



# **Comparaison des échelles de mesure de l'utilisabilité**

par

**Armand DAYANG NANA**

Essai présenté au CeFTI

en vue de l'obtention du grade de maître en technologies de l'information  
(maîtrise en génie logiciel incluant un cheminement de type cours en technologies de  
l'information)

FACULTÉ DES SCIENCES  
UNIVERSITÉ DE SHERBROOKE

Longueuil, Québec, Canada, décembre 2019

## Sommaire

Mesurer la facilité d'utilisation d'un système informatique est l'un des enjeux importants depuis les premiers travaux en ergonomie des interfaces personne-machine. Dans le souci d'améliorer constamment la qualité des interfaces Web, les concepteurs ont recours aux échelles de mesure de l'utilisabilité pour collecter les avis des utilisateurs et mesurer leur satisfaction. Très vite, et en complément de mesures objectives, des échelles de mesure de l'utilisabilité perçue ont été créées afin de systématiser et encadrer les évaluations.

L'utilisabilité est définie comme la capacité d'un système à être facilement utilisé par une personne donnée pour réaliser une tâche pour laquelle il a été conçu. Mesurer l'utilisabilité permet de se confronter à une situation d'utilisation proche des conditions réelles, et constitue une activité permettant des mesures de l'efficacité, de l'efficience et de la satisfaction perçue lors de la réalisation d'une tâche donnée. Plusieurs outils de mesure de l'utilisabilité ont été conçus et certains diffèrent notamment par leur nombre de questions et leur temps de passation.

Choisir l'échelle de mesure de l'utilisabilité adaptée à interaction quelconque est une étape fondamentale. Les échelles de mesure de l'utilisabilité sont hautement subjectives, car elles ne dépendent pas seulement de la qualité du système, mais sont aussi largement influencées par les facteurs liés à l'utilisateur ou au contexte d'interaction. Plusieurs études ont été réalisées sur des échelles d'utilisabilité proposant des questionnaires de type court (environ quatre à dix questions), contrairement aux échelles de mesure de l'utilisabilité proposant des questionnaires de type long (nombreuses questions). Dans ce dernier cas de figure, il est pertinent de se questionner si ces échelles peuvent être soumises à des biais de passation (effet de halo, tendance à l'acquiescement).

L'essai se base sur une étude de Kraig Finstad [1] qui précise que les résultats observés lors de tests utilisant les échelles de mesure de l'utilisabilité UMUX et SUS sont corrélés, même si les échelles et les questions sont différentes. Mais peut-on penser que le fait qu'un questionnaire soit plus long qu'un autre à administrer va donner des résultats tout aussi similaires ?

La problématique est de vérifier si l'échelle de mesure de l'utilisabilité DEEP, à 19 questions, donne des résultats similaires à l'échelle UMUX, à quatre questions.

La méthodologie utilisée se décline en plusieurs étapes. Dans un premier temps, deux sites Web distincts mais proposant une interaction semblable ont été identifiés. Ensuite des sondages en ligne auto-administrés, proposant d'une part le questionnaire long DEEP et d'autre part le questionnaire court UMUX, sont envoyés à deux groupes de participants. Chaque groupe évalue ensuite les mêmes deux sites Web pour lesquels une tâche spécifique leur a été demandée. Nous obtenons ainsi des évaluations de l'utilisabilité perçue pour deux sites Web, d'une part provenant d'utilisateurs ayant répondu à un long questionnaire (DEEP) et d'autre part d'utilisateurs ayant répondu un court questionnaire (UMUX). Nous avons collecté un total de 40 réponses.

Notre hypothèse est que le score final de l'échelle de mesure DEEP (19 questions) du premier groupe d'utilisateurs aura une variation de moins de 5 % par rapport au score final de l'échelle UMUX (4 questions) du deuxième groupe d'utilisateurs. Ceci validerait que les échelles de mesure de l'utilisabilité se valent, indépendamment de leur longueur et des phénomènes d'effet de halo et de tendance à l'acquiescement qui affectent parfois les longs questionnaires.

Les résultats collectés et analysés de notre recherche confirment cette hypothèse, avec des variations en deçà de 5 %.

## Remerciements

Réaliser un essai pour une maîtrise demande un énorme sacrifice et un volume important de travail. Pour finaliser cet essai, cela s'est fait avec l'appui de plusieurs personnes. Je ne saurais commencer ce travail sans toutefois remercier l'ensemble du corps enseignant du Centre de formation en technologies de l'information (CeFTI) qui a contribué à la réalisation de mon essai.

Tout d'abord, je remercie le Seigneur Dieu de m'avoir renouvelé le souffle de vie en m'accordant la santé et la force pour que je puisse réaliser cet essai.

Un grand merci à mon Directeur académique M. Martin Dozois pour sa disponibilité, ses précieux conseils qui m'ont permis non seulement de trouver mon sujet de recherche mais également de comprendre concrètement les concepts du domaine dans lequel je m'aventurais. Il m'a de plus guidé durant toute la phase de rédaction de cet essai. Merci infiniment pour votre professionnalisme et vos éclairages.

Je remercie le Directeur du CeFTI, M. Claude Cardinal, pour sa disponibilité, son écoute et ses conseils qui ont facilité mon intégration à l'Université.

Je tiens également à remercier Mme Yasmine Lee pour sa disponibilité et son souci de toujours nous faciliter la tâche, bien que ce n'était pas souvent évident.

J'ai une pensée pour Lynn Legault, qui m'a prêté son attention et m'a prodigué plusieurs conseils dans le cadre de ma recherche.

Je remercie M. Vincent Echelard de m'avoir fait comprendre les grands concepts de la recherche.

Je dis merci à Michel Hébert de m'avoir fait comprendre les notions sur les fondements de l'essai et de m'avoir guidé et conseillé dans les choix à faire pour mener à bien ce travail.

Je ne saurais terminer sans remercier ma famille et ceux qui m'ont soutenu de près ou de loin pour que je puisse atteindre l'objectif que je me suis fixé dès le départ.

## Table des matières

Sommaire .....	i
Remerciements.....	iii
Table des matières .....	iv
Liste des tableaux.....	vii
Liste des figures.....	viii
Glossaire .....	xi
Liste des sigles, des symboles et des acronymes.....	xii
Introduction.....	1
Chapitre 1 Mise en contexte .....	4
1.1 Interface personne – machine .....	4
1.2 L'utilisabilité du web.....	6
1.2.1 La notion d'utilité.....	8
1.2.2 La notion d'utilisabilité.....	8
• Efficacité.....	9
• Efficience.....	9
• Satisfaction.....	9
1.3 Les échelles de mesure de l'utilisabilité.....	10
1.3.1 La méthode DEEP.....	13
1.3.2 La méthode UMUX .....	16
1.3.3 La méthode SUS .....	17
1.4 L'échelle de Likert .....	18
1.4.1 Effet de halo .....	19
1.4.2 Tendances à l'acquiescement .....	19
Chapitre 2 Revue de la littérature.....	20
2.1 Méthodologie de recherche .....	20
2.2 Questionnaire exploratoire.....	20
2.2.1 Fondements théoriques.....	20

2.2.2	Longueur d'un questionnaire .....	21
2.2.3	Éviter les biais .....	22
2.3	Les modèles de l'utilisabilité .....	23
2.3.1	Le modèle de Shackel (1991) .....	23
2.3.2	Le modèle de Nielsen (1993) .....	24
2.3.3	Le modèle de la norme ISO 9241 – 11 (2018) .....	25
2.4	Tests utilisateurs .....	27
Chapitre 3	Problématique .....	30
3.1	Description .....	30
3.1.1	Objectifs et hypothèses .....	30
3.1.2	Limites .....	33
3.1.3	Type de recherche .....	33
Chapitre 4	Approche proposée .....	34
4.1	Échantillon .....	34
4.2	Description de l'approche .....	34
4.2.1	Panel utilisateur .....	35
4.2.2	Les sites Web à évaluer .....	35
4.3	Mise en œuvre .....	36
4.3.1	Les questionnaires de sondage .....	36
4.3.2	Les mesures .....	38
4.3.3	Approche de validation des résultats .....	39
4.4	Résultats attendus .....	39
Chapitre 5	Analyse des résultats .....	43
5.1	Analyse préliminaire .....	43
5.1.1	Source des données .....	43
5.1.2	Analyse des données : cas de l'échelle DEEP .....	45
5.1.3	Analyse des données : cas de l'échelle UMUX .....	52
5.1.4	Analyse comparative des résultats .....	55
5.2	Retour sur les hypothèses .....	67
5.3	Résultats obtenus .....	68
5.3.1	Tableaux des données .....	68
5.3.2	Différence entre les résultats obtenus .....	72
Conclusion	.....	73
Liste des références	.....	75

Bibliographie .....	78
Annexe I Formulaire du questionnaire UMUX .....	79
Annexe II Formulaire du questionnaire DEEP .....	82
Annexe III Tableau de collecte des données UMUX .....	90
Annexe IV Tableau de collecte des données DEEP .....	91
Annexe V Sites Web retenus pour les tests .....	93



## Liste des tableaux

Tableau 2.1.1 Principaux biais dans les réponses à un questionnaire.....	22
Tableau 2.2.3 Mesures de l'utilisabilité préconisées par la norme ISO 9241-11 (2018).....	26
Tableau 4.2.2 Sites Web évalués dans le cadre de l'essai.....	36
Tableau 4.3.1 : Correspondance des scores aux items inversés par rapport aux items normaux.....	38
Tableau 5.1.1 Différents groupes de participants au sondage.....	44
Tableau 5.1.2.3 : Correspondance des scores aux items inversés par rapport aux items normaux.....	49
Tableau 5.1.4.1 : Tableau comparatif des résultats des différents scores moyens en pourcentage avec le questionnaire UMUX sur le site Web Aldo.....	55
Tableau 5.1.4.2 : Tableau comparatif des résultats des différents scores moyens en pourcentage avec le questionnaire DEEP sur le site Web Aldo.....	58
Tableau 5.1.4.3 : Tableau comparatif des résultats des différents scores moyens en pourcentage avec le questionnaire DEEP sur le site Web Rubino.....	61
Tableau 5.1.4.4 : Tableau comparatif des résultats des différents scores moyens en pourcentage avec le questionnaire UMUX sur le site Web Rubino.....	64
Tableau 5.3.1.1 : Tableau comparatif des différents scores moyens en pourcentages des échelles UMUX et DEEP obtenus sur le site Web Aldo.....	68
Tableau 5.3.1.2 : Tableau comparatif des scores en pourcentage des échelles DEEP et UMUX sur le site Web Rubino.....	70
Tableau 5.3.1.3 : Tableau comparatif des résultats des différents scores moyens en pourcentage.....	71
Tableau 5.3.2 : Différence des résultats.....	72

## Liste des figures

Figure 1.1 Interface personne - machine.....	5
Figure 1.2 Les critères d'utilité et d'utilisabilité doivent être réunis pour obtenir une interface ergonomique.....	7
Figure 1.2.2 Les principales dimensions de l'utilisabilité.....	10
Figure 1.3 Les principales échelles d'utilisabilité.....	12
Figure 1.3.1 Présentation et traduction libre du DEEP.....	14
Figure 1.3.1.1 En établissant un lien entre la convivialité perçue et les dimensions clés de la conception Web, DEEP résout le problème d'alignement d'évaluation-conception.....	15
Figure 1.3.2 Présentation et traduction libre de UMUX.....	16
Figure 1.3.3 Présentation et traduction libre du SUS.....	17
Figure 1.4 Exemple des échelles de Likert utilisées dans quelques types de sondages.....	18
Figure 2.2.1 : L'utilisabilité reliée aux autres concepts d'après Shackel (1991) .....	23
Figure 2.2.2 : L'utilisabilité reliée aux autres concepts de Nielsen (1993) .....	25
Figure 2.3 : Le plan de test utilisateur est un outil de communication.....	28
Figure 3.1.1 : Schéma conceptuel de la recherche.....	32
Figure 4.3.1 : Procédure utilisée pour développer et valider le questionnaire DEEP.....	37
Figure 4.4 : Représentation des scores pour l'échelle DEEP, sous forme graphique.....	40
Figure 4.5 : Exemple de calcul du score au SUS.....	41
Figure 4.6 : Échelle d'interprétation des scores total au SUS (adaptée de Bangor et al., 2008) .....	41
Figure 5.1.1 Courriel d'invitation pour administrer le questionnaire DEEP.....	44

Figure 5.1.2 Courriel d'invitation pour administrer le questionnaire UMUX.....	45
Figure 5.1.2.1.1 : Tâche à effectuer sur le site Web Aldo.....	46
Figure 5.1.2.1.2 : Scénario à réaliser et questionnaire DEEP.....	47
Figure 5.1.2.2 : Questionnaire d'évaluation de la méthode DEEP.....	48
Figure 5.1.2.3 : Recodage des scores 12 et 15 du questionnaire DEEP.....	49
Figure 5.1.2.4.1 : Calcul des scores moyens du questionnaire DEEP.....	50
Figure 5.1.2.4.2 : Calcul des scores moyens en pourcentage du questionnaire DEEP.....	51
Figure 5.1.3.1 : Tâche à effectuer sur le site Web Rubino.....	52
Figure 5.1.3.2 : Questionnaire d'évaluation de la méthode UMUX.....	53
Figure 5.1.3.3 : Données collectés et calculés de la méthode UMUX.....	54
Figure 5.1.4.1.1 : Représentation par sexe pour l'évaluation du site Web Aldo avec le questionnaire UMUX.....	56
Figure 5.1.4.1.2 : Représentation par tranche d'âge pour l'évaluation du site Web Aldo avec le questionnaire UMUX.....	56
Figure 5.1.4.1.3 : Diagramme des scores par utilisateur de l'échelle UMUX du site Web Aldo.....	57
Figure 5.1.4.2.1 : Représentation par sexe pour l'évaluation du site Web Aldo avec le questionnaire DEEP.....	59
Figure 5.1.4.2.2 : Représentation par tranche d'âge pour l'évaluation du site Web Aldo avec le questionnaire DEEP.....	59
Figure 5.1.4.2.3 : Diagramme des scores par utilisateur de l'échelle DEEP du site Web Aldo.....	60
Figure 5.1.4.3.1 : Représentation par sexe pour l'évaluation du site Web Rubino avec le questionnaire DEEP.....	62
Figure 5.1.4.3.2 : Représentation par tranche d'âge pour l'évaluation du site Web Rubino avec le questionnaire DEEP.....	62
Figure 5.1.4.3.3 : Diagramme des scores par utilisateur de l'échelle UMUX du site Web Rubino.....	63

Figure 5.1.4.4.1 : Représentation par sexe pour l'évaluation du site Web Robino avec le questionnaire UMUX.....	65
Figure 5.1.4.4.2: Représentation par tranche d'âge pour l'évaluation du site Web Rubino avec le questionnaire UMUX.....	65
Figure 5.1.4.4.3: Diagramme des scores par utilisateur de l'échelle UMUX du site Web Rubino.....	66
Figure 5.3.1.1 : Comparaison des scores en pourcentage des échelles UMUX et DEEP sur le site Web Aldo.....	69
Figure 5.3.1.2 : Comparaison des scores en pourcentage des échelles DEEP et UMUX sur le site Web Rubino.....	71

## Glossaire

Effet de Halo	Effet de contamination ou effet de contagion, est le biais cognitif qui affecte la façon dont les gens interprètent les informations [2]. C'est une interprétation et une perception sélective d'informations allant dans le sens d'une première impression. Inventé par Edward Thorndike en 1920, Bagozzi approfondit ce terme en 1996 en soutenant qu'il s'agit d'un phénomène où les attitudes globales influencent les jugements sur des qualités d'objets spécifiques.
Échelle de Likert	Encore appelée échelle de satisfaction, est une échelle d'attitude comprenant le plus souvent cinq à sept points par laquelle on demande à un individu d'exprimer son degré d'accord ou désaccord relatif à une affirmation [3]. Créées par le sociologue américain Rensis Likert, les échelles de Likert sont très répandues car elles constituent l'un des outils les plus fiables de mesure des comportements, des avis et des perceptions.
Tendance à l'acquiescement	Est une attirance du sujet pour les réponses positives (OUI, VRAI, D'ACCORD). Le psychologue Cronbach (1942) a démontré que dans les tests de connaissances, il est généralement plus facile de répondre « OUI » que « NON » lorsqu'on est incertain de la réponse à donner à une question [4].

## Liste des sigles, des symboles et des acronymes

ACM	Association of Computing Machinery
ISO	International Organization for Standardization
SUS	System Usability Scale
UMUX	Usability Metric for User Experience
DEEP	Design-oriented Evaluation of Perceived Usability
WAMMI	Web Analysis and Measurement Inventory
QUIS	Questionnaire for User Interface Satisfaction
SUMI	Software Usability Measurement Inventory
CSUQ	Computer Usability Satisfaction Questionnaire

## Introduction

L'informatique prend une place sans cesse grandissante dans notre vie quotidienne. À la maison, au bureau, dans la rue, les ordinateurs nous accompagnent dans de nombreuses activités.

C'est aussi l'ère de la virtualisation des produits, de l'essor des sites Web, des applications mobiles et autres interfaces personne-machine. Nous utilisons chaque jour des interfaces qui nous permettent de communiquer, de commander des produits sur le Web, de faire des transactions électroniques ou d'effectuer des appels téléphoniques.

Pour faire face à la concurrence de plus en plus présente en ligne, les entreprises investissent aujourd'hui beaucoup sur le Web afin de rejoindre leur clientèle et d'augmenter leur part de marché.

Dans le but d'atteindre un plus grand nombre de personnes via ces applications Web, les entreprises souhaitent offrir des services Web de qualité afin de se distinguer de leurs concurrents. De ce fait, nous assistons de jour en jour à une augmentation exponentielle des offres d'applications mobiles et de sites Web permettant aux consommateurs d'effectuer des transactions dans des conditions les plus conviviales possibles.

Le succès commercial d'un produit informatique n'est pas uniquement lié à sa capacité technique mais également à sa facilité d'utilisation. Un logiciel ou une application facilement utilisable permettra de réaliser rapidement la tâche prévue, sans perdre de temps et avec moins de stress.

L'un des objectifs des applications Web est de répondre aux besoins et exigences des utilisateurs. Plusieurs ouvrages ont été rédigés afin d'être des guides pratiques pour faciliter la conception des sites Web. Ces différents ouvrages ne suffisent toutefois pas à modéliser l'ensemble des facteurs qui font qu'une interface est facile à utiliser ou non.

Pour s'assurer de créer de bonnes interfaces, il faut pouvoir évaluer l'interaction entre la personne et la machine. C'est le thème principal de ce travail de recherche qui s'intéresse aux

mesures de l'utilisabilité. Mesurer la facilité d'utilisation d'un système constitue l'un des enjeux depuis les premiers travaux en ergonomie des interfaces personne-machine. Très vite, et en complément de mesures objectives, une panoplie d'échelles d'utilisabilité ont été créées afin de systématiser et cadrer les évaluations. Dans le contexte, est-ce qu'une échelle en particulier est à privilégier afin de mieux mesurer l'utilisabilité du système à évaluer ?

Aussi appelées questionnaires standardisés, les échelles sont des outils qui collectent l'opinion des utilisateurs sur la facilité d'utilisation, leur satisfaction, ou encore l'expérience utilisateur liée à l'interaction avec le système. Ces échelles sont des questionnaires subjectifs, puisqu'ils reflètent l'avis des utilisateurs. Ils sont généralement auto-administrés, c'est-à-dire que l'utilisateur y répond sans la présence directe de l'évaluateur. Le principal avantage des questionnaires standardisés est qu'ils offrent une passation structurée et identique pour tous les utilisateurs. Ces échelles sont scientifiquement :

- Fiables (ou fidèles) lorsqu'elles sont appliquées dans les mêmes conditions ;
- Valides parce qu'elles doivent mesurer ce qu'elles sont censées mesurer (l'utilisabilité, la satisfaction).

S'appuyer sur les échelles standardisées permet ainsi de comparer plusieurs versions d'un système entre elles dans une approche de conception itérative, de comparer différents systèmes entre eux, de comparer différents profils d'utilisateurs entre eux, selon l'âge, le sexe, ou le niveau d'expertise.

Il existe de très nombreux questionnaires standardisés pour évaluer l'utilisabilité. En termes de cibles, ils peuvent être généraux c'est-à-dire qu'ils s'appliquent à tout type de système (site Web, application, logiciel), ou encore ils peuvent être dédiés à un seul type de système. En termes de longueur, certains peuvent être longs (jusqu'à cent items), courts (deux items) ou moyens (généralement entre 15 à 20 items). Bien souvent les longs questionnaires mesurent plusieurs dimensions du système (structure de l'information, qualité des messages, facilité de lecture, facilité de navigation).

L'utilisabilité est l'un des critères de qualité des applications existants qui appartiennent à l'exigence non fonctionnelle [5].



L'utilisabilité, encore appelée la convivialité, est un aspect fondamental des sites Web. Évaluer cette convivialité, ou facilité d'utilisation du système, constitue l'un des enjeux importants depuis les premiers travaux en ergonomie des interfaces personne-machine. Les échelles de convivialité ont été conçues pour systématiser et cadrer les évaluations. Plusieurs échelles de mesure de la convivialité existent de nos jours. Ces échelles sont utilisées tant par des professionnels que des amateurs en conception ainsi que des chercheurs afin de collecter le point de vue des utilisateurs. Généralement le grand défi qui se pose est le choix d'une échelle de mesure de convivialité adaptée au système à évaluer. Le présent essai consiste à faire une comparaison d'échelles de mesure de convivialité des sites Web. Précisons que ces échelles de convivialité sont des outils standardisés de la satisfaction perçue d'un système et liée à l'interaction. La présente recherche permet de faire ressortir les points forts et les points faibles de chaque échelle de mesure de convivialité. Dans un exercice de conception d'interface, l'accès à des outils d'évaluation de qualité est bénéfique pour assurer la conception d'une interaction satisfaisante pour l'utilisateur.

Cet essai s'articulera autour de cinq chapitres, le premier chapitre portera sur la mise en contexte qui permettra d'identifier, de comprendre et d'expliquer les concepts clefs qui seront utilisés dans notre essai. Dans le chapitre deux, nous allons examiner les écrits littéraires qui abordent notre sujet afin de mieux le comprendre. Quant au chapitre trois, il permettra d'identifier le problème qui justifie notre recherche, tout ceci en illustrant le cadre conceptuel et en invoquant l'hypothèse et la question de recherche. Nous parlerons dans le chapitre quatre de notre démarche, c'est-à-dire les procédures pour mener notre expérimentation, soit la méthodologie de travail. Les résultats de notre expérimentation seront présentés au chapitre cinq. À la fin de notre essai, nous ferons une conclusion.

# Chapitre 1

## Mise en contexte

Pour se démarquer de la concurrence et améliorer leur présence sur le Web, les entreprises font face régulièrement aux défis de satisfaire les exigences et besoins des utilisateurs. L'un des enjeux importants est de mesurer la facilité d'utilisation d'un système donné. Mesurer permet en effet de se confronter à une situation d'utilisation proche des conditions réelles, et constitue par conséquent un support permettant d'évaluer l'efficacité, l'efficacité et la satisfaction lors d'une interaction.

