particulier, ces auteurs insistent sur la perception de l'utilité et de l'utilisabilité. Ces perceptions peuvent être très différentes entre elles et très différentes de l'utilisabilité effective ou de l'utilité pratique : je peux croire que ce document sera facile à prendre en main mais peu utile pour moi, alors qu'il sera en réalité difficile à prendre en main mais utile au développement de mes compétences.

Le modèle de Dillon et Morris est donc très peu détaillé, mais il semble bien prendre en compte de façon pertinente une dimension ignorée des autres modèles : la motivation des individus à utiliser les documents. Cette motivation est conçue, comme dans le modèle de Davis, comme liée à la perception de l'utilité et de l'utilisabilité.

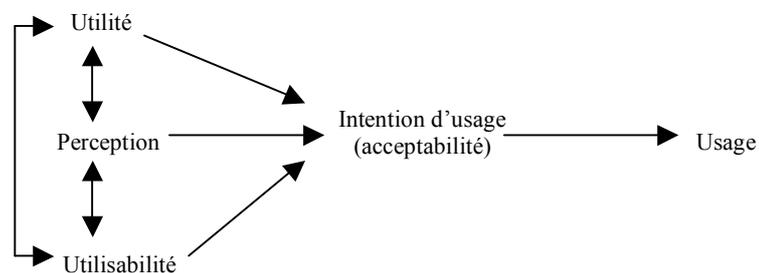


Figure 10.3. - *Le modèle de Dillon et Morris (1996)*

2.5. LE MODÈLE ACCEPTABILITÉ + UTILISABILITÉ SONT DES PRÉ-REQUIS À L'UTILITÉ

De nombreux auteurs et praticiens, par exemple dans le domaine du marketing, semblent adopter un point de vue différent de ceux qui viennent d'être présentés. Pour eux, l'acceptabilité et l'utilisabilité sont des pré-requis à l'utilité (voir par ex. Ketola & Røykkee, 2001). Une fois que le système est acceptable et utilisable, il est « prêt à être <éventuellement> utile ». Ainsi, l'utilité pratique peut devenir une dimension très floue, très peu définie, comme c'est le cas par exemple avec certains sites Web, véritables succès, dont on ne sait pas vraiment à quels besoins ils répondent, mais qui sont indéniablement acceptés et utilisés. Un des grands succès du Web en 2007 (www.votefortheworst.com) est simplement destiné à inciter les usagers à voter pour le plus mauvais chanteur d'un concours télévisé ! Ne serait-il pas plus simple de ne pas regarder le concours en question ? On pense aussi à certains *Blogs* ou à certains sites institutionnels qui présentent des informations qui concernent certes l'émetteur mais dont on se demande quel intérêt ils représentent pour les autres. Parfois l'utilité n'est pas définie à l'avance mais créée par les usages. C'est le cas célèbre des *paggers* adoptés par les adolescents il y a quelques années pour une communication de proximité, ludique et relationnelle, sans qu'aucun concepteur de *paggers* ne l'ait jamais imaginé.

2.6. LE MODÈLE DE TRICOT ET AL. (2003)

Plutôt que de trancher entre les approches présentées ci-dessus, Tricot, Plécat-Soutjis, Camps, Amiel, Lutz et Morcillo (2003) se sont demandés de quelle(s) façon(s) interpréter les relations entre trois séries de tests mesurant l'utilité, l'acceptabilité et l'utilisabilité. Imaginons le cas d'un document qui, suite à une série de tests, se révèle acceptable mais inutilisable et inutile. Comment savoir ce qu'il faut améliorer ? Doit-on croire que l'amélioration de l'utilisabilité va entraîner une amélioration de l'utilité ? Qu'il faut aussi améliorer l'utilité ? Ou que l'amélioration de l'utilité entraînera une amélioration de l'utilisabilité ? Ces questions pratiques d'interprétation des relations entre trois dimensions de l'évaluation soulèvent un problème théorique

encore débattu aujourd'hui, et qui n'est rien moins que celui de la rationalité, c'est-à-dire le problème des relations entre moyens et buts (Searle, 2001). Élaborer une rationalité c'est élaborer un modèle des relations entre moyens et buts. Mais élaborer un tel modèle ne garantit pas la valeur de ce modèle, en particulier quand il s'agit de l'exporter. On sait par exemple qu'en évaluation des interactions humains machines (IHM) un des problèmes réside dans l'exportation de critères rationnels d'un champ (ex., les bases de données) vers un autre (ex., les hypertextes). Un critère rationnel dans un champ peut ne pas l'être dans un autre (Tricot, Puigserver, Berdugo & Diallo, 1999). C'est pourquoi Tricot *et al.* (2003) proposent une approche empirique de la rationalité. Est rationnelle une relation observée entre deux états : atteinte du but et mise en œuvre des moyens.

