

Evaluer un dispositif de formation à distance : Principes et retour d'expérience

Stéphane Caro Dambreville

▶ **To cite this version:**

Stéphane Caro Dambreville. Evaluer un dispositif de formation à distance : Principes et retour d'expérience. Revue des Interactions Humaines Médiatisées (RIHM) = Journal of Human Mediated Interactions, 2008, 9 (2), pp.25-52. hal-00448500

HAL Id: hal-00448500
https://hal.science/hal-00448500

Submitted on 25 Jan 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluer un dispositif de formation à distance Principes et retour d'expérience

Evaluating a distance learning system: principles and feedback

Stéphane CARO DAMBREVILLE

Laboratoire LIMSIC, EA4177 CIMEOS, Université de Bourgogne
stephane.caro@u-bourgogne.fr

Résumé. La multiplication des dispositifs de formation à distance et la complexité de certaines plateformes utilisées au sein du système universitaire soulèvent des questions. En tout premier lieu, celle de la stabilisation des pratiques d'écriture des documents numériques. En second lieu celle de l'évaluation de ces dispositifs. C'est ce dernier point que nous souhaitons traiter dans le cadre de cet article. Qu'est-il possible d'évaluer en la matière ? Les dimensions d'évaluation sont multiples selon l'angle choisi. Nous nous limiterons à une approche centrée sur l'évaluation de l'interface personne-système. Le terrain d'expérimentation choisi est un site Internet d'apprentissage du droit du travail réalisé par le service universitaire de formation continue de l'Université de Bourgogne.
Mots-clés. Enseignement à distance, technologies d'information communication pour l'enseignement, ergonomie, évaluation ergonomique, utilité, utilisabilité, méthodes d'inspection, test utilisateur, questionnaires et entretiens.

Abstract. The multiplication of distance learning platforms and the complexity of some of those used within the university system raise questions. First of all, the stabilization of writing practices of digital documents. Secondly, the evaluation of these devices. We want to address this last point in this article. What is it possible to assess in this field? There are many dimensions of evaluation, depending on the angle chosen. We limit ourselves to an approach focused on evaluating the human-computer interface. The chosen testing ground is a website for law studies, developed by the University of Burgundy's further education section.
Keywords. Distance education, information communication technologies for education, ergonomics, usability testing, utility, usability, inspection methods, user testing, questionnaires and interviews.

1 Préliminaires

Tout d'abord, lors d'une démarche d'évaluation, il est nécessaire de définir ce que l'on souhaite évaluer. Pour cela il convient de considérer que l'apprenant se situe dans un contexte d'apprentissage « par l'intermédiaire d'un dispositif technique ». L'apprenant se trouve donc exposé à une double contrainte, celle de

devoir intégrer des connaissances nouvelles sur un domaine donné (le droit du travail dans notre cas) et en même temps, utiliser un dispositif technique pour ce faire (un site internet développé à cet effet). On comprend que cette double contrainte et cette structure de buts va rendre ce type de tâche particulièrement complexe dans le cas où le dispositif technique serait défaillant (c'est-à-dire, difficile à utiliser). Dans ce cas particulier, l'allocation de ressources cognitives « supplémentaires » à l'utilisation du dispositif technique plutôt qu'à la tâche d'apprentissage va provoquer de piètres performances en termes d'apprentissage. Dans cet article, nous nous concentrerons sur l'évaluation d'un de ces dispositifs technique à l'aide de méthodes « classiques » de l'ergonomie des interfaces personnes-système. Il s'agira donc de l'évaluation du dispositif technique plutôt que de l'évaluation du « gain » en termes d'apprentissage chez l'apprenant. Les trois méthodes choisies seront : l'évaluation par inspection (ou audit), les tests utilisateurs, les entretiens et questionnaires. Pour chacune de ces méthodes, nous donnerons des exemples de livrables. Il est peut-être utile de préciser avant tout qu'une bonne interface pour l'apprentissage est une interface qui ne consomme que très peu de ressources cognitives pour son utilisation. Une interface « transparente », discrète, va permettre à l'apprenant de concentrer ses ressources cognitives sur l'apprentissage des connaissances et non sur l'apprentissage ou la maîtrise du dispositif technique. L'objectif est donc clairement de minimiser la charge cognitive spécifiquement dédiée à l'exploitation du système (pour une présentation du concept de charge cognitive, cf. (Chanquoy *et al.,* 2007)).

1.1 Les dimensions de l'évaluation

On peut considérer que l'évaluation consiste généralement à comparer l'objet évalué à un modèle de référence afin d'en déduire des conclusions (Huart *et al.,* 2008). Bien entendu le « modèle de référence » peut être virtuel et envisagé comme une combinaison de recommandations adaptée au domaine de l'objet à évaluer. Qu'est-il possible d'évaluer dans un contexte de formation à distance ? L'évaluation doit reposer sur des critères formels et explicites. Quelles sont les dimensions d'évaluation à privilégier ? De nombreux critères d'évaluation sont mobilisables pour l'évaluation des documents numériques et dans le contexte de l'apprentissage, on peut encore en ajouter qui sont plus spécifiques. On distingue habituellement trois dimensions d'évaluation principales des systèmes interactifs (Senach, 1993) :

- l'utilité,
- l'utilisabilité,
- l'esthétique.

Tricot (Tricot e*t al.,* 2003) ajoute à ces trois dimensions l'acceptabilité (sentiment positif ou négatif à l'égard du produit, « intention » d'utilisation en fonction de nombreux paramètres : motivation, affects, culture, valeurs). La figure 1 ci-dessous présente les critères d'évaluation des trois premières dimensions.

Les notions d'utilité et d'utilisabilité sont souvent discutées dans la littérature ainsi que le périmètre qu'elles embrassent (Huart *et al.,* 2008). Dans le cas d'évaluation ergonomiques d'interfaces personnes-système, on s'intéresse à l'utilisabilité et éventuellement à certaines dimensions de l'utilité. Il n'est pas très difficile d'appréhender ce que recouvrent la plupart des critères d'utilité comme la vitesse de chargement d'une page, la présence de liens morts, la qualité du contenu (texte, image), la présence des fonctionnalités pertinentes pour l'utilisateur. L'utilisabilité ou la dimension ergonomique est la plus délicate à évaluer car les critères à respecter sont peu connus et font rarement l'objet de formation. Sur la figure 1, trois attributs principaux sont mentionnés pour l'utilisabilité : facilité

d'apprentissage, d'utilisation, qualité de la documentation. On peut aussi mentionner, selon les auteurs, l'efficacité d'utilisation, la facilité de mémorisation, l'utilisation sans erreurs, et la satisfaction de la part de l'utilisateur (Huart *et al.,* 2008). La dimension esthétique relève de compétences spécifiques mais laisse aussi place à une grande part de subjectivité. Elle est parfois considérée comme une composante de la satisfaction de la part de l'utilisateur au même titre que les préférences de l'utilisateur (Bastien et Scapin, 2001).

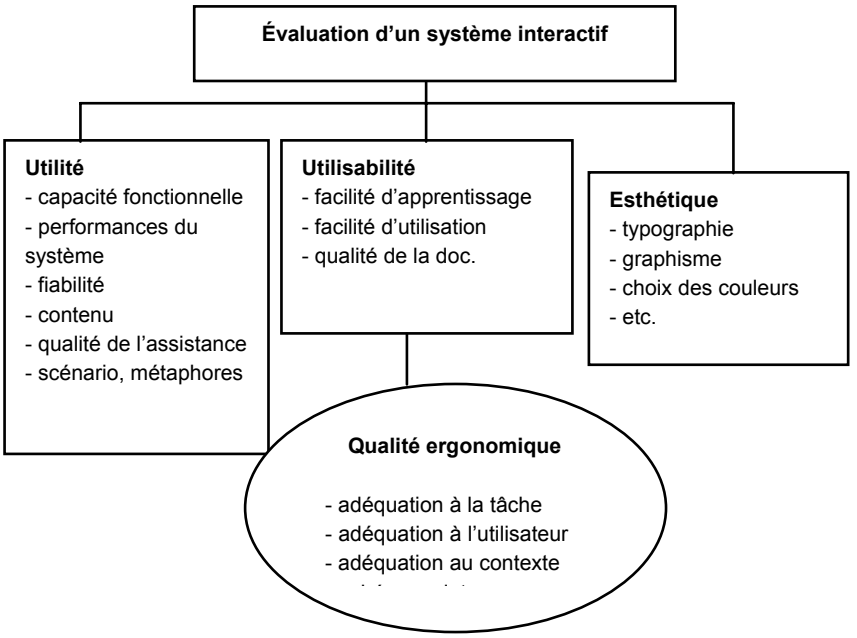


Figure 1. *Dimensions d'évaluation, adapté d'après Senach (1993)*

De ces trois dimensions, la plus importante semble être l'utilité. Une interface esthétique et facile à prendre en main, si elle n'offre pas les fonctions dont l'utilisateur a besoin (ou envie de se servir), si elle est instable techniquement, si elle présente des informations qui ne sont plus à jour, risque fort d'être délaissée. Au contraire, une interface qui offre un service dont l'utilisateur a besoin et d'une manière efficace sera utilisée même si elle est peu esthétique et quelque peu difficile à appréhender. L'interface du minitel, qui a eu son heure de gloire, en est un exemple. Ni modèle d'ergonomie ni de graphisme, les écrans du minitel ont été utilisés par l'intermédiaire d'un clavier, peu agréable de surcroît, parce que le minitel offrait des services réels (et même payants) aux usagers. Dans un contexte non-concurrentiel, une interface peut être essentiellement axée sur l'utilité. Dans un contexte concurrentiel, dans un contexte où les impératifs de sécurité ou de simplicité d'utilisation sont déterminants, la donne est tout autre et l'interface devra soigner son utilisabilité et/ou son esthétique. On emploie pour les domaines de l'utilisabilité et de l'utilité de nombreuses méthodes d'évaluation d'interfaces homme-machine. Un ouvrage francophone récent en recense pas moins d'une quinzaine assorties de fiches pratiques (Baccino *et al.,* 2005). Nous nous limiterons à des méthodes couramment utilisées pour évaluer l'utilité et l'utilisabilité du site de

formation au droit du travail. Ces méthodes sont liées à l'évaluation d'un système réel : l'inspection, les tests utilisateurs, les questionnaires et entretiens (cf. tableau 1).

