Numéro d'ordre : 16/24

Thèse de doctorat

Pour obtenir le grade de Docteur de l'Université de Valenciennes et du Hainaut-Cambrésis et de l'Université de Sfax

Sciences et Technologie, Mention : Informatique

Présentée et soutenue par Ahlem ASSILA

Le 23/09/2016, à Valenciennes

Ecole doctorale :

Sciences Pour l'Ingénieur (SPI)

Equipe de recherche, Laboratoire :

Département Informatique, Thème : INTERaction et Agents (INTERA)

Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines (LAMIH UMR CNRS 8201)

Une approche et un outil pour l'évaluation subjective et objective de l'utilisabilité des systèmes interactifs : Application à un Système d'Aide à l'Information Voyageurs

JURY

Président du jury

- GIRARD, Patrick. Professeur des universités. Université de Poitiers.

Rapporteurs

- MOUSSA, Faouzi. Professeur des universités. Université de Tunis.
- VANDERDONCKT,Jean. Professeur des universités. Université catholique de Louvain.

Examinateurs

- BASTIEN, Christian. Professeur des universités. Université de Lorraine.
- BEN AYED, Mounir. Maître de Conférences (HDR). Université de Sfax.

Invité

- KOLSKI, Christophe. Professeur des Universités. Université de Valenciennes et du Hainaut-Cambrésis.

Directeurs de thèse

- EZZEDINE, Houcine. Professeur des Universités. Université de Valenciennes et du Hainaut-Cambrésis
- BOUHLEL, Mohamed Salim. Professeur des Universités. Université de Sfax.

Co-encadrante de thèse :

- MARÇAL de OLIVEIRA, kathia. Maître de Conférences (HDR). Université de Valenciennes et du Hainaut-Cambrésis.

Université de Valenciennes et du Hainaut-Cambrésis

Numéro d'ordre : 16/24

Thèse de doctorat

Pour obtenir le grade de Docteur de l'Université de

VALENCIENNES ET DU HAINAUT-CAMBRESIS

Sciences et Technologie, Mention : Informatique

Présentée et soutenue par Ahlem ASSILA

Le 23/09/2016, à Valenciennes

Ecole doctorale :

Sciences Pour l'Ingénieur (SPI)

Equipe de recherche, Laboratoire :

Département Informatique, Thème : INTERaction et Agents (INTERA)

Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines (LAMIH UMR CNRS 8201)

Une approche et un outil pour l'évaluation subjective et objective de l'utilisabilité des systèmes interactifs : Application à un Système d'Aide à l'Information Voyageurs

JURY

Président du jury

- GIRARD, Patrick. Professeur des universités. Université de Poitiers.

Rapporteurs

- MOUSSA, Faouzi. Professeur des universités. Université de Tunis.
- VANDERDONCKT,Jean. Professeur des universités. Université catholique de Louvain.

Examinateurs

- BASTIEN, Christian. Professeur des universités. Université de Lorraine.
- BEN AYED, Mounir. Maître de Conférences (HDR). Université de Sfax.

Invité

- KOLSKI, Christophe. Professeur des Universités. Université de Valenciennes et du Hainaut-Cambrésis.

Directeurs de thèse

- EZZEDINE, Houcine. Professeur des Universités. Université de Valenciennes et du Hainaut-Cambrésis
- BOUHLEL, Mohamed Salim. Professeur des Universités. Université de Sfax.

Co-encadrante de thèse :

- MARÇAL de OLIVEIRA, kathia. Maître de Conférences (HDR). Université de Valenciennes et du Hainaut-Cambrésis.

Résumé

L'évaluation de l'utilisabilité des systèmes interactifs est l'un des plus grands intérêts de la communauté des Interactions Homme-Machine. Pour réussir cette évaluation, plusieurs méthodes et outils d'évaluation subjective et objective ont été proposés et appliqués dans l'académie et l'industrie. Dans ce contexte, les évaluateurs d'utilisabilité travaillent habituellement avec beaucoup et différents types de données. Ceux-ci sont générés à partir de diverses méthodes et outils et liés à différents critères d'utilisabilité. En général, la compréhension et l'interprétation de ces données hétérogènes ne sont pas des tâches faciles. Dans cette thèse, nous proposons d'intégrer des données d'évaluation subjective et objective afin de mieux aider les évaluateurs à la prise de décision sur la qualité des systèmes interactifs dans le domaine des systèmes de supervision du trafic. En effet, nous avons appliqué le modèle de mesure d'information défini par la norme ISO/IEC 15939 avec l'utilisation des indicateurs dans un environnement d'évaluation des systèmes interactifs. Le développement de cet environnement a débuté depuis une décennie. Ainsi, pour compléter sa mise en œuvre et le mettre en pratique, nous avons appliqué plusieurs indicateurs d'utilisabilité proposés dans l'évaluation d'un système de supervision du trafic. Ces indicateurs intègrent les résultats d'évaluation subjective et objective des questionnaires d'utilisabilité, un inspecteur des règles ergonomiques et un mouchard électronique. Ils fournissent une base pour aider à l'interprétation des résultats obtenus. De plus, ils permettent la détection des problèmes d'utilisabilité et donnent des recommandations pour améliorer les interfaces utilisateur. Enfin, cette thèse propose des perspectives de recherche sur la base des limites rencontrées.

Mots-clés. Interaction Homme-Machine, Interface Utilisateur, évaluation, utilisabilité, mesures, critères, intégration, subjective, objective.

An approach and a tool for subjective and objective evaluation of the usability of interactive systems: Application to passengers Information Assistance System

Abstract

Usability evaluation for interactive systems is one of the greatest interests for Human-Computer Interaction community. To the success of this evaluation, several subjective and objective evaluation methods and tools have been proposed and applied in academy and industry. In this context, usability evaluators usually work with a lot of and different kind of data. These data are generated from various methods and tools and related to different usability criteria. In general, understanding and interpreting these heterogeneous data are not easy tasks. In this thesis, we propose to integrate subjective and objective evaluation data to better support the quality evaluators' decision-making for traffic supervision systems domain. In fact, we applied the measurement information model defined by ISO/IEC 15939 with the use of indicators in an interactive system evaluation environment. The development of this environment has started in a decade. Thus, for completing implementation and putting it into practice, we have applied several proposed usability indicators in the evaluation of a traffic supervision system. These indicators integrate subjective and objective evaluation results from usability questionnaires, an ergonomic guidelines inspector and an electronic informer. They provide a basis for aiding the interpretation of their obtained results. Furthermore, they allow detecting usability problems and give recommendations for enhancing the user interfaces. Finally, this thesis propose research perspectives based on the encountered limits.

Keywords. Human-Computer Interaction, User interface, evaluation, usability, measures, criteria, integration, subjective, objective.

Remerciements

Je tiens à remercier tous les membres du jury :

- M. Patrick Girard, Professeur des Universités à l'Université de Poitiers pour m'avoir fait l'honneur d'être président de mon Jury.
- M. Jean Vanderdonckt, Professeur des universités à l'Université catholique de Louvain et M. Faouzi Moussa, Maître de Conférences, HDR à l'Université de Tunis pour m'avoir fait l'honneur d'être rapporteurs de ma thèse.
- M. Christian Bastien, Professeur des Universités à l'Université de Lorraine et M. Mounir Ben Ayed, Maître de Conférences, HDR à l'Université de Sfax pour m'avoir fait l'honneur d'être examinateurs de mon travail.
- M. Christophe kolski, Professeur des universités à l'université de Valenciennes pour avoir accepté de participer à ce jury de thèse. Je le remercie également pour son accueil chaleureux au sein de LAMIH depuis mon arrivée en 2012 à Valenciennes ainsi que ses nobles qualités humaines. Je suis très reconnaissante qu'il soit membre du jury.
- M. Houcine Ezzedine, Professeur des universités à l'Université de Valenciennes et M. Mohamed Salim Bouhlel, Professeur des universités à l'Université de Sfax mes deux directeurs de thèse. M. Ezzedine, je le remercie énormément de m'avoir choisi pour participer dans ce grand projet de recherche. Je tiens à le remercier pour son honnêteté intellectuelle ainsi que tous ses conseils tout au long de ces années de thèse. De même, je suis très reconnaissante à M. Bouhlel de m'avoir mis sur le bon chemin de recherche. Je les remercie tous les deux pour la confiance qu'ils m'ont accordée et de m'avoir appris à être plus autonome.
- Kathia Marcal de Oliveira, Maître de Conférences, HDR à l'Université de Valenciennes. Je lui adresse mes vifs remerciements. Elle était et restera une personne qui a pris une place particulière dans ma vie. Nous avons vécu ensemble les moments les plus durs de stress et d'inquiétude. Toutefois, nous avons partagé aussi des bons moments inoubliables. Ses critiques rigoureuses ont été très utiles pour réaliser ce travail.

Mes remerciements vont aussi à ma famille de Valenciennes et particulièrement à mes parents Sandro et Jocelene pour leurs présences, amours et leurs encouragements. J'adresse aussi mes remerciements à Taisa, Fadoua et Nadia pour leurs amitiés et leurs soutiens inconditionnels et de même pour mes frères Abdel, Amir, Salaheddine, Ali, Ahmed, José, Osama. Je remercie également mes amies Emna ben Ayed, Amira Dhouib, Shadab, Catalina et Mounira Jabnoun.

J'adresse un remerciement tout particulier à mon frère et ma sœur Aymen Bouabid et Amira Bouabid.

Merci également à mes amies de la résidence Jules Mousseron : Salma, Ihssane, zainab, nada et kawthar.

Je tiens à remercier Selem Charfi pour le temps qui m'a accordé en lisant et corrigeant mes premiers articles.

Je remercie aussi tous mes amis et collègues ainsi que le personnel du LAMIH, particulièrement Mme Rubins, Isabelle Oliveira et Marlene pour leurs bonnes humeurs et qualités humaines. J'aimerais également remercier toute les personnes et responsables du LAMIH pour prendre en charge tous mes déplacements et de m'avoir recruté comme étant un ingénieur d'étude et de recherche en 2012 ainsi qu'en qualité d'ATER pour deux ans à l'université de Valenciennes.

J'adresse également mes remerciements au Ministre de l'enseignement supérieur de la recherche scientifique en Tunisie pour m'avoir accordé deux fois une bourse d'alternance.

Je remercie, aussi, tous mes collègues de l'ISTV (université de Valenciennes) avec lesquels j'ai enseigné pendant ces deux années.

Je remercie ma copine Zohra et son mari Abdallah. J'espère que notre amitié restera tout au long de notre vie.

Enfin, j'adresse mes sincères gratitudes à toutes les personnes, de près ou de loin, qui m'ont soutenue et m'ont encouragée pendant toutes ces années.

Dédicaces

Je dédie cette thèse à :

- Mon père Mohamed,
- Ma mère Habiba,
- Mes sœurs Maysam et Balsam,
- Mon frère Yassine,
- Mon fiancé Saif,
- Ma petite ange Ghayoda
- A l'âme de Dr.Hatem Bettahar,
- A l'âme de mon amie Amna Bouhana,
- Tous ceux que j'aime, Et
- Tous ceux qui veulent partager ma joie…

Table des matières

Liste des Figures 10
Liste des Tableaux 13
Introduction Générale 15
Chapitre 1 : L'évaluation de l'utilisabilité des IHM des systèmes interactifs 19
Introduction 19
1.1 Définitions générales des concepts d'évaluation des IHM 20
1.1.1 L'évaluation des IHM 20
1.1.2 L'utilisabilité 21
1.1.3 Les mesures 22
1.1.4 Les types d'évaluation d'utilisabilité 26
1.2 Les processus d'évaluation 27
1.2.1 Les processus d'évaluation existants vus sous l'angle de standards 27
1.2.2 Les processus d'évaluation existants vus sous l'angle de l'IHM 28
1.3 Les méthodes et outils d'évaluation de l'utilisabilité des interfaces utilisateur 32
1.3.1 Les catégories de méthodes et outils d'évaluation des interfaces utilisateur 32
1.3.2 Présentation de quelques méthodes et outils d'évaluation d'utilisabilité des interfaces utilisateur 34
1.4 Les mesures d'utilisabilité pour l'évaluation des IHM 38
1.5 L'évaluation ergonomique des IHM des systèmes interactifs 40
1.6 Synthèse et conclusion 42
Chapitre 2 : Etat de l'art sur les outils d'évaluation de l'utilisabilité des IHM 43
Introduction 43
2.1 Etat de l'art sur les outils d'évaluation de l'utilisabilité des IHM basés sur une évaluation subjective ou objective 44
2.1.1 Les outils assurant une évaluation subjective 44
2.1.1.1 Etat de l'art sur les outils des questionnaires d'utilisabilité standardisés 45
2.1.1.2 Etude d'analyse des questionnaires selon des critères standards d'utilisabilité 51
2.1.2 Les outils assurant une évaluation objective 55
2.1.2.1 Les outils utilisant les règles ergonomiques 55
2.1.2.2 Les outils de recueil des données d'interaction entre l'utilisateur et le système interactif à évaluer…… 56
2.2 Les approches basées sur une combinaison des évaluations subjective et objective 57
2.2.1 Présentation des approches étudiées 57
2.2.1.1 WebUSE: approche d'évaluation d'utilisabilité des interfaces Web 57
2.2.1.2 Méthode de normalisation des mesures d'utilisabilité en un score unique (SUM) 59
2.2.1.3 Approche d'intégration de méthodes d'évaluation d'utilisabilité en utilisant des profils de problèmes d'utilisabilité 60