Les auteurs considèrent en outre que pour chaque document, il existe une représentation mentale (individuelle ou collective) de ce document (ex. de ses attributs physiques) et de sa rationalité. Ils appellent acceptabilité la valeur de cette représentation. Ils défendent la position suivante :

- quel que soit le document à évaluer, toute relation entre Utilisabilité (possibilité de mettre en œuvre les moyens), Utilité (possibilité d'atteindre le but) et Acceptabilité, est possible ;
- toute relation possible peut-être décrite de façon logique.

Chaque dimension est une variable sur laquelle on peut réaliser des mesures indépendantes des deux autres variables. Pour un même utilisateur-testeur, on peut noter ensemble les résultats obtenus selon les trois variables. On procède à une analyse des relations orientées entre ces variables, par exemple avec l'analyse en pistes causales ou l'analyse implicative (Gras, 2004) et le logiciel CHIC (Classification Hiérarchique Implicative et Cohésitive ; Couturier & Gras, 2005 et [ardm.asso.fr/CHIC.html]). Ce genre d'analyse permet de voir quel est le sens des relations entre les dimensions de l'évaluation ergonomique. En mettant à jour des relations causales, une telle analyse permet à l'ergonome d'intervenir sur les causes plutôt que sur les conséquences des problèmes repérés.

2.7. CONCLUSION À PROPOS DE LA QUALITÉ ERGONOMIQUE

Pour évaluer la qualité d'un document, l'ergonome dispose donc d'un certain nombre de variables et d'un large choix de définitions. Il

nous semble qu'il doit surtout se demander s'il connaît ou peut connaître :

- les besoins informationnels des utilisateurs ;
- leurs tâches ;
- leur situation de travail ;
- leurs caractéristiques personnelles.

Il devra ensuite se demander si chacune de ces dimensions présente une certaine homogénéité, la rendant à peu près descriptible, ou s'il dispose de techniques d'échantillonnage suffisamment fiables pour rendre compte de l'hétérogénéité des dimensions concernées. En fonction de cela, il pourra choisir une approche de l'évaluation qui se situe quelque part entre deux extrêmes :

- il ne peut rien connaître ou décrire précisément : son évaluation concernera exclusivement la qualité intrinsèque du document, son utilisabilité au sens restreint ;
- il peut tout connaître et décrire : son évaluation sera large, prenant en compte l'ensemble des dimensions de l'évaluation ergonomique des documents.

L'ergonome devra enfin se demander s'il souhaite obtenir une évaluation de l'ensemble du document évalué ou si, au contraire, il souhaite évaluer séparément différents aspects du document (voir la proposition de Brinkman, Haakma, & Bouwhuis, 2007, pour procéder à une telle analyse).

Ces résultats obtenus, l'ergonome devra les traiter pour produire un diagnostic et décider des aspects sur lesquels intervenir, recommander des améliorations. Il devra donc se demander s'il procède :

- à un traitement « à plat » des résultats, en faisant par exemple une somme des résultats obtenus pour chaque variable ;
- à un traitement de chaque variable indépendamment des autres, considérant que chaque résultat indique qu'il faut intervenir ou pas sur la dimension concernée ;
- à un traitement permettant de rendre compte de dépendances entre variables (soit à partir d'un modèle préexistant, comme celui de Nielsen, soit de façon empirique). Ce traitement pourrait lui permettre de décider quels défauts il faut traiter en priorité (par ex. si l'analyse montre que l'utilité dépend de l'utilisabilité, il faut mieux améliorer l'utilisabilité d'abord).

Nous proposons, dans la suite de ce chapitre, de retenir les définitions suivantes : l'évaluation de *l'utilité* concerne le fait que le document réponde aux besoins de l'utilisateur et celle de *l'utilisabilité*

concerne le fait que le document et l'interface qui en donne l'accès soit aisément manipulable pour permettre à l'utilisateur :

- de trouver ce document ;
- de trouver l'information pertinente à l'intérieur de ce document et de pouvoir la reconnaître comme étant pertinente ;
- de comprendre et, plus largement, d'exploiter l'information trouvée.

Le fait que le document soit utile n'est donc pas une qualité absolue, mais relative aux besoins de l'utilisateur. Il faut donc connaître précisément ces besoins pour évaluer le document. Ce qui contraint l'ergonome d'être au plus près de l'utilisateur, de sa tâche et de sa situation de travail pour réaliser correctement l'évaluation. Par exemple, pour évaluer un système documentaire utilisé par des pilotes de ligne, il faut connaître les tâches de ces pilotes, leurs conditions de travail, les informations dont ils ont besoin pour réaliser chacune de ces tâches, les raisons pour lesquelles ils cherchent ces informations, quand ils les cherchent, comment ils les cherchent, etc. Tout cela représente un long travail de terrain à réaliser par l'ergonome. L'utilité d'un document étant strictement relative aux besoins des utilisateurs, on ne produit évidemment pas de règles ergonomiques générales pour améliorer l'utilité des documents. Les méthodes en revanche sont bien connues, et depuis longtemps, puisqu'il s'agit de l'analyse des situations de travail, de la tâche et de l'activité.