1.2 Les diverses méthodes d'évaluation : classification rapide

Il existe de nombreuses possibilités de classement des méthodes d'évaluation selon que l'on s'intéresse au type d'approche (expérimentale *vs* prédictive) à l'implication d'utilisateurs, d'experts, aux aspects théoriques, situationnistes (en laboratoire, en situation réelle), au produit fini ou à des prototypes, etc. Nous présentons dans le tableau 1 une classification rapide de ces méthodes, proposée par Huart *et al.,* (2008).

	Système réel	**Système représenté**
Utilisateur réel	Méthodes d'observation (approches empiriques)	Rapports d'utilisateurs (approches empiriques)
Utilisateur représenté	Rapports de spécialistes (approches expertes)	Méthodes analytiques

Tableau 1. *Classification des méthodes d'évaluation d'interfaces personnes-systèmes d'après Huart et al., (2008)*

1.3 Présentation du dispositif évalué

Le cheminement que nous allons décrire s'inspirera d'évaluations pratiquées sur le site « Droit du travail : formation en ligne de l'institut d'AES » de l'université de Bourgogne sur une période de trois ans. Le produit testé comportait un module sur le syndicalisme. Certaines fonctions n'étaient pas implémentées ou terminées au moment de l'évaluation. Toutefois il s'agit bien du système réel et non représenté (maquette, prototype papier...) au sens de Huart *et al.,* (2008). A noter au passage qu'il n'est pas utile d'attendre qu'un produit de formation soit complètement finalisé pour l'évaluer. L'évaluation peut se concentrer sur les parties du dispositif qui sont déjà fonctionnelles. Une fois que ces parties sont évaluées et modifiées, le cas échéant, on étendra à tous les développements futurs les résultats de l'évaluation

.



Figure 2. *Ecran de choix des activités*

Le site « Droit du travail : formation en ligne de l'institut d'AES » propose différents types d'activités (voir figure 2, écran de choix des activités) :

- lecture d'un cours (mémento)
- exercices (études de cas concrets)
- tests (positionnement rapide par questions fermées pour se situer par rapport à des connaissance sur un thème)
- échanges avec d'autres apprenants, un groupe, un tuteur
- consultation de documentation liée au domaine.

A chaque type de tâche correspond une logique d'utilisation de l'interface et des outils appropriés qui peuvent donner l'impression d'une certaine complexité.

Utilisateurs

Les utilisateurs de ce dispositif sont des étudiants de licence d'AES. Toutefois une utilisation du dispositif pour la formation professionnelle est envisageable. On utilisera le terme d'utilisateur dans la suite de cet article plutôt qu'apprenant. En effet, l'évaluation décrite ci-après porte sur le dispositif en termes d'interaction et non directement sur l'acquisition des connaissances.

Contexte

Les utilisateurs travaillent avec le dispositif individuellement, à domicile ou en salle informatique en libre service. Le dispositif est pensé comme un complément à un cours classique dispensé à l'Université. Le site présente donc des liens avec le cours intitulés « annonces ».

Le site dans sa version évaluée comporte de nombreux écrans et présente des possibilités de navigation multiples d'une partie à l'autre. Les organisateurs de structure et de mise en forme (para-linguistiques : couleur, enrichissement typographique, multifenêtrage, icônes etc.) ont été utilisés abondamment afin de faciliter l'utilisation de ce dispositif de formation.

1.4 Panorama des méthodes

Les méthodes d'évaluation de l'utilisabilité sont classées généralement selon plusieurs facteurs. On peut différencier les méthodes qui nécessitent le recours à des utilisateurs et les autres. On utilisera alors le qualificatif d'observations participantes ou tests utilisateurs. Le second facteur consiste à classer les méthodes selon les outils qui les assistent. Ainsi des outils logiciels d'évaluation automatique permettent d'évaluer certaines dimensions d'utilisabilité comme la densité informationnelle par simple comptage des mots, des liens… D'autres outils informatiques assistent partiellement l'évaluateur ou permettent par exemple de suivre le parcours dans les écrans et de le reconstituer à l'aide de mouchards informatiques. Chaque méthode présente des forces et des limites et on consultera valablement l'ouvrage cité précédemment pour en avoir le détail (Baccino *et al.*, 2005). À présent, nous allons nous recentrer sur les principales méthodes qui seront mobilisées dans le cadre de cette évaluation.

2 Les méthodes d'inspection

2.1 Les méthodes d'inspection : présentation

Dans un premier temps et avant de recourir à des méthodes coûteuses comme celles qui nécessitent d'observer ou d'interviewer des utilisateurs, il convient de vérifier que l'interface respecte les normes et standards existants. Pour ce faire, on demande à un évaluateur (dont l'expertise peut varier) d'évaluer le dispositif sur la base de standards existants. Les standards utilisés peuvent être multiples comme

une norme de qualité ISO 9241-10, mais aussi des listes de critères de qualité ergonomiques comme ceux de Nielsen (1994) ou de Bastien et Scapin (1993) repris dans la norme AFNOR Z67-133-1. L'évaluateur analyse l'interface à l'aune des standards ou des critères choisis et il se familiarise avec l'interface afin de comprendre son utilisation et de prévenir les problèmes qui pourraient surgir dans ce contexte.

Concrètement l'évaluateur procède en quatre étapes distinctes qui peuvent se recouvrir temporellement.

- *La familiarisation avec le dispositif*,
- *La maîtrise de recommandations, (guidelines, grille de critères)*
- *L'application des recommandations retenues au dispositif*
- *La rédaction du rapport d'évaluation.*

La *familiarisation avec le dispositif* nécessite une bonne connaissance de la cible (public visé par le système), du contexte d'utilisation et des tâches à effectuer avec le système (résolution de problèmes, acquisition de connaissance, recherche d'information, etc.). Une exploration du système pendant une heure environ avec des commentaires de l'équipe de conception peut permettre une première prise en main dans de bonnes conditions. La formalisation de la structure de navigation (arborescence) sur un grand document est aussi utile à cette étape. L'équipe de conception n'a pas forcément de document de ce type qui soit suffisamment complet. L'évaluateur aura intérêt à s'en fabriquer un en collant des post-it sur un tableau par exemple. La numérotation des écrans sur ce documents sera utile aux étapes suivantes.

La maîtrise de recommandations, guidelines, grille de critères de qualité ergonomique sera plus ou moins importante selon l'expertise de l'évaluateur. Une bonne base de départ avec peu de critère peut consister à utiliser les critères de Nielsen qui sont peu nombreux (10) ou ceux de Bastien et Scapin (1994) qui sont également peu nombreux (18) et fournis avec des définitions précises et de nombreux exemples appliqués aux interfaces personnes-système ou aux documents numériques (Bastien *et al.,* 1998).

L'application des recommandations au dispositif est la phase d'analyse proprement dite au cœur de cette activité d'inspection. Pour un évaluateur novice, on pourra procéder à une inspection critère par critère lors du parcours dans le dispositif. L'évaluateur expert qui maîtrise et mémorise les critères d'évaluation parcourra les parties représentatives de l'application et identifiera les critères non respectés en fonction des situations d'interaction. Les problèmes détectés seront relevés aussi précisément que possible (titre, numéro de l'écran, description du problème, critère concerné et copie d'écran si nécessaire selon la complexité du problème). Pour les écrans comportant de nombreux problèmes, une copie d'écran commentée avec la liste des problèmes relevés peut être envisagée.

On pourra également, pour faciliter l'application des recommandations au dispositif (3ème étape) se construire sous forme d'écran (page Web) ou sur papier des fiches par critère (définition du critère et exemples de défauts). Ces fiches sont destinées à passer en revue tous les défauts de l'interface et à les répertorier directement « sur le critère concerné ». La figure 3 présente un dispositif logiciel de ce type conçu par les étudiants de la Licence Pro Activités et Techniques de Communication du département Services et Réseaux de Communication de l'IUT de Dijon (Julien Saurin et Rémi Chouvenc, promotion 2008). Ce dispositif peut être ouvert dans une fenêtre sur l'écran de l'évaluateur et le système à évaluer dans une autre fenêtre à proximité sur le même écran.



Figure 3. *Présentation du critère densité informationnelle (Bastien et Scapin 1998), le champ de texte permet de relever les problèmes liés à la densité informationnelle ainsi que leur localisation dans l'application. L'échelle bi-polaire permet d'attribuer une note à chaque critère*

Chaque bouton du bandeau gauche (qui représente un des 18 critères de qualité ergonomique de Bastien et Scapin) est cliquable pour permettre à l'utilisateur de revenir à n'importe quel critère lors de l'évaluation. Les boutons en bleu signifient que pour ces critères, des observations ont déjà été mentionnées et les boutons en gris sont des critères pour lesquels aucun problème n'a été signalé.

Lorsque la souris passe sur l'un des boutons, un escamot apparaît après 0,5 secondes pour préciser à quel critère correspond chaque numéro ; au bout de 3 secondes, le texte devient défilant afin de compléter la description. Si l'utilisateur clique, il est automatiquement conduit vers le critère sélectionné (cf. figure 4). Une bordure noire entoure le carré du critère courant (8 sur la copie d'écran, figure 4).

A l'issue de l'évaluation, le dernier écran présente la notation de chaque critère (échelle bi-polaire) et la liste des problèmes relevés par l'évaluateur classée par critère (Figure 5). Il convient de signaler que quand les problèmes relevés sont nombreux, une classification des problèmes par écran est plus adaptée pour l'équipe de conception.

