Observation des usages au moyen de tests utilisateurs

Esposito, N. (2017). Observation des usages au moyen de tests utilisateurs. *Actes de la journée Quels usages aujourd'hui des bibliothèques numériques ? Enseignements et perspectives à partir de Gallica.* BnF, Paris (France), 3 mai.

Nicolas Esposito

Gobelins, l'école de l'image

Résumé

Cet article décrit la manière avec laquelle le laboratoire ErgoDesign Lutin-Gobelins réalise des tests utilisateurs (contexte, méthodologie). Les outils les plus utilisés sont présentés, avec notamment des exemples d'outils conçus dans le cadre des activités du laboratoire, ainsi que trois modes de mesures oculométriques pour travailler sur l'attention visuelle. Enfin, des perspectives seront évoquées à partir d'observations récurrentes.

1 Contexte

Dans le cadre du développement des activités de recherche et d'innovation de l'école (Gobelins, l'école de l'image), le laboratoire ErgoDesign Lutin-Gobelins a été lancé en 2013 avec l'ambition d'associer la maîtrise scientifique du laboratoire lutin (Laboratoire des Usages en Technologies d'Information Numérique, ums 2809) avec le savoir-faire de l'école en conception de l'expérience utilisateur (voir par exemple, pour ce domaine, Lallemand & Gronier, 2015).

Les tests utilisateurs représentent le cœur de l'activité de ce nouveau laboratoire (voir figure 1). Ils sont réalisés pour les projets des étudiants de l'école, pour les clients et partenaires du laboratoire, et pour les projets de recherche et développement dans lesquels l'école est impliquée. Les domaines concernés sont variés : sites Web, logiciels, applications mobiles, jeux vidéo, livres et magazines numériques, objets connectés, films, photos, affiches, etc.

Les tests utilisateurs font partie des approches qui permettent l'évaluation de l'expérience utilisateur (Maguire, 2001). L'objectif est généralement de valider des intentions de conception. Cela, en évaluant l'adéquation entre le dispositif (ce qui a été conçu) et les utilisateurs auxquels il est destiné. Autrement dit, en vérifiant que ce dispositif leur réponde de manière pertinente, en s'assurant que les utilisateurs soient satisfaits de leur expérience. Un test utilisateur est une démarche expérimentale. Lors de la passation, on demande à des participants de réaliser des tâches avec le dispositif, on observe leurs actions et réactions, on réalise des mesures et l'on recueille leurs retours.

FIGURE 1 – Structuration des activités du laboratoire ErgoDesign Lutin-Gobelins

L'expertise construite au fil des ans repose sur la mise en œuvre de plus de 200 tests utilisateurs et sur un certain nombre d'ancrages théoriques. En voici quelques exemples. Pour l'expérience utilisateur : les éléments de GARRETT (2011), les niveaux de traitement de NORMAN (2004), le modèle de HASSENZAHL (2003) et la pyramide de HANCOCK *et al.*, (2005). Mais aussi : les modèles de NORMAN & DRAPER (1986), le *flow* de CSIKSZENTMIHALYI (1990), les dimensions de RUSSELL (2003) et l'énaction (LENAY, 2006).

2 Méthodologie

En conception centrée utilisateur, les approches itératives sont propices aux tests. Pour chaque itération, on peut envisager une validation par un test utilisateur (MOULE, 2012). La figure 2 présente un exemple incluant trois tests pour la première version d'un dispositif, puis un nouveau à partir de cette version 1. Avec différents niveaux de prototypage (maquette en papier, maquette interactive et prototype fonctionnel), on peut valider différents niveaux de conception, notamment : les concepts fondamentaux, l'architecture de l'information et l'ergonomie.