2.2.1.4 Evaluation d'utilisabilité des systèmes logiciels en utilisant des mesures subjectives et objectives…. 61
2.2.1.5 Approche d'aide à la prise de décision par la combinaison des méthodes d'évaluation subjectives et objectives 61
2.2.1.6 Framework RITA pour l'évaluation des interfaces utilisateur 62
2.2.2 Récapitulatif et Discussion 63
2.3 Synthèse et conclusion 66
Chapitre 3 : Approche d'intégration des résultats d'évaluation d'utilisabilité subjective et objective 67
Introduction 67
3.1 Motivation et présentation générale de l'approche 68
3.1.1 Motivation 68
3.1.2 Présentation générale de l'approche proposée 70
3.2 La construction des indicateurs d'utilisabilité des interfaces utilisateur basés sur la norme ISO/IEC 15939…. 71
3.2.1 Mapping des données basé sur les critères d'utilisabilité 71
3.2.2 Les étapes de spécification d'un indicateur d'utilisabilité 73
3.2.2.1 Etape 1: L'identification de besoin d'information 74
3.2.2.2 Etape 2: La spécification des mesures de base et dérivées 76
3.2.2.3 Etape 3: La spécification d'un indicateur d'utilisabilité 78
3.3 La spécification des modèles de mesure d'information des indicateurs d'utilisabilité des IHM basés sur la norme ISO/IEC 15939 82
3.3.1 Les modèles de mesure d'information proposés pour l'évaluation des critères ISO 9241-11 (Efficacité et Efficience) 83
3.3.2 Les modèles de mesure d'information proposés pour l'évaluation d'un ensemble des critères ergonomiques de Scapin et Bastien 86
3.3.2.1 L'évaluation de la densité informationnelle d'une interface utilisateur 87
3.3.2.2 L'évaluation de la lisibilité d'une interface utilisateur 88
3.3.2.3 L'évaluation des actions minimales des tâches réalisées 93
3.3.2.4 L'évaluation du feedback immédiat aux actions des utilisateurs sur les IHM à évaluer 95
3.4 Conclusion 97
Chapitre 4 : SOUSEval : un outil pour l'intégration d'évaluations objective et subjective de l'utilisabilité des IHM 98
Introduction 98
4.1 Environnement intégré d'aide à l'évaluation des systèmes interactifs 99
4.2 La version étendue du Mouchard électronique EISEval 103
4.3 La version étendue de l'inspecteur des règles ergonomiques 105
4.4 L'outil SOUSEval 106
4.4.1 L'outil SOUSEval comme étant un outil d'aide à la génération des questionnaires 108
4.4.1.1 Le module de préparation d'une évaluation subjective 108
4.4.1.2 Le module d'évaluation 112
4.4.1.3 Le module d'analyse des résultats d'évaluation 113
4.4.2 L'outil SOUSEval pour l'intégration des évaluations subjective et objective de l'utilisabilité des interfaces utilisateur 115
4.4.2.1 Le module de préparation de l'évaluation intégrée 116
4.4.2.2 Le module d'évaluation 116

4.4.2.2.1 L'évaluation ergonomique 116
4.4.2.2.2 L'évaluation des interactions des utilisateurs avec le système interactif à évaluer 118
4.4.2.2.3 L'évaluation subjective 120
4.4.2.3 Le module d'analyse des résultats d'évaluation 121
4.5 Synthèse et conclusion 125
Chapitre 5 : Evaluation Expérimentale 126
Introduction 126
5.1 Etude de Cas : Système d'Aide à l'Information Voyageurs 126
5.1.1 Objectif de l'étude de cas 128
5.1.2 Définition de la procédure expérimentale : Planification de l'expérimentation 128
5.1.3 Exécution de l'expérimentation en utilisant l'outil SOUSEval 131
5.2 Discussion sur les risques de validité de notre étude réalisée 147
5.3 Conclusion 149
Conclusion Générale 151
Bibliographie 159
Annexe A 172

Liste des Figures

Figure 1 : Environnement intégré d'aide à l'évaluation des systèmes interactifs (Ezzedine, 2002) …..……..….17

Figure 1. 1 : Principe de base de l'évaluation des IHM (inspiré de (Senach, 1990))……..20

Figure 1. 2 : Modèle de qualité en utilisation proposé par la norme SQuaRE (adapté de (ISO/IEC 25010, 2011))……. 22

Figure 1. 3 : Structure et relations clés de modèle de mesure d'information (ISO/IEC 15939, 2007)……24

Figure 1. 4 : Processus simplifié d'évaluation de la qualité de produit logiciel (ISO/IEC 25040, 2011) ……..28

Figure 1. 5 : Processus d'évaluation des interfaces utilisateur (inspiré de (Mariage, 2005)) …….29

Figure 1. 6 : Processus d'évaluation des interfaces utilisateur (inspiré de (Ivory et Hearst, 2001))……...31

Figure 2. 1: Modèle hiérarchique de WebUSE (inspiré de (Nikov et al., 2003)) 58

Figure 2. 2: Exemples des résultats d'évaluations objective et subjective fournis par WebUSE (Nikov et al., 2003) 58

Figure 2. 3: Exemple de résultat final d'évaluation fourni par WebUSE (Nikov et al., 2003)..... 59

Figure 2. 4: Modèle de la méthode d'évaluation d'utilisabilité SUM (Sauro et Kindlund, 2005) 60

Figure 2. 5: Approche d'intégration de méthodes d'évaluation d'utilisabilité en utilisant des profils de problèmes d'utilisabilité (Al-Wabil et Al Khalifa, 2009)..... 61

Figure 2. 6: Architecture de Framework RITA (Charfi et al., 2015)..... 63

Figure 3. 1 : Environnement intégré d'aide à l'évaluation des systèmes interactifs (Ezzedine, 2002) 68

Figure 3. 2 : Les données capturées des différents outils de l'environnement intégré d'aide à l'évaluation des systèmes interactifs 69

Figure 3. 3: Le modèle proposé de Mapping des données 72

Figure 3. 4 : Les étapes de spécification d'un indicateur d'utilisabilité 74

Figure 3. 5 : Définition des besoins d'information d'une façon évolutive 75

Figure 3. 6: Exemple générale de présentation graphique d'un indicateur avec leurs critères de décision..... 82

Figure 4. 1 : Environnement intégré d'aide à l'évaluation des systèmes interactifs..... 99

Figure 4. 2 : Environnement intégré d'aide à l'évaluation des systèmes interactifs basé sur des versions étendues d'outils existants 101

Figure 4. 3 : Architecture du mouchard électronique EISEval étendu 104

Figure 4. 4 : Aperçu de l'interface du module supplémentaire ajouté dans le mouchard électronique EISEval. 105

Figure 4. 5 : Exemple des résultats du fichier XML fourni par le module 8 105

Figure 4. 6 : La structure du design pattern MVC réalisé 107

Figure 4. 7 : L'architecture du composant de l'évaluation subjective mis en place dans l'outil SOUSEval 108

Figure 4. 8 : Aperçu de l'interface d'ajout d'une question 109
Figure 4. 9 : Aperçus des interfaces du module de préparation d'une évaluation subjective 110
Figure 4. 10 : Aperçus des interfaces de planification d'une évaluation subjective 111
Figure 4. 11 : Aperçu de l'interface du sous-module de gestion des critères d'une évaluation subjective 112
Figure 4. 12 : Aperçu de l'interface du sous-module de gestion des questions 112
Figure 4. 13 : Aperçus des interfaces du module de l'évaluation subjective 113
Figure 4. 14 : Aperçus des interfaces du module d'analyse des résultats de l'évaluation subjective 114
Figure 4. 15 : L'architecture globale de l'outil réalisé SOUSEval 115
Figure 4. 16 : Le processus adopté lors de la phase de la conduite d'une évaluation ergonomique 117
Figure 4. 17 : Aperçu de l'interface de chargement du système de sous-module d'évaluation ergonomique 117
Figure 4. 18 : Aperçu de l'interface de la liste des interfaces générées de sous-module d'évaluation ergonomique 118
Figure 4. 19 : Aperçus des interfaces du sous-module d'évaluation des interactions 119
Figure 4. 20 : Aperçu du questionnaire généré à partir du sous-module de l'évaluation subjective 120
Figure 4. 21 : Aperçu de l'interface du sous-module d'insertion des résultats du Mouchard 121
Figure 4. 22 : Aperçu d'un exemple des résultats de questionnaire lors d'une évaluation intégrée 122
Figure 4. 23 : Aperçu de l'interface d'analyse des résultats en utilisant les indicateurs d'utilisabilité 124
Figure 5. 1 : Aperçus des principales interfaces utilisateur du Système d'Aide à l'Information (SAI) 127
Figure 5. 2 : Les informations sur les utilisateurs 129
Figure 5. 3 : Plan d'une session de test d'utilisabilité 131
Figure 5. 4 : Aperçus de quelques interfaces du module de préparation relatives à notre étude de cas 132
Figure 5. 5: Aperçu des résultats de mesures capturées de l'interface Agent_Etat_Ligne_Tram 133
Figure 5. 6 : Aperçu de l'interface de la liste des tâches spécifiées via l'outil SOUSEval 134
Figure 5. 7 : Aperçu de l'interface des résultats de l'indicateur d'efficacité de réalisation des tâches du SAI 135
Figure 5. 8: Aperçu de l'interface des analyses et recommandations générées par l'indicateur d'efficacité de réalisation des tâches du système SAI 136
Figure 5. 9: Aperçu de l'interface des résultats de l'indicateur d'efficience de réalisation des tâches du SAI 137
Figure 5. 10 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur d'efficience de réalisation des tâches du système SAI 138
Figure 5. 11 : Aperçus des interfaces des résultats de l'indicateur de densité informationnelle des interfaces utilisateurs du SAI 139
Figure 5. 12 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur de densité informationnelle sur les interfaces utilisateur du SAI 140
Figure 5. 13: Aperçus des interfaces des résultats de l'indicateur des actions minimales du SAI 141
Figure 5. 14: Aperçu de l'interface des analyses et recommandations générées par l'indicateur d'actions minimales sur les tâches réalisées par SAI 141

Figure 5. 15 : Aperçus des interfaces des résultats de l'indicateur de feedback immédiat aux actions des utilisateurs dans les interfaces du SAI 142

Figure 5. 16 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur de feedback immédiat aux actions des utilisateurs sur les interfaces du SAI 142

Figure 5. 17 : Aperçus des interfaces des résultats de l'indicateur de conformité des couleurs utilisées des interfaces utilisateur du SAI 143

Figure 5. 18 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur de conformité des couleurs utilisées dans les interfaces utilisateur du SAI 144

Figure 5. 19: Aperçus des interfaces des résultats de l'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur du SAI 145

Figure 5. 20: Aperçu de l'interface des analyses et recommandations générées par l'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur du SAI 146

Figure 6. 1 : Positionnement de notre approche par rapport au cadre de référence de (Whitefield et al.,1991) .. 154

Figure 6. 2 : Positionnement de notre approche par rapport au travail de (Rohrer, 2014) 155

Liste des Tableaux

Tableau 1. 1 : Descriptions des différents concepts définis par le standard ISO/IEC 15939 (2007) 25

Tableau 1. 2: Synthèse sur quelques classifications existantes des méthodes d'évaluation des IHM des systèmes interactifs et quelques exemples 32

Tableau 1. 3: Synthèses sur quelques méthodes et outils d'évaluation de l'utilisabilité des interfaces utilisateur 35

Tableau 1. 4 : Quelques exemples de mesures proposées dans le modèle QUIM (Padda, 2003) 38

Tableau 1. 5 : Quelques exemples de mesures définies par la norme ISO/IEC 25022 (2012) 39

Tableau 1. 6 : Quelques exemples de mesures proposées dans la littérature 39

Tableau 1. 7: Critères et sous-critères de guidage proposés par (Bastien et Scapin, 1993) 41

Tableau 1. 8 : Exemples des recommandations ergonomiques relatives aux sous-critères de guidage proposés par (Bastien et Scapin, 1993) 42

Tableau 2. 1: Synthèse globale sur les caractéristiques des questionnaires standardisés existants 47

Tableau 2. 2: Classifications de questionnaires 49

Tableau 2. 3 : Les résultats d'analyse des questionnaires spécifiques d'utilisabilité 52

Tableau 2. 4 : Les résultats d'analyse des questionnaires universels d'utilisabilité 53

Tableau 2. 5 : Exemples des éléments de questionnaires d'utilisabilité en fonction des quelques critères standards 54

Tableau 2. 6: Tableau récapitulatif des approches présentées 64

Tableau 3. 1 : Exemples de résultats des Mappings par rapport à quelques critères d'utilisabilité 72

Tableau 3. 2 : Les critères d'utilisabilité de ISO 9241-11 (1998) et de (Scapin et Bastien, 1997) 74

Tableau 3. 3 : Les besoins d'information spécifiés pour les critères/ sous-critères d'utilisabilité choisis 76

Tableau 3. 4 : Les mesures utilisées pour l'évaluation des critères/ sous-critères d'utilisabilité choisis 77

Tableau 3. 5: Les mesures d'utilisabilité adaptées sur la base des mesures existantes 78

Tableau 3. 6 : Exemple des critères de décision associés à l'indicateur hypothétique 79

Tableau 3. 7: Exemple des critères de décision en spécifiant les intervalles associés à l'indicateur hypothétique 81

Tableau 3. 8 : La liste des indicateurs proposés portant sur les critères/ sous-critères d'utilisabilité choisis 82

Tableau 3. 9 : Spécification du modèle de mesure d'information de l'indicateur d'efficacité de réalisation des tâches 83

Tableau 3. 10 : Spécification du modèle de mesure d'information de l'indicateur d'efficience de réalisation des tâches 85

Tableau 3. 11 : Spécification du modèle de mesure d'information de l'indicateur de la densité informationnelle d'une/des interface(s) utilisateur 87

Tableau 3. 12 : Spécification du modèle d'information de mesure de l'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur 89

Tableau 3. 13 : Spécification du modèle de mesure d'information de l'indicateur de conformité des couleurs des interfaces utilisateur 91

Tableau 3. 14 : Spécification du modèle de mesure d'information de l'indicateur d'actions minimales des tâches réalisées 94

Tableau 3. 15: Spécification du modèle de mesure d'information de l'indicateur de feedback immédiat aux actions des utilisateurs 96

Tableau 4. 1 : Les mesures d'utilisabilité capturées des outils d'évaluation existants 100