### 1.1 Interface personne – machine

Depuis qu'existent les ordinateurs, vers 1945, la question de l'interaction personne-machine s'est posée. L'arrivée des interfaces graphiques a rendu l'informatique plus accessible à un grand nombre de personnes, et dès les années 60 les chercheurs se sont penchés sur les enjeux de la conception des interfaces personne-machine.

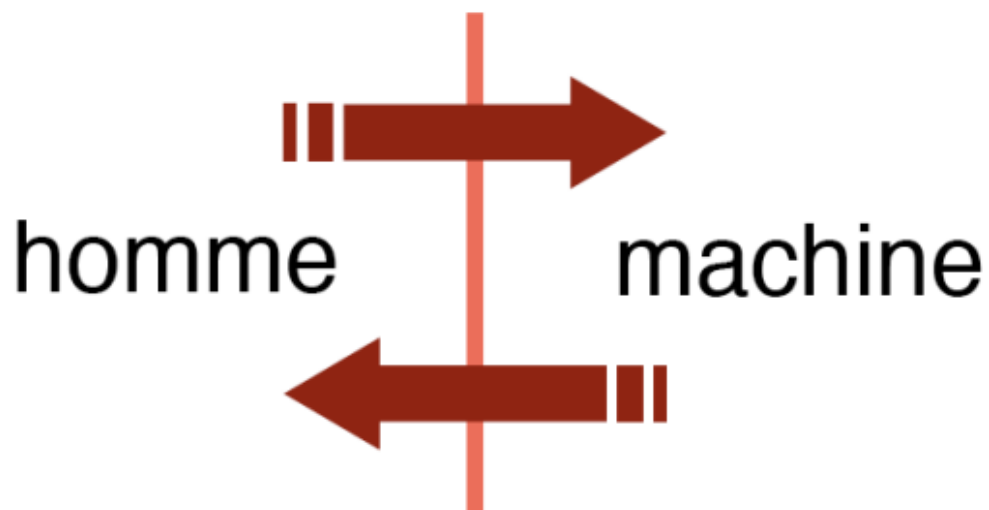
SketchPad, développé par Ivan Sutherland au début des années 1960 et publié dans sa thèse de doctorat en 1963 [6], est considéré comme étant la première interface graphique.

Plusieurs concepts sont couverts par le terme « Interaction personne-machine ». La définition donnée par l'*Association of Computing Machinery* (ACM) généralise tous les aspects du domaine de l'interaction personne-machine. Elle la définit comme « une discipline qui concerne la conception, l'évaluation et l'implémentation des systèmes interactifs de traitement de l'information destinés à être utilisés par les humains et avec l'étude des principaux phénomènes qui les entourent » [7].

Dans cette définition, tout le processus de création des systèmes interactifs est représenté par les termes conception, évaluation et implémentation. Pour commencer, le terme conception

généralise les différentes approches comme le prototypage ou la conception itérative. Ensuite, l'évaluation concerne tout processus qui permet de déterminer dans quelle mesure un système interactif englobe certains critères comme la facilité d'utilisation et d'apprentissage. Pour terminer, l'implémentation permet d'étudier les outils nécessaires pour faire une mise en œuvre d'un système interactif [8].

Le concept de l'interaction personne-machine implique trois principaux acteurs, soit : la machine ou l'ordinateur (le système informatique comprenant divers éléments dont chacun pourra affecter l'utilisateur), la personne (la personne qui va communiquer avec l'ordinateur) et l'interaction (c'est la communication entre l'utilisateur et le système). L'interface est le point central dans le dialogue entre la personne et le système.



**Figure 1.1 Interface personne – machine**

Source : <https://lipn.univ-paris13.fr/~recanati/docs/M2-InHM/Interaction.pdf>

Une technologie ou système ne répondant pas aux composants de la convivialité tel que définis par Nielsen [9] sur la convivialité à savoir l'efficacité, l'efficacéité et la satisfaction peut créer des problèmes d'utilisation aux utilisateurs.

Shackel (1991) [10], définit l'utilisabilité d'un système comme « sa capacité, en termes fonctionnels humains, à permettre une utilisation facile et effective par une catégorie donnée d'utilisateurs, avec une formation et un support adapté, pour accomplir une catégorie donnée de tâches, à l'intérieur d'une catégorie spécifique de contextes » (p. 24).

Les auteurs Jean-François Nogier et Jules Leclerc [11] définissent l'utilisabilité comme étant la capacité de l'objet à être facilement utilisé par une personne donnée pour réaliser la tâche pour laquelle il a été conçu. Cette notion englobe à la fois la performance de réalisation de la tâche, la satisfaction que procure l'utilisation de l'objet et la facilité avec laquelle on apprend à s'en servir.

## **1.2 L'utilisabilité du web**

Le concept d'utilisabilité est un aspect primordial d'un site Web. Encore appelé la convivialité, elle est un attribut de qualité qui évalue la facilité d'utilisation des interfaces utilisateurs. Le terme utilisabilité désigne aussi les méthodes permettant d'améliorer la facilité d'utilisation pendant le processus de conception. C'est également dans le même sens que Amélie Boucher [12] définit un site web ergonomique comme un site web qui satisfait les critères d'utilité et d'utilisabilité (Figure 1.2). Elle va plus loin en expliquant que pour qu'un site Web soit utilisé fréquemment, il faut que les utilisateurs y viennent pour faire quelque chose et que son utilisation soit facile dans la réalisation du besoin.



Figure 1.2 Les critères d'utilité et d'utilisabilité doivent être réunis pour obtenir une interface ergonomique

Source : Ergonomie web pour les sites Web efficaces [12]

### **1.2.1 La notion d'utilité**

La notion d'utilité concerne ce que l'interface permet de faire, c'est-à-dire ce à quoi elle sert pour l'utilisateur. Il est donc question ici d'avoir une idée des besoins et envies des visiteurs. Cela permettra de savoir ce qui leur sera utile ou non. Autrement dit, l'utilité exprime le fait que le produit répond aux besoins fonctionnels de l'utilisateur.

Amélie Boucher [12] nous fait savoir également que la notion d'utilité est multi-dimensionnelle c'est-à-dire qu'elle consiste à offrir un service aux internautes ayant un besoin auquel on peut répondre mais également elle consiste à offrir le meilleur service possible en fonction des besoins et objectifs des internautes.

### **1.2.2 La notion d'utilisabilité**

L'utilisabilité est un concept qui renvoie au sens large de la facilité avec lequel un utilisateur emploie un produit informatique (logiciel, site Web) pour réaliser une activité précise. Il s'agit donc de s'assurer que le produit final correspond bien aux attentes et aux besoins de l'utilisateur potentiel [13].

Il existe une panoplie de définitions de la notion d'utilisabilité ce qui en fait un concept difficile à appréhender, en particulier pour les développeurs de logiciels [14]. Nous allons illustrer trois définitions de la notion d'utilisabilité et, dans le chapitre suivant, nous allons associer les modèles décrits par ces définitions.

Shackel (1991) [10] définit l'utilisabilité d'un système comme « sa capacité, en terme fonctionnels humains, à permettre une utilisation facile et effective par une catégorie donnée d'utilisateurs, avec une formation et un support adapté, pour accomplir une catégorie donnée de tâches, à l'intérieur d'une catégorie de contextes » (p.24).

Nielsen (1993) [9] ne donne pas une définition explicite. Il précise que l'utilisabilité n'est pas une propriété unique et unidimensionnelle, mais l'utilisabilité comporte plusieurs composantes et est traditionnellement associée avec les cinq attributs suivants : l'apprentissage, la facilité, la mémorisation, les erreurs et la satisfaction. Selon Nielsen, l'utilisabilité fait partie d'un concept plus grand qui est l'acceptabilité.

Les rédacteurs des normes ISO 9241 – 11 [15] s'appuieront sur plusieurs définitions de l'utilisabilité afin de créer une définition plus ou moins standardisé qui la définit comme étant : « le degré auquel un système, produit ou service peut être utilisé par des utilisateurs identifiés avec efficacité, efficience et satisfaction dans le contexte d'utilisation spécifié ».

De ces différentes définitions de l'utilisabilité, trois grandes dimensions en découlent :

- **Efficacité**

L'efficacité concerne la réalisation des objectifs de l'activité par les utilisateurs. Autrement dit, c'est la précision et le degré d'achèvement avec lesquels l'utilisateur atteint des objectifs spécifiés. Cette efficacité est souvent mesurée par des taux d'erreurs ou par les actions réalisées (procédure suivie, stratégie de navigation) [13].

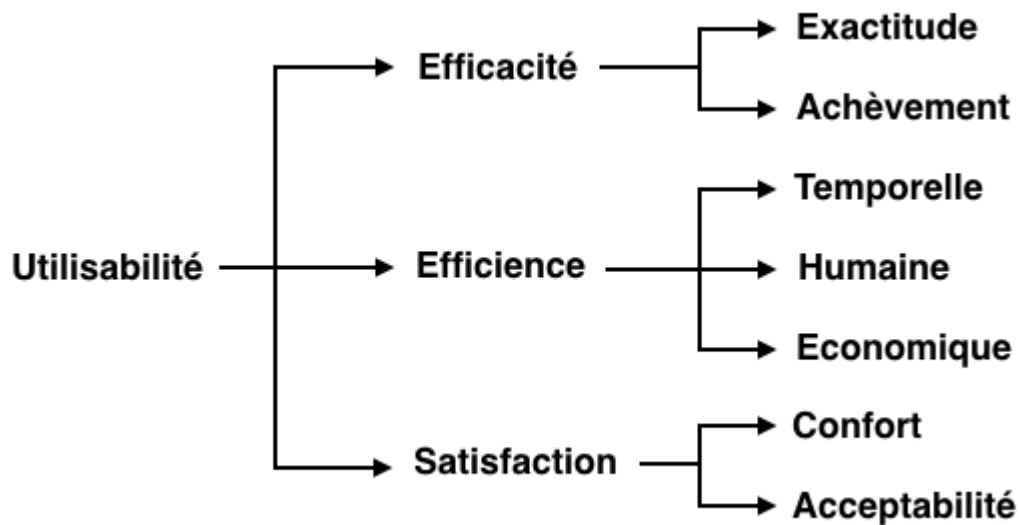
- **Efficience**

Est la capacité à utiliser une faible quantité de ressources pour atteindre les buts. C'est également le rapport entre les ressources utilisées et les résultats obtenus. Comme ressources nous pouvons ressortir le temps, l'effort humain, les coûts et les matériaux.

- **Satisfaction**

Le degré selon lesquelles les réactions physiques, cognitives et émotionnelles qui résultent de l'utilisation d'un système, produit ou service répondent aux besoins et attentes de l'utilisateur. La satisfaction correspond bien évidemment au bien-être, au confort et au plaisir ressentis par l'utilisateur lors de son activité. Elle peut également être mesurée par des questionnaires contenant des valeurs scalaires qui permettent d'estimer subjectivement le niveau global de satisfaction [13].

Une interface sera d'autant plus utilisable qu'elle facilitera la réalisation d'une activité, à la fois dans les buts poursuivis (efficacité) que dans la manière d'y parvenir (efficience) tout en procurant un sentiment de plaisir et de confort (satisfaction) à l'utilisateur.



**Figure 1.2.2 Les principales dimensions de l'utilisabilité**

Source : Selon la norme ISO 9241-11

### **1.3 Les échelles de mesure de l'utilisabilité**

Les échelles de mesure de l'utilisabilité sont des méthodes permettant de mesurer l'utilisabilité d'un site Web ou d'un système donné. L'intérêt principal des échelles de mesure de l'utilisabilité est leur format standardisé. Les questionnaires standardisés, aussi appelé « échelles », sont des outils qui recueillent l'avis des utilisateurs sur la facilité d'utilisation, la satisfaction, ou l'expérience utilisateur liée à l'interaction avec le système. Ce sont des questionnaires d'évaluation subjective, parce qu'ils reflètent l'opinion de l'utilisateur et ils sont généralement auto-administrés : l'utilisateur y répond lui-même [16].

Dès les années 1980, les premiers questionnaires d'utilisabilité sont apparus et ont séduit les concepteurs (Root & Draper, 1983). Leur principal avantage est qu'ils offrent une passation structurée et identique pour tous les utilisateurs, et pour tous les systèmes c'est-à-dire qu'une échelle standardisée étant un questionnaire reprenant un ensemble de questions prédéfinies, toujours posées dans le même ordre, et qui dispose d'une grille de réponse et de cotation identique pour tous les utilisateurs.



Plusieurs avantages plaident en faveur des questionnaires d'utilisabilité car la méthode est indirecte et ne nécessite pas d'analyse du système en lui-même pour évaluer son utilisabilité, et elle permet d'enregistrer l'opinion des utilisateurs, même à distance.

Ces échelles sont scientifiquement :

- Fiables (ou fidèles) car appliquées dans les mêmes conditions ;
- Valides parce qu'elles doivent mesurer ce qu'elles sont censées mesurer (l'utilisabilité, la satisfaction.).

Les échelles standardisées peuvent être répliquées afin de :

- Comparer plusieurs versions d'un système entre elles dans une approche de conception itérative ;
- Comparer différents systèmes entre eux, si l'on souhaite par exemple positionner son propre système par rapport à la concurrence ;
- Tester un système auprès de plusieurs catégories d'utilisateurs afin de différencier leur opinion (sénior, junior, hommes, femmes, statuts socioprofessionnels) [16].

Il existe de nos jours une panoplie de questionnaires standardisés pour évaluer l'utilisabilité. En termes de cible, ils peuvent être généraux, c'est à dire qu'ils s'appliquent à tout type de système (site Web, application, logiciel).

En termes de longueur, certaines échelles peuvent être longues (jusqu'à 100 items), courtes (deux items) ou moyennes (généralement entre 15 à 20 items).

Nom de l'échelle	Nombre d'items	Système évalué	Format d'échelle	Référence
<i>EUCS (End-User Computing Satisfaction)</i>	12	Site web	Likert à 5 points Jamais-Toujours	Doll & Torkzadeh (1988)
<i>Perceived website usability measurement scale</i>	8	Site web	Likert à 7 points Désaccord-Accord	Wang & Senecal (2007)
<i>PSSUQ v3 (Post-study System Usability Questionnaire)</i>	16	Tout type de système	Likert à 7 points Désaccord-Accord	Lewis (2002)
<i>PUTQ (Purdue Usability Testing Questionnaire)</i>	100	Tout type de système	Likert à 9 points Désaccord-Accord	Lin, Choong & Salvendy (1997)
<i>QUIS v7.0 (Questionnaire for User Interface Satisfaction)</i>	41 (version courte)	Tout type de système	Likert à 9 points Désaccord-Accord + NA	<a href="http://www.lap.umd.edu/QUIS">www.lap.umd.edu/QUIS</a> Chin, Diehl & Norman (1988)
<i>SUMI (Software Usability Measurement Inventory)</i>	50	Tout type de système	Likert à 3 points D'accord/ Ne sait pas/Pas d'accord	<a href="http://sumi.ucc.ie">http://sumi.ucc.ie</a> Kirakowski & Corbett (1993)
<i>SUPR-Q (Standardized User Experience Percentile Rank Questionnaire)</i>	8	Site web	Likert à 5 points Désaccord-Accord	Sauro (2015)
<i>SUS (System Usability Scale)</i>	10	Tout type de système	Likert à 5 points Désaccord-Accord	Brooke (1996 ; 2013)
<i>UMUX (Usability Metric for User Experience)</i>	4	Tout type de système	Likert à 7 points Désaccord-Accord	Finstad (2010)

**Figure 1.3 les principales échelles d'utilisabilité**

Source : Méthodes de design UX [16]

### 1.3.1 La méthode DEEP

La méthode d'évaluation de l'utilisabilité perçue axée sur la conception, DEEP (*Design-Oriented Evaluation of Perceived Usability*) a été développée en 2012 [17] afin de pallier un défaut des principales autres échelles, à savoir leur manque à proposer des recommandations pour résoudre les problèmes d'une interface. L'ambition du DEEP est de mesurer :

- La manifestation de l'expérience de l'utilisateur ;
- L'origine du problème dans l'interface.

Le test DEEP s'appuie sur la combinaison de 19 items extraits et adaptés d'autres échelles. Cette échelle est ainsi constituée de 19 items sous forme de phrases affirmatives et réparties en six catégories [16] :

- Contenu perçu
- Structure perçue et architecture de l'information
- Navigation perçue
- Effort cognitif perçu
- Cohérence de la mise en page perçue
- Guidage visuel perçu

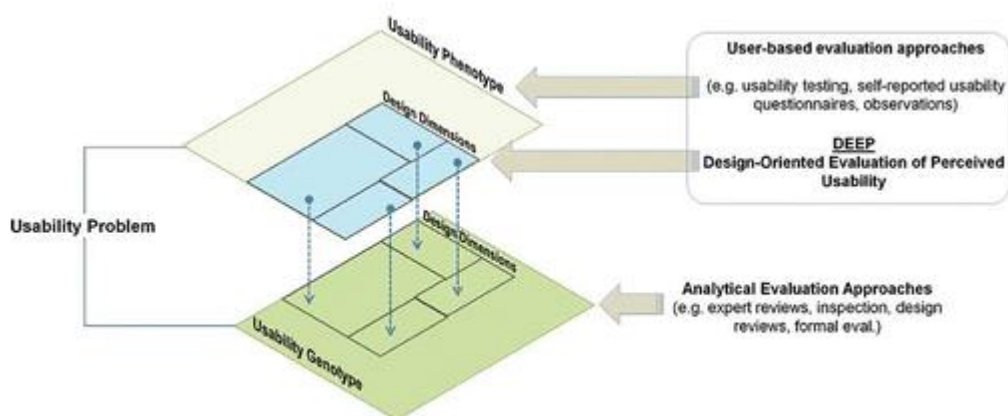
La passation d'un test DEEP par un utilisateur pour évaluer un site Web dure en moyenne entre 10 et 15 minutes [16].

DEEP (Design-oriented Evaluation of Perceived Usability)		1 = Pas du tout d'accord 5 = Tout à fait d'accord NA = Non applicable					
<b>CONTENU PERÇU</b>		1	2	3	4	5	NA
1. Le libellé du texte était clair.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Le contenu (texte, images, sons, vidéos, etc.) était facile à comprendre.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Le texte était utile.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Le texte était pertinent.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>STRUCTURE PERÇUE ET ARCHITECTURE DE L'INFORMATION</b>		1	2	3	4	5	NA
5. Je pouvais rapidement connaître la structure du site web en parcourant sa page d'accueil.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. L'organisation du site web était claire.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Dans chaque section du site web, les pages étaient bien organisées.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>NAVIGATION PERÇUE</b>		1	2	3	4	5	NA
8. Il était facile de trouver l'information dont j'avais besoin sur le site web.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Le site web m'a aidé à trouver ce que je cherchais.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. J'ai obtenu ce à quoi je m'attendais quand je cliquais sur les éléments du site web.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>EFFORT COGNITIF PERÇU</b>		1	2	3	4	5	NA
11. Utiliser ce site web s'est fait sans effort.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Utiliser ce site web m'a fatigué.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. J'ai appris à utiliser ce site web rapidement.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>COHERENCE DE LA MISE EN PAGE PERÇUE</b>		1	2	3	4	5	NA
14. La mise en page à travers tout le site web était cohérente.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. J'ai remarqué des changements soudains de mise en page à travers le site web.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. La mise en page de chaque section du site web était cohérente.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>GUIDAGE VISUEL PERÇU</b>		1	2	3	4	5	NA
17. Les couleurs m'ont aidé à distinguer les différentes sections du site web.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. Les zones mises en évidence d'une page m'ont aidé à repérer l'information dont j'avais besoin.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19. J'ai appris à connaître le contenu d'une page en parcourant les zones mises en évidence.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>Note : Les items 12 et 15 sont inversés.</i>							

**Figure 1.3.1 Présentation et traduction libre du DEEP**

Source : Méthodes de design UX [16]

Les auteurs Tao Yang, Jared Linder et Davide Bolchini nous révèlent dans leur article [18] que « D'un point de vue général, DEEP prend les piliers de la conception de l'utilisateur (le phénotype de l'utilisabilité) et rend cette perception analytique en termes d'utilisation des dimensions de conception pour décrire l'utilisabilité perçue. Cela nous permet de combler le fossé entre le génotype d'utilisabilité et la composition analytique du plan (génotype d'utilisabilité; voir figure 1.3.2) ».



**Figure 1.3.1 En établissant un lien entre la convivialité perçue et les dimensions clés de la conception Web, DEEP résout le problème d'alignement d'évaluation-conception [18]**

Source : Évaluation de l'utilisabilité perçue axée sur la conception (DEEP) [18]

La méthode DEEP peut certainement servir d'outil pour une variété d'études et de recherches en interface personne-machine. Elle peut également être utilisée par les concepteurs pour capturer les commentaires spécifiques des utilisateurs sur la convivialité perçue d'un site Web tout en répondant aux exigences de conception, alignées sur la composition conceptuelle d'une conception complexe.

### 1.3.2 La méthode UMUX

La métrique de l'utilisabilité pour l'expérience utilisateur, UMUX (*Usability Metric for User Experience*) a été développée par Finstad en 2010 [1] dans le but de chercher à réduire le plus possible le nombre d'items de la méthode SUS (*System Usability Scale*) tout en gardant et en respectant la même mesure des composants de l'utilisabilité (efficacité, efficience, satisfaction). Notons que cette méthode ne comprend que quatre items. Les utilisateurs y répondent en se positionnant sur une échelle Likert allant de 1, qui signifie (pas du tout d'accord) à 7, qui signifie (tout à fait d'accord). La méthode UMUX se veut encore plus rapide à remplir et facile à comprendre (*quick and dirty*) que la méthode SUS et répond parfaitement aux attentes des professionnels les plus pressés et aux utilisateurs les moins disponibles [16]. La passation d'un test UMUX par un utilisateur pour évaluer un site Web dure en moyenne moins de 5 minutes.

UMUX (Usability Metric for User Experience)	1 = Pas du tout d'accord 7 = Tout à fait d'accord						
	1	2	3	4	5	6	7
1. Les fonctionnalités de ce système répondent à mes exigences. (This system's capabilities meet my requirements).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Utiliser ce système est une expérience frustrante. (Using this system is a frustrating experience).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Ce système est facile à utiliser. (This system is easy to use).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Je dois passer trop de temps à corriger des choses sur ce système. (I have to spend too much time correcting things with this system).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>Note : Les items 2 et 4 sont inversés.</i>							

**Figure 1.3.2 Présentation et traduction libre de UMUX**

Source : Méthodes de design UX [16]

L'UMUX peut être administré électroniquement sous forme d'enquête ou de suivi lors des tests d'utilisabilité. Il est simple à administrer, car rapide à remplir et facile à comprendre.

### 1.3.3 La méthode SUS

La méthode SUS (*System Usability Scale*) a été l'une des premières échelles de mesure de l'utilisabilité perçue (en 1996). Elle est libre de droits et comporte un nombre relativement limité d'items (10) et facile à comprendre pour les utilisateurs. Le concepteur de cette méthode, John Brooke, explique que cette échelle a été créée avec soin en se basant sur les éléments de la norme ISO-9241-11 sur l'utilisabilité, mais quelle se voulait *quick and dirty* pour les utilisateurs, c'est-à-dire rapide à remplir et facile à comprendre [16].