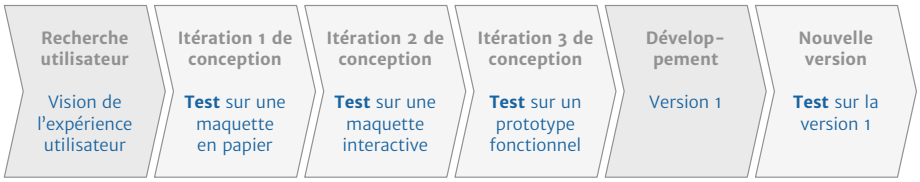

FIGURE 2 – Exemple de positionnement et de rôle pour des tests utilisateurs dans la conception d'un dispositif

La mise en œuvre d'un test utilisateur suit une démarche expérimentale. Les tests du laboratoire ErgoDesign Lutin-Gobelins sont structurés selon le processus exposé par la figure 3. La phase de préparation d'un test utilisateur donne lieu en particulier à la rédaction du protocole expérimental. Ce protocole permet de conduire la passation (deuxième phase) de manière rigoureuse avec les participants. La troisième phase est celle qui prend généralement le plus de temps : l'analyse des résultats. L'ensemble peut par exemple prendre cinq jours pour cinq ou six participants.

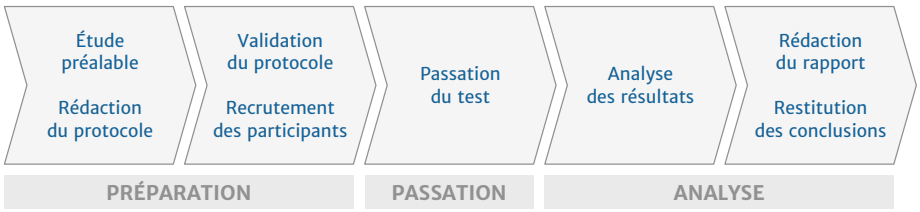

FIGURE 3 – Processus global pour un test utilisateur au laboratoire ErgoDesign Lutin-Gobelins

Un test utilisateur correspond à un cadre ouvert à de nombreux outils. On peut ainsi recueillir des données de natures diverses et les croiser pour formuler des recommandations fortement argumentées. La figure 4 montre un exemple d'articulation d'outils pour un test utilisateur simple. Pour chaque participant, on commence par l'outil questionnaire afin de mieux cerner son profil. On passe ensuite au cœur du test avec l'observation du participant réalisant les manipulations. On revient alors à l'outil questionnaire notamment pour interroger le participant sur son appréciation. Et l'on approfondit avec un entretien, avant d'exploiter les données recueillies (ex. : le nombre d'actions effectuées).

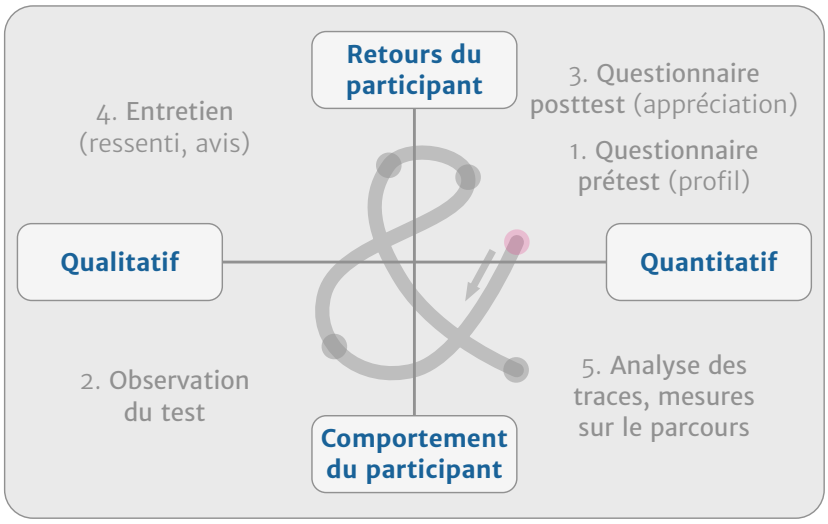

FIGURE 4 – Articulation des outils principaux pour un test utilisateur simple au laboratoire ErgoDesign Lutin-Gobelins

Le plus souvent, les tests se déroulent au laboratoire (éventuellement aménagé de manière spécifique, par exemple avec un coin salon). Ce lieu permet en particulier un contrôle scientifique élevé et il facilite l'analyse des données (BOJKO, 2013). Mais dans certains cas, le milieu usuel des utilisateurs est privilégié pour une plus grande validité écologique. Le matériel de mesure oculométrique pour travailler sur l'attention visuelle est sélectionné en fonction de la configuration retenue : barre oculométrique (placée sous l'écran ou dans un support pour appareil mobile) ou lunettes oculométriques (qui enregistre la scène du point de vue de l'utilisateur).