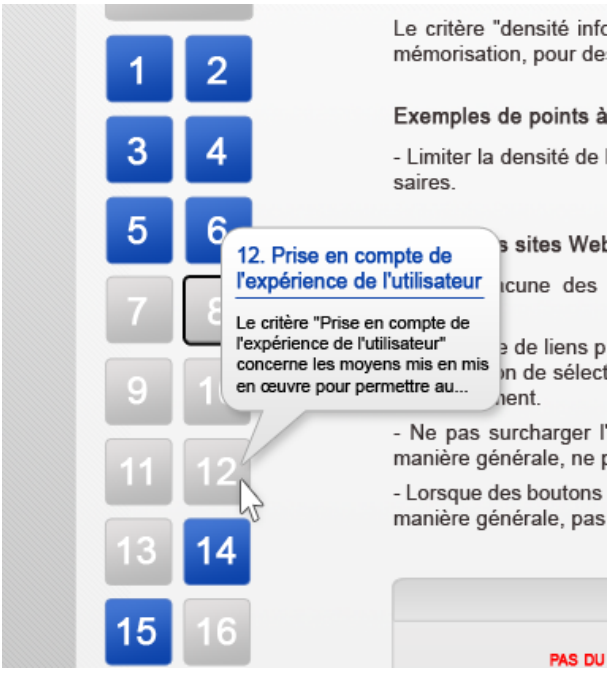


Figure 4. *Survol d'un critère avec la souris, critère prise en compte de l'expérience de l'utilisateur (Bastien et Scapin, 1998)*

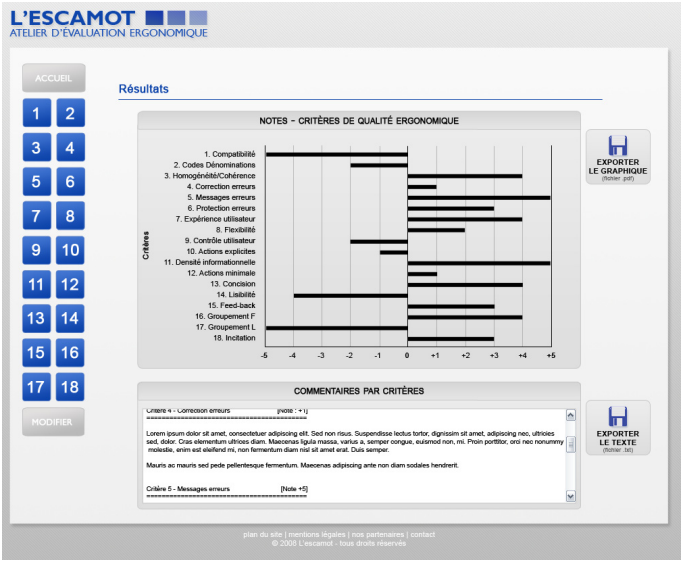


Figure 5. *Synthèse de l'évaluation, le champ de texte liste les critères non respectés, leur description ainsi que leur localisation dans l'application*

La rédaction du rapport comprend la présentation d'une courte introduction sur le dispositif évalué et du contexte de l'évaluation, la liste des critères d'évaluation choisis (les concepteurs ne connaissent pas toujours les critères mobilisés pour l'évaluation). La présentation des problèmes constatés pourra suivre l'ordre des critères mobilisés (si les défauts sont peu nombreux) ou être organisée par écran s'il y en a beaucoup sur chaque écran. Dans ce dernier cas, une copie d'écran de la partie problématique et une maquette d'écran qui solutionne les problèmes sera proposée idéalement. Le groupement de certains défauts et leur présentation globale peut être envisagée quand ces défauts reviennent sur de nombreux écrans (certains concepteurs ont des habitudes qui provoquent des problèmes répétitifs). Enfin une rubrique de remarques générales et/ou diverses peut être envisagée pour faire remonter à l'équipe de conception des informations sur des *bugs* éventuels, impressions diverses, coquilles orthographiques, problèmes esthétique... même si ces remarques ne sont pas directement liées à l'utilisabilité ou à l'utilité du produit. Une conclusion sur la perception du produit évalué et la pondération des défauts constatés (défauts majeurs, mineurs...) pourra être d'une grande utilité pour l'équipe de conception. La conclusion peut orienter l'équipe de conception vers une autre technique d'évaluation (tests utilisateurs, entretiens, questionnaires etc.) selon les résultats obtenus lors de l'inspection. La technique suivante (test utilisateur par exemple) pourra se placer après les éventuelles modifications du produit préconisées dans le rapport d'inspection.

2.2 Inspection du dispositif de formation

Dans le cadre de l'évaluation du dispositif de formation, nous avons choisi d'utiliser les critères de qualité ergonomiques de Bastien et Scapin (1993). À partir du recueil d'une synthèse de nombreuses études expérimentales sur les interfaces personne-système, Bastien et Scapin ont élaboré une grille d'analyse fondée sur une liste de critères qui permettent d'évaluer l'ergonomie des interfaces homme-machine et, par inclusion, des documents numériques. La liste de critères (voir ci-dessous) permet de classer les défauts, de façon à prévenir les problèmes d'utilisation et à faciliter la conception.

L'énumération ci-dessous présente la liste des critères d'utilisabilité ou de qualité ergonomique de Bastien et Scapin. Les 18 critères élémentaires (qui ne peuvent se subdiviser) apparaissent en caractères gras.

1. Guidage
    - 1.1 ***Prompting*** **ou incitation**
    - 1.2 Groupement/Distinction entre items
        - 1.2.1 **Groupement/Distinction par la localisation**
        - 1.2.2 **Groupement/Distinction par le format**
    - 1.3 ***Feed-back*** **immédiat**
    - 1.4 **Lisibilité**
2. Charge de travail
    - 2.1 Brièveté
        - 2.1.1 **Concision**
        - 2.1.2 **Actions minimales**
    - 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**
   5.3 **Correction des erreurs**
6. **Homogénéité / Cohérence**
7. **Signifiance des codes et dénominations**
8. **Compatibilité**

On trouvera une définition de chacun de ces critères et des exemples d'application donnés par les auteurs dans le contexte d'interfaces homme-machine dans Bastien et Scapin (1993), et des adaptations au contexte des documents numériques dans Bastien *et al.* (1998). Les listes de critères, qu'elles soient plus ou moins étendues, présentent souvent des éléments similaires comme le contrôle utilisateur. Une force de la liste de Bastien et Scapin est qu'elle a été validée expérimentalement auprès d'évaluateurs ergonomes et non-ergonomes lors d'expériences d'évaluation d'interfaces fictives comportant des défauts introduits volontairement.

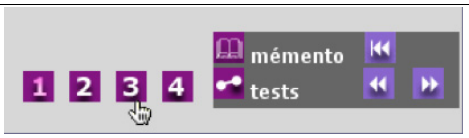
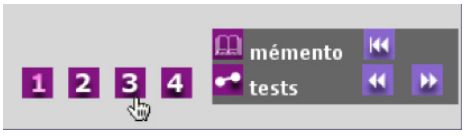
Le dispositif a été évalué par une personne expérimentée dans l'évaluation d'interfaces personnes-système. Le dispositif de formation en ligne a été présenté à l'évaluateur par son concepteur de manière informelle (pendant environ 1 heure) puis un accès personnalisé (Identifiant et mot de passe) a été fourni à l'évaluateur afin de se familiariser seul avec le dispositif. A la suite de la phase de familiarisation, l'évaluation a pu commencer directement, car l'évaluateur avait l'habitude d'utiliser la grille de critères choisie. Les temps approximatifs consacrés à cette étude peuvent se répartir selon les phases en :

- Familiarisation avec le dispositif (2 heures),
- Maîtrise de recommandations, (*guidelines*, grille de critères) (néant, évaluateur expert)
- Application des recommandations retenues au dispositif (20 heures)
- Rédaction du rapport d'évaluation de 22 pages avec copies d'écran (10 heures).

Le temps consacré à la phase 1 est relativement stable (environ 1 à 3h selon les projets et la complexité de la structure de navigation). Le temps consacré à la phase 2 dépend des compétences de l'évaluateur et/ou de sa connaissance de la grille de critères/Norme qu'on lui demande d'utiliser. Le temps consacré à la phase 3 dépend de l'aspect systématique de l'analyse (tous les écrans, une partie représentative dans les projets ambitieux) et du niveau de conseil (relevé simple de défauts et éventuellement : propositions de solutions, maquettes d'écrans « de remplacement », re-structuration de l'architecture du produit). Plus le niveau de conseil sera élevé et plus l'évaluation demandera du temps. Enfin la rédaction du rapport dépendra aussi de la complexité du projet et du niveau de conseil (solutions prototypes, maquettes etc.)

2.3 Résultat de l'inspection

Le tableau 2 présente quelques problèmes de conception relevés lors de l'inspection du site de formation au droit du travail. Les défauts sont classés selon la liste des critères de Bastien et Scapin. Ce tableau témoigne de la première partie de la phase 3 « l'application des recommandations retenues au dispositif ». Il ne s'agit dans un premier temps que de l'inventaire des problèmes qui seront repris dans le rapport avec des propositions de solutions.

Problème	Critère non respecté
Un moteur de recherche est disponible. Toutefois l'utilisateur ignore la portée de la recherche. Est-ce qu'il s'agit de la partie courante « module syndicalisme », de toute la partie sur le droit du travail, d'une recherche dans la bibliothèque ?	Guidage : *Prompting* ou incitation
 L'utilisateur ignore la portée des boutons tourne page << >>. L'utilisateur ne sait pas s'ils tournent les pages à l'intérieur de la partie tests (passage du premier test au second) ou s'ils permettent de passer d'une question à l'autre au sein d'un même test. Ceci est dû à la proximité de ces boutons avec le label « tests ».	Groupement / Distinction entre items
Les zones « cliquables » ne sont pas toujours distinguées des zones « non-cliquables ». Par exemple, si les zones qui présentent un dispositif *rollover* (activation par survol de la souris) sont parfois cliquables, parfois elles ne le sont pas. Les titres des aplats colorés de l'écran « accueil module », « choix des activités » par exemple ont un dispositif *rollover* et ne sont pas des liens.	Homogénéité, cohérence
Quand l'utilisateur clique sur certains livres de la bibliothèque et qu'il ne se passe rien, il n'a pas toujours d'indication de chargement (ou de la présence ou d'absence de contenu.)	*Feed-back* immédiat
 Des éléments graphiques ayant des aspects similaires sont parfois cliquables, parfois non-cliquables. Les chiffres qui représentent des pages sont cliquables, tandis que les boutons proches de « mémento » et de « tests » ne le sont pas.	Homogénéité, cohérence
Chaque fenêtre doit avoir un titre spécifique. Le titre « Droit du travail » ne peut donc pas convenir pour toutes les fenêtres. C'est ce qui est spécifique à une fenêtre qui doit être le titre le plus visible sur celle-ci. Par exemple « Exercice » « Test » « Activités »	Guidage : *Prompting* ou incitation

« Présentation du site » « Démonstration ».	
Dans les « exercices », comment l'utilisateur peut retrouver les notes du calepin d'une fois sur l'autre ? Est-il possible de les enregistrer ? Sont-elles perdues d'une connexion à l'autre ? Ceci n'apparaît pas clairement dans l'interface.	Guidage : *Prompting* ou incitation
Dans les tests, quand l'utilisateur revient en arrière pour refaire une question, par exemple de la question 4 à la question 1, les questions 3 et 2 défilent avant que la question 1 apparaisse. Mieux vaudrait un accès direct (non séquentiel).	Brièveté : actions minimales

Tableau 2. *Exemples de défauts relevés lors de l'évaluation par inspection. Les défauts sont décrits colonne de gauche et les critères de qualité ergonomiques non respectés sont mentionnés dans la colonne de droite*

Bien entendu, pour chaque problème constaté, il convient à l'évaluateur, si cela fait partie de la commande, de proposer une solution qui résolve le problème décrit préalablement. Pour les écrans qui comportent de nombreux défauts d'utilisabilité, il est possible de présenter l'écran évalué accompagné de la liste de problèmes et de proposer ensuite une maquette d'écran qui résolve les problèmes répertoriés (ce cas ne s'est pas présenté lors de cette évaluation par inspection).