SUS (System Usability Scale)	1 = Pas du tout d'accord 5 = Tout à fait d'accord				
	1	2	3	4	5
1. Je pense que j'aimerais utiliser ce système fréquemment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. J'ai trouvé ce système inutilement complexe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. J'ai trouvé ce système facile à utiliser.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Je pense que j'aurais besoin d'un support technique pour être capable d'utiliser ce système.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. J'ai trouvé que les différentes fonctions de ce système étaient bien intégrées.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. J'ai trouvé qu'il y avait trop d'incohérence dans ce système.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Je suppose que la plupart des gens apprendraient très rapidement à utiliser ce système.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. J'ai trouvé ce système très contraignant à utiliser.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Je me suis senti(e) très confiant(e) en utilisant ce système.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. J'ai dû apprendre beaucoup de choses avant de me sentir familiarisé(e) avec ce système.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*Note : Les items 2, 4, 6, 8 et 10 sont inversés.*

**Figure 1.3.3 Présentation et traduction libre du SUS**

Source : Méthodes de design UX [16] (p.357)

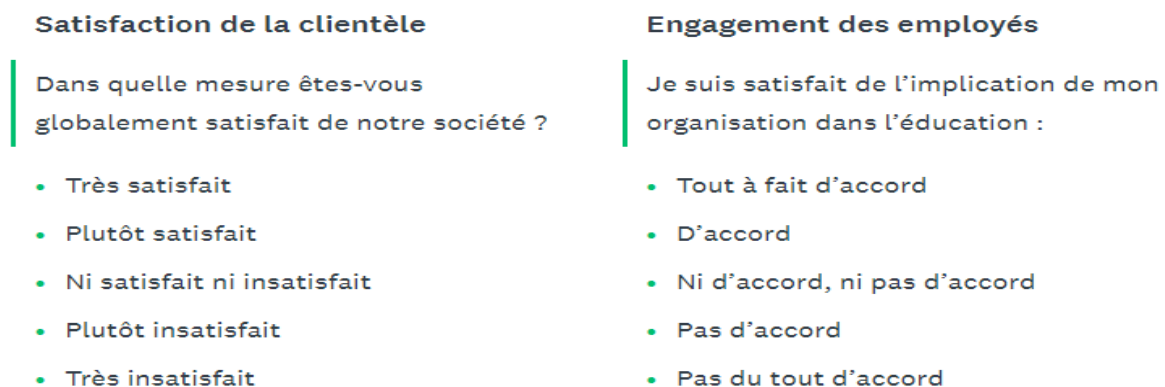
## 1.4 L'échelle de Likert

Plusieurs connaissent et observent régulièrement les populaires échelles de Likert, mais peu savent d'où vient son nom et qui en est réellement son auteur. Les échelles de Likert, du nom de leur créateur, le sociologue Rensis Likert, ont été créées en 1932 et sont très répandues parce qu'elles sont parmi les outils les plus fiables pour mesurer les opinions, les perceptions et les comportements. Les échelles de Likert sont le plus souvent utilisées dans le cadre d'études quantitatives réalisées par le biais de questionnaires.

Il existe une panoplie de définition de l'échelle de Likert, B. Bathelot [19] dans son article la définit comme étant « une échelle d'attitude comprenant le plus souvent 5 à 7 degrés par laquelle on demande à l'individu d'exprimer son degré d'accord ou désaccord relatif à une affirmation ».

SurveyMonkey [3], le site Web très populaire proposant des outils d'enquêtes de sondage, définit une échelle de Likert comme une échelle comportant cinq ou sept options de réponse allant d'une opinion/attitude extrême à une autre, telle que « Tout à fait » à « Pas du tout ». Généralement, elle inclut un point médian (opinion/attitude : modérée/neutre).

Une échelle contient pour chaque élément une graduation comprenant en général cinq ou sept choix de réponse qui permettent de nuancer le degré d'accord.



**Figure 1.4 Exemple des échelles de Likert utilisées dans quelques types de sondages**  
[3]



Dans le cadre de notre essai, nous allons nous attarder aux échelles de Likert à cinq et sept points. Diefenbach et al (1993) [20] ont toutefois constaté que les échelles à sept points surpassaient les échelles à cinq points en termes de fiabilité, de précision et de facilité d'utilisation, de même que l'analyse de Cox (1980) [21] sur les échelles de Likert a révélé que le nombre optimal d'alternatives était de sept.

Les échelles d'utilisabilité peuvent toutefois être soumises à des biais de passation (effet de halo, tendance à l'acquiescement). Il pourrait être possible que les questionnaires trop longs soient source d'ennui et que les utilisateurs soient tentés de répondre plus rapidement, questionnant la validité, afin de se débarrasser de la tâche.

#### **1.4.1 Effet de halo**

L'effet de halo précise que bien souvent, les premières réponses apportées par un participant ou utilisateur influencent ses réponses ultérieures. Cet effet est surtout observé lorsque plusieurs échelles de Likert se suivent sur un même sujet. Dans ce cas, le répondant a coché les cases qui sont sur le même « côté » [16].

#### **1.4.2 Tendance à l'acquiescement**

La tendance à l'acquiescement signifie qu'il est généralement plus facile de répondre « OUI » à une question, que « NON », qui peut apparaître comme une forme d'opposition. Il est alors préférable de formuler les questions de telle sorte que le participant ait le choix entre deux possibilités de réponse, plutôt qu'il ait à répondre par oui ou par non [16].

Ce chapitre a abordé les grands concepts clefs qui sont liés aux échelles d'utilisabilité. Cela a permis de comprendre en profondeur les notions d'utilité et d'utilisabilité et de comprendre la nature du sujet. Dans le chapitre suivant, nous allons examiner les écrits littéraires qui abordent plus précisément le sujet des outils d'évaluation des interfaces personne-machine.

## **Chapitre 2**

### **Revue de la littérature**

Cette revue de littérature s'organise autour de certains concepts fondamentaux. Dans un premier temps, nous allons présenter le concept du questionnaire exploratoire, ensuite nous allons nous intéresser au concept des échelles d'utilisabilité : leur définition, les modèles et les méthodes d'évaluation. Nous allons enfin présenter le concept de tests d'utilisabilité.

#### **2.1 Méthodologie de recherche**

La recherche des différents articles s'est faite sur Internet avec d'une part les outils offerts par le service de la bibliothèque de l'Université de Sherbrooke et d'autre part, les sites de Google Scholar et Google Recherche.

Dans le contexte des interfaces personne-machine, les recherches les plus productives se sont faites à partir des variantes telles que : « test d'utilisabilité », « tests utilisateurs », « évaluation UX », « échelle d'utilisabilité » et « mesure de l'utilisabilité ».

La liste des références et bibliographies des articles et ouvrages trouvés a été analysée afin de retenir les articles et ouvrages les plus pertinents.

#### **2.2 Questionnaire exploratoire**

##### **2.2.1 Fondements théoriques**

Depuis longtemps, le besoin de recueillir des informations sur les populations s'est fait ressentir. L'écrivain Andrea (2012) [22] relate la découverte d'un document de recensement des habitants de la Chine datant de plus de 4000 ans. Pourtant, l'utilisation d'un questionnaire comme outil de collecte de données exploratoires est beaucoup plus récente.

Le questionnaire est employé dans le domaine des interfaces personne–machine pour recueillir les données sur les opinions, les idées, les comportements, les usages, les habitudes ou les besoins des utilisateurs (Ozok, 2007) [16].

La norme ISO 16982 (2002) [13], référence le questionnaire comme l'une des 12 principales méthodes d'utilisabilité. Elle recommande d'administrer des questionnaires « fermés » c'est-à-dire qui comporte des questions auxquelles le répondant ne peut se positionner que sur une échelle quantifiable (de type Likert) ou dans une liste prédéfinie.

Pour mieux mesurer une échelle d'utilisabilité, les concepteurs de sites Web ou d'application ont recours aux indicateurs de mesure telle que les échelles de Likert. Une échelle de Likert permet d'évaluer l'attitude en mesurant l'intensité de son approbation avec une ou plusieurs affirmations.

Les questionnaires scalaires demandent à l'utilisateur de juger sur une échelle prédéfinie son opinion sur une proposition particulière. Cette échelle peut avoir seulement deux valeurs (échelle binaire – vrai/faux, oui/non) ou plusieurs valeurs (échelles multiples : cinq, sept, dix valeurs). Les échelles multiples peuvent interroger l'utilisateur sur son accord ou son désaccord avec une assertion (échelle de Likert) [13].

### **2.2.2 Longueur d'un questionnaire**

Une étude sur la longueur des questionnaires et le temps nécessaire à les remplir a été réalisée par l'équipe de *SurveyMonkey*, une plate-forme de questionnaire en ligne. L'analyse de 100 000 questionnaires comportant de 1 à 30 questions a révélé plusieurs indicateurs (Brent, 2011) [24].

- Il faut environ une minute pour remplir un questionnaire d'une question et de 9 à 10 minutes pour 26 à 30 questions.
- Plus un questionnaire est long, moins les participants consacrent de temps à chaque question (40 secondes par question en moyenne pour les questionnaires de 2 items, contre 19 secondes pour 26 à 30 items).
- Au-delà de 8 minutes de participation, le taux d'abandon passe de 5 à 20 %. En moyenne, un questionnaire ne devrait pas comprendre plus de 25 questions.

### 2.2.3 Éviter les biais

De nombreux biais peuvent affecter la qualité des réponses à un questionnaire (tableau 2.1.1). Pour l'éviter, les questions doivent être formulées de manière claire, simple et précise. Le vocabulaire et le registre lexical doivent être adaptés aux répondants. Une attention particulière doit également être portée à la formulation des questions afin qu'elles n'influencent pas les réponses.

**Tableau 2.1.1 Principaux biais dans les réponses à un questionnaire [16]**

Source: Méthodes de design UX [16] (p.112)

Effets de primauté et de récence	L'ordre de présentation des possibilités de réponse (pour les questions en éventail) peut influencer le participant, qui aura tendance à privilégier les premières propositions (effet de primauté) ou les dernières (effet de récence). L'idéal est alors de présenter les possibilités de façon aléatoire.
Désirabilité sociale	Les participants ont tendance à répondre aux questions de manière à se montrer à leur avantage, sous un angle plus « désirable » que la réalité [25]. À l'inverse, les réponses peu valorisantes sont minimisées. Les questions indirectes sont dans ce cas à privilégier pour certains sujets jugés sensibles.
Tendance à l'acquiescement	Il est généralement plus facile de répondre « oui » à une question, que « non » qui peut apparaître comme une forme d'opposition. Il est alors préférable de formuler les questions de telle sorte que le participant ait le choix entre deux possibilités de réponse, plutôt qu'il ait à répondre par oui ou non.
Effet de halo (ou effet de contagion)	Bien souvent, les premières réponses apportées par un participant influencent ses réponses ultérieures. Cet effet est surtout observé lorsque plusieurs échelles de Likert se suivent sur un même sujet. Dans ce cas, le répondant a tendance à cocher les cases qui sont sur le même « côté ». Le pré-test du questionnaire aide à détecter l'effet de halo.
Tendance à la tricherie	Certains participants indiquent des réponses non valides pour gagner du temps ou pour éviter de répondre à des questions perçues comme trop intrusives. Certaines techniques permettent de se prémunir de ces pratiques (items renversés, items de vérification, tri de données avant analyse).

## 2.3 Les modèles de l'utilisabilité

Plusieurs modèles de l'utilisabilité existent. Dans les paragraphes suivants, nous allons présenter ces différents modèles.

### 2.3.1 Le modèle de Shackel (1991)

Selon Shackel, l'utilisabilité était un facteur important pour convaincre un utilisateur d'adopter le système. Il avait également déjà compris que ce n'était pas l'unique facteur et qu'il faisait partie d'un concept plus englobant.

Dans son ouvrage [10], il explique que l'utilité, l'utilisabilité et la sympathie se combinent et doivent être considérées avec le coût du produit afin de savoir si l'utilisateur voudra acheter le produit (acceptabilité).

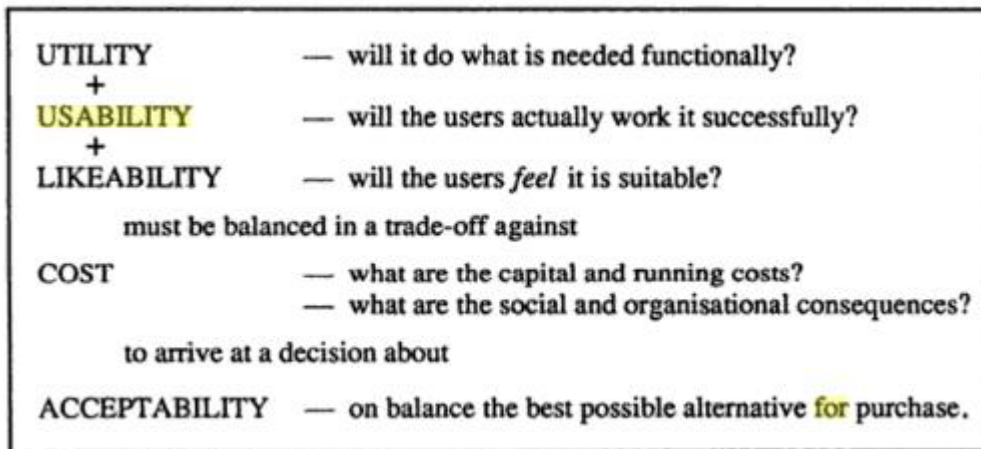


Figure 2.2.1 : L'utilisabilité reliée aux autres concepts d'après Shackel (1991)

Source: *Human factors for Informatics usability* [10]

L'utilité exprime le fait que le produit réponde aux besoins fonctionnels de l'utilisateur.

L'utilisabilité se réfère sur le succès des utilisateurs lors de l'utilisation du produit.

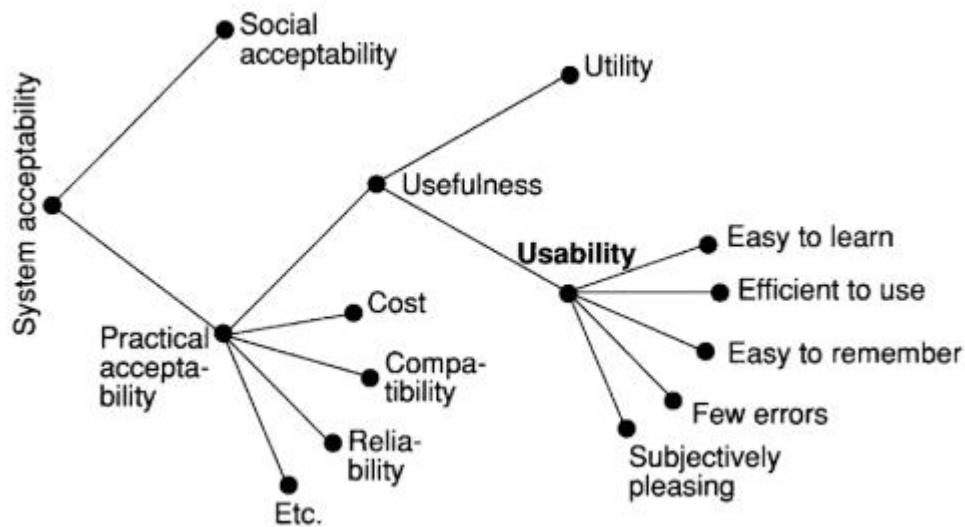
### 2.3.2 Le modèle de Nielsen (1993)

Jakob Nielsen précise dans son livre intitulé « *Usability Engineering* » que l'utilisabilité se situe dans un cadre plus large qu'il nommera « acceptabilité du système ». Pour Nielsen [9], l'acceptabilité comporte deux dimensions : l'acceptabilité pratique (considérant l'intention que le système permet d'atteindre ) et l'acceptabilité sociale (s'intéressant au contexte d'utilisation du système).

L'acceptabilité pratique est décomposée en plusieurs fonctionnalités à savoir : le coût, la comptabilité, la fiabilité et d'autres supports. Selon le modèle d'acceptabilité des systèmes de Nielsen, l'utilité et l'utilisabilité sont deux attributs qui dérivent de l'acceptabilité pratique du système.

Pour pouvoir mesurer l'utilisabilité, Nielsen (1993) a décrit trois composantes initiales dérivant de ce concept à savoir : l'efficacité, l'efficience et la satisfaction. Dans son ouvrage [9], Nielsen confirme que l'utilisabilité n'est pas une propriété unique et unidimensionnelle, mais l'utilisabilité comporte plusieurs composantes et est traditionnellement associée avec cinq attributs que sont :

- Facilité d'apprentissage (*learnability*) : « l'interface devrait être facile à apprendre pour permettre à l'utilisateur de commencer rapidement la tâche avec l'interface ».
- L'efficience : « le système devrait être efficient à utiliser afin que l'utilisateur apprenne à l'utiliser, qu'il puisse avoir un haut niveau de productivité ».
- La mémorabilité (*memorability*) : « l'utilisateur doit être capable de réutiliser l'interface après une période de non utilisation sans pour autant réapprendre de nouveau ».
- Peu d'erreurs : « l'interface devrait occasionner un faible taux d'erreurs, ce qui permet à l'utilisateur de commettre peu d'erreurs pendant l'utilisation de l'interface, et s'il arrivait qu'il fasse des erreurs, qu'il puisse facilement les récupérer et les corriger ».
- La satisfaction : « l'interface doit être plaisante et agréable à utiliser afin que les utilisateurs apprécient l'interaction avec l'interface ».



**Figure 2.2.2 : L'utilisabilité reliée aux autres concepts de Nielsen (1993)**

Source : *Usability Engineering* de Nielsen [9]

### 2.3.3 Le modèle de la norme ISO 9241 – 11 (2018)

Les normes contribuent à guider le processus de conception dans le cadre d'une meilleure utilisabilité des interfaces personne-machine mais ne la garantissent pas. Elle représente généralement les principes généraux acquis par les études empiriques. Lorsqu'on veut déterminer l'utilisabilité d'un système particulier, la conception doit s'accompagner d'une panoplie de mesures quantitatives comportementales. Ces mesures prennent souvent la forme de tests utilisateurs dont l'objectif est d'évaluer si le produit ou le prototype respecte les trois critères fondamentaux d'un système « utilisable », à savoir : l'efficacité, l'efficience et la satisfaction des utilisateurs dans le contexte de la tâche.

Les normes peuvent s'appliquer à la visibilité et à la lisibilité de l'interface et dans ce cas, elles s'accompagnent d'un ensemble de mesures qui spécifient les conditions optimales de lectures [13].

La norme ISO 9241-11 (2018), qui est la plus importante en ergonomie des écrans, a pour objectifs de spécifier les règles ergonomiques minimales pour que les écrans puissent assurer une lecture confortable, sûre et efficace dans le contexte de travail.

**Tableau 2.2.3 : Mesures de l'utilisabilité préconisées par la norme ISO 9241-11 (2018)**

Source : Mesure de l'utilisabilité des interfaces [13] (p.36)

	Mesures d'utilisabilité		
	Efficacité	Efficience	Satisfaction
Adéquation à la tâche	% de buts atteints	Temps pour réaliser la tâche	Echelle de jugement de satisfaction
Approprié pour l'utilisateur entraîné	Nombre de fonctions importantes utilisées	Efficacité relative comparée à un expert	Echelle de jugement de satisfaction
« Apprenabilité »	% de fonctions apprises	Temps d'apprentissage	Echelle de jugement de la facilité d'utilisation
Tolérance à l'erreur	% d'erreurs corrigées	Temps passé à corriger les erreurs	Echelle de jugement de facilité de correction



## 2.4 Tests utilisateurs

Le test utilisateur ou encore test d'utilisabilité est la méthode la plus efficace pour évaluer l'ergonomie d'une application. Cette phase permet de mettre en situation l'utilisateur afin d'observer la façon dont ce dernier se sert d'une application pour la réalisation de tâches prédéfinies.

Les données collectées donnent de précieuses informations sur les forces et les faiblesses du système évalué, ainsi que sur l'expérience vécue par les utilisateurs cibles. Les tests utilisateurs peuvent être réalisés au démarrage des premières maquettes du système jusqu'au produit final. Cette phase peut combiner dans une même session d'autres méthodes comme les échelles d'utilisabilité ou d'UX. Enfin, la méthode du test d'utilisabilité est également une combinaison flexible de méthodes quantitatives et qualitatives permettant d'évaluer les performances des participants, leurs comportements ou en encore leurs réactions.

Lors d'un test utilisateur, plusieurs modes de passation sont possibles, adaptés à toutes les exigences ou contraintes d'un projet : en situation contrôlée, à distance ou en mode guérilla.

La méthode du test est directement dérivée des méthodes d'expérimentation scientifique. Les tests utilisateurs peuvent être conduits de différentes manières selon les objectifs et les contraintes du projet de conception. Il existe quatre grandes catégories de tests :

- Test en situation contrôlée : il est pratiqué dans un laboratoire d'utilisabilité ou dans tout autre contexte contrôlé, ceci pour isoler le participant de ce qui pourrait le distraire de sa tâche afin de minimiser l'impact de variables externes sur les résultats observés.
- Test à contexte naturel : ici, les participants sont évalués sur le terrain c'est-à-dire sur leur lieu de travail, à la maison, dans la rue ou tout autre endroit dans lequel s'effectue d'ordinaire leur interaction avec le système.
- Test à distance : il s'agit de situation de tests où l'observateur et l'utilisateur sont géographiquement distants (Bastien, 2010) [26]. Son avantage est qu'il permet de pouvoir mobiliser rapidement un grand nombre d'utilisateurs cibles sans se soucier de leur localisation. Dans ce cas de figure, on distingue deux types de tests à distance :

- Les tests à distance synchrones : ces tests sont conduits en temps réel par l'évaluateur, qui gère et observe la passation à l'aide d'un système spécifique ou par vidéoconférence.
- Les tests à distance asynchrones : ils sont conduits par des systèmes automatiques, qui gèrent les passations individuelles sans intervention directe de l'évaluateur. Ils offrent la possibilité de faire passer simultanément un grand nombre de participants.
- Test en mode guérilla : ils prennent une place de plus en plus importante dans la pratique des professionnels [27]. Ils s'affranchissent du formalisme des tests utilisateurs classiques pour réduire les contraintes logistiques et les coûts de passation.

AUTEUR		CONTACT		COMMENTAIRES AVANT LE
<b>PRODUIT TESTÉ</b> Quel est le produit testé ? Quels sont les objectifs visés par le produit ?		<b>PARTICIPANTS</b> Combien de participants sont recrutés ? Quelles sont leurs caractéristiques ?		<b>TÂCHES ET SCÉNARIOS DU TEST</b> Quelles sont les tâches ou scénarios évalués ?
<b>OBJECTIFS DU TEST</b> Quels sont les objectifs du test ? A quelles questions cherche-t-il à répondre ? Quelles sont les hypothèses testées ?		<b>ÉQUIPEMENT</b> Quel est l'équipement nécessaire ? Comment les données seront-elles enregistrées ?		
<b>LIEU ET DATE</b> Où et quand est mené le test ?				
<b>PROCÉDURE</b> Quelles sont les principales étapes du protocole de test ?				