Au contraire, l'évaluation de *l'utilisabilité* peut être réalisée, jusqu'à un certain point, de façon assez indépendante de la situation de travail. Cela dépend toutefois des dimensions de l'utilisabilité évaluées. En effet, l'évaluation de la dimension « Compatibilité » nécessite, comme nous l'indiquions plus haut, de connaître les caractéristiques des utilisateurs, de leurs tâches et de la situation de travail des documents électroniques. *L'utilisabilité* a connu, dans le domaine de l'ergonomie des documents, de véritables succès publics avec notamment les publications de Nielsen sur l'évaluation des sites Web (par ex., Nielsen, 2000 ; Nielsen & Tahir, 2001) et celle de Nogier (2005) en France. L'évaluation de *l'utilisabilité* pouvant être conduite de façon relativement indépendante du contexte, on a progressivement pu accumuler des résultats et proposer aux concepteurs des guides de conception. Ajoutons que Nielsen (1994a) a promu très tôt une approche simple et économique de l'évaluation de *l'utilisabilité*, qui a sans doute contribué à son succès.

En résumé, évaluer l'ergonomie d'un document revient à évaluer :

- son *utilité*, c'est-à-dire sa capacité à répondre aux besoins d'information de l'utilisateur ;
- son *utilisabilité*, c'est-à-dire sa capacité à être utilisé et compris aisément pour donner accès à l'information pertinente qu'il contient.

Comme nous l'avons indiqué dans l'introduction, différentes méthodes existent pour procéder à de telles évaluations, dont certaines requièrent la participation des utilisateurs voire une connaissance approfondie de leur situation de travail : c'est le cas des tests utilisateurs. D'autres, au contraire, exploitant l'accumulation de connaissances ergonomiques, procèdent par inspection du document et application de critères : c'est le cas de l'inspection ergonomique. Nous présentons maintenant ces deux méthodes.

3. LES TESTS UTILISATEURS

Les tests utilisateurs, les questionnaires de satisfaction et les entretiens font partie des méthodes requérant la participation directe des utilisateurs (*user-based evaluations*). Généralement, lors des tests utilisateurs, une personne participe soit à l'exécution de tâches représentatives des tâches pour lesquelles le document a été conçu, soit à l'exploration libre de ce dernier. L'objectif de ces tests est d'identifier les difficultés d'usage à partir des verbalisations suscitées ou spontanées des utilisateurs ou à partir d'indices de performances variés comme le temps requis pour l'exécution d'une tâche, l'exactitude des résultats, le nombre et le type d'erreurs commises. Une fois les difficultés identifiées, des propositions d'aménagement sont faites afin d'augmenter la qualité ergonomique du logiciel ou dans le cas présent des documents électroniques.

La mise en place des tests utilisateurs est bien documentée (*cf.* par ex. Dumas & Redish, 1993). Cette mise en place passe par un certain nombre d'étapes parmi lesquelles on trouve : la définition des objectifs du test, le choix et le recrutement des participants, la sélection et l'organisation des tâches à réaliser, la création de scénarii de test, le choix des mesures d'utilisabilité qui seront effectuées, la préparation du matériel et de l'environnement de test, le choix des expérimentateurs (testeurs) et la conduite du test à proprement parler (consignes, accompagnement, intervention, etc.), le développement et

la passation des questionnaires de satisfaction, l'analyse des données et finalement la présentation des résultats. Des recherches récentes ont toutefois montré qu'une attention particulière devait être apportée à un certain nombre d'aspects des tests utilisateurs comme : le nombre de participants à tester, aux procédures de test, aux méthodes et métriques permettant de comparer des systèmes interactifs proposant les mêmes services ou fonctionnalités, aux techniques d'enregistrement vidéo et de capture d'événements utilisateurs, à l'évaluation à distance et à la rédaction des rapports d'évaluation (voir à ce propos Bastien & Scapin, 2002).

3.1. COMBIEN DE PERSONNES TESTER ?

Le nombre de personnes qui vont participer au test est choisi pour répondre au but suivant : faire en sorte qu'un problème existant dans le document soit effectivement rencontré par au moins un des testeurs. C'est donc un problème classique d'échantillonnage. On peut répondre de trois façons à ce problème :

- en constituant un groupe de testeurs représentatif de la population des utilisateurs : c'est la méthode la plus sûre, mais elle est très coûteuse ;
- en ne définissant pas le nombre de testeurs à l'avance : tout en puisant dans la population d'utilisateurs, on arrête les tests quand un nouveau testeur ne rencontre que des problèmes ayant déjà été rencontrés par d'autres ;
- en procédant de façon itérative avec de tous petits groupes (5 testeurs) ; l'argument avancé par Nielsen et Landauer (1993) est qu'il vaut mieux utiliser 15 utilisateurs en corrigeant les défauts identifiés par les 5 premiers, puis en réitérant la démarche trois fois, que d'utiliser d'emblée les 15 utilisateurs.