Le choix du nombre de participants se base sur les ordres de grandeur du domaine (NIELSEN & LANDAUER, 1993) en mettant l'accent sur trois aspects : le nombre de groupes (par exemple des débutants et des experts, ou une version A et une version B), l'exigence statistique (contexte professionnel et/ou scientifique) et l'ajustement à envisager (participants absents, problèmes techniques, etc.). La figure 5 correspond à un test pour lequel 12 personnes ont été sollicitées afin de s'assurer d'avoir des données exploitables pour au moins 10 participants (deux groupes de cinq).

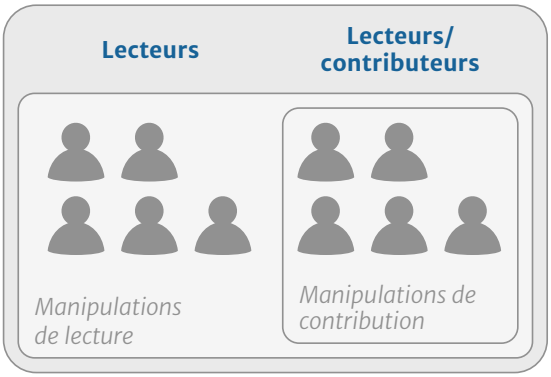

FIGURE 5 – Exemple d'objectif de nombre de participants pour un test utilisateur avec deux groupes dans un contexte professionnel

Le protocole expérimental décrit l'objet du test (dispositif, objectifs, problématiques, participants, hypothèses), la manière de réaliser la passation (mesures, modalités, questions, consignes) et certains aspects organisationnels (lieu, matériel, processus, calendrier). Lors de la rédaction d'un protocole, le laboratoire ErgoDesign Lutin-Gobelins accorde une importance particulière à la réflexion sur les biais potentiels (par exemple concernant l'ordre des questions) et à la réflexion sous l'angle des dimensions de l'expérience utilisateur (un outil est proposé pour cela dans la section suivante).

Le travail sur les données (recueil, analyse, interprétation et présentation) repose sur les pratiques du domaine (ALBERT & TULLIS, 2013) avec pour le laboratoire des modèles garantissant des mises en œuvre simples et rapides (tableaux, graphiques, rapport et format d'argumentation). Les conclusions donnent lieu à des recommandations qui sont classées selon les trois catégories suivantes : modification légère, modification plus profonde et reconception nécessaire.

3 Outils

Les cinq E de l'utilisabilité (QUESENBERY, 2003) fournissent un cadre de travail particulièrement efficace pour la conception et l'évaluation de la dimension interactive d'un dispositif. Traduits en français, nous avons ainsi : l'efficience, l'efficacité, l'engagement, la facilité d'apprentissage et la tolérance aux erreurs.

La conception de l'expérience utilisateur se caractérise notamment par la prise en compte de multiples dimensions, au-delà de l'aspect fonctionnel, par exemple : la dimension informationnelle, la dimension sociale, la dimension physique (ROBERT, 2014). Mais il convient également de penser les relations entre ces dimensions. En effet, des réussites indépendantes ne garantissent pas un succès global. Par exemple, un jeu vidéo peut disposer d'une valeur de production importante (mécaniques de jeu, esthétique et scénario) en ne remportant qu'un faible intérêt auprès du public. C'est pour répondre à cet écueil que nous proposons l'outil synthétisé sur la figure 6. Sa spécificité réside dans l'analyse des relations entre les dimensions.