Tableau 4. 2 : Les mesures d'utilisabilité mises en place dans les outils existants et les nouvelles implémentations 102

Tableau 4. 3: Le processus de l'évaluation adopté en se basant sur la norme ISO/IEC 25040 107

Tableau 5. 1 : Le scénario des tâches de l'étude de cas 130

Tableau 5. 2 : Tableau de synthèse des résultats des indicateurs d'utilisabilité pour l'évaluation globale du SAI 147

Tableau 5. 3 : Tableau de synthèse des résultats des indicateurs d'utilisabilité pour l'évaluation ergonomique du SAI 147

Tableau 5. 4 : Tableau de synthèse des résultats de l'indicateur d'actions minimales pour l'évaluation des tâches réalisées avec SAI 147

Tableau 6. 1 : Tableau comparatif de notre approche par rapport aux approches présentées dans le chapitre 2. 156

Tableau A. 1: Synthèse globale sur les échelles des questionnaires standardisés existants 172

Tableau A. 2: Aspects de qualité de questionnaires universels d'utilisabilité 174

Tableau A. 3: Aspects de qualité de questionnaires spécifiques à l'utilisabilité 175

Tableau A. 4: Exemples des éléments de questionnaires universels d'utilisabilité en fonction des critères standards 176

Tableau A. 5 : Exemples des éléments de questionnaires spécifiques d'utilisabilité en fonction des critères standards 177

Tableau A. 6 : Synthèse globale sur quelques outils d'évaluation de l'utilisabilité utilisant les connaissances et les règles ergonomiques 180

Tableau A. 7 : Synthèse globale sur quelques outils d'évaluation de recueil des données d'interaction entre l'utilisateur et le système interactif à évaluer 182

Introduction Générale

Motivations

L'**I**nteraction **H**omme-**M**achine (IHM) présente la science et la pratique des systèmes interactifs efficaces impliquant le plus souvent les opérateurs humains et les systèmes informatiques (Thimbleby, 2007). La qualité et l'acceptation d'un système interactif de n'importe quel domaine dépendent hautement de la conception de l'interface utilisateur (Zülch et Stowasse, 2000). L'évaluation de l'utilisabilité est la théorie et l'application de méthodes visant à promouvoir une meilleure qualité des systèmes interactifs (Thimbleby, 2007). Une grande variété de méthodes et outils d'évaluation d'utilisabilité ont été proposées depuis plus d'une vingtaine d'années (comme exemples, nous citons l'évaluation heuristique (Nielsen, 1993 ; Sears et Jacko, 2009), Cognitive Walkthrough (Mahatody et al., 2010), les guides ergonomiques (Vanderdonckt, 1994 ; Scapin et Bastien, 1997 ; Travis, 2014), les observations (Holzinger, 2004)). Chaque méthode à ses propres exigences, possède sa spécificité et peut être appropriée pour une ou plusieurs étapes du cycle de vie de la conception du système interactif.

Actuellement, l'intérêt est orienté vers l'utilisation de différentes méthodes et outils d'évaluation pour avoir des meilleurs résultats d'utilisabilité. Dans la littérature actuelle, divers travaux ont largement recommandé la combinaison de différentes méthodes d'évaluation de l'utilisabilité (Tran et al., 2008 ; Al-Wabil et Al-Khalifa, 2009 ; Lewis, 2014 ; Charfi et al., 2015). Quelques travaux ont étudié la combinaison entre eux (tels que (Al-Wabil et Al-Khalifa, 2009 ; Charfi et al., 2015)). La garantie d'une intégration effective entre des méthodes qui fournissent des résultats hétérogènes requiert encore beaucoup d'efforts en matière de recherche.

Afin d'atteindre cet objectif, cette thèse s'inscrit dans le cadre d'un *environnement intégré d'aide à l'évaluation des systèmes interactifs* proposé par (Ezzedine, 2002) lors du projet ISART (Intégration des Systèmes d'Aide à la Régulation dans les Transports urbains multimodaux flexibles), présenté dans la Figure 1.
Comme le montre cette Figure, cet environnement comporte quatre outils :

(i) Mouchard électronique EISEval ;
(ii) Outil d'aide à la génération de questionnaires ;
(iii) Inspecteur des règles ergonomiques ;
(iv) Outil de synthétise des résultats d'évaluation.

Les trois premiers outils concernent trois aspects hétérogènes d'évaluation, respectivement : **(i)** les interactions entre les utilisateurs et le système à évaluer ; **(ii)** les perceptions subjectives des utilisateurs sur le système ; **(iii)** la qualité ergonomique des interfaces utilisateur. Le dernier outil **(iv)** vise à intégrer ces trois aspects pour améliorer l'évaluation. Cette intégration a pour but de combiner les résultats subjectifs provenant des questionnaires avec les résultats des autres outils d'une façon unique, complémentaire et cohérente. Son objectif principal est d'aider les évaluateurs à avoir une bonne mesure d'utilisabilité en se basant sur les différents résultats hétérogènes tout en facilitant leurs analyses et leurs traitements pour détecter des problèmes précis sur l'utilisabilité des interfaces utilisateur.

Cet environnement a fait l'objet de trois thèses : la présente et deux antérieures qui ont été réalisées entre (2006-2013) au Laboratoire d'Automatique, de Mécanique et d'Informatique industrielles et Humaines (LAMIH), UMR CNRS 8201, à l'Université de Valenciennes et du Hainaut-Cambrésis (UVHC). La première thèse a été réalisée par Chi Dung Tran entre (2006-2009). Elle consiste à concevoir et mettre en œuvre l'outil du mouchard électronique EISEval. La deuxième a été effectuée par Selem Charfi entre (2009-2013). Elle a conduit à la définition et le développement d'un outil d'aide à l'évaluation des systèmes interactifs en utilisant les règles ergonomiques (i.e. l'inspecteur des règles ergonomiques (iii)).

Ces deux thèses nous ont fourni la base nécessaire pour continuer le développement complet de l'environnement intégré dans ce présent mémoire. Notre travail porte sur la conception des outils

manquants, notamment l'outil d'aide à la génération de questionnaires et l'outil de synthèse des résultats d'évaluation, ainsi que la mise en œuvre et l'exécution totale de l'environnement.

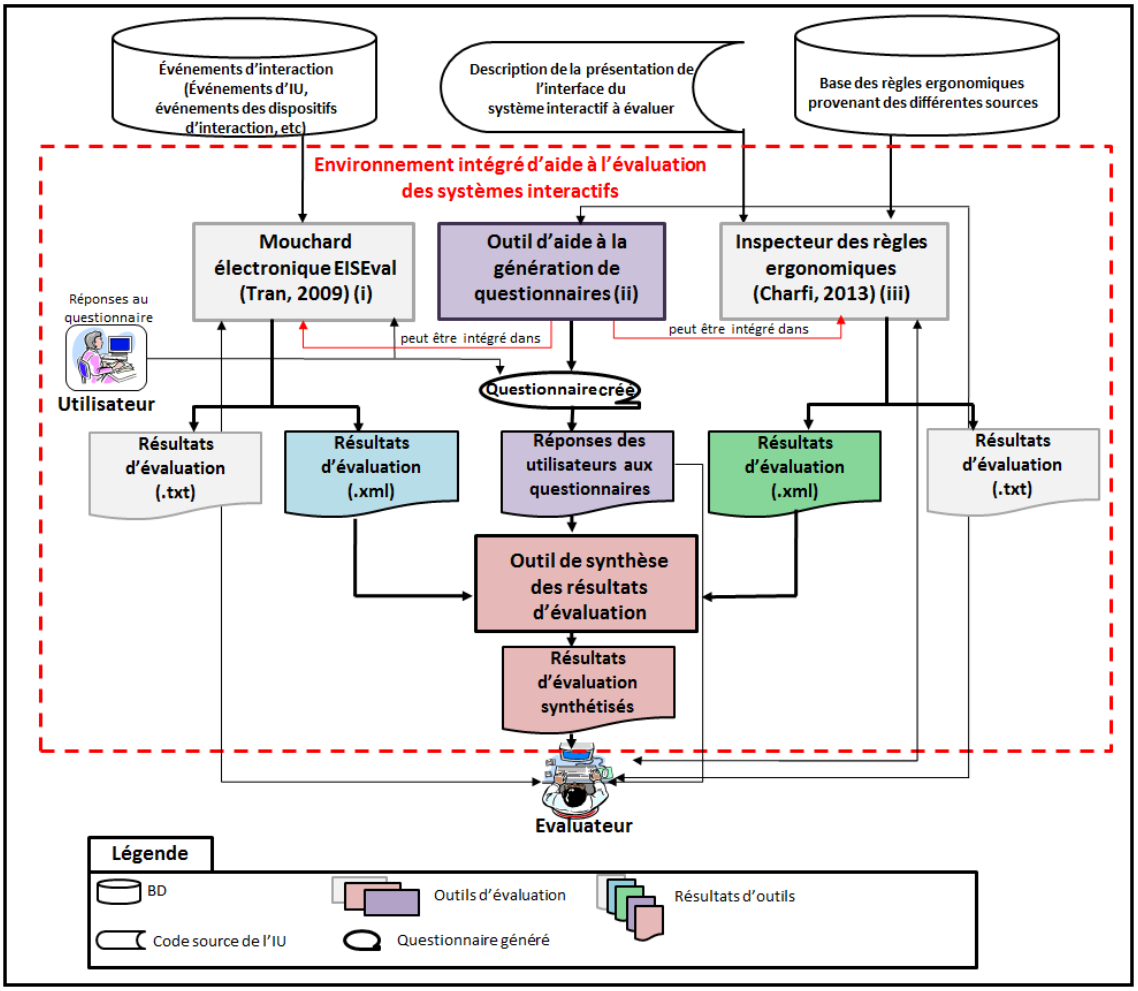


Figure 1 : Environnement intégré d'aide à l'évaluation des systèmes interactifs (Ezzedine, 2002)

Problématique

Dans ce cadre, il est important d'apporter une proposition permettant d'intégrer les résultats fournis par les travaux précédents et l'outil de génération de questionnaires pour réellement compléter la mise en œuvre de l'environnement proposé. De ce fait, la problématique abordée dans notre thèse est la suivante :

> ***Comment intégrer effectivement les résultats des évaluations subjective et objective d'une façon complémentaire et cohérente pour assurer une bonne mesure d'utilisabilité des interfaces utilisateur ?***

Objectif de cette thèse

Cette thèse propose une approche pour intégrer les résultats des évaluations subjective et objective de l'utilisabilité des systèmes interactifs (Figure 1). Pour atteindre cet objectif, un standard de mesure défini par la norme ISO/IEC 15939 a été appliqué. Notre approche fournit un cadre de base pour soutenir l'interprétation des différents résultats. Elle offre ainsi des recommandations permettant d'améliorer l'utilisabilité des systèmes interactifs. Cette approche a été mise en œuvre dans l'environnement via l'outil de synthèse des résultats d'évaluation. Celui-ci inclut un outil d'aide à la génération de questionnaires et leurs exploitations.

Organisation de la thèse

Ce mémoire est structuré autour de cinq chapitres. Dans le premier chapitre, nous introduisons un aperçu général sur l'évaluation de l'utilisabilité des systèmes interactifs. D'abord, nous commençons par la présentation des notions clés portant sur l'évaluation des interfaces utilisateur. Par la suite, nous nous intéressons aux processus de l'évaluation existants définis par les standards. Puis, nous nous intéressons aux processus d'évaluation d'utilisabilité des IHM proposés dans la littérature. Ensuite, nous exposons un panorama de méthodes et outils d'évaluation d'utilisabilité. Ces dernières permettent de recueillir un ensemble de données qui servent par la suite à des mesures. Nous présentons une collection de ces mesures d'utilisabilité existantes permettant d'assurer l'évaluation des interfaces utilisateur. Par la suite, nous spécifierons les concepts portant sur l'évaluation ergonomique des IHM. Enfin, nous achevons ce chapitre par une synthèse et conclusion.

Dans le deuxième chapitre, nous présentons un état de l'art sur les outils d'évaluation d'utilisabilité des IHM. Ce chapitre comporte deux parties : la première concerne une étude bibliographique sur les outils d'évaluation existants basés sur une évaluation subjective ou objective. La deuxième partie décrit les travaux proposés dans la littérature basés sur une combinaison des évaluations subjective et objective. L'objectif de cette seconde partie est de positionner ces travaux par rapport à notre contribution. Ce chapitre est conclu par une synthèse et conclusion sur les différents travaux apportés.

Dans le troisième chapitre, nous présentons notre approche d'intégration des résultats d'évaluation d'utilisabilité subjective et objective. D'abord, nous introduisons la motivation et une présentation générale de l'approche proposée. Ensuite, nous décrivons d'une façon détaillée notre méthodologie et les étapes à suivre en appliquant le standard de mesure ISO/IEC 15939. Par la suite, nous spécifions les différents modèles de mesure d'information proposés permettant l'évaluation des différents critères d'utilisabilité des interfaces utilisateur. Enfin, une conclusion est présentée à la fin du chapitre.

Au quatrième chapitre, nous abordons la mise en œuvre de l'approche proposée. Nous présentons l'ensemble des développements réalisés. Ceux-ci portent principalement sur l'outil de synthèse des résultats d'évaluation proposé baptisé « **S**ubjective **O**bjective **US**ability **Eval**uation » : SOUSEval. Celui-ci va former le cœur de l'environnement proposé pour parvenir à l'objectif principal de l'intégration. Il inclut également l'outil d'aide à la génération de questionnaires. En premier lieu, nous commençons par la présentation de l'environnement intégré ainsi que le positionnement de notre travail réalisé. En second lieu, nous spécifions l'ensemble des développements effectués permettant de mettre en exécution l'environnement intégré. Finalement, nous finissons ce chapitre par une conclusion portant sur le travail réalisé.