A ce type de défauts d'utilisabilité vont s'ajouter des défauts d'utilité qui pourront également être signalés (fonctionnalités manquantes par exemple, *bugs*, problèmes d'affichages, liens morts, etc.) Le tableau 3 présente un exemple de défaut relevant de l'utilité du dispositif.

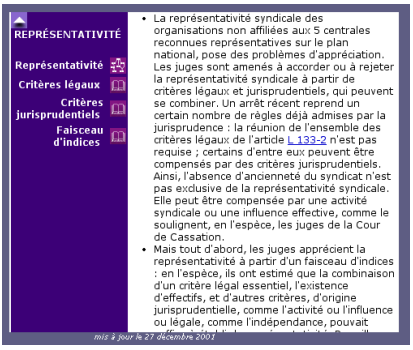
Problème	**Critère non respecté**
 Dans le module Syndicat > Docs > Représentativité, l'utilisateur ne peut faire défiler le texte faute de barre de défilement	Capacité fonctionnelle

Tableau 3. *Exemples de défauts relevés lors de l'évaluation par inspection. Le défaut est décrit colonne de gauche et le critères d'utilité non respecté est mentionné dans la colonne de droite*

2.4 Inspection : conclusion

Quand l'inspection ne relève pas de critères explicitement énoncés, on parle alors d'inspection heuristique. Selon les compétences de l'évaluateur, on pourra qualifier l'évaluation d'experte (Nogier, 2005). Les limites de l'inspection sont celles d'une situation d'évaluation quelque peu « artificielle » car l'évaluateur ne présente pas nécessairement un profil analogue aux utilisateurs réels du dispositif. D'autre part, l'inspection consiste à évaluer tout ou partie des écrans du dispositif sans nécessairement s'immerger dans un scénario à exécuter conformément à l'objet du dispositif de formation. Pour renforcer la validité écologique de l'évaluation par inspection, on peut demander à l'évaluateur de suivre un scénario déterminé qui va l'amener à adopter un parcours relativement similaire aux parcours des futurs utilisateurs du dispositif de formation (technique du parcours cognitif). La principale faiblesse de la méthode est qu'il est difficile pour l'évaluateur de prévoir les failles du système quand il s'agira d'utilisateurs réels qui interagiront avec le dispositif dans leur contexte d'utilisation habituel. Toutefois cette méthode permet d'éviter les défauts fréquents des interfaces moyennant un investissement modéré. Cette méthode est également tributaire de la qualité / de l'expérience des évaluateurs. A noter que ces derniers doivent s'imprégner du domaine de l'interface évaluée et / ou être assistés par un spécialiste de ce domaine (le droit du travail dans notre cas) (Baccino *et al.*, 2005). La liste des critères utilisée pour l'évaluation par inspection peut être complétée ou modifiée selon le contexte afin de l'adapter à l'objectif de l'interface testée (contexte d'apprentissage à distance).

3 Les tests utilisateur

3.1 Les tests utilisateurs : présentation

Il s'agit cette fois de tester le produit en phase finale, ou en phase de prototypage auprès d'un petit nombre d'utilisateurs représentatifs du public auquel le dispositif est destiné. On peut conduire un test avec différents groupes d'utilisateurs qui vont exécuter des tâches sur des versions du dispositif qui peuvent être différentes (plan de test inter-sujets). Par exemple, un groupe de cinq utilisateurs exécute les tâches avec la maquette d'un site de formation A et un groupe de cinq utilisateurs exécute les tâches avec la maquette d'un site de formation B. Bien entendu, les deux documents ne varient que sur le facteur que l'on veut étudier (deux systèmes de navigation différents par exemple). Avec un plan de test intra-sujets, le même groupe de participants teste les deux versions du document avec les mêmes tâches ou des tâches différentes pour chaque document mais qui sont analogues et comparables en difficulté. Dans ce cas, l'ordre de passation des sujets pour le test est contrebalancé, la moitié des participants commencent par tester le document A et l'autre moitié le document B

. Généralement les tests ont lieu individuellement. La participation de cinq à dix utilisateurs représentant la population ciblée par le produit permet d'identifier la grande majorité des problèmes. Généralement les tests ont lieu en individuel. La participation de 5 utilisateurs représentant la population ciblée par le produit permettrait d'identifier 85 % des problèmes (Nielsen, 2000). Toutefois des travaux plus récents ont démontré qu'à moins de 10 voire 20 utilisateurs, le risque était grand de négliger de graves problèmes d'utilisabilité (Baccino *et al.*, 2005). Il semble que 10 à 20 utilisateurs permettent de détecter respectivement 80 % et 95 % des problèmes d'utilisabilité (Faulkner, 2003). Par ailleurs, 15 à 20 utilisateurs sont

nettement préférables du point de vue de la fiabilité statistique (Baccino *et al.*, 2005). Si le produit est particulièrement défaillant, on pourra poursuivre les tests en ajoutant un utilisateur supplémentaire dans chaque groupe d'utilisateurs testés jusqu'à avoir circonscrit les problèmes posés par le produit (quand un nouvel utilisateur ne détecte pas de nouveaux problèmes). Nielsen a montré qu'il valait mieux faire 3 tests avec 5 utilisateurs à différentes étapes de la conception plutôt qu'un seul test avec 15 utilisateurs (Nielsen, 2000).

	Interface/Tâche	**Interface/Tâche**
Participant 1	Interface 1 Tâche 1,2,3,4	Interface 2 Tâche 5,6,7,8
Participant 2	Interface 2 Tâche 5,6,7,8	Interface 1 Tâche 1,2,3,4
Participant 3	Interface 1 Tâche 4,3,2,1	Interface 2 Tâche 8,7,6,5
Participant 4	Interface 2 Tâche 8,7,6,5	Interface 1 Tâche 4,3,2,1
Etc.		

Tableau 4. *Exemples de protocole avec documents/interfaces et tâches contrebalancées*

Généralement, la procédure la plus simple consiste à demander aux utilisateurs de faire ce pourquoi le dispositif est fabriqué, acquérir des connaissances sur un domaine, une organisation ou exécuter une procédure. Après un temps d'exploration libre du dispositif par l'utilisateur pendant quelques minutes, on lui demandera soit de rechercher des informations qui peuvent nécessiter de consulter plusieurs écrans, soit d'exécuter une procédure pour laquelle le document numérique est conçu (ou d'autres tâches représentatives de son usage). L'expérimentateur peut être présent à proximité du participant pour administrer la consigne générale : « Vous allez tester le site… Ce site est encore en phase de prototypage… Ce n'est pas vous qui êtes testés mais le produit… », puis les consignes (ou scénarios) spécifiques pour que le participant exécute les opérations demandées. Il est conseillé d'écrire les différentes consignes (consigne générale et consignes des tâches à exécuter) pour que les participants puissent participer au test dans les mêmes conditions.

3.2 Test utilisateur du dispositif de formation

L'évaluation du site de formation au droit du travail a été menée auprès de dix participants (Taux d'identification de défauts d'environ 80% selon certains travaux, cf. paragraphe ci-dessus). Cette expérience a fait suite à l'évaluation par inspection. Cette inspection a entraîné un certain nombre de modifications dans le prototype d'alors. C'est donc ce site, déjà en partie aménagé, qui a fait l'objet du test décrit ci-dessous. Nous donnerons une brève description de la méthode utilisée ainsi que quelques résultats.

Objectifs du test

Il s'agit d'évaluer le produit de formation dans un contexte qui soit le plus proche possible de la réalité. Il est difficile de prédire l'interaction des nombreuses variables qui entrent en jeu lors d'une tâche aussi complexe que l'analyse d'un problème juridique par le biais d'une interface homme-machine. L'objectif est d'identifier les éventuels défauts de l'interface et les manques qui auraient pu échapper à l'analyse de la première évaluation par inspection. Il s'agit aussi de sélectionner des alternatives de conception (voir l'exemple concernant la navigation

par des liens en bandeau ou par des boutons « tourne page »). Des changements d'organisation peuvent aussi être envisagés le cas échéant.

Protocole expérimental

L'évaluation consiste à observer des participants en situation d'utilisation du dispositif de formation. Les quatre tâches représentatives à effectuer sont proposées par l'expérimentateur à l'aide de consignes écrites (voir ci-après). Les participants doivent verbaliser (dire « tout haut » ce qu'ils font, ou pensent). Les participants sont filmés pour la mesure des temps, l'analyse des parcours, et l'enregistrement de leurs commentaires. Les participants passent l'expérience individuellement, en présence de l'expérimentateur.

Le fait que les utilisateurs soient observés par un tiers peut provoquer un biais, l'effet Hawthorne (les résultats positifs ou négatifs peuvent ne pas être dus qu'aux facteurs expérimentaux). Différentes techniques expérimentales existent pour s'en prémunir (salles d'expérimentation avec vitre sans tain, procédures d'observation indirectes : magicien d'Oz, évaluation coopérative... (Baccino *et al.*, 2005)). Lors d'une utilisation des tests utilisateur en milieu industriel, un éventuel biais dû à l'observation est tout à fait acceptable. Si l'observation de participants a pour objectif l'obtention de résultats dans un contexte de recherche fondamentale, on aura avantage à se prémunir de ces phénomènes.