FICHE SYNTHÉTIQUE	
Nom de l'outil	Analyse relationnelle des dimensions de l'expérience utilisateur
Principe	Analyse des dimensions de l'expérience utilisateur et de leurs relations pour concevoir ou évaluer l'orchestration de ces relations
Mise en pratique	1. Identifier les dimensions de l'expérience utilisateur du point de vue des intentions de conception 2. Analyser les relations entre les dimensions en positionnant chaque dimension par rapport aux autres selon la formulation « d_1 est au service de d_2 » 3. Réaliser une représentation graphique faisant figurer ces dimensions et leurs relations 4. Pour la conception, vérifier que les dimensions sont effectivement orchestrées selon les relations analysées 5. Pour l'évaluation, mettre en œuvre des modalités pour valider que les relations analysées sont bien effectives

EXEMPLE

À partir d'un livre enrichi pour enfant

1. Dimension interactive, dimension ludique, dimension narrative
2. La dimension interactive est au service de la dimension ludique et de la dimension narrative, la dimension ludique est au service de la dimension narrative
3. Voir ci-contre
4. Notamment : vérifier que la dimension ludique est bien au service de la dimension narrative
5. Notamment : valider que la dimension ludique ne détourne pas l'attention qui se portait sur la dimension narrative (ex. : passer une lecture pour aller au prochain jeu)

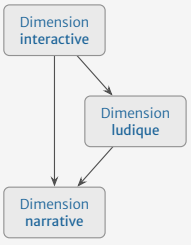

FIGURE 6 – Fiche synthétique sur un outil pour travailler sur les dimensions de l'expérience utilisateur et leurs relations

Concernant l'observation, nous utilisons une matrice difficultés/participants, connue sous le nom de Rainbow Spreadsheet (SHARON, 2016). Il s'agit d'une grille d'observation simple permettant de répertorier les difficultés rencontrées au fur et à mesure et d'exploiter très facilement ce travail lors de l'analyse. On part d'une matrice vide. Puis, à chaque fois qu'une nouvelle difficulté est observée : on la décrit dans une ligne vierge et l'on inscrit 1 dans la colonne du participant concerné. Quand cette difficulté est à nouveau rencontrée, on inscrit simplement 1 dans la colonne correspondante.

Pour recueillir les réponses aux questions fermées, l'utilisation d'échelles de LIKERT (1967) à cinq valeurs a été assez efficace au regard des besoins (choix moyen présent et deux niveaux de part et d'autre). Mais elles ont soulevé des questions variées de la part des participants, faisant alors perdre du temps. Nous

avons alors mis en place des échelles avec plusieurs modalités (voir figure 7), pour s'adapter aux différents styles cognitifs et pour permettre le croisement des modalités. Ainsi, les questions posées précédemment ont disparu et nous avons constaté que la modalité choisie pour répondre changeait effectivement d'un participant à un autre.

FIGURE 7 – Exemples d'échelles à cinq valeurs avec plusieurs modalités pour « Quelle est votre pratique [...] ? » et « Avez-vous apprécié [...] ? »

Pour recueillir des données liées aux émotions ressenties par les participants (REDET *et al.*, 2016), nous réservons les mesures physiologiques et le casque électroencéphalogramme à des projets pour lesquels les émotions sont particulièrement vives. Dans le cas contraire (le plus courant), ces types de mises en œuvre exigeantes ne nous fournissent pas des résultats suffisamment significatifs. De même pour la reconnaissance automatique des émotions à partir des images d'une caméra (confirmation des résultats de LOCKNER, 2015). À l'opposé, les retours subjectifs des participants sur leurs émotions se révèlent bien plus exploitables dans ce cas, notamment avec des approches graphiques comme la roue de Genève (SCHERER, 2005) et le Self-Assessment Manikin (BRADLEY, 1994).

Lorsque l'on recueille des données liées aux émotions, nous nous intéressons à la personnalité des participants, avec pour cela l'approche des Big Five. Un test de personnalité complet étant généralement trop long au regard du dimensionnement d'un test utilisateur, nous avons conçu un outil de positionnement rapide sur les cinq traits de personnalité des Big Five en nous basant sur les formulations en français du test NEO-PI-R (PLAISANT *et al.*, 2010). Le positionnement se fait à l'aide de l'une des échelles présentées plus haut (à cinq valeurs, avec plusieurs modalités).

Concernant le traitement statistique des données, nous utilisons notamment l'analyse de variance ANOVA, typiquement pour comparer des moyennes. Nous proposons une représentation pour ce type de comparaison. À partir des données, un programme (script en langage R) génère une représentation qui intègre la répartition des données avec un diagramme en violon (voir figure 8).