Dans le cinquième chapitre, nous nous focalisons à la fois sur la mise en pratique de notre environnement et l'étude de faisabilité de notre approche. D'abord, nous présentons les expérimentations qui correspondent à une étude de cas permettant d'évaluer l'utilisabilité des interfaces utilisateur d'un système interactif dédié à la supervision de transport du Trafic urbain. Ce système est baptisé SAI « **S**ystème d'**A**ide à l'**I**nformation Voyageurs » conçu dans le cadre du projet SART (Système d'Aide à la Régulation du Trafic). Par la suite, nous présentons une discussion qui porte sur les risques de validité de notre travail mené. Enfin, nous achevons ce chapitre par une conclusion.

A la fin de ce présent mémoire, nous présentons une conclusion générale portant sur le travail apporté, son positionnement par rapport aux approches existantes et les perspectives de recherche sur la base des limites rencontrées.

Chapitre 1 : L'évaluation de l'utilisabilité des IHM des systèmes interactifs

"Usability evaluation is an important part of the overall user interface design process, which ideally consists of iterative cycles of designing, prototyping, and evaluating"
(MELODY Y. IVORY & MARTI A. HEARST, 2001)

Introduction

Après plus d'une vingtaine d'années de recherche et de pratique, il est indubitable que l'évaluation de l'utilisabilité des interfaces utilisateur est une activité essentielle dans la production des systèmes interactifs de qualité (Nielsen, 1993 ; Zhang et al., 1999 ; Rubin et Chisnell, 2008 ; Charfi, 2013). L'évaluation présente l'un des plus grands intérêts de la communauté d'Interaction Homme-Machine. De ce fait, la garantie d'une haute qualité des systèmes interactifs dans des conditions favorables offre la réussite dans l'industrie et la satisfaction de l'être humain (utilisateur). Néanmoins, l'obtention d'une excellente qualité requiert la connaissance, l'exploitation et la mise en œuvre des concepts, méthodes et outils d'évaluation des interfaces utilisateur qui ont été proposés dans ce contexte par la communauté d'IHM. La distinction entre les différents concepts (tels que les critères d'évaluation, les mesures à utiliser, les processus à adopter, etc.) est importante pour assurer l'évaluation et promouvoir la qualité des systèmes interactifs. De ce fait, nous proposons à travers ce premier chapitre d'introduire les éléments de base constituant ce mémoire de thèse.

En effet, ce chapitre est structuré autour de six sections. La première introduira un ensemble des notions sur les concepts généraux d'évaluation d'utilisabilité des IHM. La seconde portera sur les

processus d'évaluation proposés dans ce contexte. Ensuite, la troisième section présentera une revue de littérature des méthodes et outils d'évaluation existants d'utilisabilité des interfaces utilisateur. Compte tenu de notre cible d'évaluation (*Cf. Chapitre 3*) la quatrième section sera consacrée à la présentation d'un panorama sur les mesures d'utilisabilité dédiées à l'évaluation des IHM. La cinquième portera sur l'évaluation ergonomique des interfaces utilisateur. Finalement, ce chapitre sera clôturé par une synthèse et une conclusion.

1.1 Définitions générales des concepts d'évaluation des IHM

Dans cette section, nous présentons les notions des concepts généraux portant sur l'évaluation des interfaces utilisateur.

1.1.1 L'évaluation des IHM

L'évaluation des *Interfaces utilisateur*[1] est un domaine de recherche très actif visant principalement à produire des systèmes interactifs optimaux et d'une haute qualité (Nielsen, 1993 ; Mariage et al, 2004 ; Tran , 2009). Au fil des ans, diverses définitions de l'évaluation des IHM ont été proposées, néanmoins elles convergent toutes vers le même principe qui consiste à *vérifier* et à *valider* les interfaces utilisateur, dans le but de garantir un fonctionnement optimal des systèmes interactifs (Moussa et al., 2006 ; Brangier, 2007 ; Charfi, 2013). La vérification consiste à montrer si les interfaces correspondent aux spécifications issues de la définition des besoins. Quant à la validation, elle permet de montrer leurs correspondances aux besoins en respectant les contraintes du domaine d'application. Dans le cas contraire, les insuffisances dégagées du système par rapport à des critères identifiés a priori doivent être mises en évidence (Grislin et Kolski, 1996). Dans certains travaux de recherche, l'évaluation des IHM est présentée comme étant une comparaison entre un modèle d'objet à évaluer par rapport à un modèle de référence permettant d'établir des conclusions (Senach, 1990 ; kolski et Ezzedine, 2003), comme le montre la Figure 1.1.

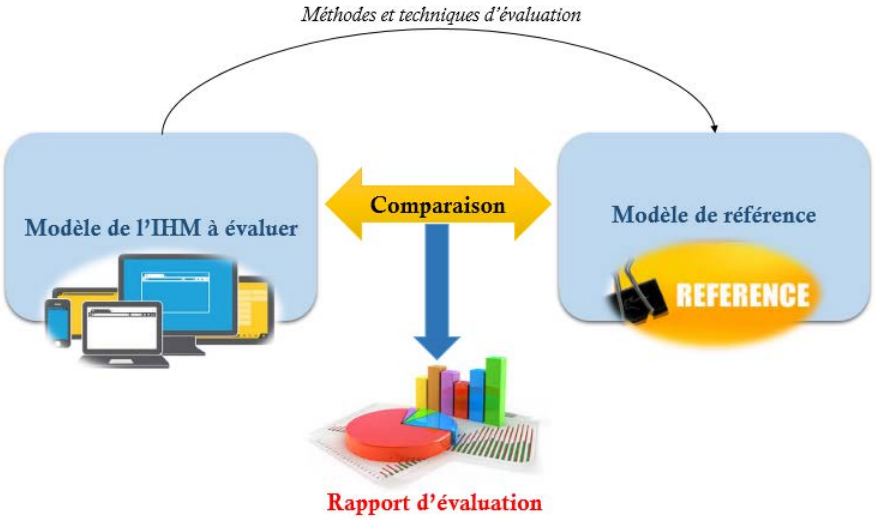


Figure 1. 1 : Principe de base de l'évaluation des IHM (inspiré de (Senach, 1990))

L'évaluation de qualité des IHM vise à accomplir un ensemble d'objectifs dont nous citons :

- L'amélioration de l'acceptabilité des systèmes interactifs (Zhang et al., 1999) ;
- La facilité de la détection de problèmes des interfaces utilisateur portant sur différents aspects de qualité tels que la lisibilité et le guidage (Tran et al., 2008) ;

[1]Une interface utilisateur (aussi connue sous le nom Interface Homme-Machine (IHM)) est l'ensemble des moyens avec lesquels les utilisateurs interagissent avec un système (une machine particulière, dispositif, programme informatique ou d'un autre outil complexe) (Sodiya, 2009).

- La réduction de risques d'erreurs que les utilisateurs peuvent commettre (Wickens et al., 2004 ; Olsen et Dan, 2007 ; Tran, 2009) ;
- La mesure du confort et de la satisfaction des utilisateurs et l'accroissement de la productivité (Rubin et chisnell, 2008 ; Tran, 2009) ;
- La production des systèmes interactifs de qualité (Nielsen, 1993 ; Mariage et al., 2004) ;
- La mise en relief sur les défauts de la conception afin d'aider les concepteurs à améliorer les systèmes interactifs déjà conçus (Nielsen, 1993, Tran, 2009, Assila et al., 2013a) ;
- La vérification de la capacité des utilisateurs à bien accomplir leurs tâches lors de l'utilisation du système interactif et qu'il répond à leurs besoins (Ezzedine et al., 2008).

Afin de promouvoir une meilleure qualité des systèmes interactifs, l'utilisabilité est un aspect primordial à prendre en compte pendant l'évaluation de leurs interfaces utilisateur.

1.1.2 L'utilisabilité

L'évaluation des IHM des systèmes interactifs porte principalement sur le concept de l'utilisabilité comme étant un facteur de qualité important (Nielsen, 1993 ; Bastien et Scapin, 2001 ; Ezzedine, 2002 ; Seffah et al., 2006 ; Rubin et Chisnell, 2008). Ce concept a été largement défini dans la littérature et les normes internationales (IEEE Std.610.12, 1990 ; ISO 9241-11, 1998 ; ISO/IEC 9126-1, 2001).

Parmi les définitions proposées par les standards, l'utilisabilité a été définie comme suit :

> *"La facilité avec laquelle un utilisateur peut apprendre à faire fonctionner, préparer des entrées, et interpréter les résultats d'un système ou d'un composant"* **(IEEE Std.610.12, 1990)**

> *" Le degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié."* **(ISO 9241-11, 1998)**

> *~~L~~a capacité du produit logiciel à être compris, appris, utilisé et attractif pour l'utilisateur, lorsqu'il est utilisé dans des conditions spécifiées."* **(ISO/IEC 9126-1, 2001)**

Différentes autres définitions d'utilisabilité ont été proposées par les normes et les travaux de recherche. Ces définitions ont porté sur des critères plus spécifiques d'utilisabilité. A titre d'exemples, Shackel (1991) s'est focalisé sur cinq critères (l'efficacité, la facilité d'apprentissage, la rétention, l'erreur et l'attitude). Schneiderman (1992) a mis l'accent sur l'efficacité et l'efficience. Nielsen (1993) a pris en considération l'efficacité, la facilité d'apprentissage, la mémorisation, les erreurs, la sécurité et la satisfaction.

La norme ISO 9241-11 (1998) a identifié l'efficacité, l'efficience et la satisfaction dans un contexte d'utilisation spécifié, comme étant les principaux attributs d'utilisabilité. D'autres standards tels que le cas de l'ISO/IEC 9126 (2001) ont défini l'utilisabilité en termes de l'intelligibilité, la facilité d'apprentissage, l'opérabilité et l'attractivité. Cette norme a récemment évolué sous la norme de SQUARE (Systems and software Quality Requirements and Evaluation) (ISO/IEC 25010, 2011), y compris les critères suivants : la protection contre l'erreur de l'utilisateur, l'esthétique de l'interface utilisateur et l'accessibilité. Cette norme a proposé un modèle de qualité en utilisation qui a également

intégré les trois critères d'utilisabilité définis par la norme ISO 9241-11 (1998), comme le montre la Figure 1.2.

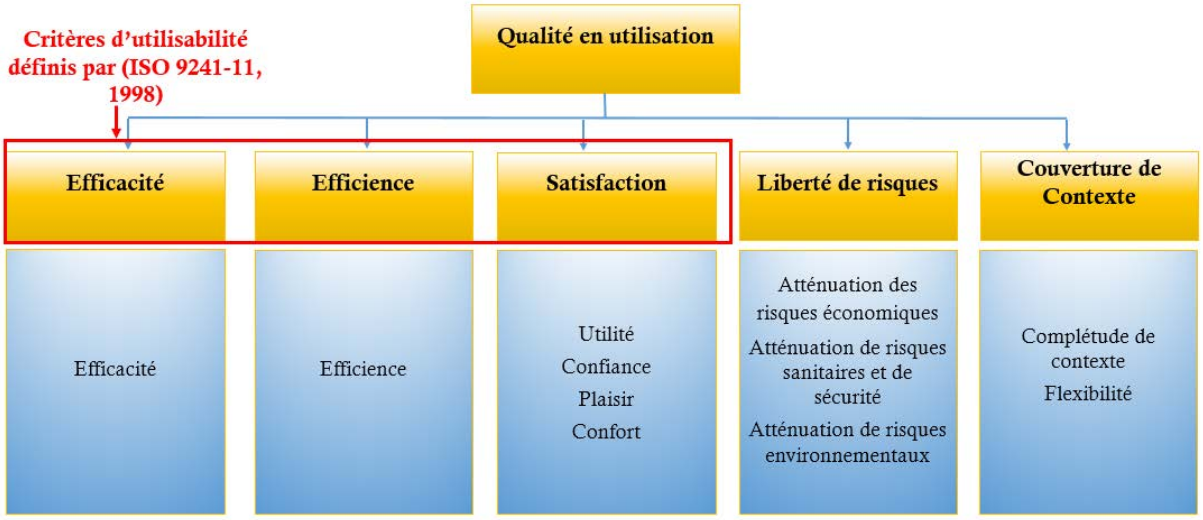


Figure 1. 2 : Modèle de qualité en utilisation proposé par la norme SQuaRE (adapté de (ISO/IEC 25010, 2011))

De nombreux autres travaux se sont focalisés sur la définition des critères d'utilisabilité portant sur l'aspect ergonomique de l'interface utilisateur (voir Section 1.5). Par exemple, Vanderdonckt (1994) a défini huit critères ergonomiques : (*la compatibilité, la cohérence, la charge de travail, l'adaptabilité, le contrôle de dialogue, la représentativité, le guidage et la gestion d'erreur*). En outre, Scapin et Bastien (1997) ont proposé huit critères ergonomiques (*le guidage, la charge de travail, le contrôle explicite, l'adaptabilité, la gestion des erreurs, l'homogénéité/Cohérence, la signifiance des codes et dénominations et la compatibilité*), incluant au total 13 sous-critères visant à évaluer l'utilisabilité des interfaces utilisateur. Ces critères ont été largement utilisés (Filho et al., 2015 ; Castro et al., 2015) pour différents types d'interfaces tels que les interfaces utilisateur graphique (GUI) comme WIMP, les interfaces 3D, les interfaces utilisateur Web (WUI).

Cette hétérogénéité au niveau des définitions d'utilisabilité existantes dans la littérature a été abordée par certains travaux de recherche citons comme exemples (Abran et al., 2003 ; Seffah et al., 2006 ; Lewis, 2014). Différentes discussions ont été établies à ce propos. La plus récente est celle conduite par Lewis (2014) ayant révélé une des leçons, importante à apprendre sur l'utilisabilité. Selon (Lewis, 2014): "*There is yet to be a clear and generally accepted definition*". Par ailleurs, toutes ces définitions s'entendent généralement sur les mesures d'efficacité, d'efficience, de la satisfaction ou l'absence de problèmes d'utilisabilité (Lewis, 2014). En outre, selon d'autres travaux (Nielsen, 1993 ; Dix, 1998 ; Hornbæk, 2006), l'utilisabilité ne peut pas être mesurée directement. Néanmoins, elle est généralement liée à plusieurs critères qui peuvent être mesurés.