**Figure 2.3 : Le plan de test utilisateur est un outil de communication**

Source : Méthodes de design UX [16]

Dans ce chapitre, nous avons examiné la majeure partie des articles scientifiques qui traitent des échelles d'utilisabilité. Au regard de ces articles, plusieurs définitions de l'utilisabilité ont été définies de manières différentes mais l'idée générale demeure la même. Le prochain chapitre permettra d'identifier la problématique, de définir la question de recherche et d'expliquer l'hypothèse de la recherche. Les connaissances qui soutiennent le cadre conceptuel de la recherche seront également présentées.

## Chapitre 3

### Problématique

Les auteurs Yang, Linder et Bolchini, dans leur article « DEEP: Design-Oriented Evaluation of Perceived Usability » [18], révèlent que lors de l'inspection sur la facilité d'utilisation par les experts, la convivialité perçue reflète les sentiments subjectifs des utilisateurs à l'égard d'un produit et évite les biais introduits par les experts en convivialité. D'où l'un des moyens les plus populaires de mesurer la convivialité perçue, qui consiste à demander aux utilisateurs de remplir un questionnaire leur permettant de fournir leurs propres évaluations en ce qui concerne les aspects prédéfinis de la convivialité d'un produit.

#### 3.1 Description

L'article *Interaction Design Foundation* [28] révèle que l'utilisabilité, aussi appelée la convivialité, fait partie du terme général « expérience utilisateur » et fait référence à la facilité d'accès et / ou d'utilisation d'un produit ou d'un site Web. Une conception n'est pas utilisable ou inutilisable en soi; ses fonctionnalités, ainsi que le contexte de l'utilisateur (ce que l'utilisateur veut en faire et l'environnement de l'utilisateur), déterminent son niveau de convivialité.

Les interfaces sont de plus en plus présentes dans notre quotidien; et pour faire face à la concurrence grandissante, les entreprises sont aussi très actives sur le Web afin d'augmenter leur part de marché et conserver leur clientèle. Elles ont donc besoin d'interfaces conviviales, adaptées aux besoins des utilisateurs. Les échelles d'utilisabilité vont jouer un rôle primordial pour mesurer l'utilisabilité d'un site Web ou d'un système donné.

##### 3.1.1 Objectifs et hypothèses

Concevoir un système donné présentant une bonne expérience utilisateur est l'un des principaux objectifs de tout projet numérique. Choisir l'échelle de mesure de la convivialité

adaptée à un besoin quelconque est une étape fondamentale. Les échelles d'évaluation de la convivialité sont hautement subjectives, car elles ne dépendent pas seulement de la qualité du système, mais sont aussi largement influencées par les facteurs liés à l'utilisateur ou au contexte d'interaction.

Bien qu'il existe plusieurs travaux sur les échelles d'utilisabilité des sites Web, plusieurs recherches sont basées sur les échelles d'utilisabilité de questionnaires de type courts, c'est-à-dire des questionnaires avec un nombre de questions réduit (environ quatre à dix questions) et également basées sur les critères de Likert de sept points et de cinq points. Peu de recherches concernent les échelles d'utilisabilité de questionnaire de type longs, c'est-à-dire des questionnaires à plusieurs questions.

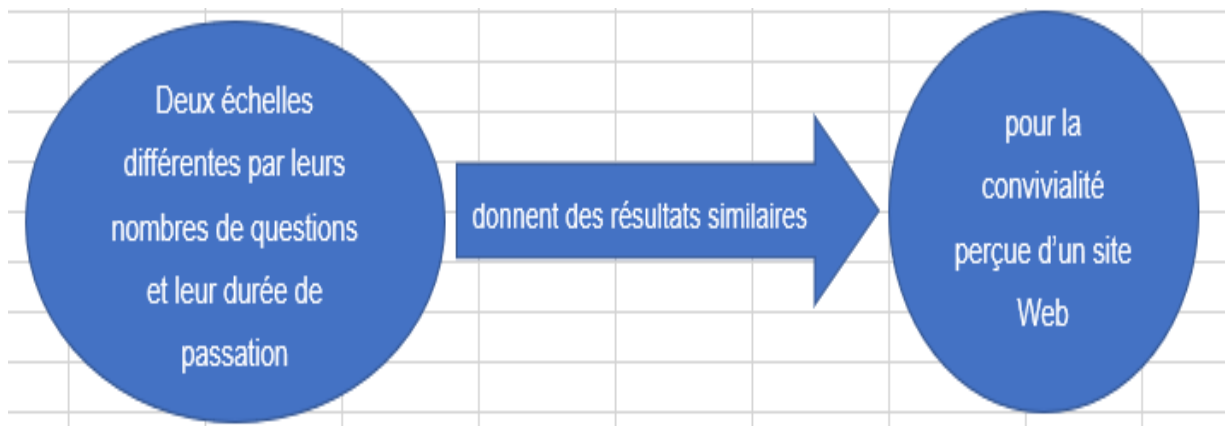
Tel qu'illustré précédemment, les échelles d'utilisabilité peuvent être soumises à des biais de passation (effet de halo, tendance à l'acquiescement). Nous rappelons que l'effet de halo précise que bien souvent, les premières réponses apportées par un participant ou utilisateur influencent ses réponses ultérieures. De la même manière, il pourrait être possible que les questionnaires trop longs soient source d'ennui et que les utilisateurs soient tentés de répondre plus rapidement afin de se débarrasser de la tâche [16].

L'auteur Kraig Finstad nous révèle toutefois dans son article « *The Usability Metric for User Experience* » [1] que toutes les échelles de notre étude sont valides. Mais peut-on penser que le fait qu'un questionnaire soit plus long qu'un autre à administrer va donner des résultats tout aussi similaires ?

Est-ce possible, par exemple, que des utilisateurs vont sembler apprécier davantage un site Web s'ils répondent à un long questionnaire par rapport aux utilisateurs qui répondront à un court questionnaire ?

La question de recherche est donc : **Le score final de l'échelle de mesure de l'utilisabilité DEEP (19 questions) a-t-il une variation de moins de 5 % par rapport au score final de l'échelle UMUX (4 questions) dans le cadre de la mesure de la satisfaction d'une interaction sur un site Web transactionnel ?**

Autrement dit, cette recherche vise à mesurer et vérifier si l'échelle d'utilisabilité UMUX, très courte (4 items), offre des résultats corrélés à une autre échelle, DEEP, très longue (19 items) étant donné que les échelles se valent dans le fond.



**Figure 3.1.1 : Schéma conceptuel de la recherche**

L'objectif de l'essai étant de comparer deux sites Web différents et de les évaluer avec une échelle très courte (UMUX) de quatre questions et une autre échelle très longue (DEEP) de dix-neuf questions.

En examinant la littérature actuelle, Kraig Finstad [1] l'auteur de l'article « *The Usability Metric for User Experience* » révèle que les résultats observés lors des tests entre UMUX et SUS sont corrélés, même si les échelles et les questions sont différentes.

Les auteurs Diefenbach et al. (1993) [20] ont toutefois constaté que les échelles à sept points surpassaient les échelles à cinq points en termes de fiabilité, de précision et de facilité d'utilisation, tandis que l'analyse de Cox (1980) [29] des échelles de Likert a révélé que le nombre optimal d'alternatives était de sept. Des tests seront donc effectués pour vérifier si les échelles UMUX et DEEP sont corrélées, sachant que UMUX utilise notamment une échelle Likert à sept points et DEEP une échelle à cinq points. D'où notre hypothèse que : **oui, les échelles d'utilisabilité UMUX et DEEP offrent des résultats similaires avec une variation de moins de 5 %.**

Advenant que les résultats de notre expérimentation ne soient pas corrélés (c'est-à-dire avec une variation dépassant 5 %), nous vérifierons si les résultats sont davantage positifs (les utilisateurs disant avoir apprécié plus positivement les sites) lors du test avec l'échelle DEEP, auquel cas les effets de halo ou de tendance à l'acquiescement pourraient expliquer la différence.

### **3.1.2 Limites**

Pour bien mener notre étude, les utilisateurs seront soumis à la réalisation de tâches sur deux sites Web distincts, un premier groupe évaluant le site avec le long questionnaire (DEEP) et un deuxième groupe avec un court questionnaire (UMUX). Ceci permettra d'obtenir des comparaisons entre les échelles sur plusieurs sites susceptibles d'obtenir naturellement des résultats différents. La méthodologie sera présentée au chapitre suivant.

### **3.1.3 Type de recherche**

Le but de notre essai étant de valider ou de réfuter des résultats quantifiés, le type de recherche quantitative semble ici être approprié. Les données collectées à l'aide des questionnaires utilisent des éléments de recherche quantitative. Les indicateurs de mesure utilisés font ainsi appel aux critères d'échelle de Likert.

## **Chapitre 4**

### **Approche proposée**

Pour faire face à la concurrence et aux nouvelles attentes des utilisateurs, les entreprises ont tout intérêt à évaluer l'utilisabilité de leur site Web. Les échelles de mesure de l'utilisabilité deviennent des outils incontournables que les grandes entreprises souhaiteront s'approprier. Les auteurs Thierry Baccino, Catherie Bellino et Teresa Colombi nous révèlent dans leur ouvrage [13] (p.27) qu'« évaluer est probablement l'activité-clé de l'ergonomie car elle va lui permettre de repérer les erreurs, les redondances ou les situations illogiques qu'il faudra peu à peu régler pour améliorer l'utilisabilité des interfaces ».

#### **4.1 Échantillon**

La population cible de cette étude est constituée d'étudiants du Centre de formation en technologies de l'information (CeFTI) ainsi que de contacts variés sollicités pour participer à l'étude. Des données sociodémographiques (âge, sexe) seront récoltées afin de valider la représentativité de l'échantillon.

#### **4.2 Description de l'approche**

Afin de s'adapter aux besoins des utilisateurs, le développement des applications Web positionne aujourd'hui les utilisateurs au centre de la conception. Évaluer un site Web du point de vue de son utilisateur devient primordial pour les entreprises. Les échelles d'utilisabilité et questionnaires standardisés apparaissent alors comme des outils importants que les entreprises souhaiteront s'approprier.



#### **4.2.1 Panel utilisateur**

Les utilisateurs étant des acteurs importants dans les processus d'évaluation des sites Web, les méthodes de conception et d'évaluation des sites Web ont été développées en grande partie pour favoriser l'implication des utilisateurs. Notre enquête a ainsi sollicité la participation de deux groupes de trente utilisateurs.

Carine Lallemand et Guillaume Gronier révèlent dans leur ouvrage [16] que « Dans le domaine académique, la limite basse se situe aux alentours de 30 participants, souvent décrite comme la taille d'échantillon minimale requise pour réaliser des statistiques inférentielles ». Les études de l'auteur Brent convergent également dans ce sens [24].

D'autres travaux ont montré que pour diminuer le risque de négliger certains graves problèmes d'utilisabilité, un nombre d'évaluateurs allant de 10 à 20 utilisateurs permettrait de détecter respectivement 80 % et 95 % des problèmes d'utilisabilité [30].

Nous allons donc constituer deux groupes de 30 utilisateurs chacun, un groupe par échelle d'évaluation. Ceci dans le but d'avoir des profils variés et représentatifs en termes de population (sexe, âge). Le nombre final d'évaluateurs sera précisé lorsque les résultats auront été filtrés (questionnaires incomplets, données extrêmes ou incohérentes).

Les tests se feront à distance de manière asynchrones [16], c'est-à-dire conduits par des systèmes automatiques qui gèrent les passations individuelles sans l'intervention directe de l'administrateur du test. Les utilisateurs seront invités à effectuer des tâches sur des sites de commerce en ligne.

#### **4.2.2 Les sites Web à évaluer**

Dans le cadre de l'essai, il s'agira d'évaluer deux sites Web. Les deux sites Web seront des sites de commerce en ligne. Cette démarche est semblable à celle de l'étude de Kraig Finstad [1] dans laquelle ont été étudiés deux sites Web, dont l'un était une application d'entreprise de services contractuels dont la convivialité avait été jugée médiocre et une application de téléconférence dotée d'une bonne convivialité.

Afin de recueillir les données de notre expérimentation, nous allons demander à nos utilisateurs d'effectuer une tâche sur deux sites Web différents : la tâche consistera à effectuer une

recherche de nouvelles chaussures chics sur les sites Web suivants : <https://www.aldoshoes.com/ca/fr> et <https://rubinoshoes.com/fr>. Plus précisément, nous allons demander aux utilisateurs de chaque groupe de trouver de nouvelles chaussures chics à leur convenance, de consulter les détails du produit retenu, de choisir la taille et de le mettre dans les paniers d'achat respectifs des sites Web en question.

**Tableau 4.2.2 : Sites Web évalués dans le cadre de l'essai**

Site Web à évaluer	Site Web à évaluer
<a href="https://www.aldoshoes.com/ca/fr">https://www.aldoshoes.com/ca/fr</a>	<a href="https://rubinoshoes.com/fr">https://rubinoshoes.com/fr</a>
Site de e-commerce	Site de e-commerce

## 4.3 Mise en œuvre

### 4.3.1 Les questionnaires de sondage

Les données collectées lors d'un sondage jouent un rôle de premier rang dans l'échantillonnage. Les résultats de ce sondage permettent ainsi d'identifier certaines caractéristiques de la population cible. La cueillette des données s'est faite via les questionnaires standardisées DEEP et UMUX.

Le questionnaire standardisé DEEP est un format composé de 19 questions catégorisées en six dimensions, à savoir : le contenu perçu, la structure perçue et l'architecture de l'information, la navigation perçue, l'effort cognitif perçu, la cohérence de la mise en page perçue et le guidage visuel perçu. Ce questionnaire est coté sur une échelle de Likert à cinq points. De cette manière, il sera aisé de calculer les scores moyens de convivialité perçue à trois niveaux : un score global de convivialité perçue, un score pour chaque dimension de conception et un score pour chaque élément.

En examinant la littérature actuelle, trois types de questionnaires ont été développés pour mesurer la convivialité perçue : les questionnaires d'utilisabilité perçue universels, les questionnaires d'utilisabilité perçue pour les sites Web et les questionnaires d'utilisabilité perçue pour les applications mobiles. Le questionnaire DEEP s'appuie quant à lui sur la combinaison d'éléments extraits et adaptés de plusieurs autres échelles, à savoir : PHUE, CSUQ, QUIS, SUS, PUTQ, USE et WAMMI (voir figure 4.3.1).

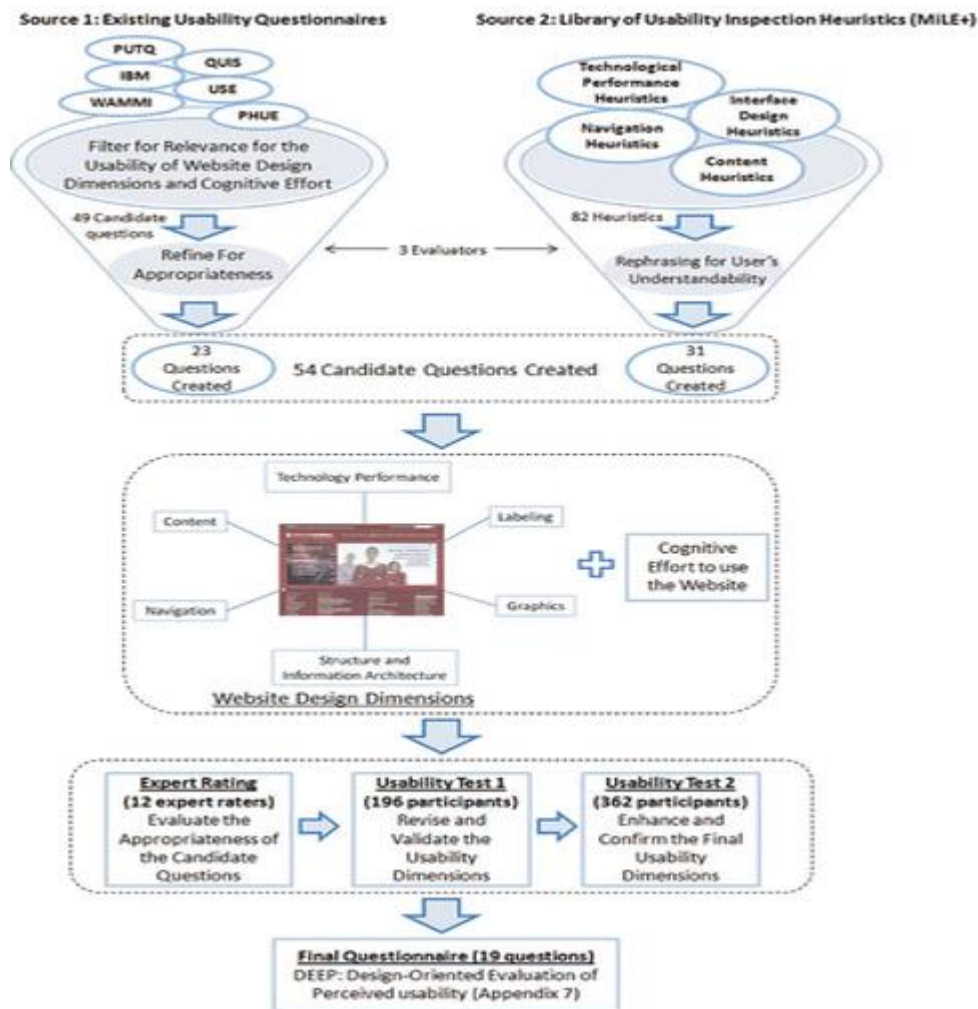


Figure 4.3.1 : Procédure utilisée pour développer et valider le questionnaire DEEP

Source : *International Journal of Human-Computer Interaction* [18]

Pour chaque élément d'un questionnaire, il faut initialement faire la moyenne des scores donnés par les utilisateurs. Carine Lallemand et Guillaume Gronier [16] nous révèlent que des moyennes peuvent également être calculées pour chacune des six catégories du questionnaire DEEP, afin d'établir un profil général des problèmes d'utilisabilité du système. Il est important de noter que les items 12 et 15 du questionnaire DEEP sont inversés : pour que les scores soient cohérents avec ceux des autres items, il faut les recoder pour qu'ils correspondent aux scores des items non inversés (tableau 4.3.1).

**Tableau 4.3.1 : Correspondance des scores aux items inversés par rapport aux items normaux [16]**

<b>Item normal</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>item inversé</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>

Toutes les moyennes inférieures ou égales à trois peuvent être considérées comme valeurs seuils à partir desquelles un problème d'utilisabilité doit être corrigé.

Le questionnaire standardisé UMUX quant à lui est une échelle à quatre éléments cotés sur Likert à sept points. Cette échelle est également utilisée pour l'évaluation subjective de la convivialité perçue d'une application ou d'un site Web.

### **4.3.2 Les mesures**

Le but ici sera d'assigner une tâche à chaque participant, répartis en deux groupes.

Les participants au sondage recevront un courriel avec des consignes précises. Ces consignes donnent les indications sur la procédure pour compléter les échelles de l'utilisabilité. Il est question de s'approprier les sites Web et de les évaluer avec les échelles proposées. Le contexte proposé est celui-ci : « *Imaginez que vous êtes invité(e) à une soirée chic et que vous devez acheter de nouvelles chaussures. Rendez-vous sur le site Web <https://rubinshoes.com/fr> (ou [https://www.aldoshoes.com/ca/fr selon le groupe](https://www.aldoshoes.com/ca/fr_selon_le_groupe)), afin de trouver une paire de chaussures à votre goût. Consultez les détails du produit retenu,*

*choisissez la taille, puis voyez comment vous pouvez l'acheter (sans compléter l'achat). Par la suite, commentez votre expérience en répondant au questionnaire suivant : [Questionnaire] ».*

Après avoir effectué les différentes tâches sur les deux sites Web, les participants sont invités à remplir les questionnaires, soit DEEP ou UMUX, afin d'évaluer leur expérience sur chaque site en question. L'évaluation avec l'échelle de mesure DEEP a une durée anticipée de dix à quinze minutes, tandis que l'évaluation avec l'échelle UMUX dure moins de trois minutes.

### **4.3.3 Approche de validation des résultats**

Les auteurs Fricker et Schonlau révèlent dans leur ouvrage [31] que les questionnaires en ligne obtiennent généralement un meilleur taux de réponse que le format papier. Il est généralement conseillé de proposer aux utilisateurs de répondre aux questionnaires à l'aide d'outils en ligne plutôt que par formulaire papier. Plusieurs avantages dérivent de ce type de mode, à savoir : économie du papier, recueil des scores facilement importable dans les logiciels de traitement statistique. La qualité de conception des questionnaires en format électronique préparés dans le cadre cet essai sera par ailleurs validée par le directeur académique, en tant qu'expert du domaine.

La collecte des données et les analyses statistiques ont été réalisées à l'aide de l'application Google Forms. Les données collectées par questionnaires pourront faire l'objet de plusieurs traitements statistiques, mais avant d'analyser les résultats, il sera nécessaire de faire un tri des données. Cela permettra de vérifier par exemple qu'aucun utilisateur n'a répondu de manière systématique, ou que des résultats incomplets ou incohérents soient considérés dans l'analyse.

## **4.4 Résultats attendus**

Avec une échelle de mesure de l'utilisabilité, les scores recueillis sont généralement présentés de manière graphique. Les auteurs Carine Lallemand et Guillaume Gronier révèlent par ailleurs dans leur ouvrage [16] que les échelles plus longues, comme DEEP, recueillent des données plus fines sur les problèmes d'utilisabilité rencontrés par les utilisateurs.

### Exemple de représentation des scores pour le DEEP

Il est parfois utile de donner une représentation graphique des scores d'un test d'utilisabilité. Non seulement les représentations restituent en un coup d'œil les scores des items ou des catégories, mais elles facilitent aussi la comparaison entre plusieurs évaluations (figure 23-6).