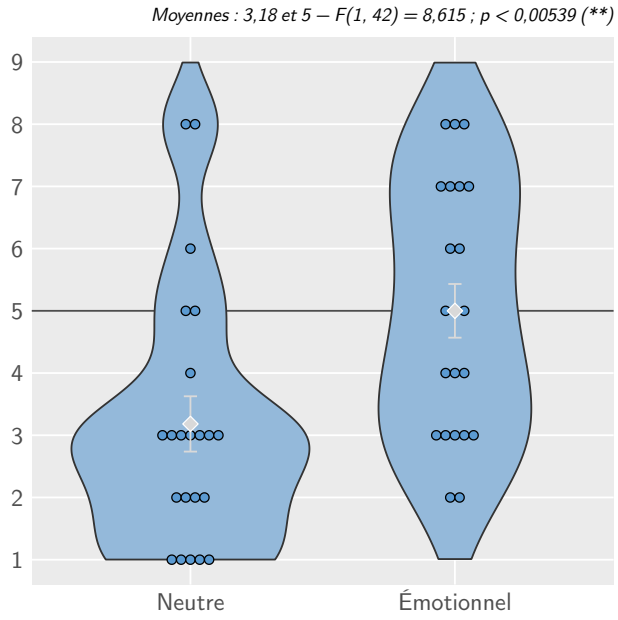

FIGURE 8 – Représentation de la répartition de données subjectives (éveil pour deux versions de la page d'accueil d'un site Web)

Nous proposons également, pour communiquer de manière générale sur l'ergonomie d'une interface, un indicateur temporel basé sur la durée moyenne des manipulations pour les participants et sur la durée de référence obtenue avec une personne maîtrisant déjà le dispositif. Le facteur qui en résulte nous renseigne sur le positionnement global de l'ergonomie. Des facteurs de références (à moduler selon le temps consacré aux consignes) ont été établis grâce aux données de plusieurs dizaines de tests adoptant la même démarche (voir figure 9).

INDICATEUR TEMPOREL	
Durée moyenne	8,5 min
Durée de référence	3,5 min
Facteur	**2,43**

FACTEURS DE RÉFÉRENCE*		
2	+	Ergonomie fluide
3	~	Ergonomie à fluidifier
4	-	Structure de l'interface à revoir
5	--	Structure de l'interface à refaire
6	---	Approche globale à revoir

*Avec peu de consignes

FIGURE 9 – Exemple d'indicateur simple (facteur entre une valeur de référence et la moyenne constatée)

4 Oculométrie

Les tests utilisateurs menés au laboratoire ErgoDesign Lutin-Gobelins prennent en compte (sauf exception) l'attention visuelle, information particulièrement utile pour suivre les actions des participants (BOJKO, 2013). Pour cela, on enregistre les mouvements des yeux avec un oculomètre.

Si le participant utilise un ordinateur, on place une barre oculométrique sous l'écran. Cela peut par exemple concerner un site Web (voir figure 10), un logiciel, un jeu vidéo, un film, des photos, des affiches, etc.

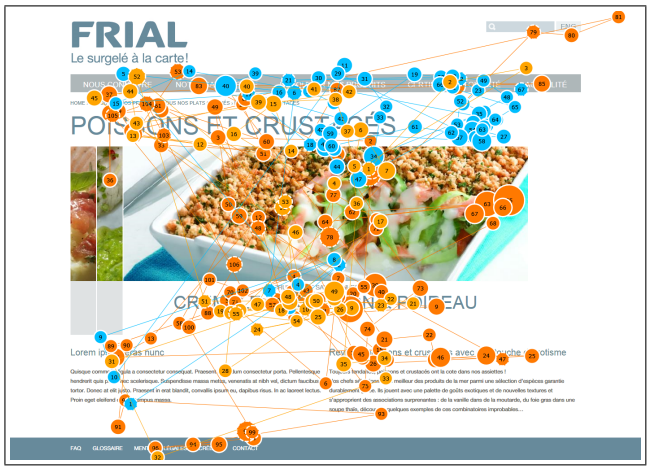

FIGURE 10 – Trajectoires oculaires des participants n'ayant pas trouvé la deuxième recette de poisson (accès via la zone non regardée à gauche)

Si le participant utilise un smartphone, une tablette, une liseuse, un livre ou un magazine (papier), on place cette barre dans un support pour appareil mobile. Une caméra intégrée à ce support filme alors l'écran (voir figure 11). Cela peut par exemple concerner la version mobile d'un site Web, une application mobile, un jeu mobile, un livre ou un magazine numérique, etc.