Dans notre contexte, nous considérons l'utilisabilité en termes d'attributs et des critères à prendre en compte dans une évaluation. De ce fait, l'évaluation de ces critères requiert l'utilisation d'un ensemble des mesures portant sur les interfaces utilisateur (Hornbæk et Law, 2007).

1.1.3 Les mesures

L'utilisation des mesures pour l'aide à l'évaluation de la qualité est un sujet largement exploré en génie logiciel. Plusieurs méthodologies ont été proposées pour soutenir la définition des mesures telles que Software Quality Measure (Suetendael et Elwell, 1991), Goal-Question-Metric (Basili et al., 1994), Goal-Question-Indicator-Metric (Park et al., 1996), Software Measurement (Morasca, 2001) et Practical Software Measurement (McGarry et al., 2002). D'autres méthodologies relatives à la

spécification des mesures dédiées à l'évaluation des applications web ont été aussi proposées (par exemple (Olsina et al., 2008)).

Concernant l'évaluation de l'utilisabilité des interfaces utilisateur, l'utilisation des mesures a été largement exploitée dans la littérature (*Cf. Section 1.4*). De nombreuses mesures ont été proposées par les standards (tels que ISO 9126, (2001) ; ISO/IEC 25022, (2012)) et les modèles d'évaluation très connus comme le modèle MUSIC (Macleod et al., 1997) et le modèle QUIM (Seffah et al., 2006)).

A travers ces travaux et standards, nous avons distingué une hétérogénéité en terminologies adoptées. Certains standards (tels que (ISO/IEC 14598 (1999) et ISO 9126 (2001)) se réfèrent au terme « Métrique », d'autres standards comme ISO/IEC 25022 (2012) adoptent plutôt le terme « mesure ». Certains autres travaux portant sur l'évaluation des interfaces utilisateur ont considéré les deux termes à la fois (Bevan et Macleod, 1994). De ce fait, la spécification des concepts de base qui correspondent aux notions de « mesure et métrique » est cruciale afin de bien l'exploiter.

La distinction entre les différentes terminologies des mesures a été bien explorée dans le domaine du génie logiciel. En effet, le terme « Métrique » a été publié dans le deux séries de standards ISO/IEC 14598 (1999–2001) et ISO/IEC 9126 (2001–2004) qui correspondent aux premiers consensus internationaux sur la terminologie pour les caractéristiques de qualité pour l'évaluation du produit logiciel.
Selon ces deux standards : une métrique est définie comme suit :

> *" Une échelle de mesure et la méthode utilisée pour la mesurer"* (**ISO/IEC 14598-1, 1999**)

> *"La méthode de mesure définie et l'échelle de mesure"* (**ISO/IEC 9126-2, 2003**)

IEEE Std. 1061 (IEEE Standard for a Software Quality Metrics Methodology), ancienne norme des méthodologies de mesure de la qualité logicielle, a défini une métrique comme suit :

> *" Une fonction dont les entrées sont les données du logiciel et dont la sortie est une valeur numérique unique qui peut être interprété comme le degré auquel le logiciel possède un attribut donné qui affecte sa qualité"*
>
> (**IEEE Std 1061, 1998**)

Afin d'harmoniser les différentes normes d'ingénierie des systèmes, l'ISO (the International Organization for Standardization) et l'IEC (the International Electrotechnical Commission) ont créé un comité technique mixte (JTC1: «Technologies de l'information», www.jtc1.org) (Garcia et al., 2006 ; ISO/IEC 25000, 2014 ; ISO/IEC 25022, 2012). Celle-ci a proposé une nouvelle génération de normes de qualité du produit logiciel, nommée (SQuaRE – ISO/IEC 25000). Cette série de normes a remplacé les deux séries de standards ISO/IEC 14598 et ISO/IEC 9126. Elle a inclus explicitement dans ses directives, les procédures qui garantissent l'uniformité et la cohérence entre ses normes (ISO/IEC 25000, 2014). De plus, elle a établi des critères pour la spécification des exigences de qualité des systèmes et du produit logiciel, leurs mesures et leurs évaluations (ISO/IEC 25000, 2014).

Dans ces nouvelles séries de normes, SQuaRE a évité l'utilisation de terme « métrique » adopté dans les normes ISO/IEC 9126 et ISO/IEC 14598 et elle l'a remplacé par le terme « mesure », conformément à la norme ISO/IEC 15939 (2007) (Abran et al., 2006).

Celle-ci a été adoptée dans les séries de normes SQuaRE (telles que (ISO/IEC 25000, 2014 ; ISO/IEC 25021, 2012 ; ISO/IEC 25022, 2012)). Ce standard a défini une mesure comme étant:

> *"Une variable à laquelle une valeur est assignée comme résultat de mesure."*
>
> (**ISO/IEC 15939, 2007**)

Par ailleurs, selon la norme NIST (National Institut of Standards and Technology), les termes —métrique et mesure" ont un certain chevauchement (Schofield, 2005). Le terme *mesure* est utilisé pour les attributs les plus concrets. Quant au terme métrique, il s'adresse aux attributs les plus abstraits au plus haut niveau. Néanmoins, aucune distinction entre les deux termes n'a été fournie par les standards d'ISO/IEC. De notre part, nous adoptons le terme « mesure » en se référant à la norme SQuaRE.

En outre, le standard ISO/IEC 15939 (2007) a classifié les mesures selon trois catégories comme suit :

Mesure de Base
"Une mesure définie en fonction d'un attribut et la méthode pour la quantifier."
(ISO/IEC 15939, 2007)

Mesure dérivée
"Une mesure définie en fonction de deux ou plusieurs valeurs de mesures de base."
(ISO/IEC 15939, 2007)

Indicateur
"Une mesure qui fournit une estimation ou une évaluation des attributs spécifiés dérivés d'un modèle compte tenu des besoins d'information définis. Les indicateurs sont la base d'analyse et la prise de décision."
(ISO/IEC 15939, 2007)

En effet, un indicateur est constitué par un ensemble de mesures (de base et dérivées). La sélection de mesures utilisées qui spécifie un indicateur est basée sur leur besoin d'information (l'objectif à atteindre en utilisant l'indicateur). La mesure d'indicateur a été structurée autour d'un modèle de mesure d'information qui schématise les concepts de base à adopter pour la construction d'un indicateur. En outre, ce modèle présente une structure reliant les besoins aux entités concernées qui caractérisent les attributs d'intérêt comme l'illustre la Figure 1.3.

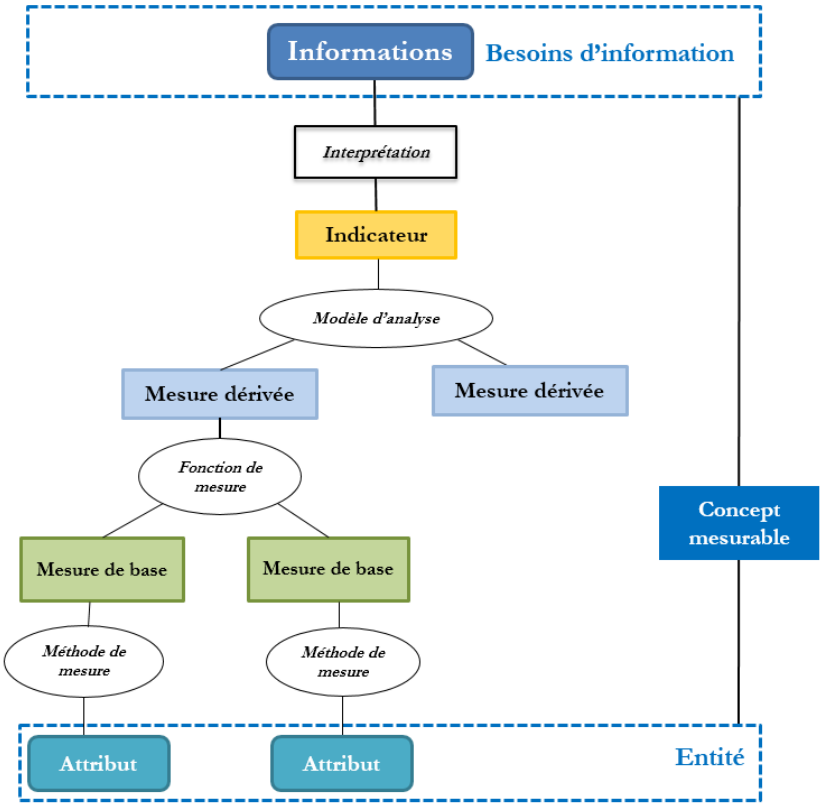


Figure 1. 3 : Structure et relations clés de modèle de mesure d'information (ISO/IEC 15939, 2007)

Dans ce qui suit, nous présentons dans le Tableau 1.1, les notions de base portant sur les concepts des mesures définis par ce standard.

Tableau 1. 1 : Descriptions des différents concepts définis par le standard ISO/IEC 15939 (2007)

Concepts	**Définitions**
Besoins d'information	Connaissances nécessaires pour gérer les objectifs, les buts, les risques et les problèmes
Informations	Un ou plusieurs indicateurs et leurs interprétations associées qui répondent à un besoin d'information
Concept mesurable	Relation abstraite entre les attributs des *entités* et des besoins d'information
Entité	Objet qui doit être caractérisé par la mesure de ses *attributs*. Une entité peut être un processus, produit, projet ou une ressource.
Attribut	Propriété ou particularité d'une entité qui peut être distinguée d'une façon quantitative ou qualitative par des moyens humains ou automatisés.
Mesure de base	Mesure définie en fonction d'un attribut et la méthode pour la quantifier.
Mesure dérivée	Mesure définie en fonction de deux ou plusieurs valeurs de mesures de base.
Méthode de mesure	Séquence logique des opérations, décrite de manière générique, utilisée pour quantifier un attribut par rapport à une *échelle* spécifiée. Le type de méthode de mesure dépend de la nature des opérations utilisées pour quantifier un attribut. Deux types de méthode peuvent être distingués : - **Subjective :** la quantification en impliquant un jugement humain ; - **Objective :** la quantification basée sur des règles numériques
Echelle	Ensemble ordonné de valeurs, continues ou discrètes, ou un ensemble de catégories à laquelle l'attribut est mappé. Le type de l'échelle dépend de la nature de la relation entre les valeurs de l'échelle. Quatre types d'échelle sont généralement définis : - **Nominale :** les valeurs de mesure sont catégoriques. - **Ordinale :** les valeurs de mesure sont classées et ordonnées. - **Intervalle :** les valeurs de mesure ont les mêmes distances correspondant à des quantités égales de l'attribut (la valeur de zéro est impossible). - **Ratio :** les valeurs de mesure ont les mêmes distances correspondant à des quantités égales de l'attribut (la valeur de zéro est incluse). Une *unité de mesure* est souvent associée à une échelle.
Unité de mesure	Quantité particulière, définie et adoptée par convention, à laquelle d'autres grandeurs de même nature sont comparées afin d'exprimer leur ampleur par rapport à la quantité.
Fonction de mesure	Algorithme ou calculs effectués pour combiner deux ou plusieurs mesures de base.
Indicateur	Mesure qui fournit une estimation ou une évaluation des attributs spécifiés dérivé d'un *modèle* compte tenu des besoins d'information définis
Modèle d'analyse	Algorithme ou calculs combinant une ou plusieurs mesures de base et / ou dérivées avec des *critères de décision* associés.
Critères de décision	Seuils, cibles, ou modèles utilisés pour déterminer la nécessité d'une action ou d'une enquête plus approfondie, ou pour décrire le niveau de confiance dans un résultat donné.
Valeur de l'indicateur	Résultat numérique ou catégorielle attribué à un indicateur.
Interprétation	Description de la façon dont l'indicateur devrait être interprété.

Le modèle d'analyse relatif à un indicateur requiert la spécification de types de données des mesures avec lesquelles les critères de décisions seront associés.

Dans la littérature, nous distinguons généralement deux types de données :

- *Les données continues et discrètes :*

> *" Une donnée est dite continue si elle est mesurée le long d'un continuum en tout endroit au-delà du point décimal. Les données continues peuvent être mesurées en unités entières ou fractionnées quantifier."*
>
> (**Privitera, 2015**)

> *" Une donnée est dite discrète si elle est mesurée en unités entières ou des catégories qui ne sont pas distribuées le long d'un continuum."*
>
> (**Privitera, 2015**)

- *Les données qualitatives et quantitatives :*

> *" Une donnée quantitative varie par la quantité. Elle présente une donnée numérique et souvent collectée par la mesure et le comptage."*
>
> (**Privitera, 2015**)

> *" Une donnée qualitative varie par classe. Elle est souvent représentée comme un label et décrit les aspects non numériques."*
>
> (**Privitera, 2015**)

En effet, tous ces concepts de mesures peuvent être adoptés pour assurer l'évaluation de l'utilisabilité des interfaces utilisateur. Tout comme mentionné dans le Tableau 1.1, la définition d'une mesure dépend de la méthode utilisée qui peut être de deux types : subjective ou objective (ISO/IEC 15939, 2007). Par conséquent, le choix du type de méthode de mesure peut dépendre du type d'évaluation à adopter.

Dans ce qui suit, nous nous intéressons aux principaux types d'évaluation d'utilisabilité existants.

1.1.4 Les types d'évaluation d'utilisabilité

Généralement, nous distinguons différents types d'évaluation. En effet, l'évaluation de l'utilisabilité des interfaces utilisateur est souvent réalisée pour collecter des informations (i) avant et/ou après le développement d'un système ; (ii) pour améliorer et affiner sa qualité ; (iii) pour inspecter les jugements des utilisateurs ou (iv) les données de l'interface relatives à la qualité d'un système (Russ-Eft et Preskill, 2001).