Ces trois propositions ne sont pas contradictoires entre elles : il vaut mieux procéder de façon itérative, en utilisant l'échantillon le plus représentatif possible selon les moyens humains, temporels et financiers dont on dispose, tout en étant particulièrement attentif à ce qu'un véritable seuil de redondance des erreurs détectées soit atteint.

Ce problème de la détermination du nombre de testeurs à solliciter date des années 1980 (Al-Awar, Chapanis, & Ford, 1981) et fait toujours l'objet d'études et d'écrits (*cf.*, par ex., Lewis, 2006 ; Lewis & Sauro, 2006). En 2001, une étude de Spool et Schroeder indiquait

que huit testeurs (chiffre évoqué par les études de Nielsen) n'étaient vraiment pas assez pour évaluer un site Web. Dans leur étude, quarante-neuf personnes devaient réaliser des achats sur 4 sites de commerce. L'analyse des résultats a indiqué qu'avec les 5 premiers participants, seulement 35% des problèmes étaient détectés. Par ailleurs, des problèmes sérieux ne sont apparus qu'avec le 13^{ème} et 15^{ème} participant. Le type d'interface et les tâches demandées aux testeurs semblent pouvoir expliquer, en partie, les résultats obtenus. Cette question du nombre de participants n'est donc définitivement pas réglée.

3.2. QUELLES PROCÉDURES METTRE EN ŒUVRE ET DANS QUEL TYPE D'ENVIRONNEMENT ?

Si l'évaluation ergonomique d'un document peut intervenir à tout moment de son cycle de vie, c'est pendant le développement du document qu'elle est la plus importante. C'est pendant cette phase que l'on pourra identifier et corriger les erreurs à moindre coût. Quel que soit le moment de l'évaluation et le mode de présentation du document électronique (maquette, prototype, version finale, *etc.*) il importe de porter une attention particulière à la procédure, au protocole de test (protocole à mesures répétées, groupes indépendants, *etc.*) et à la nature des tests à réaliser. S'agira-t-il :

- d'un test exploratoire (réalisés tôt dans le cycle de développement et visant plus particulièrement l'évaluation de l'efficacité des premiers concepts. Ces premiers tests peuvent être réalisés à l'aide de maquettes ou de prototypes) ;
- d'un test d'évaluation (réalisés, comme les précédents, tôt dans le cycle et une fois que la conception détaillée est plus avancée. L'objectif est ici de fournir des données précises sur différents aspects de l'utilisabilité) ;
- d'un test de validation (qui surviennent tardivement dans le cycle de développement, généralement juste avant la « livraison » du produit. L'objectif est ici d'évaluer la qualité ergonomique du système en référence à des objectifs d'utilisabilité) ;
- d'un test de comparaison (l'objectif étant ici de comparer des alternatives de conception dans le cadre d'un projet donné ou de comparer différents produits proposant les mêmes services ou les mêmes fonctionnalités).

Barendregt, Bekker, Bouwhuis et Baauw (2006) mettent en exergue le « temps d'utilisation » visé par l'application lors de la conception des tests. Il prennent l'exemple des jeux vidéos, mais l'on peut transférer leur démarche dans le domaine des documents. Un document qui sera utilisé longtemps ou fréquemment ne devra pas être testé de la même façon d'un document destiné à être utilisé une seule fois. Pour un document destiné à être utilisé plusieurs fois (par ex. un intranet), on aura tout intérêt à répéter les tests auprès des mêmes testeurs, ou même à procéder à des tests après une utilisation effective du document testé.

Lors des tests, le participant peut, selon la procédure mise en place, se trouver seul ou accompagné du « testeur », les deux alternatives présentant des avantages et des inconvénients (cf., Dumas & Redish, 1993 ; Rubin, 1994). Parfois, plutôt que des passations individuelles, on optera pour des passations par paires (*paired-user testing*). Lors de ces tests, deux participants sont invités à réaliser des tâches ensemble et sur un même ordinateur, ce qui a souvent pour conséquence de susciter des discussions et des justifications sur la manière de faire. Dans ces conditions, les interactions entre participants prennent le pas sur les interactions avec le « testeur » (cf. à ce propos O'Malley, Draper, & Riley, 1984 ; Westerink, Rankin, Majoor, & Moore, 1994 ; Wildman, 1995 ; Wilson & Blostein, 1998). Ces procédures seront à choisir en fonction de leurs avantages et inconvénients et des objectifs de l'évaluation.

Le test utilisateur se déroule généralement dans un environnement spécifique, plus ou moins contrôlé et plus ou moins outillé. Plusieurs configurations physiques sont possibles (salle unique, salle double, avec ou sans salle d'observation, laboratoire mobile, *etc.*). Les équipements comprennent généralement des caméras vidéo, des moniteurs, des caméscopes, des microphones, *etc.* (par ex., Fath, Mann, & Holzman, 1994 ; Lund, 1994 ; Neugebauer & Spielmann, 1994 ; Nielsen, 1994b ; Sazegari, 1994).