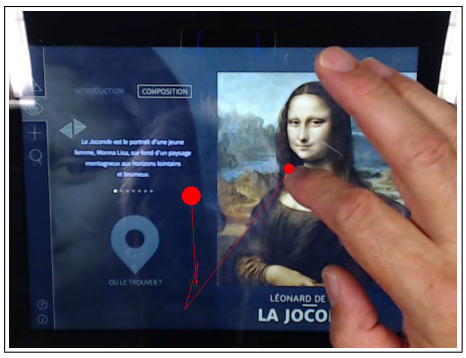

FIGURE 11 – Capture d'une vidéo obtenue lors du test utilisateur de l'application *Louvrissime*

Enfin, quand l'objet à tester repose sur la géolocalisation ou quand il est important que le participant ait une autre posture qu'être assis devant une table (par exemple assis sur un canapé), on utilise des lunettes oculométriques. Une caméra intégrée à ces lunettes filme alors ce que voit le participant (voir figure 12).

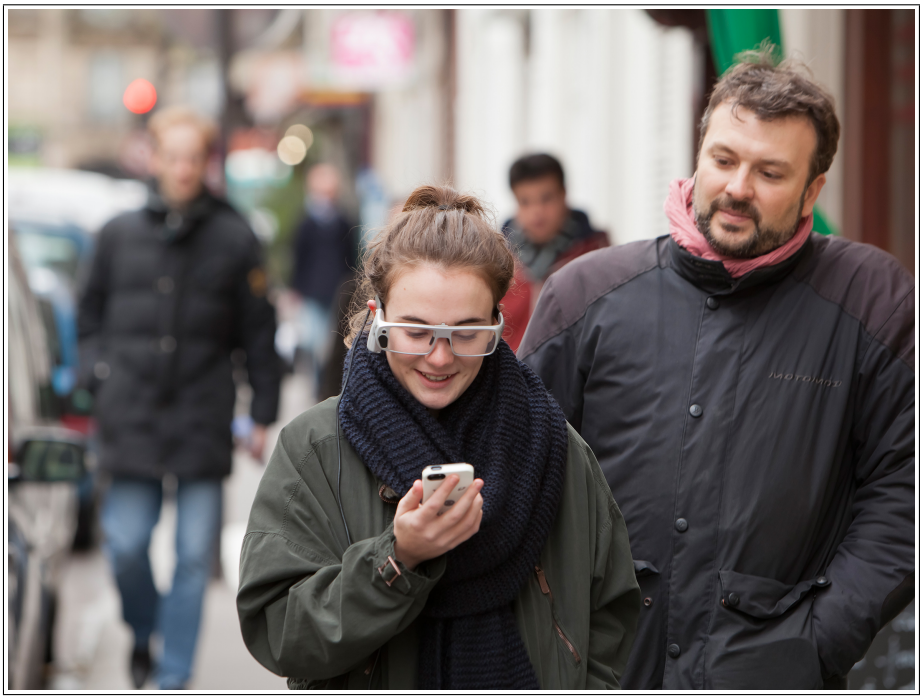

FIGURE 12 – Utilisation de lunettes oculométriques pour un test utilisateur d'une application impliquant le déplacement dans les rues

5 Perspectives

Les tests utilisateurs dont il est question dans cet article ont été l'occasion de confirmer certains résultats connus et de soulever de nouvelles hypothèses. Nous avons par exemple pu apprécier sur le plan pratique la pertinence des principes de NORMAN (1988) et des critères de BASTIEN & SCAPIN (1993).

Nous avons également identifié un certain nombre de pièges à éviter lors de la conception. Par exemple : le fait de ne pas suffisamment prendre en compte la grande variété de comportements qui peut être rencontrée chez des utilisateurs appartenant à un même profil (stratégies oculaires, mouvements de la souris, expressions faciales, etc.). Ces observations sont accompagnées d'hypothèses qui pourront faire l'objet d'expérimentations spécifiques.