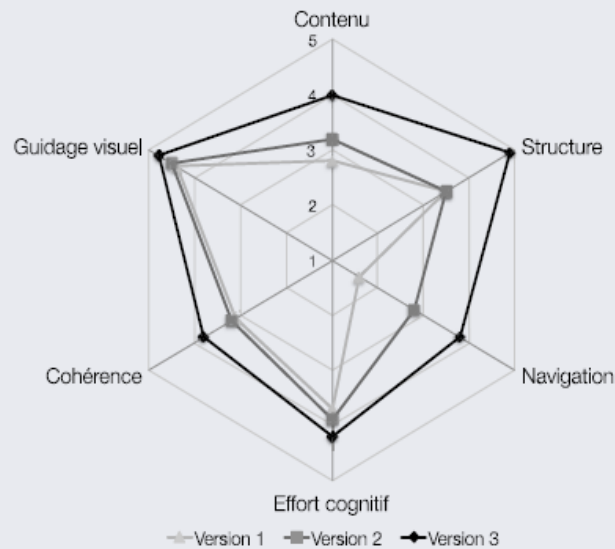


Figure 23-6

Exemple de représentation graphique en radar des scores du DEEP par catégorie, pour trois versions d'un système

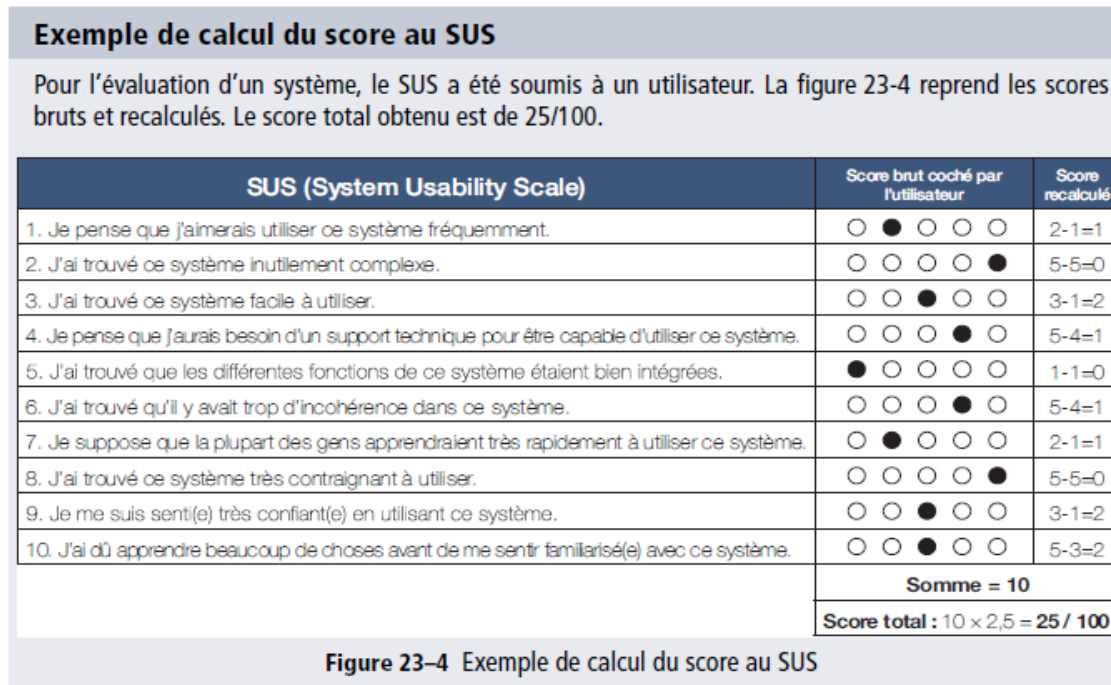
### Figure 4.4 : Représentation des scores pour l'échelle DEEP, sous forme graphique

Source : Méthodes UX de design [16] (p. 362)

L'échelle SUS utilise un système de cotation particulier qui permet de calculer les scores en quatre étapes, c'est-à-dire [16]:

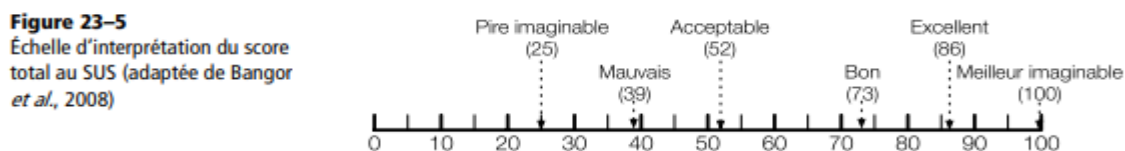
- Pour les éléments impairs, nous devons soustraire un point au score brut coché par l'utilisateur ;
- Pour les éléments pairs, nous calculerons cinq points moins le score brut coché par l'utilisateur ;
- Nous allons faire la somme des scores ;
- Et enfin, nous allons multiplier la somme des scores recalculés par 2,5. Nous obtiendrons ainsi le score total compris entre 0 et 100.

Une fois que nous aurons obtenu chacun des scores totaux de chaque utilisateur, nous calculerons la moyenne de tous ces scores obtenus, cela représentera l'évaluation générale de la tâche effectuée sur le site Web.



**Figure 4.5 : Exemple de calcul du score au SUS**

Source : Méthodes de design UX [16]



Dans son guide SUS, Jeff Sauro ([www.measuringu.com/products/SUSpack](http://www.measuringu.com/products/SUSpack)) (2011) a établi que le score moyen du SUS est de 68/100. Ce score de référence peut vous aider à situer votre système.

**Figure 4.6 : Échelle d'interprétation des scores total au SUS (adaptée de Bangor et al., 2008)**

Source : An Empirical Evaluation of the system Usability Scale [32]

La collecte des données avec l'outil UMUX, utilisée dans cet essai, doit être correctement recodée avec une méthode empruntée au SUS. L'article de Finstad [1] révèle que les éléments impairs sont notés comme [score - 1], et les éléments pairs sont notés comme [7 - score]. Comme avec le SUS, cela supprime la saisie positive / négative des éléments et permet un score minimum de zéro. Chaque élément UMUX individuel a une plage de 0 à 6 après décodage. Pour atteindre la parité avec la plage de 0 à 100 fournie par le SUS, le score UMUX du participant est la somme des quatre items divisés par 24, puis multipliée par 100. Ce calcul remplace la méthode antérieure de pondération des items par un multiplicateur de 2,5. Dans le prochain chapitre, les résultats collectés seront exploités et analysés. Ces scores des participants sont ensuite moyennés pour trouver un score UMUX moyen.



## Chapitre 5

### Analyse des résultats

Au chapitre précédent, nous avons détaillé l'approche utilisée pour répondre à la problématique énoncée au chapitre 3. Dans ce chapitre, nous allons présenter les résultats permettant de confirmer ou d'infirmer l'hypothèse de la recherche, en réponse à la question : « **Le score final de l'échelle de mesure de l'utilisabilité DEEP (19 questions) a-t-il une variation de moins de 5 % par rapport au score final de l'échelle UMUX (4 questions) dans le cadre de la mesure de la satisfaction d'une interaction sur un site Web transactionnel ?** »

#### 5.1 Analyse préliminaire

##### 5.1.1 Source des données

Pour mener à bien notre expérimentation, deux groupes constitués respectivement d'une trentaine d'évaluateurs ont été sollicités par courriel pour participer à une brève enquête qui avait pour objet d'évaluer deux sites web à l'aide d'échelles de mesure de l'utilisabilité. Pour y participer, chaque participant devait évaluer une tâche précise sur deux sites Web retenus. Par la suite, chaque participant devait évaluer l'expérience de la tâche effectuée avec l'une des échelles de mesure de l'utilisabilité proposée (voir figure 5.1.1 et figure 5.1.2).

Des 60 évaluateurs au total ayant été sollicités (30 par groupe), un certain nombre n'ont pas complété la totalité des questionnaires, ou ont fourni des réponses incohérentes ou extrêmes. Ces données ont été retirées des fichiers à analyser. Ainsi les données de 18 participants ont été retenues pour le groupe ayant évalué les sites avec DEEP, et les données de 21 participants ont été retenues pour le groupe ayant évalué les sites avec UMUX.

**Tableau 5.1.1 : différents groupes de participants au sondage**

Groupe A	Groupe B
18 participants	21 participants
Échelle DEEP	Échelle UMUX
<b>site Web à évaluer</b>	<b>site Web à évaluer</b>
Aldo	Aldo
Rubino	Rubino

Invitation à participer à un projet de recherche (CeFTI) 

**Armand Dayang** <dayang.armand@gmail.com>

À cci : sekehal.nadjib, cci : infokarim, cci : Anique.Thibault, cci : giovanna.diaz.gonzales, cci : Simon.Gingras4, cci : wuradclan, cci : David.Arsenault.Landry, cci : martine, cci : Drissa.Konate, c

Chers étudiantes et étudiants,

Dans le cadre de ma maîtrise en génie logiciel au CeFTI (UdeS), je sollicite votre participation à une brève enquête qui a pour objet d'évaluer les échelles de mesure d'utilisabilité.

Votre avis nous intéresse, et grâce à vos réponses nous obtiendrons des données importantes qui permettront de valider notre recherche.

Pour y participer, répondez aux **deux** questionnaires suivants :

Perception du site web Rubino :

<https://forms.gle/CY5ur5V9q2e887mk7>

Perception du site web Aldo :

<https://forms.gle/wFD7u15Ey7kvAMz89>

Temps requis estimé : 6-7 minutes par sondage.

Les données recueillies sont anonymes et ne seront pas associées ni à votre adresse courriel ni à votre ordinateur.

Le questionnaire est disponible du 5 au 12 octobre.

Merci de votre précieuse collaboration,

Très cordialement,

Armand Dayang Nana

Étudiant à la maîtrise, CeFTI, Université de Sherbrooke

[dayang.armand@gmail.com](mailto:dayang.armand@gmail.com)

**Figure 5.1.1 Courriel d'invitation pour administrer le questionnaire DEEP**

Invitation à participer à un projet de recherche (CeFTI) 

**Armand Dayang** <dayang.armand@gmail.com>

À cci : Lahoucine, cci : Destin, cci : Marcel, cci : boudrika, cci : evens.midy, cci : davidkapanga, cci : Sory, cci : Djamilia, cci : Mamadou.Bassir.Diallo, cci : Mehdi.Hazourli, cci : Jean.Christophe.Nguiamba,

Chers étudiantes et étudiants,

Dans le cadre de ma maîtrise en génie logiciel au CeFTI (UdeS), je sollicite votre participation à une brève enquête qui a pour objet d'évaluer les échelles de mesure d'utilisabilité.

Votre avis nous intéresse, et grâce à vos réponses nous obtiendrons des données importantes qui permettront de valider notre recherche.

Pour y participer, répondez aux **deux** questionnaires suivants :

Perception du site web Rubino :

<https://forms.gle/S8qHwTPtztzjSE2E7A>

Perception du site web Aldo :

<https://forms.gle/gd8RxxvvgwkcCS8LCA>

Temps requis estimé : 2-3 minutes par sondage.

Les données recueillies sont anonymes et ne seront pas associées ni à votre adresse courriel ni à votre ordinateur.

Le questionnaire est disponible du 5 au 12 Octobre.

Merci de votre précieuse collaboration,

Très cordialement,

Armand Dayang Nana

Étudiant à la maîtrise, CeFTI, Université de Sherbrooke

[dayang.armand@gmail.com](mailto:dayang.armand@gmail.com)

## Figure 5.1.2 : Courriel d'invitation pour administrer le questionnaire UMUX

### 5.1.2 Analyse des données : cas de l'échelle DEEP

#### 5.1.2.1 Tâche à effectuer : évaluation avec l'échelle DEEP

Un groupe constitué d'une trentaine de participants a été invité à effectuer une tâche précise sur deux sites Web pour ensuite répondre aussitôt au questionnaire d'évaluation DEEP (voir figure 5.1.2.1.2). Pour rappel, les participants étaient invités à s'approprier les sites Web dans un contexte où ils devaient magasiner de nouvelles chaussures.

Pour un temps limité, profitez de la livraison gratuite sur toutes les commandes + des retours gratuits... toujours!

Mon Compte  CAD (FR) 

Femmes  Hommes  Enfants  Soldes 

**ALDO**

Rechercher   0 article

Hommes / Chaussures / Chaussures habillées / Lacées chics



#### À propos

Les richelieus hyper élégants conçus avec du cuir fin sont dotés d'un talon en bois et

Chaussure oxford

## Gerawen

125 \$

[Évaluer et écrire un avis](#)

Couleur / Noir



Taille

Tableau des tailles

7	7.5	8	9	9.5
10	10.5	11	11.5	

 Trouvez votre vraie taille 

Ajouter au panier

Disponibilité en magasin

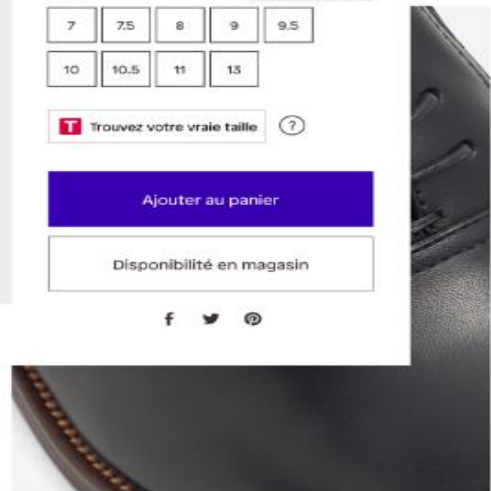


Figure 5.1.2.1.1 : Tâche à effectuer sur le site Web Aldo

## Questionnaire d'évaluation - Site web Aldo

Description du formulaire

---

**Scénario**

Imaginez que vous êtes invité(e) à une soirée chic et que vous devez acheter de nouvelles chaussures.

Rendez-vous sur le site web <https://www.aldoshoes.com/ca/fr> afin de trouver une paire de chaussures à votre goût.

Consultez les détails du produit retenu, choisissez la taille, puis voyez comment vous pouvez l'acheter (sans compléter l'achat).

Par la suite, commentez votre expérience en répondant au questionnaire suivant :

---

**CONTENU PERÇU**

Description (facultative)

---

1. Le libellé du texte était clair. \*

1 2 3 4 5

Pas du tout d'accord      Tout à fait d'accord

**Figure 5.1.2.1.2 : Scénario à réaliser et questionnaire DEEP**

### 5.1.2.2 Questionnaire d'évaluation : cas de l'échelle DEEP

Le questionnaire DEEP, constitué de 19 questions, a été soumis aux participants du sondage pour leur permettre d'évaluer leur expérience lors de la tâche sur le site Web. La figure 5.1.2.2 illustre la conception et la rédaction du questionnaire sur l'application Google Forms. Notons ici que les polarités des questions 12 et 15 ont été inversées pour éviter les effets de halo, comme le décrivent les auteurs Carine Lallemand et Guillaume Gronier [16].

12. Utiliser ce site web m'a fatigué. \*

Pas du tout d'accord  1  2  3  4  5 Tout à fait d'accord

13. J'ai appris à utiliser ce site web \*

Pas du tout d'accord  1  2  3  4  5 Tout à fait d'accord

COHÉRENCE DE LA MISE EN PAGE PERÇUE

Description (facultative)

14. La mise en page à travers tout le site web était cohérente. \*

Pas du tout d'accord  1  2  3  4  5 Tout à fait d'accord

15. J'ai remarqué des changements soudains de mise en page à travers le site \*

Pas du tout d'accord  1  2  3  4  5 Tout à fait d'accord

Figure 5.1.2.2 : Questionnaire d'évaluation de la méthode DEEP

### 5.1.2.3 Données collectées : cas de l'échelle DEEP

Une fois le questionnaire rempli et soumis par les utilisateurs, Google Forms a permis de collecter les données pour ensuite les analyser. Les items 12 et 15 étant inversés dans le questionnaire DEEP, il est important de recalculer les scores de ces deux items pour assurer la cohérence des résultats avec les autres items (voir figure 5.1.2.3).

Pour rappel, voici la correspondance des scores aux items inversés par rapport aux items normaux.

**Tableau 5.1.2.3 : Correspondance des scores aux items inversés par rapport aux items normaux [16]**

<b>Item normal</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>item inversé</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>

N	O	P	Q	R	S
12. Utiliser ce site web	Score inversé	13. J'ai appris à utiliser ce	14. La mise en page à tra	15. J'ai remarqué des cha	Score inversé
2	4	4	4	4	2
2	4	4	3	3	3
3	3	3	3	2	4
1	5	5	5	1	5
2	4	3	4	5	1
2	4	4	4	2	4
1	5	5	4	1	5
2	4	4	4	3	3
1	5	4	5	1	5
1	5	5	5	1	5
1	5	4	4	3	3
4	2	4	4	1	5
1	5	4	3	1	5
1	5	5	5	1	5
1	5	5	5	1	5
1	5	5	5	1	5
1	5	5	5	1	5
1	5	1	4	4	2

**Figure 5.1.2.3 : Recodage des scores 12 et 15 du questionnaire DEEP**

### 5.1.2.4 Calcul des scores : cas de l'échelle DEEP

Pour chaque élément du questionnaire de mesure DEEP, il faut d'abord faire la moyenne des scores donnés par les participants (voir figure 5.1.2.4).

=AVERAGE(C2:M2;O2;Q2;S2;W2)										
N	O	P	Q	R	S	T	U	V	W	X
12. Utiliser ce site web	m Score inversé	13. J'ai appris à utiliser ci	14. La mise en page à tra	15. J'ai remarqué des cha	Score inversé	16. La mise en page de ci	17. Les couleurs m'ont air	18. Les zones mises en è	19. J'ai appris à connaître	Score moyen
2	4	4	4	4	2	3	3	4	4	3,526315789
2	4	4	3	3	3	5	5	4	5	4,263157895
3	3	3	3	2	4	3	4	4	3	3,526315789
1	5	5	5	1	5	5	5	5	5	5
2	4	3	4	5	1	4	3	4	4	3,736842105
2	4	4	4	2	4	4	4	4	4	3,789473684
1	5	5	4	1	5	4	2	4	4	4,315789474
2	4	4	4	3	3	4	4	4	4	3,894736842
1	5	4	5	1	5	5	4	5	5	4,578947368
1	5	5	5	1	5	5	5	5	5	5
1	5	4	4	3	3	4	3	4	4	4,157894737
4	2	4	4	1	5	4	4	4	4	4,052631579
1	5	4	3	1	5	4	4	4	4	3,789473684
1	5	5	5	1	5	5	5	5	5	4,894736842
1	5	5	5	1	5	5	1	1	1	4,368421053
1	5	5	5	1	5	5	4	4	5	4,894736842
1	5	5	5	1	5	5	5	5	5	5
1	5	1	4	4	2	4	3	3	2	3,421052632
										4,233918129
										4,233918129

Figure 5.1.2.4.1 : Calcul des scores moyens du questionnaire DEEP

Formule :

$$\text{Moyenne des scores} = \frac{\text{Sommes des scores}}{\text{Nbre Total du questionnaire}}$$

Les résultats étant en score, il faut les convertir en pourcentage pour facilement vérifier leur corrélation avec les résultats d'autres échelles.

$$\text{Moyenne de score en pourcentage} = \frac{\text{Score moyen} \times 100}{5}$$



$= (X2 * 100) / 5$

M	V	W	X	AA
11. Utiliser ce site web s'e	18. Les zones mises en é	19. J'ai appris à connaître	Score moyen	Score moyen en %
4	4	4	3,526315789	70,52631579
4	4	5	4,263157895	85,26315789
3	4	3	3,526315789	70,52631579
5	5	5	5	100
4	4	4	3,736842105	74,73684211
4	4	4	3,789473684	75,78947368
5	4	4	4,315789474	86,31578947
4	4	4	3,894736842	77,89473684
5	5	5	4,578947368	91,57894737
5	5	5	5	100
4	4	4	4,157894737	83,15789474
5	4	4	4,052631579	81,05263158
3	4	4	3,789473684	75,78947368
5	5	5	4,894736842	97,89473684
5	1	1	4,368421053	87,36842105
5	4	5	4,894736842	97,89473684
5	5	5	5	100
4	3	2	3,421052632	68,42105263

Figure 5.1.2.4.2 : Calcul des scores moyens en pourcentage du questionnaire DEEP

## 5.1.3 Analyse des données : cas de l'échelle UMUX

### 5.1.3.1 Tâche à effectuer : cas de l'échelle UMUX

Un deuxième groupe constitué d'une trentaine de participants a été invité à effectuer une tâche équivalente au premier groupe, sur les mêmes deux sites Web mais pour ensuite répondre au questionnaire d'évaluation UMUX (voir figure 5.1.3.2). Pour rappel, les participants étaient invités à s'appropriier les sites Web dans un contexte où ils devaient magasiner de nouvelles chaussures.

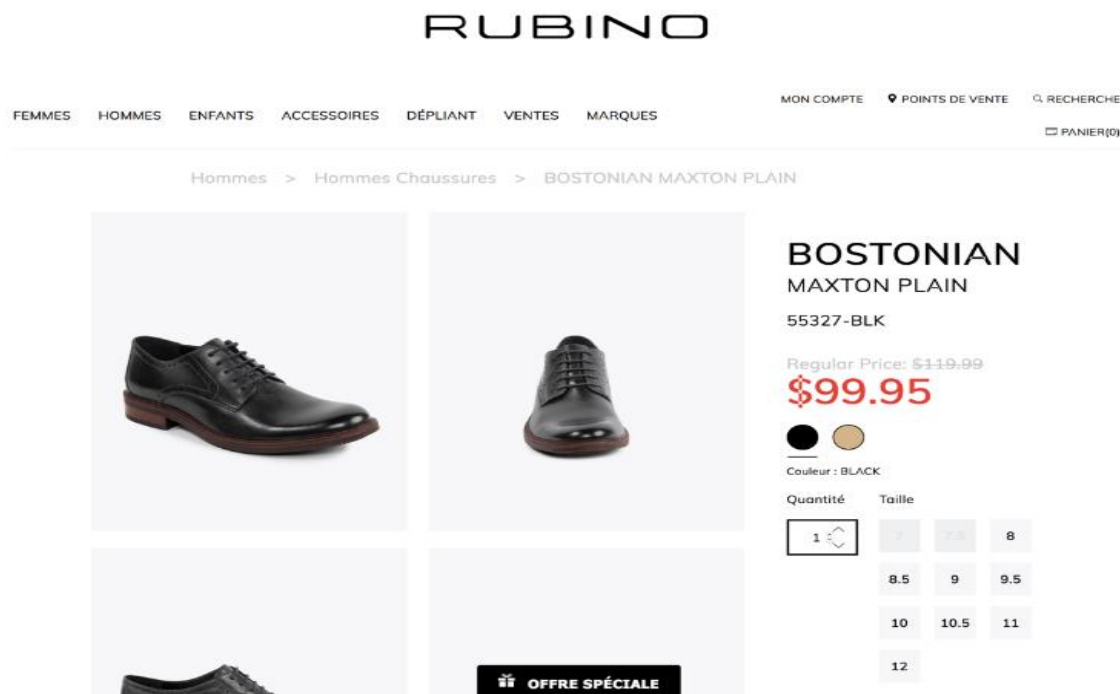


Figure 5.1.3.1 : Tâche à effectuer sur le site Web Rubino

### 5.1.3.2 Questionnaire d'évaluation : cas de l'échelle UMUX

Le questionnaire UMUX, constitué de quatre éléments, a été soumis aux participants pour leur permettre d'évaluer leur expérience lors de la tâche sur le site Web. La figure ci-dessous illustre la conception du questionnaire sur l'application Google Forms. (Voir figure 5.1.3.2)

Questions Réponses 22

## Questionnaire d'évaluation - Site web Rubino

Description du formulaire

### Scénario

Imaginez que vous êtes invité(e) à une soirée chic et que vous devez acheter de nouvelles chaussures.

Rendez-vous sur le site web <https://rubinoshoes.com/fr>, afin de trouver une paire de chaussures à votre goût.

Consultez les détails du produit retenu, choisissez la taille, puis voyez comment vous pouvez l'acheter (sans compléter l'achat).