Les ergonomes ont par ailleurs à leur disposition de nombreux outils :

- des outils d'enregistrement automatique des interactions sur le Web (ex., Morae) qui permettent d'enregistrer les adresses des pages visitées, le texte saisi au clavier, d'illustrer sur l'enregistrement vidéo de la session la position des clics gauches et droits et d'incruster sur cette vidéo le visage de l'utilisateur,

- des outils de codage des enregistrements vidéo (ex., The Observer de la société Noldus) qui permettent d'obtenir, à la fin du codage, des données quantitatives sur les comportements observés,
- ou encore des systèmes d'oculométrie qui permettent de suivre les parcours oculaires des participants sur un document, *etc.* (cf. le chapitre de Stahm, Kicka, & Baccino, cet ouvrage).

Tous ces outils sont plus ou moins faciles à mettre en œuvre. L'évaluateur devra donc préciser avant toute chose les questions à poser et choisir ensuite les outils de mesure les plus appropriés. Il devra surtout ne pas succomber à la tentation d'enregistrer tout ce qui est possible. Cette stratégie, permet en effet d'accumuler des montagnes de données. Mais la plupart du temps, celles-ci sont sous exploitées, pour des raisons de ressources et de temps.

Suite à des interactions avec les documents électroniques, des questionnaires et des entretiens permettent de recueillir de données subjectives relatives aux attitudes, aux opinions et à la satisfaction des utilisateurs. Ces données sont généralement utilisées pour compléter les données objectives recueillies lors des tests d'utilisation. En ce qui concerne les questionnaires, ces derniers ne sont pas très récents, du moins ceux qui s'intéressent plutôt aux systèmes interactifs. Le lecteur pourra trouver sur le site Web de l'*Association for Computing Machinery* [www.acm.org/perlman/question.html] la liste des questionnaires de satisfaction les plus courants. Il s'agit généralement d'énoncés (ex., j'aimerais utiliser ce système fréquemment) à propos desquels l'utilisateur doit dire s'il est entièrement d'accord ou non sur une échelle en 5 ou 7 points. Seul le *Website Analysis and Measurement Inventory* (WAMMI, [www.wammi.com/index.html]) est un questionnaire de satisfaction qui porte spécifiquement sur le Web.

3.3. COMMENT PRÉSENTER LES RÉSULTATS DES TESTS ?

Afin de faciliter les échanges entre consultants mais aussi entre consultants et universitaires sur les données de tests utilisateurs et afin que les consommateurs soient mieux informés sur la qualité ergonomique des logiciels interactifs, le National Institute of Standards and Technology a mis sur pied un projet intitulé « the Industry Usability Reporting project » (IUSR, www.nist.gov/iusr) (Scholtz & Morse, 2002). Pour ce faire, le groupe a défini un format

de présentation (*The Common Industry Format, CIF*) des tests utilisateurs. Ce format précise notamment les points qui doivent être abordés dans les rapports de tests utilisateurs à savoir : les tâches proposées aux utilisateurs, l'environnement de test (matériel et logiciel), le protocole de test, les méthodes de recueil de données, les techniques d'analyse mise en œuvre et les résultats de l'évaluation.

4. L'INSPECTION ERGONOMIQUE

L'inspection ergonomique fait partie d'une autre catégorie de méthodes, celles s'appliquant aux caractéristiques des interfaces. Cette catégorie comporte par ailleurs les modèles, méthodes et langages formels, le recours à l'expert et les outils d'évaluation automatique. Les méthodes d'inspection sont sans doute les plus utilisées dans cette catégorie. Ici aussi, des recherches récentes attirent l'attention des évaluateurs sur un certain nombre de points, notamment sur les performances de ce type d'évaluation et sur les moyens de les augmenter.

Les méthodes d'inspection de l'utilisabilité (*usability inspection methods*) regroupent un ensemble d'approches faisant appel au jugement d'évaluateurs, que ces derniers soient experts ou non en utilisabilité (Cockton, Lavery, & Woolrych, 2003 ; Nielsen & Mack, 1994 ; Virzi, 1997). Bien que toutes ces méthodes aient des objectifs différents, elles visent généralement la détection des aspects des documents pouvant entraîner des difficultés d'utilisation ou alourdir le travail des utilisateurs. Les méthodes d'inspection se distinguent les unes des autres par la façon dont les jugements des évaluateurs sont dérivés et par les critères d'évaluation à la base de leurs jugements.

Parmi les méthodes d'inspection, celles qui sont les plus utilisées sont : l'analyse de la conformité à un ensemble de recommandations (*guideline reviews*), l'analyse de la conformité à des normes (*standards inspection*), l'utilisation d'heuristiques et l'utilisation de critères.