Dispositif d'observation

La figure 6a présente la configuration du système d'observation en situation. L'expérimentateur est à proximité de la personne observée[1]. Dans le dispositif de recueil présenté ci-dessous, un convertisseur transforme l'image de l'écran en signal vidéo. Cette image est intégrée à un plan d'ensemble de l'espace de travail du participant enregistré par une caméra. Cette vue permet de distinguer les temps de « lecture des consignes » (données sur papier) des temps de consultation des écrans. Elle permet également d'enregistrer les réactions des participants et les commentaires sonores qui sont formulés pendant l'expérience (verbalisations simultanées). Lors du mixage des deux images, une incrustation de l'image du plan d'ensemble est faite en haut à droite de l'image de l'interface. Cette zone de l'écran présente en effet peu d'interaction généralement (cf. figure 6b). L'expérimentateur donne les consignes aux participants, recueille les observations de ces derniers et note les problèmes que les participants rencontrent. Le dispositif d'enregistrement vidéo du parcours de l'utilisateur peut permettre une analyse fine et *a posteriori* des parcours des participants au sein du dispositif de formation. La durée de l'enregistrement vidéo a oscillé entre 45 minutes et 1 heure 30 environ pour chaque sujet selon leur rapidité. Les vidéos ont été dépouillées manuellement lors d'un visionnement systématique. Les variables sélectionnées avec l'équipe de conception (cf. ci-après le Tableau 5) sont enregistrées lors du visionnement dans le logiciel de statistique associé au laboratoire d'observation (Logiciel *The Observer* de la société Noldus). L'enregistrement des variables d'intérêt se fait par appui de touches dédiées au clavier d'un ordinateur (par exemple la touche « a » du clavier est enfoncée chaque fois que le sujet est sur la page d'accueil, la touche « c » lorsque le sujet lit la consigne écrite, la touche « r » quand il répond à une question, etc.). Cet enregistrement permet d'obtenir des statistiques qui sont présentées au paragraphe

1 L'expérimentateur et les éventuels observateurs peuvent également suivre le déroulement de l'expérience depuis un autre local sur écran ou à travers une glace sans tain. Cela permet de faire varier le nombre d'observateurs sans que le participant en ait conscience. L'équipe de conception peut ainsi venir observer quelques utilisateurs.

3.3. Le coût du visionnement est important eu égard au retour sur investissement qu'il procure (éléments formels d'appréciation de durées et de nombre d'occurrence d'événements). Il ne se justifie que pour des projets ambitieux où la mesure du temps ou l'utilisation de fonctionnalités par exemple doivent être quantifiés formellement (comparaison entre deux systèmes de navigation par comptage du nombre de clics sur ces dispositifs de navigation par exemple). Il est rare de s'apercevoir, lors du visionnement, d'éléments que l'expérimentateur n'aurait pas pu déceler pendant les observations. Par contre, les enregistrements vidéo sont déterminants pour la présentation des résultats à l'équipe de conception qui a parfois du mal a admettre qu'il se produise des phénomènes problématiques en dehors d'éléments chiffrés et de séquences vidéo. Au-delà de l'aspect statistique, l'enregistrement vidéo a donc parfois valeur de preuve.

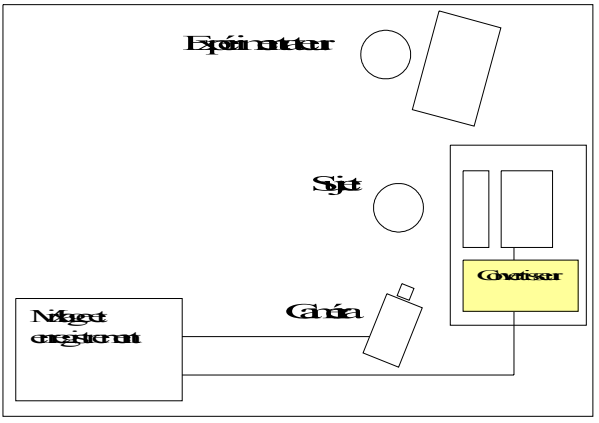


Figure 6a. Plan du dispositif



Figure 6b. Image enregistrée

Méthodologie

Participants

Deux variables indépendantes peuvent faire l'objet d'analyses : le genre (masculin, féminin) et la compétence en droit du travail (expert, novice). Nous ne retiendrons que la compétence dans le cadre de cette étude.

Les participants se répartissent en deux groupes de cinq. Sur les dix participants, cinq sont considérés comme « experts » dans le domaine juridique et cinq sont considérés comme « novices » mais disposent de compétences qui leur

permettent d'utiliser le dispositif (étudiants familiers des interfaces Web et des TIC). La répartition expert novice s'est faite sur la base du nombre d'heures de formation en droit du travail et de la pratique professionnelle. Aucun sujet n'a été écarté dans le cadre de cette expérience. Les sujets ont été recrutés sur la base du volontariat. Les participants considérés comme « experts » ont bénéficié récemment d'une solide formation en droit du travail ou occupent un poste dans le domaine des ressources humaines. Les participants « novices » sont des étudiants adultes en formation continue, sans formation juridique pour trois d'entre eux (niveau bac+2 minimum) et deux étudiants de formation initiale en DUT Information-Communication ayant suivi peu de cours de droit du travail pendant leur cursus.

Matériel

Un certain nombre de pré-requis ont été vérifiés concernant l'utilisation du site. Les messages d'aide sont disponibles, même si leur contenu est provisoire. L'application fonctionne normalement en général, même si toutes les commandes réelles ne sont pas implémentées (fonction « rechercher » par exemple). Dans le cas où les commandes ne fonctionneraient pas complètement, l'expérimentateur indique au participant le comportement futur de l'application. La consigne écrite de départ est la suivante.

Consignes

Les consignes données à lire sur un support papier ont été les suivantes :

L'objectif de la séance est de valider un nouvel outil de formation en ligne de l'Université de Bourgogne. Ce site internet est destiné à l'apprentissage de différents modules de droit du travail.

Je vais vous demander d'effectuer certaines tâches. Au fur et à mesure que vous utiliserez le logiciel pour exécuter les tâches indiquées, je pourrais vous poser des questions sur ce que vous voyez ou sur ce que vous attendez des fonctions que vous utilisez. Sentez-vous libre de donner des observations pendant la session. Nous vous demandons de « penser tout haut » en indiquant les actions que vous voulez faire et les informations que vous recherchez. Il n'y a pas de question stupide ou de mauvaise réponse. Ce produit est un prototype, ne soyez pas étonné qu'il puisse réagir d'une façon inattendue. C'est le produit de formation que nous testons et non vos capacités à répondre aux différentes questions.

L'étude comporte trois questions ponctuelles et un cas pratique. A l'issue de cette étude, nous répondrons à vos différentes questions et prendrons note de vos remarques éventuelles.

Vous allez essayer le site : Droit du travail en ligne.

À la suite de la lecture de la consigne quatre tâches sont à réaliser. En voici deux exemples :

Tâche 1

Répondre à la question suivante en la justifiant : Un salarié ayant un an d'ancienneté conteste le fondement de son licenciement. Il s'est vu notifier son licenciement le jour même de l'entretien préalable. Cette précipitation de l'employeur peut-elle faire l'objet d'une sanction par les juges ? Illustrer avec une décision jurisprudentielle. Saisissez votre réponse dans la fenêtre Word ouverte à l'écran.

Tâche 4

Répondre par écrit au maximum possible des 5 questions figurant dans le cas pratique du module « licenciement ». Saisissez votre réponse dans la fenêtre Word ouverte à l'écran.

Trois tâches sont assez modestes en temps de travail nécessaire (quinze minutes environ) et en nombre d'écrans à parcourir (tâches 1, 2, 3). La dernière tâche (4) est la résolution du cas pratique du module « Licenciement » et nécessite une interaction prolongée avec le dispositif ainsi que le parcours de nombreux

écrans au sein de l'application (environ une heure). L'ordre d'exécution des tâches 1, 2, et 3 est contrebalancé pour éviter les effets d'ordre.

La tâche quatre est donnée à tous les participants après l'exécution des 3 premières tâches.

Variables quantitatives recueillies

Les temps consacrés à seize types d'événements sont mesurés. Le choix des événements mesurés se fait en concertation avec l'équipe de conception. Un événement correspond à la consultation d'un type d'écran, le choix d'une fonction ou l'exécution de tâches annexes liées à l'activité (lecture de la consigne, rédaction de la réponse). Le temps est mesuré en secondes. Le détail des événements pris en compte est présenté dans le tableau 5 (le nom des variables dans la deuxième colonne correspond au nom affecté à chaque variable dans le logiciel de dépouillement).

Elément	**Nom Variable**	**Définition**	**Type de variable**
1	Accueil	Temps passé sur la page d'accueil	État
2	Carte	Temps passé sur un écran « carte conceptuelle »	État
3	Choix	Temps passé sur l'écran « choix d'activités »	État
4	Plan	Temps passé sur l'écran « plan du site »	État
5	NavDiv	Temps passé sur d'autres écrans intermédiaires de navigation	État
6	Réponse	Temps passé à la saisie de la réponse dans un document Word	État
7	Recherch[e]	Temps passé à rechercher un mot	État
8	Texte	Temps passé à la lecture d'un texte à l'écran	État
9	Surligna[ge]	Clic sur le bouton de la fonction « surlignage » d'un paragraphe	Événement ponctuel (sans durée)
10	Consigne	Temps passé à la lecture de la consigne de l'exercice (sur papier)	État
11	Aide	Temps passé à lire l'aide en ligne	État
12	Calepin	Temps passé sur le « calepin » pour y consigner des annotations	État
13	Document	Accès à un document par la bibliothèque « documentation juridique » (seulement pour la tâche 1)	État
14	Etape1	Clic sur le bouton Etape 1 (seulement pour la tâche 4)	Événement ponctuel (sans durée)
15	PlanHier	Clic sur le plan hiérarchique (seulement pour la tâche 1)	Événement ponctuel (sans durée)
16	Tournpag	Clic sur un bouton « tournepage » (seulement pour la tâche 1)	Événement ponctuel (sans durée)

Tableau 5. *Événements enregistrés*

Il est possible de mesurer les temps nécessités pour l'exécution des différentes tâches, mais aussi le nombre d'écrans parcourus. Les événements correspondent à des états (affichage d'un écran, lecture de la consigne, lecture d'un texte à l'écran

etc.), ou à des actions ponctuelles (surlignage d'un paragraphe, clic sur un bouton ou une zone de l'écran). Le tableau 5 ci-dessus indique la liste des variables choisies pour analyser les enregistrements vidéo.