6 Bibliographie

ALBERT, W. & TULLIS, T. (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics* Morgan Kaufmann, 2e éd.

BOJKO, A. (2013). *Eye Tracking the User Experience: A Practical Guide to Research*. Rosenfeld Media.

BASTIEN, C. & SCAPIN, D. (1993). *Critères ergonomiques pour l'évaluation d'interfaces utilisateurs*. Technical report 156, INRIA.

BRADLEY, M. M. & LANG, P. J. (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59.

CSIKSZENTMIHALYI, M. (1990). *Flow: The Psychology of Optimal Experience*. Cambridge University Press.

GARRETT, J. J. (2011). *The Elements of User Experience: User Centered Design for the Web and Beyond*. Pearson, 2e éd.

HANCOCK, P. A., PEPE, A. & MURPHY, L. L. (2005). Hedonomics: The power of positive and pleasurable ergonomics. *Ergonomics in Design*, 13(1), 8–14.

HASSENZAHL, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. *In* BLYTH, M. A., MONK, A. F., OVERBEEKE, K. & WRIGHT, P. C. (éd.), *Funology: From Usability to Enjoyment*. Kluwer Academic Publishers, 31–42.

LALLEMAND C. & GRONIER, G. (2015). *Méthodes de design UX – Trente méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Eyrolles.

LENAY, C. (2006). Énaction, externalisme et suppléance perceptive. *Intellectica*, 43, 27–52.

LIKERT, R. (1967). The Method of Constructing an Attitude Scale. *In* FISCHBEIN, M. (éd.), *Attitude Theory and Measurement*. John Wiley & Sons, 90–95.

LOCKNER, D. (2015). *Évaluation émotionnelle des interfaces – Vers une proposition de* patterns *de conception améliorant le caractère empathique de la relation humain-machine*. Thèse de doctorat de l'Université d'Aix-Marseille.

MAGUIRE, M. (2001). Methods to Support Human-Centred Design. *International Journal of Human-Computer Studies*, 55(4), 587–634.

MOULE, J. (2012). *Killer UX Design: Create User Experiences to Wow Your Visitors*. SitePoint.

NIELSEN, J. & LANDAUER, T. K. (1993). A Mathematical Model of the Finding of Usability Problems. *Proceedings of ACM InterCHI '93 Conference*. Amsterdam (Pays-Bas), 24–29 avril, 206–213.

NORMAN, D. A. & DRAPER, S. W. (1986). *User Centered System Design: New Perspectives on Human-Computer Interaction*. Lawrence Erlbaum Associates.

NORMAN, D. A. (1988). *The Design of Everyday Things*. Doubleday.

NORMAN, D. A. (2004). *Emotional Design*. Basic Books.

PLAISANT, O., COURTOIS, R., RÉVEILLÈRE, C., MENDELSOHN, G. A. & JOHN, O. P. (2010). Validation par analyse factorielle du Big Five Inventory français (BFI-Fr) – Analyse convergente avec le NEO-PI-R. *Annales médico-psychologiques, revue psychiatrique*, 168(2), 97–106.

QUESENBERY, W. (2003). The Five Dimensions of Usability. *In* ALBERS, M. J. & MAZUR, B. (éd.), *Content and Complexity: Information Design in Technical Communication*. Lawrence Erlbaum Associates, 82–102.

REDET, J., VIAN, M., ESPOSITO, N. & TIJUS, Ch. (2016). Comparaison d'outils adaptés au terrain pour l'évaluation du ressenti à l'usage des IHM. *Actes de Ergo'IA 2016 (Conférence internationale en ergonomie et informatique avancée)*. Bidart-Biarritz (France), 6–8 juillet.

ROBERT, J. M. (2014). Defining and Structuring the Dimensions of User Experience with Interactive Products. *International Conference on Engineering Psychology and Cognitive Ergonomics*, 272–283.

RUSSELL, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, 110(1), 145–172.

SCHERER, K. R. (2005). What Are Emotions? And How Can They Be Measured? *Social Science Information*, 44(4), 693–727.

SHARON, T. (2016). *Validating Product Ideas: Through Lean User Research*. Rosenfeld Media.