Par la suite, commentez votre expérience en répondant au questionnaire suivant :

1. Les fonctionnalités de ce système sont conformes à mes \*

1 2 3 4 5 6 7

Pas du tout d'accord        Tout à fait d'accord

2. Utiliser ce système est une expérience \*

1 2 3 4 5 6 7

Pas du tout d'accord        Tout à fait d'accord

3. Ce système est facile à \*

1 2 3 4 5 6 7

Pas du tout d'accord        Tout à fait d'accord

Figure 5.1.3.2 : Questionnaire d'évaluation de la méthode UMUX

### 5.1.3.3 Données collectées : cas de l'échelle UMUX

Une fois le questionnaire rempli et soumis par les utilisateurs, Google Forms permet de collecter et analyser les données. Pour rappel, les éléments impairs sont calculés comme [score – 1], et les éléments pairs sont calculés comme [7 – score]. Comme avec le SUS, cela supprime la saisie positive / négative des éléments et permet un score minimum de zéro. Chaque élément UMUX individuel a une plage de 0 à 6 après recodage, ce qui donne un maximum préliminaire de 24 points à l'échelles de quatre éléments. Pour atteindre la parité avec la plage de 0 à 100 fournie par le SUS, le score UMUX du participant est la somme des quatre items divisés par 24, puis multipliée par 100 (voir figure 5.1.3.3).

	B	C	D	E	F	G	H	I	J	K
	1. Les fonctionnalités de	Score recalculé	2. Utiliser ce système est	Score recalculé	3. Ce système est facile	Score recalculé	4. Je dois passer trop de	Score recalculé	somme score	Score en %
	6	5	2	5	5	4	2	5	19	79,16666667
	6	5	2	5	6	5	6	1	16	66,66666667
	7	6	1	6	7	6	2	5	23	95,83333333
	6	5	1	6	6	5	1	6	22	91,66666667
	5	4	1	6	4	3	2	5	18	75
	5	4	3	4	6	5	3	4	17	70,83333333
	7	6	1	6	7	6	1	6	24	100
	4	3	4	3	4	3	5	2	11	45,83333333
	7	6	1	6	7	6	1	6	24	100
	5	4	4	3	4	3	4	3	13	54,16666667
	5	4	2	5	5	4	2	5	18	75
	7	6	1	6	7	6	1	6	24	100
	7	6	1	6	7	6	1	6	24	100
	5	4	2	5	5	4	4	3	16	66,66666667
	6	5	1	6	6	5	6	1	17	70,83333333
	4	3	3	4	5	4	3	4	15	62,5
	7	6	1	6	6	5	1	6	23	95,83333333
	7	6	1	6	7	6	1	6	24	100
	6	5	2	5	6	5	2	5	20	83,33333333
	7	6	2	5	5	4	1	6	21	87,5
	7	6	1	6	7	6	2	5	23	95,83333333

Figure 5.1.3.3 : Données collectés et calculés de la méthode UMUX

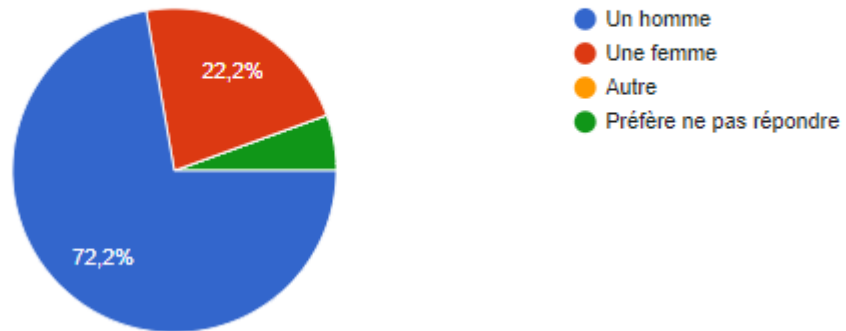
### 5.1.4 Analyse comparative des résultats

Tableau 5.1.4.1: Tableau comparatif des résultats des différents scores moyens en pourcentage avec le questionnaire UMUX avec le site Web Aldo

1. Les fonctionnalités de ce système sont conformes à mes exigences.	2. Utiliser ce système est une expérience frustrante.	3. Ce système est facile à utiliser.	4. Je dois passer trop de temps à corriger les choses sur ce système.	Somme scores	Scores-en %	Vous êtes?	Quel âge avez-vous?	Site Web - ALDO
5	5	4	5	19	79,1666667	Un homme	35 à 44 ans	UMUX
5	5	5	1	16	66,6666667	Un homme	45 à 54 ans	
6	6	6	5	23	95,8333333	Un homme	35 à 44 ans	
5	6	5	6	22	91,6666667	Un homme	35 à 44 ans	
4	6	3	5	18	75	Un homme	45 à 54 ans	
4	4	5	4	17	70,8333333	Un homme	35 à 44 ans	
6	6	6	6	24	100	Un homme	35 à 44 ans	
6	6	6	6	24	100	Une femme	25 à 34 ans	
4	5	4	5	18	75	Un homme	25 à 34 ans	
6	6	6	6	24	100	Une femme	Préfère ne pas répondre	
6	6	6	6	24	100	Une femme	45 à 54 ans	
5	6	5	1	17	70,8333333	Une femme	25 à 34 ans	
3	4	4	4	15	62,5	Une femme	Préfère ne pas répondre	
6	6	5	6	23	95,8333333	Un homme	Préfère ne pas répondre	
6	6	6	6	24	100	Préfère ne pas répondre	Préfère ne pas répondre	
5	5	5	5	20	83,3333333	Un homme	35 à 44 ans	
6	5	4	6	21	87,5	Un homme	35 à 44 ans	
6	6	6	5	23	95,8333333	Un homme	25 à 34 ans	

Vous êtes?

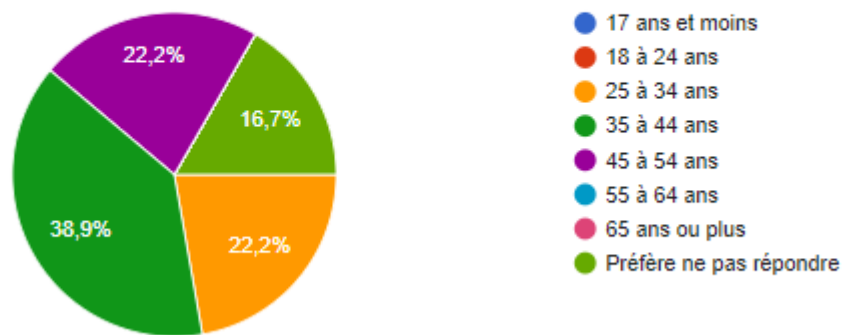
18 réponses



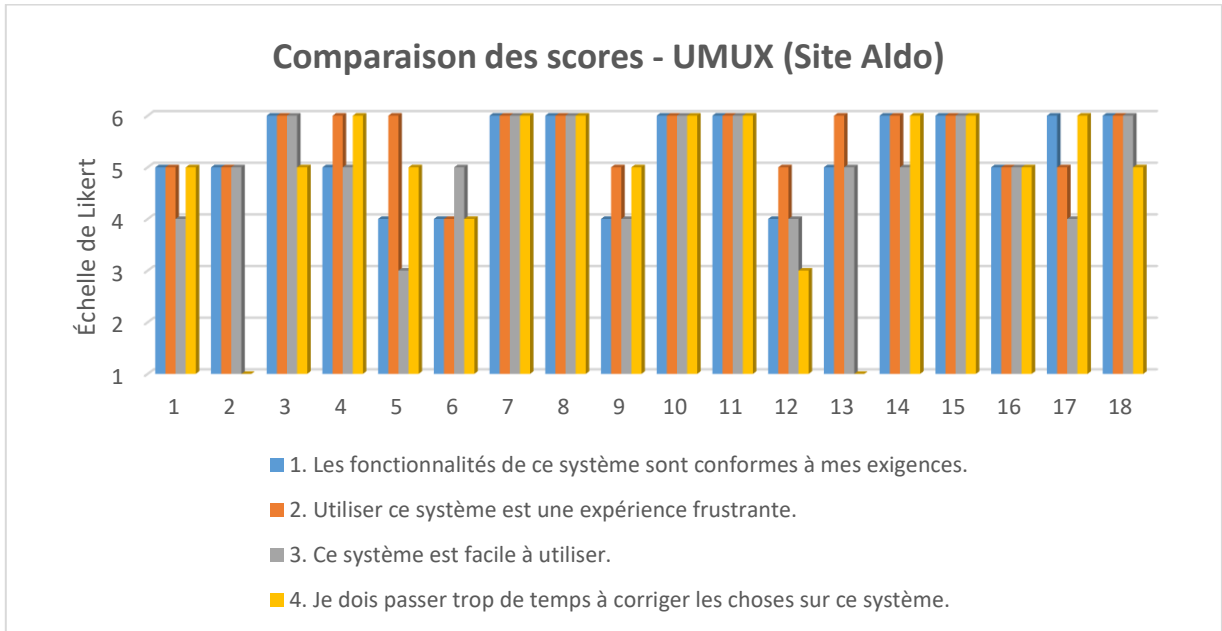
**Figure 5.1.4.1.1 : Représentation par sexe pour l'évaluation du site Web Aldo avec le questionnaire UMUX**

Quel âge avez-vous?

18 réponses



**Figure 5.1.4.1.2 : Représentation par tranche d'âge pour l'évaluation du site Web Aldo avec le questionnaire UMUX**



**Figure 5.1.4.1.3 : Diagramme des scores par utilisateur de l'échelle UMUX du site Web Aldo**

**Tableau 5.1.4.2 : Tableau comparatif des résultats des différents scores moyens en pourcentage  
avec le questionnaire DEEP sur le site Web Aldo**

16. La mise en page de chaque section du site web était cohérente.	17. Les couleurs m'ont aidé à distinguer les différentes sections du site web.	18. Les zones mises en évidence d'une page m'ont aidé à repérer l'information dont j'avais besoin.	19. J'ai appris à connaître le contenu d'une page en parcourant les zones mises en évidence.	Score moyen	Vous êtes?	Quel âge avez-vous?	Score moyen en %	Site Web - Aldo	
3	3	4	4	3,526315789	Un homme	35 à 44 ans	70,52631579	<b>DEEP</b>	
5	5	4	5	4,263157895	Un homme	55 à 64 ans	85,26315789		
3	4	4	3	3,526315789	Un homme	45 à 54 ans	70,52631579		
5	5	5	5	5	Un homme	Préfère ne pas répondre	100		
4	3	4	4	3,736842105	Un homme	35 à 44 ans	74,73684211		
4	4	4	4	3,789473684	Une femme	Préfère ne pas répondre	75,78947368		
4	2	4	4	4,315789474	Une femme	25 à 34 ans	86,31578947		
4	4	4	4	3,894736842	Un homme	35 à 44 ans	77,89473684		
5	4	5	5	4,578947368	Une femme	45 à 54 ans	91,57894737		
5	5	5	5	5	Un homme	35 à 44 ans	100		
4	3	4	4	4,157894737	Un homme	25 à 34 ans	83,15789474		
4	4	4	4	4,052631579	Une femme	55 à 64 ans	81,05263158		
4	4	4	4	3,789473684	Une femme	25 à 34 ans	75,78947368		
5	5	5	5	4,894736842	Un homme	25 à 34 ans	97,89473684		
5	1	1	1	4,368421053	Une femme	65 ans ou plus	87,36842105		
5	4	4	5	4,894736842	Un homme	55 à 64 ans	97,89473684		
5	5	5	5	5	Une femme	25 à 34 ans	100		
4	3	3	2	3,421052632	Une femme	25 à 34 ans	68,42105263		



Vous êtes?

18 réponses

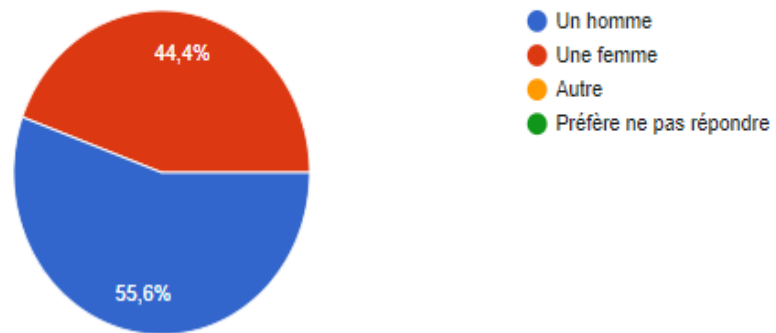


Figure 5.1.4.2.1 : Représentation par sexe pour l'évaluation du site Web Aldo avec le questionnaire DEEP

Quel âge avez-vous?

18 réponses

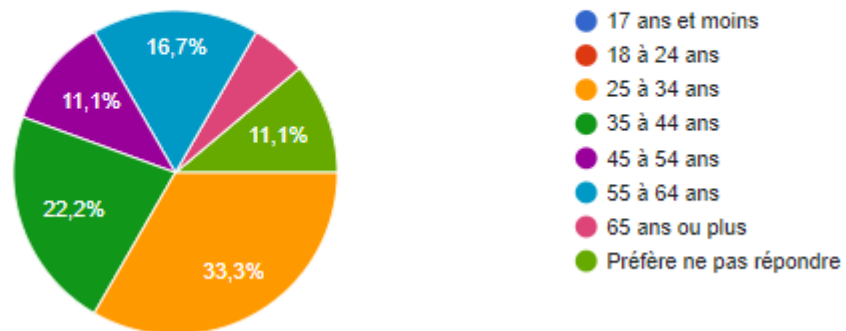
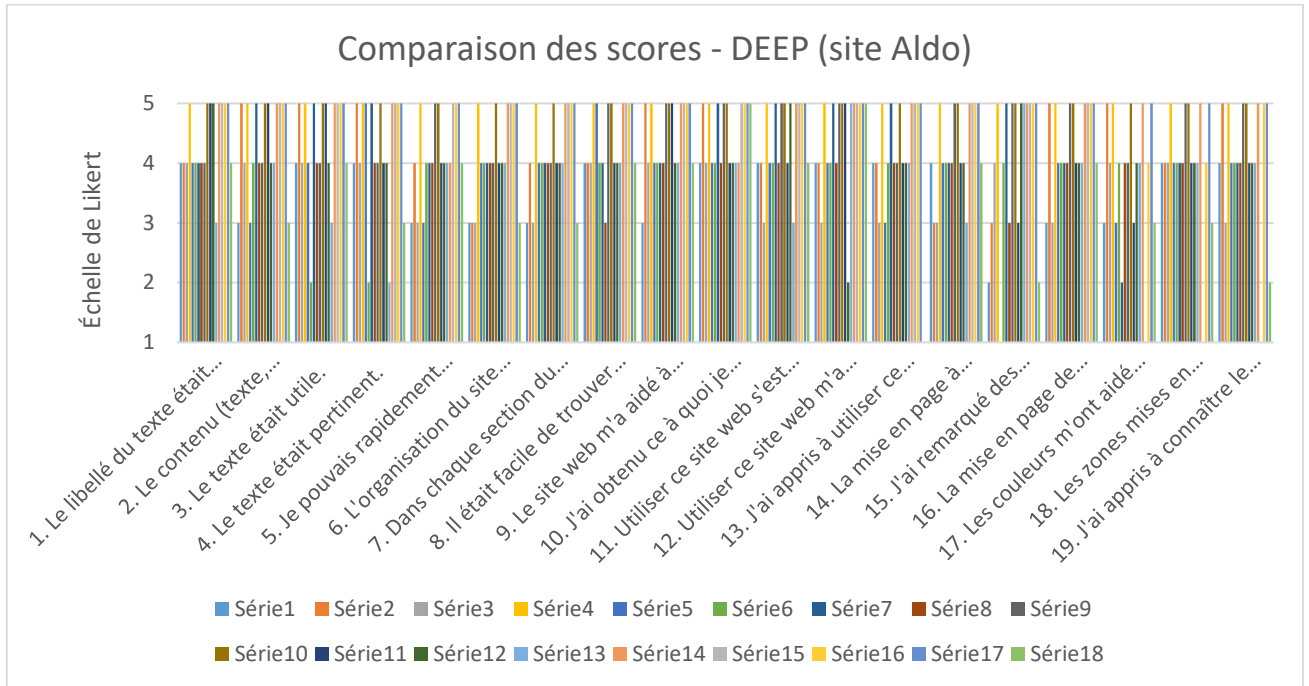


Figure 5.1.4.2.2 : Représentation par tranche d'âge pour l'évaluation du site Web Aldo avec le questionnaire DEEP



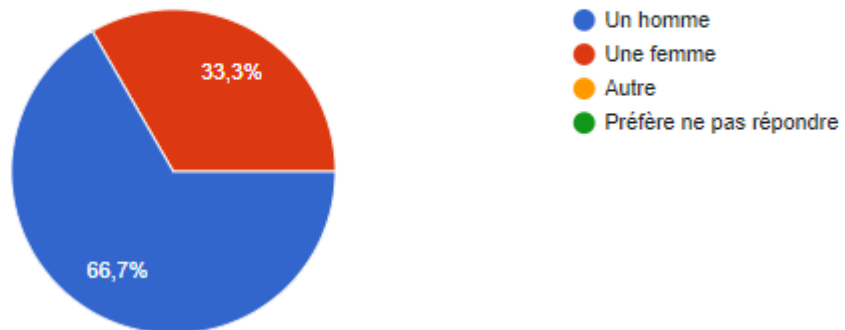
**Figure 5.1.4.2.3 : Diagramme des scores par utilisateur de l'échelle DEEP du site Web Aldo**

**Tableau 5.1.4.3 : Tableau comparatif des résultats des différents scores moyens en pourcentage  
avec le questionnaire DEEP sur le site Web Rubino**

16. La mise en page de chaque section du site web était cohérente.	17. Les couleurs m'ont aidé à distinguer les différentes sections du site web.	18. Les zones mises en évidence d'une page m'ont aidé à repérer l'information dont j'avais besoin.	19. J'ai appris à connaître le contenu d'une page en parcourant les zones mises en évidence.	Score moyen	Vous êtes?	Quel âge avez-vous?	Score moyen en %	Site Web - Rubino
4	4	5	5	4,57894737	Un homme	35 à 44 ans	91,5789474	<b>DEEP</b>
3	2	2	3	3,73684211	Un homme	55 à 64 ans	74,7368421	
3	4	3	3	4	Un homme	45 à 54 ans	80	
5	4	5	5	4,42105263	Un homme	35 à 44 ans	88,4210526	
4	2	4	4	3,68421053	Un homme	17 ans et moins	73,6842105	
3	1	2	3	2,84210526	Une femme	Préfère ne pas répondre	56,8421053	
1	1	3	1	2,42105263	Une femme	25 à 34 ans	48,4210526	
4	3	4	3	3,78947368	Un homme	35 à 44 ans	75,7894737	
4	2	3	3	3	Une femme	45 à 54 ans	60	
5	5	5	4	3,89473684	Un homme	35 à 44 ans	77,8947368	
4	2	3	3	3,63157895	Un homme	35 à 44 ans	72,6315789	
4	4	4	4	4,47368421	Un homme	35 à 44 ans	89,4736842	
3	3	4	3	3,31578947	Un homme	25 à 34 ans	66,3157895	
5	3	3	3	4,26315789	Un homme	25 à 34 ans	85,2631579	
4	4	4	4	4,42105263	Un homme	25 à 34 ans	88,4210526	
4	3	3	3	3,36842105	Une femme	25 à 34 ans	67,3684211	
5	3	4	4	4,68421053	Un homme	25 à 34 ans	93,6842105	
4	1	1	1	3,47368421	Une femme	65 ans ou plus	69,4736842	
2	2	2	2	1,73684211	Un homme	55 à 64 ans	34,7368421	
2	2	2	2	2,05263158	Une femme	25 à 34 ans	41,0526316	
5	4	5	2	4,52631579	Une femme	25 à 34 ans	90,5263158	

Vous êtes?

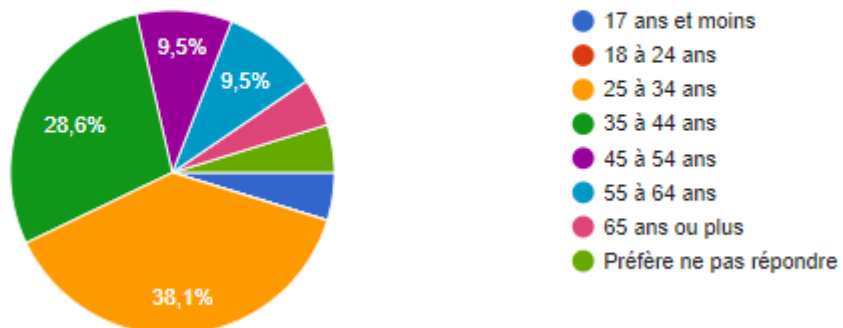
21 réponses



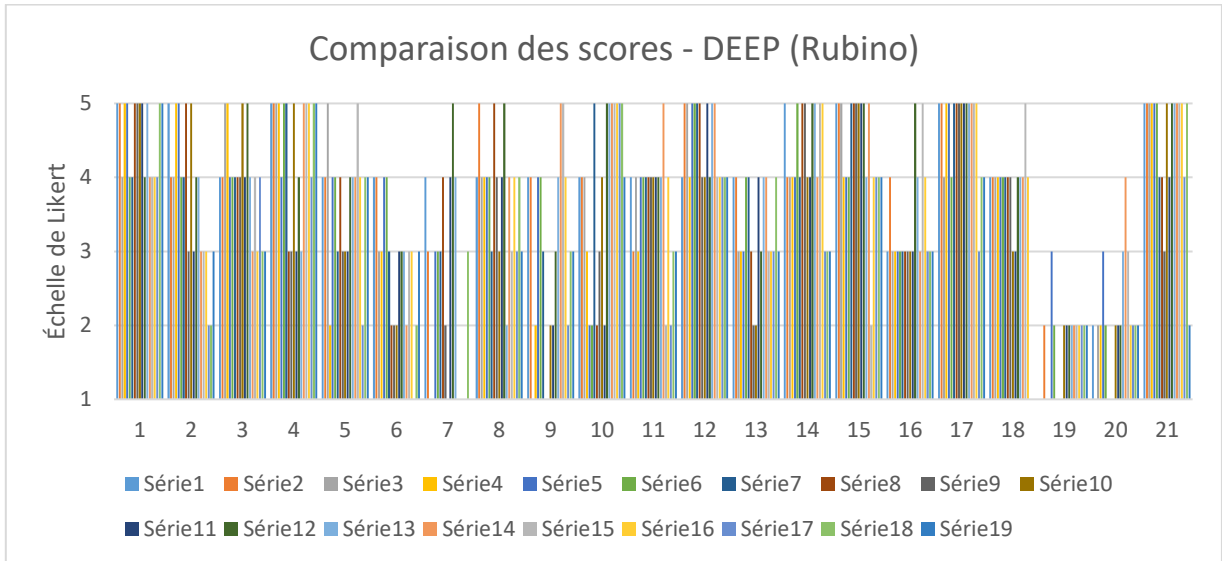
**Figure 5.1.4.3.1 : Représentation par sexe pour l'évaluation du site Web Rubino avec le questionnaire DEEP**

Quel âge avez-vous?