Un certain nombre de remarques sont notées par l'expérimentateur pendant le test. Les questions et remarques soulevées pendant le test seront retranscrites lors de l'analyse des tâches.

3.3 Résultats du test utilisateur

Les participants sont placés en situation d'utilisation du site sans formation spécifique à l'utilisation du dispositif et sans documentation papier (seules les aides en ligne sont disponibles). On peut postuler que dans des conditions moins difficiles (après une courte présentation du logiciel), lors de la deuxième ou troisième utilisation, ou après une formation, les résultats seraient différents.

Il est possible de restituer pour chaque participant les traces des événements enregistrés pour chaque tâche demandée. C'est ce que présente la figure 7 ci-dessous obtenue à l'aide du logiciel *The Observer*. Chaque barre de l'histogramme présente le parcours d'un participant différent (dix participants dans notre expérience). La figure 7 donne les temps consacrés aux événements observés et leur moment d'observation. La légende indique les événements observés pour chaque participant. Les zones blanches accompagnées de traits pointillés sont des moments d'interruption de l'observation et du chronométrage pour différentes raisons (accès réseau interrompu, plantage du système, discussion avec l'expérimentateur). Les cinq premières barres de la figure en partant du bas représentent les temps des participants « experts ». Chaque couleur représente un événement. Les événements ponctuels (sans durée) sont représentés par des traits plus fins et plus hauts, variables « PlanHier », « Tournpag », « Surligna ».

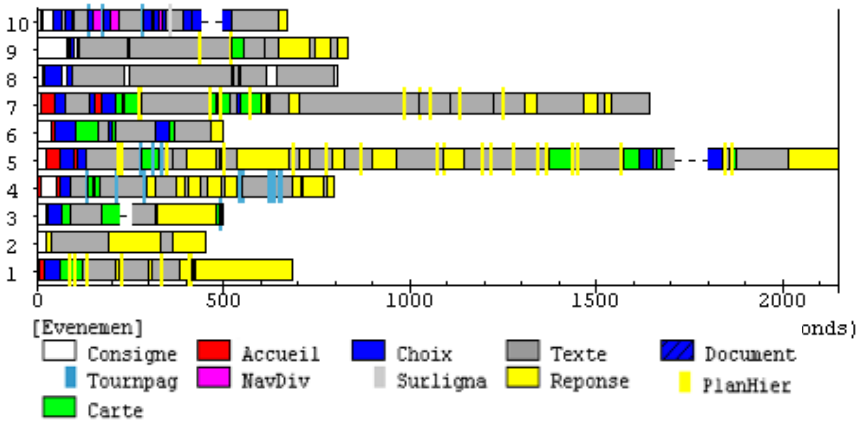


Figure 7. *Analyse des parcours pour la tâche 1*

Le tableau 6 présente les principaux indicateurs statistiques concernant cette tâche. Le nombre d'occurrences de chaque événement, la durée des états enregistrés peuvent donner lieu à des analyses statistiques. Il est intéressant par exemple pour une équipe de conception de savoir quels sont les outils de navigation utilisés, quelles sont les fonctions inutilisées ou appréciées, etc. Il arrive qu'une fonction soit appréciée des utilisateurs mais qu'ils aient du mal à la retrouver quand ils en ont

besoin. La fonction « surlignage » par exemple n'a été utilisée qu'une seule fois par l'un des dix participants.

Élément	**Freq**	**Durée Totale en sec.**	**Durée Totale en %**	**Moyenne**	**StdDev**	**StdErr**	**Durée Mini**	**Durée Maxi**
Accueil	15	174.80	1.970	11.653	10.885	2.810	0.64	37.36
Carte	21	656.28	7.396	31.251	19.400	4.233	3.72	61.20
Choix	29	711.20	8.014	24.524	15.648	2.906	1.04	58.64
NavDiv	3	55.28	0.623	18.427	10.075	5.817	7.72	27.72
Reponse	38	1842.24	20.760	48.480	53.663	8.705	4.28	257.00
Texte	91	5024.80	56.624	55.218	68.168	7.146	0.28	318.08
Surligna	1	-	-	-	-	-	-	-
Consigne	20	399.68	4.504	19.984	18.686	4.178	1.88	80.64
Document	3	9.68	0.109	3.227	1.217	0.703	1.84	4.12
PlanHier	40	-	-	-	-	-	-	-
Tournpag	18	-	-	-	-	-	-	-
Total	**279**	**8873.96**	**100.00**	**40.336**	**52.582**	**3.545**	**0.28**	**318.08**

Tableau 6. *Principaux indicateurs statistiques pour les dix participants, tâche 1*

Le tableau 6 a été obtenu automatiquement depuis le logiciel *The Observer* a l'issue du dépouillement des vidéos. Les participants mettent en moyenne 15 minutes à accomplir la tâche 1. Deux participants mettent nettement plus de temps que les autres (traces 5 et 7 sur la figure 7). On peut expliquer cela par le fait que ces deux participants ont commencé l'expérience par la tâche 1. Ces deux participants découvrent donc le fonctionnement du site lors de la tâche 1. 56,62% du temps total est consacré à l'étude des textes juridiques proprement dite et un peu plus de 7% du temps aux cartes conceptuelles. 20,76% du temps est alloué à la rédaction de la réponse. Près de 11% du temps est dévolu à la navigation, ce qui peut probablement être diminué en aménageant l'interface. Il ne semble pas y avoir de différence significative entre participants experts et novices dans les temps consacrés à la tâche.

Le concepteur du dispositif voulait connaître la fréquence d'utilisation de la zone de navigation située en bandeau à gauche dans les écrans. Chaque clic sur un lien dans cette zone a été enregistré (variable « PlanHier » du tableau 6). Une comparaison avec les possibilités de navigation séquentielle, analogues à la navigation dans un ouvrage papier a été proposée (boutons « tourne page », variable « Tournpag » du tableau 6). La figure 8 présente pour chaque participant le nombre d'occurrence de clic dans le bandeau de navigation.

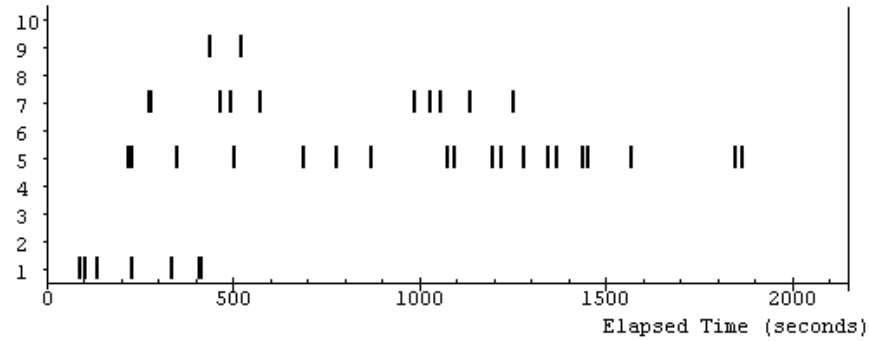


Figure 8. *Clics dans le plan hiérarchique, tâche 1*

La navigation via le plan hiérarchique semble privilégiée par rapport aux boutons tourne page (40 occurrences contre 18). La figure 9 présente les occurrences de clics sur les boutons « tourne page ». Le type de navigation (« tourne page » *versus* plan « hiérarchique ») ne semble pas dépendre de l'expertise puisque l'on retrouve les deux modes de navigation chez les participants experts et chez les novices. Seul un participant (trace 5) utilise les deux modes de navigation.

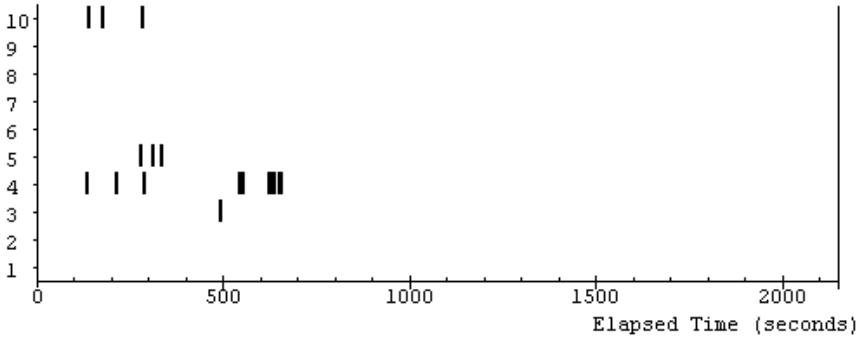


Figure 9. *Clics sur les boutons tourne page, tâche 1*

Remarques et observations recueillies pendant la tâche 1

Différents aspects ont été observés lors de la tâche 1 (collectés lors de l'expérience ou pendant la visualisation des vidéos). En voici quelques-uns, qui relèvent tantôt de l'utilité (possibilité d'imprimer les textes juridiques pour les étudier sur papier) tantôt de l'utilisabilité (longueur des écrans à faire défiler).

Ces phrases sont des commentaires de l'analyste (elles ont été ré-organisées et versées au rapport d'évaluation).

La longueur des écrans de la partie analyse de jurisprudence désoriente parfois les utilisateurs. Une segmentation en plusieurs écrans serait peut-être préférable.

L'icône de globe dans les cartes de concept de la partie mémento fait penser à un accès réseau (type d'icône que l'on peut trouver dans certains navigateurs pour le web par exemple) et ne convient pas au concept représenté.

L'icône « I<< » (retour à l'écran choix d'activité) est souvent utilisée pour revenir en arrière et les utilisateurs sont surpris de voir leur position dans l'arborescence remonter d'un niveau. Ils pensent retourner au premier écran du niveau courant. De plus il arrive parfois qu'étant dans le module licenciement, en cliquant sur cette icône, l'utilisateur arrive sur l'écran choix d'activité du module CDD Intérim.

L'icône jurisprudence semble ambiguë (remplacer ou doubler d'une infobulle plus proche de l'icône).

Les participants cliquent sur les icônes qui servent de légende dans les cartes conceptuelles (globe et texte) celles-ci ayant également l'aspect de boutons.

Dans les cartes conceptuelles, les « pop-up » (fenêtres ponctuelles ou escamots) ne se désactivent pas toujours après consultation du lien, de ce fait ils masquent une partie de la carte et les utilisateurs n'arrivent pas toujours à les faire disparaître. En fait il faut ouvrir un autre pop-up pour faire disparaître le précédent.