Dans ce contexte, nous distinguons quatre types d'évaluation qui peuvent s'articuler autour de deux catégories comme suit :

- ***Évaluation subjective versus Evaluation objective***

L'évaluation subjective représente typiquement les opinions personnelles des utilisateurs, expert ou non, portant sur la qualité des interfaces utilisateur du système à évaluer (Bernsen et Dybkjær, 2009). Son objectif porte sur la détection des problèmes d'utilisabilité à partir d'un point de vue subjectif (Nishiuchi et Takahashi, 2015). Tandis que l'évaluation objective est généralement effectuée par des évaluateurs internes ou par une équipe d'évaluation externe. Son but est d'obtenir des résultats qui sont indépendants des opinions personnelles des utilisateurs ou des sujets (Bernsen et Dybkjær, 2009).

- ***Évaluation formative versus Evaluation sommative***

En général, en pédagogie, une évaluation est dite *formative* lorsqu'elle est censée améliorer les apprentissages en cours en détectant les difficultés de l'apprenant (diagnostic) afin de lui venir en aide (remédiation) (Allal, 1991). Elle est dite *sommative* lorsqu'elle est destinée à la certification en visant à sanctionner la réussite des apprentissages (Gerard, 2013).

Concernant l'évaluation de l'utilisabilité des IHM, une évaluation *formative* (parfois nommée interne) est une évaluation portant sur l'inspection des problèmes d'utilisabilité qui doivent être résolus au cours de la phase de conception du prototype avant la phase de la finalisation (Hartson et al., 2003). Elle vise à identifier et diagnostiquer les problèmes, ensuite mettre en œuvre des recommandations, puis apporter des améliorations à la conception du système évalué et l'évaluer à nouveau (Albert et Tullis, 2013 ; Lewis, 2014). Quant à l'évaluation *sommative*, elle se focalise sur l'évaluation d'efficacité de la conception finale d'un système. Son but est de déterminer à quel point un système permet de répondre à ses objectifs (Albert et Tullis, 2013). Cette évaluation porte sur la comparaison du niveau d'utilisabilité réalisé dans une conception d'un système interactif. Elle peut être utilisée aussi dans le but de comparer les alternatives concurrentes de conception en termes d'utilisabilité (Hartson et al., 2003).

Selon (Ivory et Hearst, 2001), l'évaluation de l'utilisabilité est elle-même un processus qui comporte plusieurs activités en fonction de la méthode adoptée. Dans la section suivante, nous allons aborder les processus d'évaluation les plus utilisés et référencés dans la littérature.

1.2 Les processus d'évaluation

Plusieurs processus d'évaluation ont été proposés. Dans cette section, nous présentons quelques uns parmi les plus connus et les plus référencés dans la littérature. Ceux-ci portent d'une part sur les processus d'évaluation de produits logiciels existants vu sous l'angle des standards et d'autre part sur les processus d'évaluation d'utilisabilité proposés dans le contexte d'IHM.

1.2.1 Les processus d'évaluation existants vus sous l'angle de standards

Dans la littérature, différents standards (tels que ISO 14598-5 (1998) et ISO/IEC 25040, (2011)) ont porté sur la définition d'un formalisme général sur le processus d'évaluation standardisé de la qualité des produits logiciels.
En s'inspirant de la norme ISO/IEC 25040 (2011), le processus d'évaluation s'articule généralement autour de cinq étapes essentielles, comme l'illustre la Figure 1.4 :

1. **Etablir les exigences d'évaluation** : cette étape consiste à établir les objectifs d'évaluation ; identifier les exigences de qualité de produits logiciels ; identifier les parties du produit à inclure dans l'évaluation ; identifier les intervenants et définir la rigueur de l'évaluation.
2. **Spécifier l'évaluation** : cette étape consiste à sélectionner les mesures de qualité à adopter qui dépendent des exigences spécifiées d'évaluation et du contexte de l'évaluation ; établir les critères de décisions qui correspondent aux mesures de qualité et ensuite identifier les critères de décisions concernant l'évaluation.
3. **Concevoir l'évaluation** : Il s'agit de spécifier un plan d'évaluation à décrire dans un cahier des charges de l'évaluation. Différents attributs doivent être spécifiés tels que l'objectif de l'évaluation, l'environnement de l'évaluation, les méthodes et outils d'évaluation impliqués, les critères de décision relatifs aux mesures de qualité, etc.
4. **Exécuter l'évaluation** : Cette étape dépend principalement de plan d'évaluation spécifié. Elle consiste d'abord à collecter les valeurs des mesures de qualité adoptées. Ensuite, il s'agit d'appliquer les critères de décision portant sur les mesures de qualité et celles sur l'évaluation.

5. **Conclure l'évaluation** : Cette étape consiste à analyser les résultats de l'évaluation et élaborer le rapport d'évaluation qui inclut une liste de critiques portant sur les problèmes détectés pour améliorer la qualité du logiciel.

Chaque étape peut porter sur une ou plusieurs questions :

- *La spécification des exigences d'évaluation* peut répondre aux questions suivantes :
  - Pourquoi évaluer ?
  - Qui va procéder à l'évaluation ?
  - Qu'est ce qu'on va évaluer ?
- *La spécification de l'évaluation* peut répondre aux deux questions :
  - Où procéder à l'évaluation ?
  - Quelles sont les méthodes et les critères à adopter ?
- *La conception de l'évaluation* qui permet de produire un plan d'évaluation répond à la question : Quand évaluer ?
- *L'exécution de l'évaluation* porte sur la question : Comment se déroule l'évaluation ?
- *La conclusion de l'évaluation* porte sur la question : Quels sont les résultats de l'évaluation ?

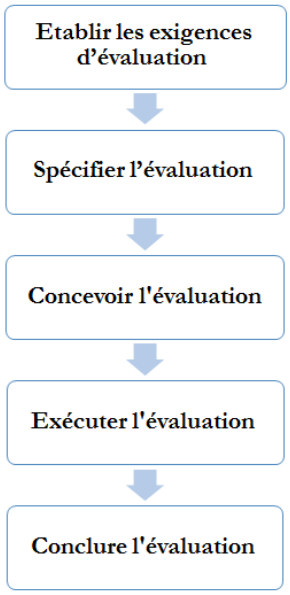


Figure 1. 4 : Processus simplifié d'évaluation de la qualité de produit logiciel (ISO/IEC 25040, 2011)

1.2.2 Les processus d'évaluation existants vus sous l'angle de l'IHM

Dans le contexte de l'évaluation de la qualité des IHM, il n'existe pas de standards spécifiques qui ont formalisé le processus d'évaluation. Néanmoins, de nombreux processus d'évaluation ont été proposés dans la littérature (Ivory et Hearst, 2001 ; Mariage, 2005 ; Seffah et Habieb-Mammar, 2009). Chaque processus est constitué d'un ensemble d'activités qui comprennent en commun généralement les trois activités suivantes :

- *La capture* concerne la collecte de données de l'utilisabilité ;
- *L'analyse* concerne le traitement et l'interprétation des données d'utilisabilité dans le but d'identifier les problèmes liés à l'interface ;
- *La critique* concerne la suggestion des solutions ou des améliorations pour atténuer les problèmes.

En s'inspirant des travaux de Mariage (2005), le processus d'évaluation s'articule généralement autour de quatre étapes, comme l'illustre la Figure 1.5. Les trois activités de capture, analyse et critique sont introduites dans la deuxième étape « *Conduite* ». Chaque étape vise certains objectifs. L'ensemble de ces étapes concerne :

1. **La planification.** Cette phase consiste à assurer l'évaluation et à spécifier :
   - Les objectifs de l'évaluation en tenant compte des intervenants de l'évaluation (concepteur, évaluateur, utilisateur, etc.) et de type des résultats désirés.
   - Le contexte d'utilisation du système à évaluer est composé du triplet : Utilisateur, Plate-forme et Environnement physique de travail.
   - Le choix de la (s) méthode(s) à appliquer en se basant sur la spécification des objectifs de l'évaluation.
   - Le protocole d'évaluation qui comporte les différents points déjà cités suivi par le scénario de l'évaluation s'il existe (défini par l'évaluateur).
2. **La conduite.** Cette étape est fondée sur les objectifs de l'évaluation. Elle vise à découvrir les problèmes relatifs à l'interface à évaluer en suivant les étapes du scénario d'évaluation. Elle se focalise sur les trois activités de collecte de données, analyse et critique. Elle est précédée par une étape de préparation afin de mettre en place les outils nécessaires pour la conduite de l'évaluation.
3. **La finalisation.** Cette étape consiste à élaborer le rapport final de l'évaluation contenant les problèmes de l'interface détectés durant la phase de la conduite de l'évaluation. Par la suite, une communication entre les différents intervenants est élaborée pour interpréter les résultats obtenus.
4. **Le suivi de l'évaluation**. Suite à la communication entre les intervenants, différents conseils et/ou critiques concernant la qualité des interfaces évaluées peuvent être spécifiés. Ces derniers peuvent approuver la bonne qualité des interfaces ou indiquer le suivi de l'évaluation afin d'améliorer la qualité des interfaces.

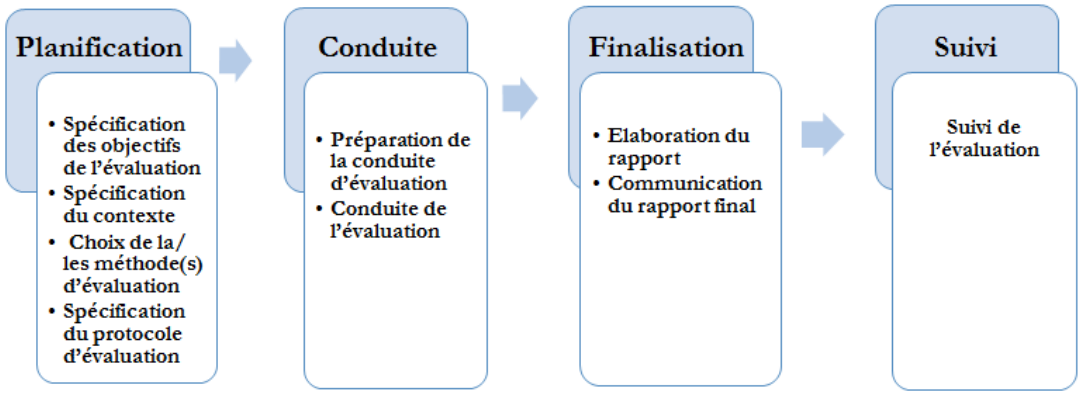


Figure 1. 5 : Processus d'évaluation des interfaces utilisateur (inspiré de (Mariage, 2005))

A propos de l'évaluation de l'utilisabilité des interfaces utilisateur, Ivory et Hearst (Ivory et Hearst, 2001) ont proposé un processus qui s'articule autour de 12 étapes. Comme l'illustre la Figure 1.6, les trois activités de capture, analyse et critique sont introduites respectivement dans les étapes 8, 9 et 10.

1. **Spécifier les objectifs d'évaluation d'utilisabilité.** Cette phase consiste à définir clairement les objectifs visés par l'évaluation qui peuvent porter à titre d'exemples sur la précision des exigences de l'interface utilisateur, l'identification des problèmes d'utilisabilité spécifiques, l'amélioration de la performance de l'interface, etc. La spécification des objectifs influence généralement d'autres aspects d'évaluation tels que le choix de la méthode d'évaluation.

2. **Déterminer les aspects d'interface utilisateur à évaluer.** Dans cette étape, il s'agit d'indiquer l'ensemble des aspects d'interface utilisateur à considérer lors de l'évaluation. Ces aspects doivent correspondre et être compatibles avec les objectifs d'évaluation définis.
3. **Identifier les utilisateurs cibles.** L'identification des utilisateurs cibles par l'évaluation est requise. Cette phase porte sur la spécification des caractéristiques et les profils des utilisateurs à impliquer dans l'évaluation.
4. **Sélectionner les mesures d'utilisabilité.** Dans cette étape, la sélection des mesures d'évaluation est essentielle pour choisir celles qui correspondent le mieux aux objectifs de l'évaluation et qui révèlent un maximum de détails de l'utilisabilité pour l'interface utilisateur à évaluer. Dans la littérature, de nombreuses mesures d'utilisabilité portant sur une variété des aspects d'interface utilisateur (tels que l'efficacité, l'efficience, la densité, etc.) ont été proposées (Voir *Section 1.4*).
5. **Sélectionner la/les méthode(s) d'évaluation.** Le choix d'une ou plusieurs méthodes d'évaluation de l'utilisabilité est une étape primordiale du processus d'évaluation. Il existe de nombreuses méthodes permettant la conduite d'évaluation. Chacune dépend d'un contexte particulier et porte sur ses propres objectifs d'évaluation (Farenc et al., 2001). Elles se différencient généralement les unes des autres sur plusieurs points, nous citons à titre d'exemples : son applicabilité, la nature des données recueillies, les critères à évaluer, ses résultats. De plus, ces méthodes visent à inspecter différents types de problèmes d'utilisabilité (Nielsen, 1993 ; Ivory et Hearst, 2001). Une étude de ces méthodes sera présentée lors de la section suivante (*Section 1.3*).
6. **Identifier les tâches.** Dans cette étape, il s'agit de spécifier les tâches appropriées qui correspondent aux aspects de l'interface utilisateur à évaluer, les utilisateurs cibles et la méthode d'évaluation adoptée (Nielsen, 1993 ; Shneiderman, 1998 ; Ivory et Hearst, 2001).
7. **Concevoir les expériences.** Une fois que les étapes décrites précédemment sont achevées, l'évaluateur peut avoir besoin de concevoir des expériences qui requièrent l'intervention des utilisateurs afin de collecter des données d'utilisabilité. La spécification de types d'expériences (par exemples, les tests d'utilisabilité, les enquêtes, etc.) dépend de la méthode d'évaluation adoptée. De ce fait, l'évaluateur doit spécifier certaines informations permettant de bien mener les expérimentations, telles que le nombre de participants (évaluateurs et utilisateurs), la procédure d'évaluation (qui dépend largement de la méthode d'évaluation adoptée), les données relatives à l'environnement et le système à évaluer.
8. **Capturer les données d'utilisabilité.** Cette étape consiste à collecter les données d'utilisabilité sur la base de(s) méthode(s) d'évaluation adoptée(s). Ces données peuvent concerner les tâches et/ou les interfaces utilisateur. Elles peuvent être quantitatives (citons comme exemples, le taux de réalisation des tâches utilisateurs et le nombre des tâches) ou qualitatives (à titre d'exemple, les opinions et avis des utilisateurs) (Witte & Witte, 2009).