21 réponses



**Figure 5.1.4.3.2 : Représentation par tranche d'âge pour l'évaluation du site Web Rubino avec le questionnaire DEEP**



**Figure 5.1.4.3.3 : Diagramme des scores par utilisateur de l'échelle DEEP du site Web Rubino**

**Tableau 5.1.4.4 : Tableau comparatif des résultats des différents scores moyens en pourcentage  
avec le questionnaire UMUX sur le site Web Rubino**

1. Les fonctionnalités de ce système sont conformes à mes exigences.	2. Utiliser ce système est une expérience frustrante.	3. Ce système est facile à utiliser.	4. Je dois passer trop de temps à corriger les choses sur ce système.	Somme score	Scores-en %	Vous êtes?	Quel âge avez-vous?	Site Web - Rubino
4	1	5	1	11	45,8333333	Un homme	35 à 44 ans	<b>UMUX</b>
6	6	6	1	19	79,1666667	Un homme	45 à 54 ans	
5	5	5	5	20	83,3333333	Un homme	35 à 44 ans	
6	6	6	5	23	95,8333333	Un homme	35 à 44 ans	
4	6	5	5	20	83,3333333	Un homme	35 à 44 ans	
5	5	5	5	20	83,3333333	Un homme	45 à 54 ans	
3	3	3	2	11	45,8333333	Un homme	35 à 44 ans	
6	6	6	6	24	100	Un homme	35 à 44 ans	
4	5	5	5	19	79,1666667	Un homme	35 à 44 ans	
4	4	5	2	15	62,5	Une femme	25 à 34 ans	
4	3	3	3	13	54,1666667	Un homme	35 à 44 ans	
4	4	4	4	16	66,6666667	Un homme	25 à 34 ans	
5	5	5	5	20	83,3333333	Une femme	Préfère ne pas répondre	
5	6	6	6	23	95,8333333	Un homme	25 à 34 ans	
4	5	3	5	17	70,8333333	Une femme	25 à 34 ans	
3	4	2	2	11	45,8333333	Une femme	Préfère ne pas répondre	
5	6	6	6	23	95,8333333	Un homme	Préfère ne pas répondre	
6	6	6	6	24	100	Préfère ne pas répondre	Préfère ne pas répondre	
5	5	5	6	21	87,5	Un homme	35 à 44 ans	
5	5	5	1	16	66,6666667	Un homme	35 à 44 ans	
4	5	5	3	17	70,8333333	Un homme	25 à 34 ans	

Vous êtes?

21 réponses

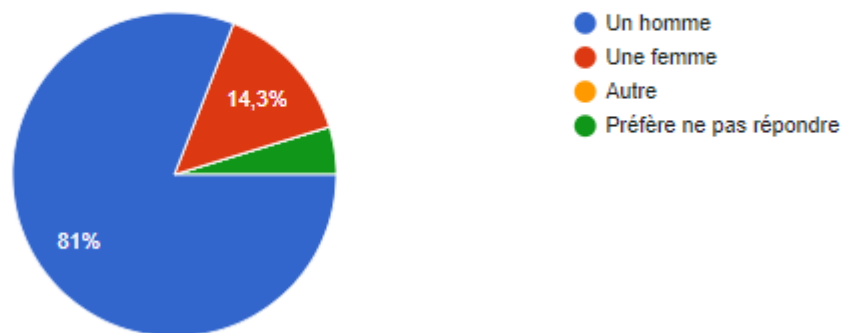


Figure 5.1.4.4.1 : Représentation par sexe pour l'évaluation du site Web Robino avec le questionnaire UMUX

Quel âge avez-vous?

21 réponses

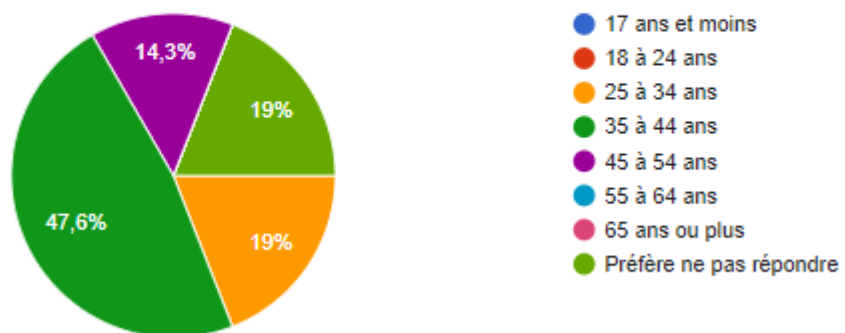
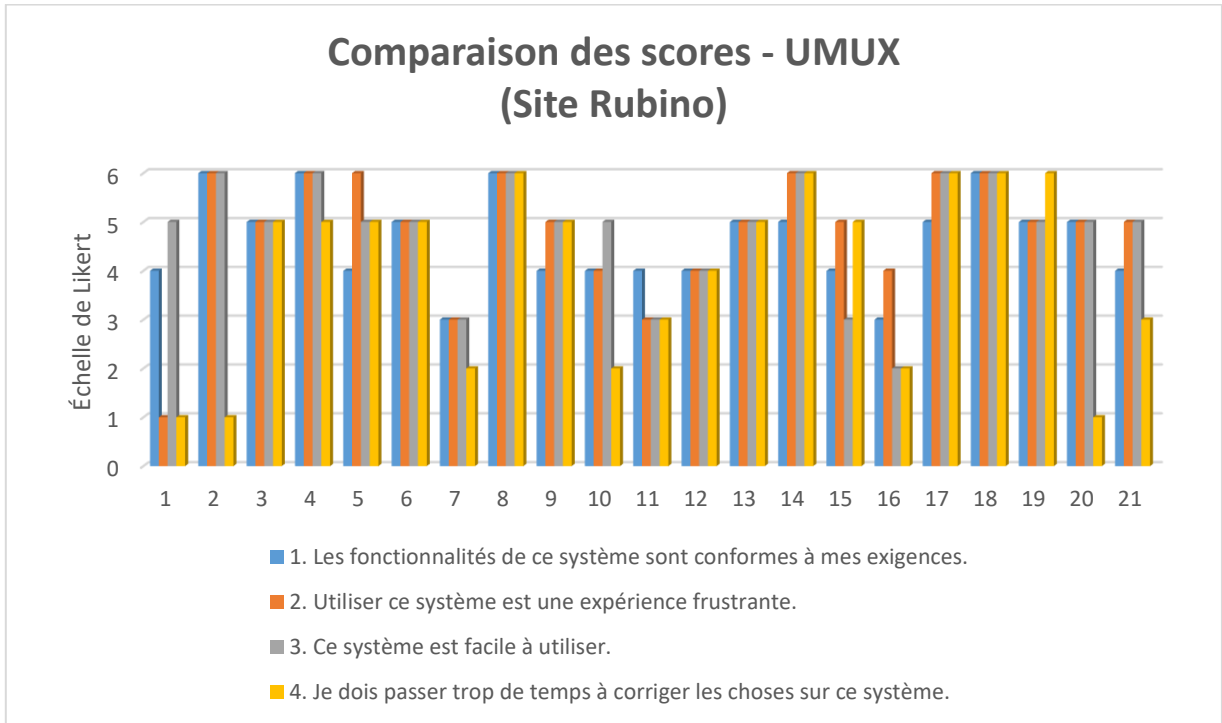


Figure 5.1.4.4.2: Représentation par tranche d'âge pour l'évaluation du site Web Rubino avec le questionnaire UMUX



**Figure 5.1.4.4.3: Diagramme des scores par utilisateur de l'échelle UMUX du site Web Rubino**



## 5.2 Retour sur les hypothèses

Kraig Finstad [1], auteur de « *The Usability Metric for User Experience* », révèle que les résultats observés lors des tests entre UMUX et SUS sont corrélés, même si les échelles et les questions sont différentes. D'où l'hypothèse de départ qui propose que les résultats de l'échelle de mesure de l'utilisabilité DEEP (19 questions) soient similaires à l'échelle UMUX (quatre questions) dans le cadre de la mesure de la satisfaction d'une interaction sur un site Web transactionnel.

Pour vérifier notre hypothèse de départ, c'est-à-dire la confirmer ou la rejeter, une comparaison a été faite d'une part, entre les scores moyens collectés lors des évaluations des sites Web Aldo et Rubino avec l'échelle DEEP et d'autre part, grâce aux données collectées lors des évaluations des sites Web d'Aldo et Rubino avec l'échelle UMUX.

Selon les résultats observés, les scores des deux échelles sont consistants de part et d'autre. Le site le plus apprécié (Aldo) est mesuré avec les deux échelles, et le site le moins apprécié (Rubino) est mesuré avec les deux échelles (Tableau 5.3.1.3. p.71). L'échelle de mesure de l'utilisabilité DEEP semble toutefois illustrer un plus grand écart entre les différents sites Web évalués, et propose des moyennes d'appréciation plus basses que les sites évalués avec UMUX. La différence entre les résultats obtenus avec UMUX et ceux obtenus avec DEEP est toutefois inférieure à 5 %, tant pour l'évaluation du site Aldo que du site Rubino. Les données détaillées sont présentées à la section suivante.

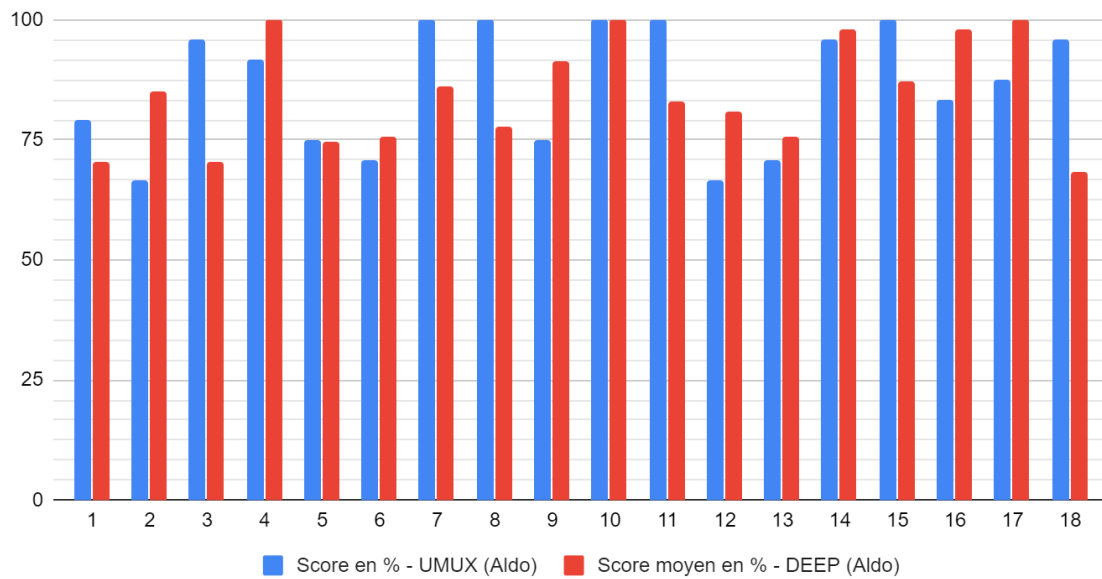
## 5.3 Résultats obtenus

### 5.3.1 Tableaux des données

**Tableau 5.3.1.1 : Tableau comparatif des différents scores moyens en pourcentage des échelles UMUX et DEEP obtenus sur le site Web Aldo**

Nbre de participants	Score en % - UMUX (Aldo)	Score moyen en % - DEEP (Aldo)
1	79,16666667	70,52631579
2	66,66666667	85,26315789
3	95,83333333	70,52631579
4	91,66666667	100
5	75	74,73684211
6	70,83333333	75,78947368
7	100	86,31578947
8	100	77,89473684
9	75	91,57894737
10	100	100
11	100	83,15789474
12	66,66666667	81,05263158
13	70,83333333	75,78947368
14	95,83333333	97,89473684
15	100	87,36842105
16	83,33333333	97,89473684
17	87,5	100
18	95,83333333	68,42105263

### Comparaison des scores en pourcentage des échelles UMUX et DEEP sur le site Web Aldo



**Figure 5.3.1.1 : Comparaison des scores en pourcentage des échelles UMUX et DEEP sur le site Web Aldo**

**Tableau 5.3.1.2 : Tableau comparatif des scores en pourcentage des échelles DEEP et UMUX sur le site Web Rubino**

Nombre de participants	Score moyen en % - DEEP (Rubino)	Score moyen en % - UMUX(Rubino)
1	91,57894737	45,83333333
2	74,73684211	79,16666667
3	80	83,33333333
4	88,42105263	95,83333333
5	73,68421053	83,33333333
6	56,84210526	83,33333333
7	48,42105263	45,83333333
8	75,78947368	100
9	60	79,16666667
10	77,89473684	62,5
11	72,63157895	54,16666667
12	89,47368421	66,66666667
13	66,31578947	83,33333333
14	85,26315789	95,83333333
15	88,42105263	70,83333333
16	67,36842105	45,83333333
17	93,68421053	95,83333333
18	69,47368421	100
19	34,73684211	87,5
20	41,05263158	66,66666667
21	90,52631579	70,83333333

Comparaison des scores en pourcentage des échelles DEEP et UMUX sur le site Web Rubino.

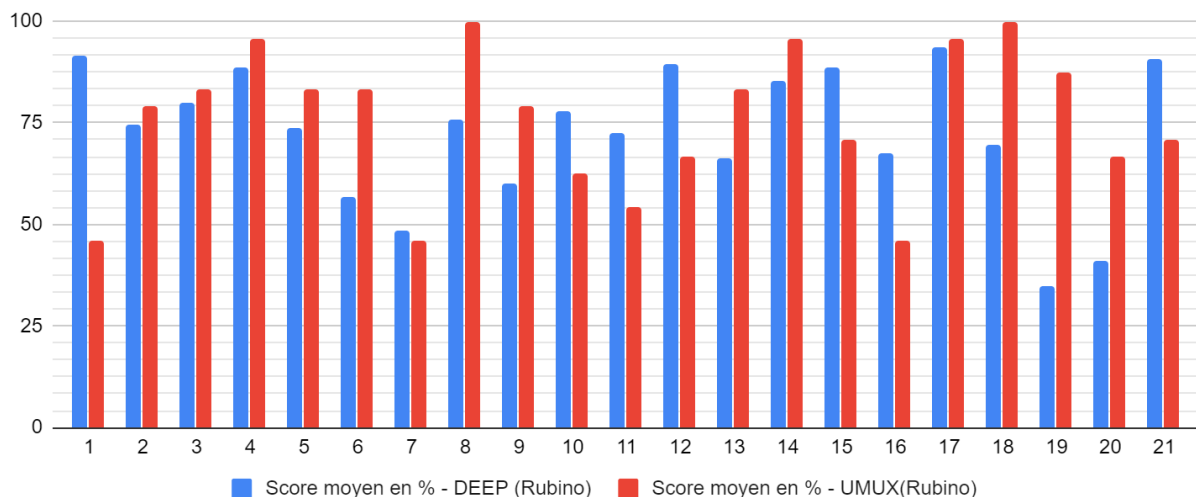


Figure 5.3.1.2 : Comparaison des scores en pourcentage des échelles DEEP et UMUX sur le site Web (Rubino)

Tableau 5.3.1.3 : Tableau comparatif des résultats des différents scores moyens en pourcentage

UMUX	Site Web - Aldo	Score moyen en %	86,34259259
DEEP	Site Web - Aldo	Score moyen en %	84,67836257
UMUX	Site Web - Rubino	Score moyen en %	75,99206349
DEEP	Site Web - Rubino	Score moyen en %	72,68170426

### 5.3.2 Différence entre les résultats obtenus

Tableau 5.3.2. : Différence des résultats

	<b>UMUX</b>	<b>DEEP</b>	<b>Différence entre les résultats</b>
<b>Aldo</b>	<b>86,34 %</b>	<b>84,68 %</b>	<b>86,34 % - 84,68 % = 1,66 %</b>
<b>Rubino</b>	<b>75,99 %</b>	<b>72,68 %</b>	<b>75,99 % - 72,68 % = 3,31 %</b>

## Conclusion

L'objectif fixé par cet essai est de répondre à la question de recherche suivante : « Le score final de l'échelle de mesure de l'utilisabilité DEEP (19 questions) a-t-il une variation de moins de 5 % par rapport au score final de l'échelle UMUX (4 questions) dans le cadre de la mesure de la satisfaction d'une interaction sur un site Web transactionnel ? ».

Pour ce faire, cette étude a débuté par une problématique qui nous a permis de comprendre le concept de l'utilisabilité des interfaces Web, dont l'indicateur de mesure était ici la satisfaction de l'utilisateur, mesurée avec des échelles différentes par leur nombre de questions.

L'étude des échelles de mesure de l'utilisabilité nous a permis de vérifier la praticabilité des trois principales composantes de l'utilisabilité que sont : l'efficacité, l'efficience et la satisfaction. Il a également été établi que les échelles à sept points surpassaient les échelles à cinq points en termes de fiabilité, de précision et de facilité d'utilisation.

La revue littéraire a permis de répertorier les différentes recherches effectuées ces dernières années et de s'apercevoir qu'il existe un grand nombre d'échelle de mesure de l'utilisabilité, souvent différentes par leur nombre de questions. Pour cet essai, deux méthodes réputées d'échelles de mesure de l'utilisabilité ont été retenues, à savoir : l'échelle UMUX (4 questions) et l'échelle DEEP (19 questions). Soixante participants, répartis en deux groupes de trente participants (un groupe pour chaque échelle), ont été invités à participer à une brève enquête en ligne. Trente-neuf participants au total ont répondu aux sondages qui consistaient à évaluer deux site Web différents.

L'approche proposée consistait à inviter les participants à effectuer une tâche prédéfinie sur les deux sites Web, pour ensuite les inviter à répondre à un questionnaire d'évaluation attribué à chacun des deux sites. Ce choix d'approche a permis de vérifier s'il y a une corrélation entre les deux échelles en étude, ceci pour confirmer ou infirmer notre hypothèse de départ.

Lors de la collecte des données, nous avons obtenu les réponses de 18 participants pour l'échelle DEEP et de 21 participants pour l'échelle UMUX. À partir de cet ensemble de données,

les résultats ont été analysés en utilisant les méthodes et techniques de calculs issues de la revue littéraire.

Dans un premier temps, pour la méthode DEEP, nous avons recodé les scores des questions 12 et 15 (polarités inversés) afin d'observer les éventuels phénomènes d'effet de halo ou de tendance à l'acquiescement. Après avoir recodé les scores en question selon la méthode de Lallemand et Gronier [16], nous avons calculé les moyennes des scores obtenus par les participants. Comme ces valeurs sont représentées en score, nous les avons converties en pourcentage afin de les comparer aux valeurs de l'échelle UMUX. L'échelle DEEP a donc fourni finalement un score moyen de 85 % pour le site web Aldo, et de 73 % pour le site web Rubino.

Dans un second temps, pour la méthode UMUX, nous avons recodé les données avec une méthode empruntée à la méthode SUS, c'est-à-dire les éléments impairs sont calculés [score - 1], et les éléments pairs sont calculés [7 - score], cela supprime la saisie positive / négative des éléments et permet un score minimum de zéro. Ceci a permis d'avoir une plage de 0 à 6 après recodage pour chaque élément UMUX individuel, ce qui donne un maximum préliminaire de 24 à l'échelle à quatre éléments. Pour atteindre la parité avec la plage de 0 à 100 fournie par le SUS, le score UMUX du participant est la somme des quatre items divisés par 24, puis multiplié par 100. Les résultats nous donnent au final un score moyen de 86 % pour le site Web Aldo, et de 76 % pour le site web Rubino.

Ces résultats nous permettent de conclure que les deux échelles à l'étude (DEEP et UMUX), différentes par leur nombre de questions et le temps requis pour compléter leur questionnaire, offrent des résultats similaires. Les possibles effets de halo ou de tendance à l'acquiescement, parfois observés lors de longs questionnaires, n'ont pas été observés de façon significative. Plus précisément, nous avons observé : une différence de 1,66 % entre les résultats obtenus d'une part avec UMUX et d'autre part avec DEEP lors de l'appréciation du site web Aldo ; et une différence de 3,31 % entre les résultats de ces deux échelles pour le site web Rubino. Ces différences sont inférieures à 5 % et nous permettent de valider notre hypothèse.

Nous recommandons à d'autres chercheurs de reproduire ces expériences avec des sites web de différentes natures et un nombre plus élevé de participants afin de solidifier nos constats. Ceci permettrait à tout un chacun de sélectionner des échelles d'évaluation de l'utilisabilité, qu'elles soient longues ou courtes à administrer, en étant confiant de la validité de leur résultat.



## Liste des références

- [1] K. Finstad, « The Usability Metric for User Experience », *Interact. Comput.*, vol. 22, n° 5, p. 323-327, sept. 2010.
- [2] « Halo effect », *Wikipedia*. 30-nov-2019.
- [3] « L'échelle de Likert pour les réponses aux questionnaires d'évaluation | SurveyMonkey », *[French] SurveyMonkey*. [En ligne]. Disponible sur: <https://fr.surveymonkey.com/mp/likert-scale/>. [Consulté le: 30-mars-2019].
- [4] K. A. Hounnou, « L'expérience utilisateur (UX) des sites Web par le design émotionnel », Université de Sherbrooke, p. 107, 2018.
- [5] R. S. Pressman, *Software Engineering: A Practitioner's Approach*. Palgrave Macmillan, 2005.
- [6] I. E. Sutherland, « Sketchpad a Man-Machine Graphical Communication System », *SIMULATION*, vol. 2, n° 5, p. R-3, mai 1964.
- [7] T. T. Hewett *et al.*, « ACM SIGCHI Curricula for Human-Computer Interaction », ACM, New York, NY, USA, 1992.
- [8] B. Chikhaoui, *Vers une évaluation analytique des interfaces homme machine développées dans le contexte des habitats intelligents*. [S.l. : s.n.], 2009, 2009.
- [9] J. Nielsen, *Usability Engineering*. Elsevier, 1994.
- [10] B. Shackel et S. J. Richardson, *Human Factors for Informatics Usability*. Cambridge University Press, 1991.
- [11] J.-F. Nogier et J. Leclerc, *UX Design et ergonomie des interfaces*. 2016.
- [12] A. Boucher, *Ergonomie web: pour des sites web efficaces*. Paris: Eyrolles, 2010.
- [13] T. Baccino, C. Bellino et T. Colombi, *Mesure de l'utilisabilité des interfaces*. Hermes Science Publications, 2004.
- [14] A. Seffah et E. Metzker, « The obstacles and myths of usability and software engineering », *Commun. ACM*, vol. 47, n° 12, p. 71-76, déc. 2004.
- [15] « ISO 9241-11:2018(fr) Ergonomie de l'interaction homme-système — Partie 11: Utilisabilité — Définitions et concepts ». ISO.org, 2018.