Sur un grand écran, un des participant (P7) déplore de n'utiliser qu'une petite partie de l'espace disponible.

Un participant aimerait pouvoir imprimer les écrans du mémento.

Lors d'une première utilisation, sur l'écran choix d'activité, les participants cliquent dans les aplats colorés en vain pour faire apparaître les menus avant de se rendre compte qu'il faut cliquer (survoler du moins) seulement les titres de « colonnes ». Peut-être que les différents choix des sous-parties pourraient être affichés en permanence.

Le participant 6 éprouve des difficultés pour revenir à l'écran d'accueil et de ce fait utilise la fonction « précédent » du navigateur pour revenir en arrière. Il conviendrait de distinguer davantage ce bouton des autres boutons de navigation. De plus la zone d'information sur la signification des boutons ne se rafraîchit pas toujours correctement (conserve parfois le label du dernier bouton survolé). Il arrive ainsi que le commentaire soit « Accès au mémento » au lieu de « Retour à l'accueil » ou « Retour à l'écran de choix des modules ».

Conclusion tâche 1

Le temps alloué à la navigation pourrait vraisemblablement être diminué en améliorant la navigation ou certaines icônes afin de consacrer la plus grande part du temps à la résolution du problème juridique (lecture des textes et des cartes conceptuelles par exemple). Il semble pertinent de maintenir les deux modes de navigation (« plan hiérarchique » et « tourne page »), ceux-ci étant utilisés par les participants, de plus l'accès à des modalités de navigation différentes respecte le critère de qualité ergonomique « flexibilité ». Le temps consacré à la lecture de textes juridiques à l'écran est trop important (56,62 % du temps total) et cette activité est jugée désagréable. Il conviendrait de pouvoir imprimer ces textes pour les analyser sur support papier.

3.4 Tests utilisateur : conclusion

Le biais potentiel de l'« observation » par un tiers est en fait peu préjudiciable en général dans le cadre de ce type d'expérience. La précision des mesures effectuées et le type de données collectées en priorité (détection de problèmes pour l'essentiel) supporte un éventuel biais de ce type. La méthode des tests utilisateurs, quoique coûteuse (environ 60 heures pour cette étude), permet de relever les problèmes posés dans le cadre de l'interaction du système et de l'activité du participant dans une situation proche de la réalité. Les prédictions de ces problèmes sont difficiles même par le biais d'une évaluation experte ou d'un parcours cognitif (*cognitive walkthrough*, expert qui tente de reproduire le cheminement qu'effectuera

l'utilisateur). On gagnera à y recourir au maximum, même si pour cela on doit se passer de moyens d'enregistrement vidéo si la mesure du temps, du nombre d'occurrence d'événements et la valeur de preuve des enregistrements ne sont pas indispensables. L'analyste, avec un peu d'habitude, enregistrera la plupart des problèmes rencontrés par les utilisateurs au fur et à mesure, à l'aide d'une simple prise de note. Les résultats qualitatifs (observations en direct, commentaires et impressions de participants recueillies pendant l'observation) sont tout aussi significatifs en termes d'efficacité opératoire et de décisions de modification du dispositif que les mesures quantitatives (nombres d'écrans parcourus, de clics dans une zone etc.) En effet, c'est souvent sur la base de ces résultats « qualitatifs » que les propositions de modifications du dispositif sont fondées et précises (changements d'icônes inappropriées, modalités d'accès plus directes aux documents, possibilités d'imprimer certaines parties du site, etc.) C'est très certainement une des méthodes d'évaluation des plus efficace avec une grande validité écologique et une vocation opératoire des plus directe. Parmi les autres méthodes présentées ici (inspection et questionnaires / entretiens), les tests utilisateurs permettent souvent à l'observateur d'avoir des « intuitions » fondées sur l'observation concernant les solutions qui permettront de fixer les défauts de l'interface.

Il est judicieux d'administrer à la fin d'une expérience de test utilisateur un questionnaire d'évaluation ou de satisfaction. A cette occasion, on pourra aussi s'entretenir de manière informelle avec le participant à propos des éventuels problèmes rencontrés lors de son parcours dans le dispositif (débriefing).

4 Les questionnaires et entretiens

4.1 Les questionnaires et entretiens : présentation

Les questionnaires et les entretiens complètent valablement les tests avec des utilisateurs. Il conviendra d'adapter les questions à chaque cas d'étude et en particulier à l'état d'avancement du produit. On peut trouver en ligne des questionnaires d'évaluation (questionnaire WAMMI par exemple). Toutefois, certains de ces questionnaires mêlent les dimensions d'évaluation (utilité, utilisabilité, esthétique) et sont rarement adaptés aux enjeux de l'évaluation d'une interface spécifique (comme la formation à distance).

Les questionnaires à eux seuls ne permettent souvent pas d'obtenir des recommandations précises en termes de changements à apporter à l'interface. Par contre ils donnent la mesure la plus compréhensive de la satisfaction générale de l'utilisateur (Huart *et al.,* 2008). De plus, ils peuvent permettre de dégager des tendances sur des notions subjectives, mais importantes d'agrément d'usage. Ainsi des expériences ont montré que lire des textes avec une longueur de ligne de 100 cpl (caractères par ligne) sur écran était plus rapide qu'avec des lignes de 50 cpl. Ceci est dû au moindre recours aux ascenseurs qui font défiler le texte à l'écran. Si l'on s'en tient là, on pourrait généraliser cette observation et la transformer en recommandation. Des questionnaires administrés en fin d'expérience ont démontré que les participants préféraient les lignes de 50 cpl pour lesquelles ils trouvaient la lecture plus fluide, bien que, dans l'absolu, les lignes de 100 cpl entraînent des temps de lecture plus courts.

4.2 Questionnaire sur le dispositif de formation

Après l'utilisation du dispositif de formation, (dans le cadre du test utilisateur, à la fin du test), un questionnaire papier est administré et rempli par les participants.

Nous présentons quelques réponses à deux questions ouvertes posées aux participants à propos du dispositif qu'ils ont utilisé pendant une à deux heures environ :

Comment trouvez vous ce site ?

Avez-vous des suggestions d'amélioration du site ?

Les remarques des participants sont données quelle que soit leur pertinence par rapport à l'interface et sont rapportées pratiquement dans leur formulation initiale.

Réponses à la question « Comment trouvez vous ce site ? »

P1 (Participant 1) : Ce site est un bon outil pour approcher la matière qu'est le droit du travail. Il est complet et surtout, il est ordonné, clair et compréhensible. Cependant, l'outil internet ne remplacera pas tout de suite la matière papier. En clair, pour moi, ***travailler sur un support papier reste encore essentiel. De fait, il m'est difficile de me concentrer pendant une longue période sur l'ordinateur (comme par exemple la résolution d'un cas pratique).*** *Sur la recherche d'un document ou d'une partie de cours, le site me semble être un excellent outil d'analyse et de compréhension. On arrive bien à trouver rapidement des documents.*

P2: Je trouve le site très intéressant et surtout très utile.

P3 : Site intéressant et pratique notamment pour les révisons, car il englobe les thèmes étudiés en cours. Les cas pratiques permettent de s'entraîner pour résoudre des problèmes ***(pour examens)****.*

P4 : Ce site est très fonctionnel et permet une approche plus simple et pratique du cours notamment ***en période d'examen****.*

P5 : Réservé à des étudiants en droit. Difficile d'appréhension pour un novice. Le graphisme est agréable.

P6 : Intéressant pour en savoir plus sur droits et devoirs salariés/employeurs. Précis dans le vocabulaire juridique utilisé (pas assez décomposé ou simplifié pour des non initiés). ***Lecture de certains textes rendue difficile par le défilement répété.***

P7 : Au début de la consultation ***je n'ai pas immédiatement réussi à me repérer dans le site****. Le vocabulaire employé semble accessible.*

P8 : Convivial, étoffé. Si ***la navigation dans le site peut paraître un peu difficile dans un premier temps,*** *elle devient ensuite beaucoup plus aisée. Les possibilités d'obtenir une réponse semblent tellement nombreuses que* ***cela peut laisser croire que l'on a jamais vraiment exploré l'ensemble d'un domaine,*** *peut-être cibler les domaines d'une manière plus précise.*

P9 : ***La première utilisation n'est pas évidente, il faut quelques minutes pour se familiariser et se repérer.*** *Par contre une fois que l'on s'est imprégné de l'architecture la navigation est assez facile. Les modules sont bien guidés.*

*P10 : Le site est présenté d'une façon tout à fait attrayante pour un domaine aussi rébarbatif que le droit du travail dans ses pires moments. La navigation est assez aisée. Le seul problème est que l'****on a du mal à repérer tous les éléments et les liens qui peuvent nous aider à progresser dans la navigation. Certains composants comme les questions du cas pratique gagneraient à êtres plus visibles.***

Les réponses à ces questions font apparaître des éléments constatés lors de l'observation, mais également des considérations nouvelles. Parmi les éléments déjà constatés lors du test utilisateur, la lisibilité des textes juridiques (longs et peu structurés pour le support numérique) à l'écran est jugée désagréable (P1 et P6). La possibilité d'imprimer tous les textes nécessaires à la résolution d'un cas pratique semble indispensable dans ce contexte. La prise en main du produit n'est pas aisée

(première utilisation) P7, P8 et P9. Par ailleurs, des problèmes de repérage, de visibilité de certains éléments sont également évoqués (P9 et P10) qui conduisent au sentiment de ne jamais être très assuré d'avoir exploré l'ensemble d'un domaine (P10). L'élément nouveau est que certains participants considèrent que cet outil est un dispositif intéressant de préparation aux examen (P3 et P4). De fait, cela confirme le positionnement du produit comme complément à un cours « classique ».