L'utilisation et l'application de ces documents dans le cadre de l'évaluation pose un certain nombre de difficultés (voir à ce propos Bastien 2004). En effet, des études ont montré par exemple, que les recommandations posaient un certain nombre de difficultés et ce, en dépit du fait qu'elles soient généralement bien accueillies et parfois

même souhaitées. Certaines recommandations sont : difficiles à comprendre et à interpréter ; trop générales ; difficiles à traduire sous forme de règles et donc difficiles à implanter ; peu utilisées pour établir des spécifications et encore moins pour évaluer des choix de conception ; les exemples qui les accompagnent sont parfois préférés aux recommandations elles-mêmes. Leur utilisation pour l'évaluation entraîne donc des performances assez faibles, c'est-à-dire qu'elles permettent la détection d'un nombre assez restreint de défauts de conception. Dans certains cas, les recommandations font considérer à tort de bons choix de conception comme étant des défauts.

La détection des défauts de conception n'est pas chose facile. En effet, des évaluateurs détectent en moyenne de 20 % à 60 %, selon les études, des défauts de conception qu'un document interactif peut comporter. Cette variabilité des performances se double d'une variabilité interindividuelle : le pourcentage d'erreurs détectées peut varier, selon les individus, de 0 % (Molich & Nielsen, 1990) à 74,5 % (Nielsen, 1990). Aussi, il semble que les défauts liés à certaines recommandations telles « *Fournir de bons messages d'erreurs* » et « *Prévenir les erreurs* » sont plus difficiles à détecter que les défauts liés aux autres. Encore plus difficiles à détecter seraient les défauts liés à des recommandations telles que « *Fournir des moyens explicites de contrôle du dialogue* » (Nielsen, 1992a). Des résultats similaires ont été observés pour l'usage des normes et des critères ergonomiques (Bastien & Scapin, 1992, 1995 ; Bastien, Scapin et Leulier, 1999).

Une étude récente (Hvannberg, Law, & Lárusdóttira, 2007) a comparé l'approche de Nielsen avec celle de Gerhardt-Powals (1996). Elle montre que les deux approches produisent des résultats comparables (elles détectent autant de défauts), mais que la seconde produit une variation plus grande des résultats, ce qui permet aux auteurs de conclure que cette dernière implique que les évaluateurs soient plus formés qu'avec la méthode de Nielsen. Hvannberg et ses collègues montrent aussi que des outils informatiques d'aide à l'analyse par inspection ne sont pas forcément plus efficaces que l'utilisation de simples supports papiers. Les résultats de cette étude conduisent les auteurs à plaider pour le développement des analyses comparatives de l'efficacité des méthodes d'évaluation par inspection. Ils préconisent des plans d'expériences où les deux méthodes comparées le sont sur deux types de d'outils différents.

Ces résultats montrent de manière générale que la tâche d'évaluation n'est pas une mince affaire. Pour pallier ces difficultés diverses stratégies sont adoptées. Certains auteurs proposent des formats structurés de présentation des résultats (Cockton, Woolrych, & Hindmarch, 2004), qui auraient pour effet de modifier le comportement des évaluateurs en les amenant à réfléchir sur leurs processus de décision. D'autres auteurs proposent des outils permettant d'assister les évaluateurs. C'est le cas de SUIT (Ardito, Lanzilotti, Buono, & Piccinno, 2006) un outil Internet qui se présente sous forme de formulaire et qui permet à l'évaluateur de documenter son évaluation et de la partager via Internet avec d'autres évaluateurs. Le formulaire permet à l'évaluateur de décrire le défaut, de préciser l'endroit où il a été détecté, le principe ou recommandation qui n'est pas respecté, le degré de sévérité du défaut et éventuellement de proposer une solution. R-IDE est un autre outil Web qui permet aussi à l'évaluateur de documenter son évaluation (Kemp & Setungamudalige, 2006).

5. CONCLUSION

L'évaluation ergonomique des documents est une pratique déjà bien balisée. Elle recouvre un ensemble de dimensions, que nous avons présentées dans le chapitre 2, dont certaines sont très dépendantes de la situation de travail (évaluation de l'utilité, de la compatibilité, de l'acceptabilité) et d'autres relativement indépendantes de la situation (évaluation de l'utilisabilité dans un sens restreint). La littérature sur les méthodes prenant en compte cette seconde catégorie de dimensions est foisonnante depuis 20 ans. Ce qui a parfois eu pour effet malheureux de pousser au second plan les méthodes prenant en compte la situation de travail, et conduit certains évaluateurs naïfs, prenant conscience de la vacuité de leurs résultats, à s'écrier « *usability doesn't work!* ».

Les méthodes d'évaluation ergonomique des documents sont, soit fondées théoriquement, soit fondées sur la « compilation de pratiques » ou la compilation de critères (Hvannberg, Law, & Lárusdóttira, 2007). Une littérature est aussi consacrée à la validation empirique de ces méthodes. Cependant, Lindgaard (2006) est extrêmement critique vis à vis des études qui ont pour but de valider

une méthode d'évaluation ergonomique. Selon cet auteur, les critères qu'utilisent les chercheurs pour valider une méthode (sa capacité à identifier tous problèmes ergonomiques existants) ne correspondent pas aux objectifs des praticiens de l'ergonomie (identifier le maximum de problèmes avec un minimum de ressources – par ex. temporelles – et de façon itérative).