- [16] C. Lallemand et G. Gronier, *Méthodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Paris: Eyrolles, 2016.
- [17] A. M. Aladwani et P. C. Palvia, « Developing and validating an instrument for measuring user-perceived web quality », *Inf. Manage.*, vol. 39, n° 6, p. 467-476, mai 2002.
- [18] T. Yang, J. Linder, et D. Bolchini, « DEEP: Design-Oriented Evaluation of Perceived Usability », *Int. J. Human-Computer Interact.*, vol. 28, n° 5, p. 308-346, mai 2012.
- [19] « Définition : Echelle de Likert » Définitions marketing ». [En ligne]. Disponible sur: <https://www.definitions-marketing.com/definition/echelle-de-likert/>. [Consulté le: 14-déc-2019].
- [20] M. A. Diefenbach, N. D. Weinstein, et J. O'Reilly, « Scales for assessing perceptions of health hazard susceptibility », *Health Educ. Res.*, vol. 8, n° 2, p. 181-192, juin 1993.
- [21] E. P. Cox, « The Optimal Number of Response Alternatives for a Scale: A Review », *J. Mark. Res.*, vol. 17, n° 4, p. 407-422, 1980.
- [22] L. Andres, *Designing and Doing Survey Research*. SAGE, 2012.
- [23] « ISO/TR 16982 Ergonomie de l'interaction homme-systeme -- Méthodes d'utilisabilité pour la conception centrée sur l'opérateur humain ». 2002.
- [24] « How long should a survey be? What is the ideal survey length? », *SurveyMonkey*. [En ligne]. Disponible sur: [https://www.surveymonkey.com/curiosity/survey\\_completion\\_times/](https://www.surveymonkey.com/curiosity/survey_completion_times/). [Consulté le: 19-mars-2019].
- [25] J. A. Krosnick, « Survey Research », *Annu. Rev. Psychol.*, vol. 50, n° 1, p. 537-567, 1999.
- [26] J. M. C. Bastien, « Usability testing: a review of some methodological and technical aspects of the method », *Int. J. Med. Inf.*, vol. 79, n° 4, p. e18-e23, avr. 2010.
- [27] C. M. MacDonald et M. E. Atwood, « Changing perspectives on evaluation in HCI: past, present, and future », p. 10.
- [28] « What is Usability? », *The Interaction Design Foundation*. [En ligne]. Disponible sur: <https://www.interaction-design.org/literature/topics/usability>. [Consulté le: 07-avr-2019].
- [29] E. P. Cox, « The Optimal Number of Response Alternatives for a Scale: A Review », *J. Mark. Res.*, vol. 17, n° 4, p. 407-422, 1980.
- [30] L. Faulkner, « Beyond the five-user assumption: Benefits of increased sample sizes in usability testing », *Behav. Res. Methods Instrum. Comput.*, vol. 35, n° 3, p. 379-383, août 2003.

- [31] R. D. Fricker et M. Schonlau, « Avantages et inconvénients des enquêtes de recherche sur Internet: données probantes tirées de la littérature », *Field Methods*, vol. 14, n° 4, p. 347-367, nov. 2002.
- [32] A. Bangor, P. T. Kortum, et J. T. Miller, « An Empirical Evaluation of the System Usability Scale », *Int. J. Human-Computer Interact.*, vol. 24, n° 6, p. 574-594, juill. 2008.

## Bibliographie

- « ISO 9241-210:2010(fr) "Ergonomie de l'interaction homme-système – partie 210 : conception centrée sur l'opérateur humain pour les système interactifs" », [En ligne] disponible <https://www.iso.org/obp/ui/fr/#iso:std:iso:9241:-210:ed-1:v1:fr>. [Consulté en Mars - 2019].
- « SurveyMonkey "Comment analyser les données d'un sondage", [En ligne] disponible sur : <https://fr.surveymonkey.com/mp/how-to-analyze-survey-data/> ». [Consulté en Novembre - 2019].
- « M. Jon Gitlin, "Épuration des données de sondage : sept points à vérifier avant de commencer une analyse", recherche google ». [En ligne] disponible sur : <https://fr.surveymonkey.com/curiosity/epuration-des-donnees-de-sondage-sept-points-a-verifier-avant-de-commencer-une-analyse/>. [Consulté en Novembre - 2019].
- « M. AurelienSalomon, "Évaluation de l'interface utilisateur d'un système d'aide à l'apprentissage de l'économie et conception d'une nouvelle interface", p.3 à 7 ». [En ligne] disponible : [https://publications.polymtl.ca/1521/1/2014\\_Aur%C3%A9lienSalomon.pdf](https://publications.polymtl.ca/1521/1/2014_Aur%C3%A9lienSalomon.pdf). [Consulté en Février-2019].
- « Mr Mehmet Ilker Berkman, "Re-assessing the usability metric for user experience (UMUX) scale", recherche google scholar ». [En ligne] disponible sur : <https://dl.acm.org/citation.cfm?id=2993221>. [Consulté en Mai-2019].
- « Usabilis, "Définition utilisabilité ou usabilité", recherche google ». [En ligne] disponible sur : <https://www.usabilis.com/definition-utilisabilite-usabilite/>. [Consulté en Février-2019].
- « Wang, J., & Senecal, S. (2007). Measuring perceived Website Usability. *Journal of Internet Commerce*, 6(4), 97 – 112. ».
- « Lewis, J., Utesch, B., & Maher, D. (2013). UMUX-LITE: when there's no time for the SUS. *Proc. of CHI 2013*, 2099–2102. »
- « Catherine Recanati, "Interaction homme-machine" ». [En ligne] disponible sur : <https://lipn.univ-paris13.fr/~recanati/docs/M2-InHM/Interaction.pdf>.
- « Koffi Anoumou HOUNNOU "L'expérience utilisateur (UX) des sites Web par le design émotionnel" ». [En ligne] disponible sur : [https://www.usherbrooke.ca/cefti/fileadmin/sites/cefti/documents/Essais/Koffi-Anoumou-HOUNNOU-Essai-Final\\_26\\_juin\\_2018.pdf](https://www.usherbrooke.ca/cefti/fileadmin/sites/cefti/documents/Essais/Koffi-Anoumou-HOUNNOU-Essai-Final_26_juin_2018.pdf)



2. Utiliser ce système est une expérience \*

	1	2	3	4	5	6	7	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

---

3. Ce système est facile à \*

	1	2	3	4	5	6	7	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

---

4. Je dois passer trop de temps à corriger les choses sur ce \*

	1	2	3	4	5	6	7	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

Vous êtes? \*

- Un homme
- Une femme
- Autre
- Préfère ne pas répondre

Quel âge avez-vous? \*

- 17 ans et moins
- 18 à 24 ans
- 25 à 34 ans
- 35 à 44 ans
- 45 à 54 ans
- 55 à 64 ans
- 65 ans ou plus
- Préfère ne pas répondre

















Vous êtes?

\*

- Un homme
- Une femme
- Autre
- Préfère ne pas répondre

Quel âge avez-vous?

\*

- 17 ans et moins
- 18 à 24 ans
- 25 à 34 ans
- 35 à 44 ans
- 45 à 54 ans
- 55 à 64 ans
- 65 ans ou plus
- Préfère ne pas répondre

## Annexe III

### Tableau de collecte des données UMUX

Questionnaire d'évaluation - Aldo (réponses) ☆

Fichier Édition Afficher Insertion Format Données Outils Formulaire Modules complémentaires Aide Toutes les modifications ont été enregistrées...

100% € % .0 .00 123 Par défaut 8 B I A

	C	D	E	F	G	H	I	J	K	L
1	Score recalculé	2. Utiliser ce système est une ex	Score recalculé	3. Ce système est facile à utilis	Score recalculé	4. Je dois passer trop de temps	Score recalculé	somme score	Score en %	Vous êtes?
2	5	2	5	5	4	2	5	19	79.16666667	Un homme
3	5	2	5	6	5	6	1	16	66.66666667	Un homme
4	6	1	6	7	6	2	6	23	95.83333333	Un homme
5	5	1	6	6	5	1	6	22	91.66666667	Un homme
6	4	1	6	4	3	2	5	18	75	Un homme
7	4	3	4	6	5	3	4	17	70.83333333	Un homme
8	6	1	6	7	6	1	6	24	100	Un homme
9	6	1	6	7	6	1	6	24	100	Une femme
10	4	2	5	5	4	2	5	18	75	Un homme
11	6	1	6	7	6	1	6	24	100	Une femme
12	6	1	6	7	6	1	6	24	100	Une femme
13	4	2	5	5	4	4	3	16	66.66666667	Un homme
14	5	1	6	6	5	6	1	17	70.83333333	Une femme
15	6	1	6	6	5	1	6	23	95.83333333	Un homme
16	6	1	6	7	6	1	6	24	100	Préfère ne pas répondre
17	5	2	5	6	5	2	5	20	83.33333333	Un homme
18	6	2	5	5	4	1	6	21	87.5	Un homme
19	6	1	6	7	6	2	5	23	95.83333333	Un homme



## Annexe IV

### Tableau de collecte des données DEEP

Questionnaire d'évaluation - Rubino (réponses) ☆ ■

Fichier Édition Afficher Insertion Format Données Outils Formulaire Modules complémentaires Aide [Toutes les modifications ont été enregistrées.](#)

100% € % .0 .00 123 Par défaut ... 9 B I S A

fx 1. Le libellé du texte était clair.

	B	C	D	E	F	G	H
1	1. Le libellé du texte était clair.	2. Le contenu (texte, images	3. Le texte était utile.	4. Le texte était pertinent.	5. Je pouvais rapidement co	6. L'organisation du site web	7. Dans chaque section du s
3	5	4	4	4	5	5	4
4	4	4	4	5	5	4	4
5	5	5	5	5	5	4	5
6	4	4	4	5	2	4	4
7	4	4	4	3	3	4	4
8	4	3	3	1	1	3	3
9	4	5	5	4	4	4	4
10	4	4	4	1	2	4	4
11	4	4	4	4	3	2	2
12	4	3	3	4	3	4	4
13	4	5	5	5	4	5	5
14	4	4	4	3	3	3	4
15	5	4	4	4	4	4	5
16	5	5	5	5	4	4	4
17	3	4	4	3	3	3	3
18	5	5	5	4	5	5	4
19	4	4	4	4	4	4	4
20	1	2	2	1	1	3	2
21	2	1	1	2	2	3	2



# Questionnaire d'évaluation - Rubino (réponses)

Fichier Édition Afficher Insertion Format Données Outils Formulaire Modules complémentaires Aide [Toutes les modifications ont été enregistrées...](#)

100% € % .0 .00 123 Par défaut 9 B I A

	I	J	K	L	M	N	O	P
1	8. Il était facile de trouver l'in	9. Le site web m'a aidé à tro	10. J'ai obtenu ce à quoi je n	11. Utiliser ce site web s'est	12. Utiliser ce site web m'a fatigué.	Score inversé	13. J'ai appris à ut	14. La mise en page
3	5	3	5	3	2	4	4	3
4	4	4	5	4	1	5	4	3
5	3	3	5	3	2	4	3	5
6	4	3	3	3	2	4	4	4
7	2	2	2	3	3	3	3	2
8	4	2	1	4	1	5	4	1
9	5	4	3	4	1	5	2	4
10	1	1	2	2	3	3	4	5
11	2	3	4	2	1	5	5	5
12	4	4	4	4	2	4	4	5
13	5	4	4	5	2	4	5	5
14	3	2	2	4	3	3	4	4
15	5	5	4	4	1	5	5	4
16	5	5	5	5	1	5	4	5
17	3	3	3	3	1	5	4	3
18	5	5	5	5	1	5	5	5
19	4	4	3	3	2	4	4	4
20	1	1	2	2	4	2	2	2
21	1	1	2	2	4	2	3	4

## Annexe V

### Sites Web retenus pour les tests

#### 1- Sites Web retenus

Aldo: <https://www.aldoshoes.com/ca/fr>

Rubino: <https://rubinoshoes.com/fr>

#### 2- Formats de courriel d'invitation au sondage

##### Exemple de format d'invitation pour l'échelle UMUX.

Chers étudiantes et étudiants,

Dans le cadre de ma maîtrise en génie logiciel au CeFTI (UdeS), je sollicite votre participation à une brève enquête qui a pour objet d'évaluer les échelles de mesure d'utilisabilité.

Votre avis nous intéresse, et grâce à vos réponses nous obtiendrons des données importantes qui permettront de valider notre recherche.

Pour y participer, répondez aux **deux** questionnaires suivants :

Perception du site web Rubino :

<https://forms.gle/S8qHwTPptzjSE2E7A>

Perception du site web Aldo :

<https://forms.gle/gd8RxwvgwkcCS8LCA>

Temps requis estimé : 2-3 minutes par sondage.

Les données recueillies sont anonymes et ne seront pas associées ni à votre adresse courriel ni à votre ordinateur.

Le questionnaire est disponible du 5 au 12 Octobre.

Merci de votre précieuse collaboration,

Très cordialement,

Armand Dayang Nana

## Exemple de format d'invitation pour l'échelle DEEP.

Chers étudiantes et étudiants,

Dans le cadre de ma maîtrise en génie logiciel au CeFTI (UdeS), je sollicite votre participation à une brève enquête qui a pour objet d'évaluer les échelles de mesure d'utilisabilité.

Votre avis nous intéresse, et grâce à vos réponses nous obtiendrons des données importantes qui permettront de valider notre recherche.

Pour y participer, répondez aux **deux** questionnaires suivants :

Perception du site web Rubino :

<https://forms.gle/S8qHwTPptzjSE2E7A>

Perception du site web Aldo :

<https://forms.gle/gd8RxwvgwkcCS8LCA>

Temps requis estimé : 6-7 minutes par sondage.

Les données recueillies sont anonymes et ne seront pas associées ni à votre adresse courriel ni à votre ordinateur.

Le questionnaire est disponible du 5 au 12 Octobre.

Merci de votre précieuse collaboration,

Très cordialement,

Armand Dayang Nana

Étudiant à la maîtrise, CeFTI, Université de Sherbrooke

**Exemple de formulaire des questions - réponses des participants au sondage dans l'application Google Forms avec l'échelle UMUX.**

Les réponses ne peuvent pas être modifiées.

## Questionnaire d'évaluation - Site web Rubino

\*Obligatoire

### Scénario

Imaginez que vous êtes invité(e) à une soirée chic et que vous devez acheter de nouvelles chaussures.

Rendez-vous sur le site web <https://rubinoshoes.com/fr>, afin de trouver une paire de chaussures à votre goût.

Consultez les détails du produit retenu, choisissez la taille, puis voyez comment vous pouvez l'acheter (sans compléter l'achat).

Par la suite, commentez votre expérience en répondant au questionnaire suivant :

1. Les fonctionnalités de ce système sont conformes à mes exigences. \*

	1	2	3	4	5	6	7	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

2. Utiliser ce système est une expérience frustrante. \*

	1	2	3	4	5	6	7	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Tout à fait d'accord

**Exemple de formulaire du résumé des questions - répondues des participants au sondage dans l'application Google Forms avec l'échelle UMUX.**

1. Les fonctionnalités de ce système sont conformes à mes exigences.

---

	1	2	3	4	5	6	7	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Tout à fait d'accord

---

7 réponses

---

	1	2	3	4	5	6	7	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

---

7 réponses

---

	1	2	3	4	5	6	7	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Tout à fait d'accord

---

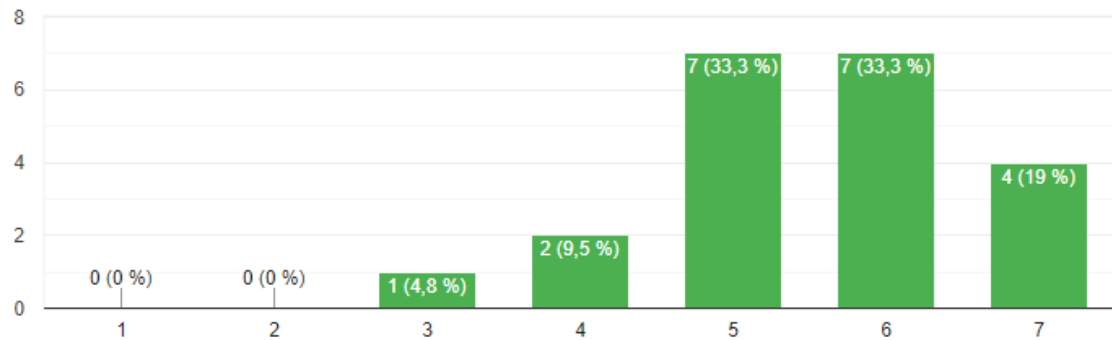
4 réponses

## Exemple de formulaire du résumé des réponses des participants au sondage dans l'application Google Forms avec l'échelle UMUX.

1. Les fonctionnalités de ce système sont conformes à mes exigences.

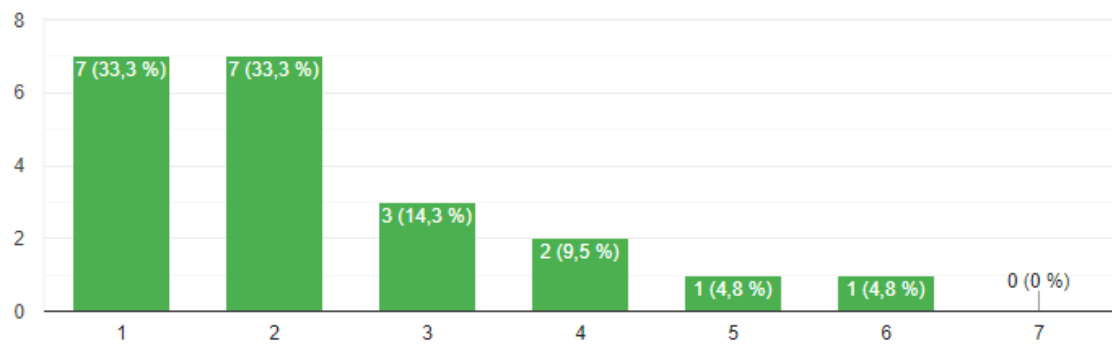


21 réponses



2. Utiliser ce système est une expérience frustrante.

21 réponses

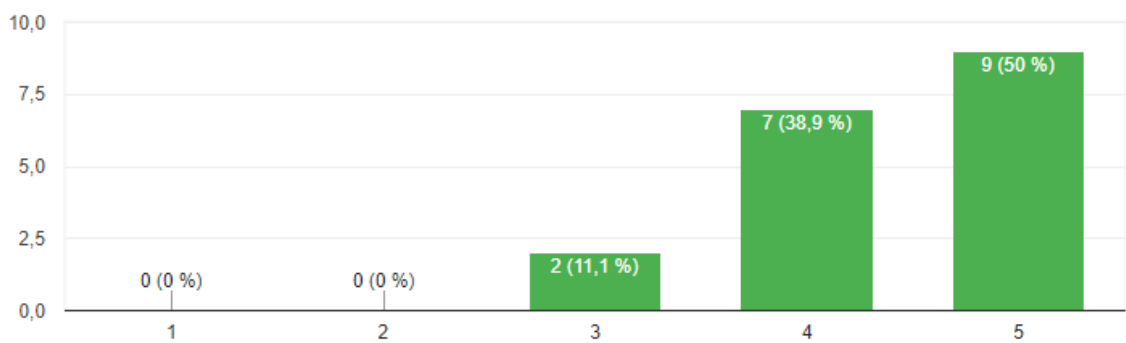


## Exemple de formulaire du résumé des réponses des participants au sondage dans l'application Google Forms avec l'échelle DEEP.

### EFFORT COGNITIF PERÇU

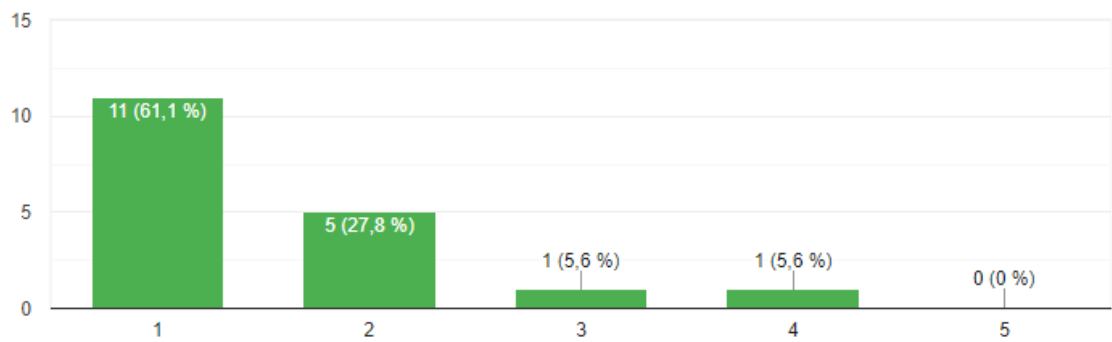
11. Utiliser ce site web s'est fait sans effort.

18 réponses



12. Utiliser ce site web m'a fatigué.

18 réponses



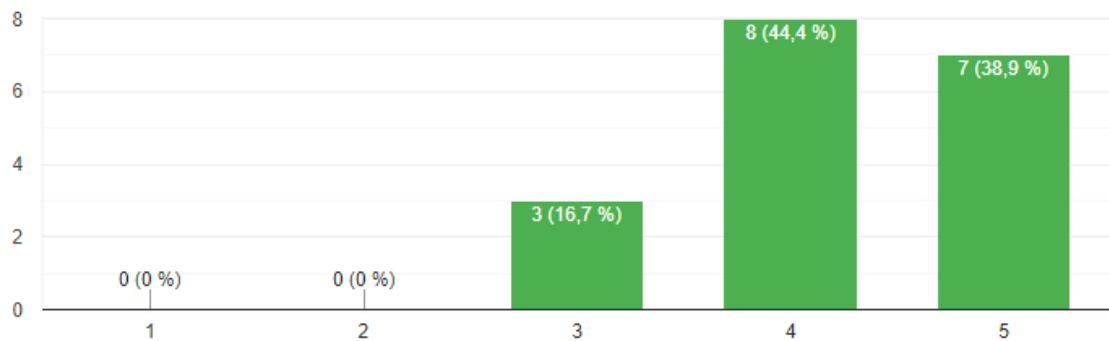


## Exemple de formulaire du résumé des réponses des participants au sondage dans l'application Google Forms avec l'échelle DEEP.

### COHÉRENCE DE LA MISE EN PAGE PERÇUE

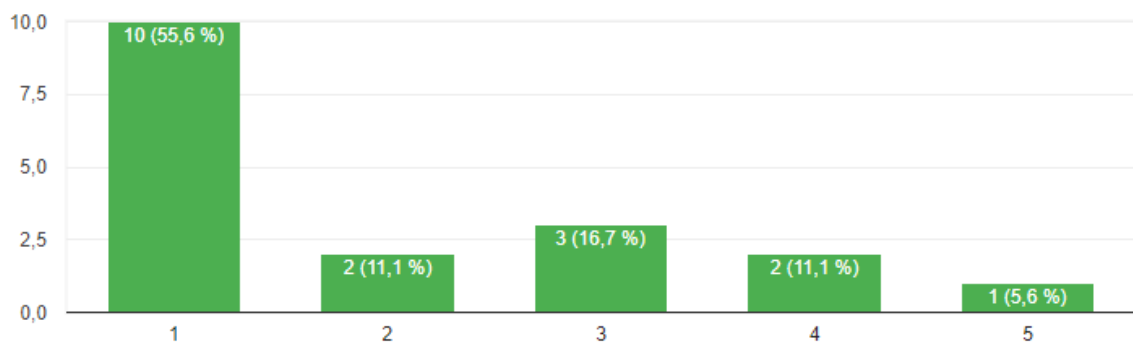
14. La mise en page à travers tout le site web était cohérente.

18 réponses



15. J'ai remarqué des changements soudains de mise en page à travers le site web.

18 réponses



**Exemple de formulaire du résumé des questions - répondues des participants au sondage dans l'application Google Forms avec l'échelle DEEP.**

1. Le libellé du texte était clair.

---

	1	2	3	4	5	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Tout à fait d'accord

---

9 réponses

---

	1	2	3	4	5	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Tout à fait d'accord

---

8 réponses

---

	1	2	3	4	5	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

---

1 réponse