Réponses à la question « Avez vous des suggestions d'amélioration du site ? »

P1 : Pourrait-on peut-être ***avoir la possibilité d'imprimer*** *les documents qui sont sur le site (surtout pour les cas pratique).*

P5 : Manque de visibilité dans le cas pratique. Liens non actifs dans le bandeau. ***Icônes peu signifiantes.***

P6 : ***Icônes plus claires*** *(distinctes, par exemple « maison » pour retour page d'accueil pas très lisible), plus précises, situation dans l'espace page. Attention à la juxtaposition d'une image animée avec une phrase ou une question, c'est peu lisible du coup (par exemple, dans l'exercice cas pratique licenciement). Prévoir ascenseurs pour lectures des décisions des différentes cours de justice.*

P7 : Dans le cas pratique, les consignes des différentes étapes ne sont pas attractives. ***Le contraste entre la couleur du fond de cartouche et la couleur des caractères n'est pas approprié.*** *Possibilité d'avoir un lexique pour expliquer les termes techniques.*

P9 : Certains ***fonds bleus étaient trop foncés par rapport au texte.***

P10 : Pour les questions du cas pratique il est peut être nécessaire de ***changer la taille des caractères et la casse, ainsi que leur couleur.***

La plupart des éléments signalés dans les réponses ont été constatés pendant le test utilisateur. Sans revenir sur ce qui a été signalé lors de la question précédente (possibilité d'imprimer P1), on peut signaler que les propositions d'amélioration concernent la signification des codes utilisés et en particulier des icônes qui sont parfois ambiguës (P5, P6). Par ailleurs, des problèmes de lisibilité liés à un contraste insuffisant entre couleur de fond et couleur de texte apparaissent très nettement (P7, P9, P10).

4.3 Questionnaires et entretiens : conclusion

Les questionnaires et entretiens sont des méthodes « économiques » qui permettent de prendre la température concernant la perception d'utilisateurs sur un dispositif donné. Il n'est pas souhaitable qu'il y ait un grand nombre de questions dans le questionnaire, surtout si celui-ci est administré après un test utilisateur qui a déjà duré près d'une heure par exemple. Des questions fermées peuvent être envisagées pour mesurer des dimensions précises relatives au dispositif. Des questions ouvertes sur l'agrément général, des suggestions d'amélioration et/ou les principaux problèmes rencontrés peuvent permettre d'ouvrir un espace d'expression pour l'utilisateur. Le point de vue de l'utilisateur est à pondérer à l'aune des résultats scientifiques car les préférences de certains utilisateurs ne sont pas toujours en adéquation avec les bonnes pratiques qui ont fait l'objet de validations expérimentales. Les questionnaires et entretiens utilisés seuls n'apportent que des éléments d'appréciation sur les problèmes les plus saillants. De plus, ces éléments sont abordés de manière générale, ce qui ne permet pas, bien souvent, d'interpréter ces remarques de manière à modifier précisément l'interface. Par exemple, une remarque comme « on a du mal à repérer tous les éléments et les liens qui peuvent nous aider à progresser dans la navigation » est difficilement interprétable d'un

point de vue opératoire pour l'équipe de conception faute de précision suffisante. Aussi des entretiens complémentaires peuvent consister à demander aux participants de préciser tous ces points : « Icônes peu signifiantes », pouvez-vous m'indiquer lesquelles ? Pouvez-vous en dessiner une qui soit plus signifiante pour vous ? L'objectif pour les concepteurs est de convertir les remarques en propositions de modifications si les remarques sont pertinentes et récurrentes. Le rapport coût efficacité en termes opérationnels est faible par rapport aux autres techniques, mais les entretiens sont une méthodes intéressante pour recueillir les impressions d'utilisateurs sur un dispositif et décider de poursuivre l'évaluation si nécessaire avec une technique plus lourde. Pour certains aspects précis et déjà identifiés (problèmes de compréhension d'icônes par exemple), les questionnaires permettent d'avoir un retour rapide des utilisateurs.

5 Conclusion générale sur l'évaluation du dispositif de formation

Après avoir évalué le dispositif de formation avec les trois méthodes présentées ci-dessus, la conclusion de l'évaluation pour ce dispositif était de cet ordre :

Le dispositif évalué répond à son objectif de complément à un cours « classique ». Ce type de produit de formation convient bien pour des tâches courtes avec peu de lecture à l'écran. Lorsque le problème à traiter est complexe et demande de naviguer plus fréquemment ainsi que de lire de nombreux textes sur écran, plusieurs problèmes se posent. Le maintien en mémoire des consignes et des questions est difficile lorsque l'on doit se concentrer sur la navigation et la recherche d'information. La lecture de longs textes sur écran est jugée fatigante. Les utilisateurs demandent donc de minimiser les déplacements dans les écrans en rendant la consultation des documents plus directe (au plus près de la question posée) et pour les longs textes, ils souhaitent pouvoir « imprimer ». L'usage de certaines fonctions est à développer ou améliorer (recherche, surlignage, cartes conceptuelles) et d'autres fonctions n'ont pas été utilisées (calepin) ou été peu utiles en l'état actuel (plan du site et aide en ligne). La partie cas pratique, jugée intéressante nécessite le plus d'aménagement à cause de la multiplicité des difficultés rencontrées. Difficultés à identifier la question générale, à activer les sous-questions, à faire le lien entre sous-questions et documents suggérés à la lecture, à accéder à certains documents suggérés. Toutefois les modifications qui permettraient une prise en main plus rapide du site ne remettent pas en cause son organisation actuelle. Ces modifications visent essentiellement à améliorer

- la navigation (longueur des pages, boutons de navigations),

- le repérage (en évitant à l'utilisateur de quitter un espace courant pour consulter des informations en rapport à cet espace),

- la signifiance des codes utilisés (icônes et termes choisis pour désigner certains concepts),

- la lisibilité (combinaisons de couleur texte/fond).

6 Conclusion sur l'évaluation

Cet article est essentiellement destiné à donner un aperçu rapide de ce qu'il est possible d'obtenir en évaluant l'interface d'un dispositif de formation à distance d'un point de vue ergonomique. Au cours de la phase de développement, le concepteur peut utiliser la grille de Bastien et Scapin (1993) afin d'auto-évaluer sa production, mais un vrai travail d'évaluation relève souvent d'un spécialiste. Lorsque l'enjeu est important notamment, le mieux est d'avoir recours à un ergonome et ceci, dès le début de la conception (une correction coûte entre dix et cent fois plus cher lorsqu'elle intervient en phase de développement plutôt qu'en phase de conception (Baccino *et al.*, 2005)). On peut signaler aussi que des tests de

perception peuvent être mis en œuvre dès le début du projet sur les maquettes, prototypes papier.

De nombreuses autres méthodes d'évaluation existent dont certaines nécessitent des outils logiciels (évaluation assistée, analyse des *log-files*, etc.) Certaines méthodes d'évaluation peuvent être très sophistiquées (comme le suivi des mouvements oculaires). Un des objectifs de cet article aura été de montrer que l'association de plusieurs méthodes d'évaluation est souvent nécessaire et s'inscrit dans la perspective de pratiques intégrées comme la méthodologie SUE *Systematic Usability Evaluation* (Garzotto *et al.,* 1997) qui associe méthode d'inspection et tests empiriques (Huart *et al.*, 2008).

A l'évaluation de l'interface, il convient aussi d'ajouter la question de la performance du dispositif en termes d'apprentissage. Des méthodes spécifiques destinées à évaluer cet aspect pourront être couplées aux méthodes d'évaluation des interactions avec le dispositif que nous venons de décrire. Cette fois, ce ne sera pas la médiation technique qui sera évaluée, mais davantage la qualité des contenus, leur bonne adaptation au niveau de compréhension des utilisateurs visés et leur progression pédagogique. On pourra pour ce faire s'appuyer sur des principes énoncés pour l'évaluation de dispositifs hypermédia de formation comme les sept qualités de l'apprentissage de Jonassen, 1995 (constructif, actif, coopératif, conscient, contextuel, (qui permet le) transfert, réflexif). Toutefois ces dimensions rendent compte de processus de haut niveau qui nécessitent pour certains d'entre eux (comme le transfert) un arsenal méthodologique conséquent eu égard au cadre de l'évaluation de dispositifs de formation à distance.

Bibliographie

Baccino, T., Bellino, C. & Colombi, T. (2005). *Mesure de l'utilisabilité des interfaces.* Paris, Hermès Lavoisier éditions.

Bastien, J.M.C. & Scapin, D.L. (1993). *Ergonomic criteria for the evaluation of human-computer interfaces.* Rapport technique, n° 156, INRIA, Rocquencourt.

Bastien, J.M.C., Leulier, C. & Scapin, D.L. (1998). L'ergonomie des sites web. In *Cours INRIA Créer et maintenir un service web*, ADBS, INRIA (Eds.), ADBS éditions, Paris, 111-173.

Bastien, J.M.C. & Scapin, D.L. (2001). Evaluation des systèmes d'information et critères ergonomiques. In Kolski C. (dir.), *Environnements évolués et évaluation de l'IHM, Interaction homme-machine pour les SI 2*, Hermès Sciences Publications, Paris, 53-80.

Chanquoy, L., Sweller, J. & Tricot, A. (2007). *La charge cognitive.* Paris, Armand Colin.

Faulkner, L. (2003). Beyond the five-user assumption : Benefits of Increased sample sizes in usability testing. *Behavior Research Methods, Instruments & Computers*, 35(3), 2003, 379-383.

Garzotto, F. & Matera, M. (1997). A Systematic Method for Hypermedia Usability Inspection. *The new review of hypermedia and multimedia*, Vol. 3, p. 39-65.

Huart, J., Kolski, C. & Bastien, C. (2008). L'évaluation de documents multimédias. Etat de l'art. In Leleu-Merviel (dir.), *Objectiver l'humain ? Volume 1, Qualification, quantification,* Paris, Hermès Sciences Publications.

Jonassen, D. (1995). Supporting Communities of Learners with Technology: A vision for Integrating Technology with Learning in Schools, *Educational Technology*, 35(4), 60-63, Jul.-Aug. 1995.

Nielsen, J. (1994). Heuristic evaluation. In *Usability Inspection Methods*. In Nielsen, J., Mack, R.L. (Eds.), John Wiley & Sons éditions, New York, 25-62.

Nielsen, J., (2000). Why you only need to test with 5 users, Alertbox, 19 mars 2000, en ligne : http://www.useit.com/alertbox/20000319.html.

Nogier, J.F. (2005). *Ergonomie du logiciel et design web : le manuel des interfaces utilisateur*. Paris, Dunod éditions.

Senach, B. (1993). L'évaluation ergonomique des interfaces homme-machine : une revue de la littérature. In Sperandio, J.C. (Ed.), *L'ergonomie dans la conception des projets informatiques,* Toulouse, Octares éditions, 69-122.

Tricot, A., Plégat-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 Desmoulins, C., Marquet, P., Bouhineau, D. (Eds.), *Environnements informatiques pour l'apprentissage humain,* ATIEF / INRP éditions, Paris, 391-402.