L'évaluation ergonomique de documents comme pratique professionnelle est donc au plus près de terrain, de la situation de travail, de la tâche et de l'activité. Elle passe par l'élaboration d'un protocole particulier à chaque situation, en tentant de répondre de façon optimale à une demande dans le cadre de contraintes. Elle ne doit pas oublier que son but est de produire des recommandations qui permettront d'améliorer l'utilité et l'utilisabilité du document. Dans ce domaine comme dans d'autres, les relations entre la recherche et la pratique ergonomiques nous semblent devoir être conçues comme un dialogue, avec suffisamment de distance.

Après 20 années de recherche dans le domaine, il nous semble que la recherche a produit suffisamment de résultats et de méthodes pour qu'aujourd'hui le praticien puisse réaliser un choix à peu près rationnel parmi un ensemble de possibles.

6. BIBLIOGRAPHIE

- Al-Awar, J., Chapanis, A., & Ford, R. (1981). Tutorials for the first-time computer user. *IEEE Transactions on Professional Communication*, 24, 30-37.
- Amiel, A., Tricot, A., & Mariné, C. (2004). Quels facteurs peuvent influencer l'engagement dans une formation à distance? Étude exploratoire auprès de prescripteurs de formation en milieu industriel. *Les Dossiers des Sciences de l'Education*, 12, 65-78.
- Ardito, C., Lanzilotti, R., Buono, P., & Piccinno, A. (2006, May 23-26). *A tool to support usability inspection*. Paper presented at the AVI'06, New York NY.
- Barendregt, W., Bekker, M.M., & Bouwhuis, D.G., & Baauw, E. (2006). Identifying usability and fun problems in a computer game during first use and after some practice. *International Journal of Human-Computer Studies*, 64, 830-846
- Bastien, J. M. C. (2004). L'inspection ergonomique des logiciels interactifs : intérêts et limites. In J.-M. Hoc & F. Darses (Eds.), *Psychologie ergonomique : tendances actuelles* (pp. 49-70). Paris : PUF.

- Bastien, J. M. C., & Scapin, D. L. (1992). A validation of ergonomic criteria for the evaluation of human-computer interfaces. *International Journal of Human-Computer Interaction*, 4, 183-196.
- Bastien, J. M. C., & Scapin, D. L. (1995). Evaluating a user interface with ergonomic criteria. *International Journal of Human-Computer Interaction*, 7, 105-121.
- Bastien, J. M. C., & Scapin, D. L. (2002). *Les méthodes ergonomiques : de l'analyse à la conception et à l'évaluation*, Communication présentée à Ergonomie et Informatique Avancée (Ergo-IA'2002) (pp. 127-143). Biarritz : I.D.L.S.
- Bastien, J. M. C., Scapin, D. L., & Leulier, C. (1999). The Ergonomic Criteria and the ISO 9241-10 Dialogue Principles : A pilot comparison in an evaluation task. *Interacting with Computers*, 11, 299-322.
- Brinkman, W.P., Haakma, R., & Bouwhuis, D.G. (2007). Towards an empirical method of efficiency testing of system parts : A methodological study. *Interacting with Computers*, 19, 342-356.
- Cockton, G., Lavery, D., & Woolrych, A. (2003). Inspection-based evaluations. In J. A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook. Fundamentals, Evolving Technologies, and Emerging Applications* (pp. 1118-1138). London: Lawrence Erlbaum Associates.
- Cockton, G., Woolrych, A., & Hindmarch, M. (2004, April 24-29). *Reconditioned merchandise : extended structured report formats in usability inspection*. Paper presented at the CHI'2004, New York, NY.
- Couturier, R., & Gras, R. (2005). Chic : traitement de données avec l'analyse implicite. *Actes d'Extraction et gestion des connaissances (EGC'2005)*, 679-684.
- Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319-340.
- Dillon, A., & Morris, M. (1996) User acceptance of new information technology - theories and models. In M. Williams (Ed.), *Annual Review of Information Science and Technology* (pp. 3-32). Medford : Information Today.
- Dumas, J. S., & Redish, J. C. (1993). *A practical guide to usability testing*. Norwood, NJ : Ablex.
- Fath, J. L., Mann, T. L., & Holzman, T. G. (1994). A practical guide to using software usability labs : Lessons learned at IBM. *Behaviour & Information Technology*, 13(1 & 2), 94-105.
- Gerhardt-Powals, J., (1996). Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human-Computer Interaction*, 8, 189-211.
- Gras, R. (2004). L'analyse implicite : ses bases, ses développements. *Educação Matemática Pesquisa*, 4, 11-48.