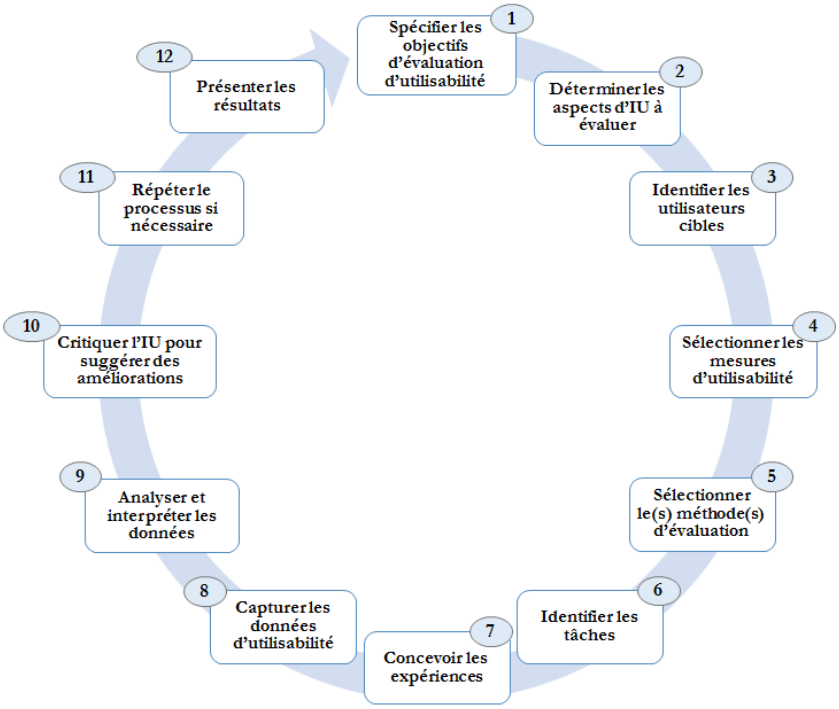


Figure 1. 6 : Processus d'évaluation des interfaces utilisateur (inspiré de (Ivory et Hearst, 2001))

9. **Analyser et interpréter les données.** Cette étape concerne l'analyse et l'interprétation des données d'utilisabilité recueillies. D'une part, l'analyse vise principalement à synthétiser les résultats d'une manière qui facilite le traitement et l'interprétation. Ces analyses peuvent porter sur des résultats statistiques qui correspondent aux objectifs de l'évaluation. De plus, elles peuvent établir l'ensemble des problèmes d'utilisabilité détectés. Quant à l'interprétation, son objectif principal est de tirer des conclusions sur la base des analyses inspirées sur la qualité des interfaces utilisateur évaluées. Par exemple, une bonne qualité ergonomique est indiquée dans le cas où les composants graphiques de l'interface sont conformes aux règles ergonomiques.

10. **Critiquer l'interface utilisateur pour suggérer des améliorations.** Les analyses et les interprétations des données d'évaluation permettent en général la détection des failles dans la conception de l'interface utilisateur (Ivory et Hearst, 2001). Par conséquent, une phase de critique est requise dans le but de suggérer des améliorations des interfaces évaluées et permettant d'éliminer les problèmes engendrés.

11. **Répéter le processus si nécessaire.** La phase de critique peut indiquer parfois la nécessité de répéter le processus d'évaluation d'utilisabilité. Cela peut être justifié à titre d'exemple par l'identification d'autres aspects de l'interface qui nécessitent une évaluation ou des modifications de l'interface utilisateur (Ivory et Hearst, 2001).

12. **Présenter les résultats.** Il s'agit de la dernière étape de processus d'évaluation. Elle consiste à illustrer les résultats et leurs interprétations d'une façon facile à comprendre, selon différents formats (par exemple, graphique, textuel, tableau, etc.).

Comme mentionné dans ce processus d'évaluation portant sur la qualité des produits logiciels ou l'utilisabilité des interfaces utilisateur, l'identification de la/les méthode(s) et/ou outils d'évaluation est cruciale pour l'élaboration du processus. La section suivante sera consacrée pour introduire un ensemble des méthodes et/ou outils d'évaluation permettant d'évaluer différents aspects d'un système interactif et d'identifier les problèmes spécifiques.

1.3 Les méthodes et outils d'évaluation de l'utilisabilité des interfaces utilisateur

Dans la littérature, nous distinguons une grande variété de méthodes et outils d'évaluation de l'utilisabilité. Le choix d'une ou plusieurs méthodes et/ou outils est une étape essentielle du processus d'évaluation.

Diverses classifications de méthodes et outils d'utilisabilité des interfaces utilisateur des systèmes interactifs ont été proposées. Nous présentons dans le tableau suivant (Tableau 1.2) quelques exemples de classifications, parmi les plus connues dans la littérature, portant sur différentes catégories. Pour chaque catégorie, nous introduisons une liste d'exemples des méthodes et/ou outils existants.

Tableau 1. 2: Synthèse sur quelques classifications existantes des méthodes d'évaluation des IHM des systèmes interactifs et quelques exemples

Références	Catégories de méthodes et outils	Exemples des méthodes et outils
(Holzinger, 2005)	Méthodes d'inspection de l'utilisabilité (sans utilisateurs finaux)	Evaluation heuristique ; Cognitive Walkthrough ; Analyse d'action
	Méthodes de Tests d'utilisabilité (avec utilisateurs finaux)	Thinking-Aloud ; Observations (Field observation) ; Questionnaires
(Ivory et Hearst, 2001)	Tests d'utilisabilité	Protocole Thinking-Aloud ; Protocol Question-Asking ; Méthode de Shadowing ; Méthode de Coaching ; Mesure des performances ; Analyses de fichiers log
	Enquête	Enquêtes contextuelles ; Observations ; Focus groups ; Interviews ; Questionnaires
	Inspection d'utilisabilité	Evaluation heuristique ; Cognitive Walkthrough ; Walkthrough pluraliste ; Guides ergonomiques ; Tests Formels d'utilisabilité ; Inspection de la cohérence
	Modélisation analytique	Analyse GOMS ; Analyse UIDE ; Analyse de tâche cognitive
	Simulation	Modélisation de réseau de Pétri ; Modélisation d'algorithme génétique
(Nishiuchi et Takahashi, 2015)	Méthodes d'évaluation subjective	Protocole Thinking-Aloud ; Interviews ; Questionnaires
	Méthodes d'évaluation objective	Occulométrie (Eye tracking) ; Brain waves ; Capture video ; keystroke-level model ; Recording user operation logs.

Sans être exhaustif, nous introduisons dans ce qui suit un aperçu sur quelques catégories d'évaluation d'utilisabilité. Par la suite, nous présenterons un récapitulatif dans le Tableau 1.3 de quelques méthodes et outils parmi les plus utilisés et référencés dans la littérature.

1.3.1 Les catégories de méthodes et outils d'évaluation des interfaces utilisateur

- Le Test d'utilisabilité : Il s'agit d'une méthode d'évaluation fondamentale (Nielsen, 1993 ; Shneiderman, 1998 ; Ivory et Hearst, 2001). Elle permet à l'évaluateur d'observer et d'enregistrer les interactions des utilisateurs avec le produit à évaluer. Le test d'utilisabilité fournit à l'évaluateur des informations directes sur la façon avec laquelle les utilisateurs exécutent certaines tâches. Ceci a pour objectif de détecter les problèmes d'utilisabilité des interfaces utilisateur évaluées. Pendant les tests d'utilisabilité, les participants utilisent le système ou le prototype à évaluer pour accomplir un ensemble prédéterminé de tâches. Un testeur/un logiciel est utilisé pour enregistrer les résultats des utilisateurs et peut générer un

ensemble des mesures portant sur l'exécution de leurs tâches, telles que le nombre d'erreurs et le temps d'exécution des tâches. Comme exemples, nous citons : le protocole Thinking-Aloud (Jaaskelainen, 2000 ; Olmsted-Hawala et al., 2010), les mesures des performances (Nielsen, 1993 ; Bevan et Macleod, 1994 ; Hartson et al., 2003 ; Seffah et al., 2006), l'analyse de fichiers log (Zettlemoyer et al. 1999).

- L'inspection d'utilisabilité : Il s'agit d'une méthode d'évaluation avec laquelle un évaluateur examine les aspects de l'utilisabilité d'une interface utilisateur par rapport à sa conformité à un ensemble de règles et/ou heuristiques (Ivory et Hearst, 2001 ; Holzinger, 2005 ; Sears et Jacko, 2009). Contrairement à d'autres méthodes d'utilisabilité, les inspections se fondent uniquement sur le jugement de l'évaluateur. De nombreux guides et lignes directrices d'utilisabilité des interfaces utilisateur ont également été proposés. A titre d'exemple, le standard ISO 9241 (1998) a fourni des lignes directrices pour différents types d'interactions (e.g. textuelle, tactile, haptique) avec des systèmes informatiques. Comme exemples, nous citons : l'évaluation heuristique (Nielsen, 1993 ; Bowman et al., 2002), Cognitive Walkthrough (Lewis et al., 1990c ; Mahatody et al., 2010), les guides ergonomiques (Vanderdonckt, 1994 ; Farenc et Palanque, 1999).
- Les enquêtes : Tout comme les approches de tests d'utilisabilité, les méthodes d'enquête requièrent l'intervention des utilisateurs et sont souvent utilisées lors des tests d'utilisabilité. Toutefois, l'accent n'est pas mis sur l'étude des tâches spécifiques ou des mesures de la performance. Plutôt, l'objectif de ces méthodes est de recueillir les impressions, les préférences ou les opinions subjectives des utilisateurs sur divers aspects d'une interface utilisateur (Ivory et Hearst, 2001 ; Seffah et Metzker, 2008 ; Sears et Jacko, 2009). A titre d'exemples des méthodes et outils, nous mentionnons : les observations (Holzinger, 2004 ; Bentley et al., 1992 ; Ivory et Hearst, 2001 ; Farenc et al., 2001), les questionnaires (Kirakowski et Corbett, 1993 ; McNamara et Kirakowski, 2011 ; Lewis et al., 2013 ; Lewis et al., 2015), le Focus groups (Rosenbaum et al., 2002) et les entretiens (interviews) (Kuusela et Paul, 2000). Généralement, ces méthodes et outils peuvent être utilisés par les évaluateurs afin de recueillir des données supplémentaires d'une part au début du processus de conception et d'autre part après la réalisation du système ; Ces données sont utiles pour l'amélioration de l'interface pour obtenir des futures versions.
- Modélisation analytique : Il s'agit d'une méthode complémentaire aux méthodes d'évaluation traditionnelles, comme les tests d'utilisabilité (Ivory et Hearst, 2001). La modélisation analytique permet à l'évaluateur de prédire l'utilisabilité à un faible coût sur la base d'une certaine représentation du modèle de l'interface utilisateur et/ou l'utilisateur (Ivory et Hearst, 2001 ; Sears et Jacko, 2009). Nous introduisons quelques exemples tels que, le modèle GOMS (Goals, Operators, Methods, and Selection rules) (John et Kieras, 1996) et CTA (Cognitive Task Analysis) (May et Barnard, 1994). De Haan et al. (1992) ont classé les approches de modélisation en quatre catégories (Ivory et Hearst, 2001) :

    - *Modèles pour l'analyse de l'environnement de la tâche* : permettent à l'évaluateur d'évaluer la correspondance entre les objectifs de l'utilisateur et les tâches de l'interface utilisateur.
    - *Modèles pour analyser les connaissances des utilisateurs* : permettent à l'évaluateur d'utiliser les grammaires formelles pour représenter et évaluer les connaissances requises pour l'utilisation de l'interface.

- *Les modèles de la performance de l'utilisateur :* permettent à l'évaluateur de prédire le comportement de l'utilisateur, nous citons par exemple le temps d'exécution d'une tâche.
- *Les modèles de l'interface utilisateur* : permettent à l'évaluateur de représenter la conception de l'interface utilisateur à plusieurs niveaux d'abstraction (tels que les niveaux syntaxiques et sémantiques) et évaluer cette représentation.

- Simulation : Cette catégorie supporte de manière intrinsèque l'analyse automatisée. Il s'agit des programmes informatiques qui simulent l'utilisateur d'interagir avec l'interface en utilisant des modèles de l'utilisateur et/ou de la conception de l'interface. Ces programmes présentent également les résultats de cette interaction, sous différentes formes telles que les mesures de performance (Ivory et Hearst, 2001). Comme exemples, nous pouvons nous référer à la modélisation de l'algorithme génétique (Kasik et George, 1996) et la modélisation du réseau de Pétri (telle que la méthode AMME (Rauterberg et Aeppili, 1995)).

1.3.2 Présentation de quelques méthodes et outils d'évaluation d'utilisabilité des interfaces utilisateur

Comme mentionné précédemment, des nombreuses méthodes et outils d'évaluation d'utilisabilité des interfaces utilisateur ont été proposés. En effet, chaque méthode (ou outil) d'évaluation a ses propres exigences, possède certains avantages et présente aussi quelques inconvénients. Ces méthodes et outils divergent généralement vers des problèmes d'utilisabilité bien déterminés.