- Hvannberg, E.T., Law, E.L.C., & Lárusdóttir, M.K. (2007). Heuristic evaluation : Comparing ways of finding and reporting usability problems. *Interacting with Computers*, 19, 225-240
- International Organisation for Standardization. (2006). ISO 9241-110. *Ergonomie de l'interaction homme-système - Partie 110 : principes de dialogue*. Paris, France: AFNOR.
- Kemp, E., & Setungamudalige, D. T. (2006, July 6-7). *A resource support toolkit (R-IDE) : supporting the DECIDE framework*. Paper presented at Chinz'06, New York, NY.
- Ketola, P., & Røykkee, M. (2001). *The three facets of usability in mobile handsets Nokia*. Paper presented at CHI'01 Conference on Human Factors in Computing Systems, Seattle, 31 March-5 April.
- Lewis, J. R. (2006). Sample sizes for usability tests : mostly math, not magic. *interactions*, XIII(6), 29-33.
- Lewis, J. R., & Sauro, J. (2006). When 100% really isn't 100% : improving the accuracy of small-sample estimates of completion rates. *Journal of Usability Studies*, 1(3), 136-150.
- Lindgaard, G. (2006). Notions of thoroughness, efficiency and validity : Are they valid in HCI practice? *International Journal of Industrial Ergonomics*, 36, 1069-1074
- Lund, A. M. (1994). Ameritech's usability laboratory : From prototype to final design. *Behaviour & Information Technology*, 13(1 - 2), 67-80.
- Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM*, 33(3), 338-348.
- Neugebauer, C., & Spielmann, N. (1994). The ergonomics lab : A practical approach. *Behaviour & Information Technology*, 13(1 - 2), 45-50.
- Nielsen, J. (1990, 27-31 August). *Paper versus computer implementations as mockup scenarios for heuristic evaluation*. Paper presented at the IFIP TC 13 Third International Conference on Human-Computer Interaction : INTERACT'90, Amsterdam.
- Nielsen, J. (1992, May 3-7). *Finding usability problems through heuristic evaluation*. Paper presented at CHI'92 Conference on Human Factors in Computing Systems, New York.
- Nielsen, J. (1993). *Usability engineering*. Boston : Academic Press.
- Nielsen, J. (1994a). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25-62). New York : Wiley.
- Nielsen, J. (1994b). Usability laboratories. *Behaviour & Information Technology*, 13, 3-8.
- Nielsen, J. (2000). *Designing Web usability : The practice of simplicity*. Indianapolis : New Riders.

- Nielsen, J., & Landauer, T. K. (1993, 24-29 April). *A mathematical model of the finding of usability problems*. Paper presented at the INTERCHI'93 Conference on Human Factors in Computing Systems, New York, NY.
- Nielsen, J., & Mack, R. L. (Eds.). (1994). *Usability inspection methods*. New York : Wiley.
- Nielsen, J., & Tahir, M. (2001). *Homepage usability : 50 websites deconstructed*. Indianapolis : New Riders.
- Nogier, J.-F. (2005). *Ergonomie du logiciel et design Web : le manuel des interfaces utilisateur*. Paris : Dunod.
- O'Malley, C. E., Draper, S. W., & Riley, M. S. (1984, 4-7 September). *Constructive interaction : A method for studying human-computer-human interaction*. Paper presented at the IFIP INTERACT'84, London, U.K.
- Rubin, J. (1994). *Handbook of usability testing. How to plan, design, and conduct effective tests*. New York : Wiley.
- Sazegari, S. (1994). Designing a usability lab : A case study from Taligent. *Behaviour & Information Technology*, 13, 20-24.
- Scapin, D. L., & Bastien, J. M. C. (1997). Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology*, 16, 220-231.
- Scholtz, J., & Morse, E. (2002, May/June). A new usability standard and what it means to you. *SIGCHI Bulletin*, 10-11.
- Searle, J. (2001). *Rationality in action*. Cambridge : MIT Press.
- Spool, J., & Schroeder, W. (2001, 31 March - 5 April). *Testing Web sites : five users is nowhere near enough*. Paper presented at CHI'01 Conference on Human Factors in Computing Systems, New York, NY.
- Tricot, A., Plé gat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G., & Morcillo, A. (2003). Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH. In C. Desmoulins, P. Marquet & D. Bouhineau (Eds). *Environnements informatiques pour l'apprentissage humain* (pp. 391-402). Paris : ATIEF - INRP.
- Tricot, A., Puigserver, E., Berdugo, D., & Diallo, M. (1999). The validity of rational criteria for the interpretation of user-hypertext interaction. *Interacting with Computers*, 12, 23-36.
- Virzi, R. A. (1997). Usability inspection methods. In M. G. Helander, T. K. Landauer & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 705-715). Amsterdam : Elsevier.
- Westerink, J. H. D. M., Rankin, P. J., Majoor, G. M., & Moore, P. S. (1994). *A new technique for early user evaluation of entertainment product interfaces*. Paper presented at the Human Factors and Ergonomics Society 38th Annual Meeting.

- Wildman, D. (1995, july). Getting the most from paired-user testing. *Interactions*, 2, 21-27.
- Wilson, C., & Blostein, J. (1998, April). Usability techniques. Pros and cons of co-participation in usability studies. *Usability Interface*, 4.