Dans ce qui suit, nous introduisons un tableau récapitulatif (Tableau 1.3) représentant une synthèse de quelques méthodes et outils d'évaluation en se référant à certains critères, parmi ceux-ci, nous citons :

- **L'objectif de l'évaluation :** Prédire les problèmes d'utilisabilité, dégager les perceptions des utilisateurs, etc.
- **Principe de la méthode ou l'outil :** Concerne la description de la méthode ou l'outil d'évaluation des interfaces utilisateur.
- **Phase d'application lors du cycle de développement :** Spécifie la phase d'application de la méthode ou l'outil lors du cycle de développement des systèmes interactifs, telles que la phase de test, la phase de début du cycle de développement.
- **Avantages et inconvénients :** Décrits les principaux avantages et Inconvénients de la méthode ou outil d'évaluation.
- **Intervenants dans le processus d'évaluation :** Indique les intervenants lors de l'évaluation et l'exécution des tâches : expert, évaluateur, concepteur, utilisateurs, etc.

Tableau 1. 3: Synthèses sur quelques méthodes et outils d'évaluation de l'utilisabilité des interfaces utilisateur

Méthode et outil d'évaluation	Objectif d'évaluation	Principe d'évaluation	Phase(s) d'application lors du cycle de développement	Avantages et inconvénients	Intervenants
Evaluation heuristique (Nielsen et Molich, 1990 ; Nielsen, 1993 ; Ivory et Hearst, 2001 ; Bowman et al., 2002)	Evaluer l'utilisabilité d'une IHM en se basant sur un ensemble de critères et heuristiques[2] dans le but d'identifier les problèmes d'utilisabilité potentiels	Dans une évaluation heuristique un ou plusieurs évaluateurs (experts en utilisabilité) évaluent indépendamment une interface en utilisant une liste d'heuristiques. Tous les résultats sont ensuite combinés et classés pour prioriser la re-conception itérative de chaque problème d'utilisabilité découvert.	- au début du cycle de développement. - phase de test (à la fin du cycle de développement) - pour un prototype ou système	Avantages : Méthode rapide, non coûteuse, très rentable et facile à utiliser (Nielsen, 1993; Ivory et Hearst, 2001 ; Bowman et al., 2002). Très souple et ne nécessite pas une planification avancée (Kantner et Rosenbaum, 1997). Inconvénients : dépend de la compétence et l'expérience d'évaluateurs (Nielsen, 1992). Détecte seulement les violations majeures des règles sans suggestion des solutions aux problèmes (Nielsen, 1992, 1993). Difficulté de reproduction du contexte d'usage (Nielsen et Mack, 1994).	Un ou plusieurs évaluateurs (experts en utilisabilité)
Cognitive Walkthrough (Lewis et al. 1990c ; Ivory et Hearst, 2001 ; Mahatody et al., 2010)	Evaluer la compréhensibilité et la facilité d'apprentissage du système à évaluer et prédire les problèmes rencontrés par l'utilisateur	L'évaluateur simule le comportement d'un utilisateur lors de l'exécution des tâches avec le système. A chaque étape d'une tâche, cette simulation permet à l'évaluateur de prédire si l'utilisateur peut réussir ou échouer l'accomplissement de l'étape. A la fin de l'évaluation, l'évaluateur produit une documentation complète sur cette analyse.	- au début du cycle de développement. - phase de test - pour un prototype ou système	Avantages : Permet d'évaluer un système au début du développement avec un coût relativement faible (Lewis et al., 1990c) Inconvénients : approche trop lourde et fastidieuse à employer (Rieman et al., 1991). Pas facile à apprendre ou à appliquer (Nielsen et Mack, 1994 ; Zhang et al., 1999). Nécessite de documentations intensives (Ivory et Hearst, 2001).	Un ou plusieurs évaluateurs, les experts en utilisabilité, les développeurs de logiciels
Guides ergonomiques (Vanderdonckt, 1994 ; Farenc et Palanque, 1999)	Evaluer la qualité ergonomique d'une interface utilisateur en se basant sur la conformité des éléments graphiques par rapport aux règles et guides ergonomiques afin de détecter les incohérences de l'interface.	Cette méthode consiste à inspecter et juger la conformité des éléments graphiques de l'interface à évaluer par rapport à un ensemble de recommandations et de guides ergonomiques. Nombreuses règles et guides ont été proposés dans la littérature (Nielsen, 1994 ; Scapin et Bastien, 1997 ; Schneiderman, 1998). A la fin de l'évaluation, l'ensemble des incohérences ergonomiques est produit.	- au début du cycle de développement. - phase de test - pour un prototype ou système	Avantage : en général, les règles et les guides sont représentés en langage naturel ce qui rend simple leurs compréhensions (Keith, 2005) Inconvénients : le nombre important des règles existantes rendent leur utilisation fastidieuse et complexe (Mariage, 2005). Absence d'uniformité dans la présentation des recommandations entrainent des difficultés lors de la recherche d'une règle bien précise (Bastien et Scapin, 2001)	Evaluateur

[2] Les heuristiques ne présentent pas des lignes directrices spécifiques, mais sont plutôt des règles générales qui décrivent les propriétés communes de bonnes conceptions d'utilisabilité. Il n'existe aucun ensemble standard des heuristiques. Molich et Nielsen (Nielsen et Molich, 1990) ont développé une petite liste des heuristiques avec neuf principes d'utilisabilité après des années d'expérience en utilisabilité. Nielsen a affiné la liste et a défini une série de 10 heuristiques (Nielsen, 1994).

Tableau 1.3 (suite) : Synthèses sur quelques méthodes et outils d'évaluation de l'utilisabilité des interfaces utilisateur

Méthode et outil d'évaluation	Objectif d'évaluation	Principe d'évaluation	Phase(s) d'application lors du cycle de développement	Avantages et inconvénients	Intervenants
Think-Aloud Protocol (Jaaskelainen, 2000 ; Ivory et Hearst, 2001 ; Olmsted-Hawala et al., 2010)	Obtenir une meilleure compréhension de ce que pense l'utilisateur lors de l'interaction avec l'interface au cours d'un test d'utilisabilité.	Pendant une évaluation, il s'agit d'observer l'utilisateur lors de l'interaction avec le système en l'exigeant au fur et à mesure à verbaliser ses pensées et réflexions à haute voix. A la fin de chaque session, ces pensées sont enregistrées dans le but d'être étudiées et analysées ultérieurement.	- phase de test du système à évaluer	Avantages : Méthode rapide, non coûteuse se fonde sur l'observation de l'utilisateur final, facile à mettre en place, combine à la fois l'interview et l'observation (Jaaskelainen, 2000 ; Olmsted-Hawala et al., 2010) Inconvénients : la difficulté d'interprétation et d'analyse des données collectées. La réflexion à haute voix de l'utilisateur, au fur et à mesure qu'il interagit avec le système, peut le perturber.	Evaluateur Utilisateur
Mouchard électronique (Ezzedine, 2002 ; Trabelsi et Ezzedine, 2013; Tran et al., 2013)	Evaluer les interactions des utilisateurs avec le système en capturant leurs actions en temps réel au cours d'un test d'utilisabilité.	Pendant une évaluation, le mouchard électronique capture et analyse en temps réel les actions des utilisateurs, ainsi que leurs interactions avec le système évalué. A la fin de l'évaluation, un ensemble des données et des mesures est fourni permettant de détecter les problèmes d'utilisabilité.	- phase de test du système à évaluer	Avantages : la transparence du mouchard et leur utilisation d'une manière discrète ne perturbe pas l'utilisateur pendant le test d'utilisabilité (Ezzedine, 2002). Fournit la collecte automatique et une phase d'analyse des informations, ce qui facilite le processus d'évaluation. Inconvénients : difficile à mettre en place, l'évaluateur doit interpréter lui-même les résultats qui peuvent être très nombreux et difficile à comprendre et analyser (Trabelsi et Ezzedine, 2013).	Evaluateur Utilisateur
Mesure des performances (Nielsen, 1993 ; Bevan et Macleod, 1994 ; Ivory et Hearst, 2001 ; Hartson et al., 2003 ; Seffah et al., 2006 ; ISO/IEC 25022, 2012)	Evaluer la performance des utilisateurs lors de réalisation de leurs tâches par la capture d'un ensemble de mesures (tel que le temps de réalisation d'une tâche) au cours d'un test d'utilisabilité.	Pendant l'évaluation, l'évaluateur choisit les mesures des performances à collecter sur la base des attributs d'utilisabilité à évaluer. Ces mesures peuvent être capturées à travers des outils automatiques lors de l'interaction des utilisateurs avec le système. A la fin de l'évaluation, les évaluateurs peuvent utiliser les résultats pour évaluer si les objectifs d'utilisabilité ont été atteints ainsi que pour établir une analyse concurrentielle avec des résultats antérieurs.	- phase de test du système à évaluer	Avantages : fournit des résultats fiables et valides (Bevan et Macleod, 1994). Quantifier les interactions des utilisateurs. (Nielsen, 1993 ; Hartson et al., 2003). Considère les mesures subjective et objective des performances (Bowman et al., 2002) Inconvénients : la difficulté de capture des mesures des performances. Ceux-ci requièrent la mise en place des outils automatiques pour faciliter le processus d'évaluation (Bevan et Macleod, 1994 ; Ivory et Heart, 2001). Difficulté d'interprétation et d'analyse des résultats des mesures.	Evaluateur Utilisateur

Tableau 1.3 (suite) : Synthèses sur quelques méthodes et outils d'évaluation de l'utilisabilité des interfaces utilisateur

Méthode et outil d'évaluation	Objectif d'évaluation	Principe d'évaluation	Phase(s) d'application lors du cycle de développement	Avantages et inconvénients	Intervenants
Questionnaires (Kirakowski et Corbett, 1993 ; McNamara et Kirakowski, 2011, Lewis et al.,2013 ; Lewis et al., 2015a)	Evaluer la satisfaction des utilisateurs sur la base d'une liste des questions afin de détecter les problèmes d'utilisabilité des interfaces utilisateur du système interactif (Nielsen, 1993 ; Yang et al., 2012).	Lors d'une évaluation, il s'agit de questionner les utilisateurs sur la qualité d'utilisabilité perçue relative au système interactif à évaluer. A la fin de l'évaluation, l'ensemble d'appréciations et perceptions des utilisateurs sont collectées afin d'identifier les problèmes rencontrés.	-phase de test du système à évaluer -phase de conception	Avantages : facile à utiliser et à mettre en place. Une bonne méthode pour la collecte de données subjectives (Bowman et al., 2002). Inconvénients : l'analyse difficile des réponses des utilisateurs (Charfi, 2013). Les questions sont spécifiques et souvent fermées (Ivory et Hearst, 2001). Difficulté d'analyse des questions ouvertes.	Evaluateur Utilisateur
Observations (Holzinger, 2004 ; Bentley et al., 1992 ; Ivory et Hearst, 2001 ; Farenc et al., 2001)	Evaluer la qualité d'un système interactif en observant l'utilisateur lors de son interaction avec le système évalué dans le but de détecter les problèmes rencontrés.	Durant une évaluation, cette méthode permet à l'évaluateur d'observer l'utilisateur lors de leur travail. Cela lui permet de comprendre leur façon d'utilisation du système pour accomplir leurs tâches. Elle peut s'effectuer d'une façon directe, ou via des enregistrements (audio, vidéo).	-phase de test du système à évaluer	Avantages : implique l'utilisateur final (Farenc, 1997). Fournit des informations directes sur le comportement des utilisateurs (Holzinger, 2004). Facile à mettre en place. Inconvénients : Variation des jugements d'un évaluateur à un autre (Molich et al., 1999). Nécessite des observateurs expérimentés, qualifiés et bien formés. Peut être difficile à interpréter les comportements (Farenc et al., 2001).	Evaluateur Utilisateur
Focus groups (Rosenbaum et al., 2002)	Evaluer la qualité d'un système sur la base d'une discussion entre plusieurs utilisateurs.	Il s'agit d'une réunion d'un groupe des utilisateurs qui discutent par rapport à la qualité du système. L'évaluateur joue le rôle de modérateur et recueille les informations nécessaires de la discussion.	-phase de test du système à évaluer	Avantages : Efficace pour obtenir des diverses informations et profondes dans un court temps (Rosenbaum et al., 2002) Inconvénients : Peut être difficile à analyser les réponses ; nécessite un bon modérateur (Rosenbaum et al., 2002).	Evaluateur Un groupe d'utilisateurs

1.4 Les mesures d'utilisabilité pour l'évaluation des IHM

Les méthodes et les outils d'évaluation des interfaces utilisateur proposés sont intimement liés à l'objectif de la mesure de l'utilisabilité afin de détecter les problèmes ponctuels des interfaces. L'utilisabilité ne peut pas être mesurée directement, néanmoins sa notion a été entourée par différents critères qui peuvent être mesurés (Nielsen, 1993 ; Hornbæk, 2006). La définition des mesures portant sur des critères d'utilisabilité a été largement exploitée dans la littérature par les standards et les chercheurs (Hornbæk, 2006 ; Seffah et al., 2006 ; Hornbæk et Law, 2007 ; ISO / IEC 25021, 2012 ; ISO/IEC 25022, 2012 ; Albert et Tullis, 2013 ; Zen et Vanderdonckt, 2014). Par conséquent, des nombreuses et diverses mesures ont été proposées (Hornbæk et Law, 2007). Shneiderman et Plaisant (2005), par exemple, ont identifié cinq mesures d'utilisabilité : le temps d'apprentissage, la rapidité d'exécution, le taux d'erreurs commises par les utilisateurs, le maintien au cours du temps et la satisfaction subjective. Seffah et al. (2006) ont développé un modèle de synthèse d'utilisabilité (appelé ***Q**uality in **U**se **I**ntegrated **M**easurement* [QUIM]) basé sur les travaux existants qui incluent les modèles classiques (par exemple celles de la norme ISO 9126 (2001) et ISO 9241-11 (1998)) et les modèles conceptuels (tel que *Metrics for Usability Standards in Computing* [MUSIC] (Bevan, 1995 ; Macleod et al., 1997)). Leur modèle QUIM intègre plus de 127 mesures spécifiques portant sur 10 facteurs qui sont décomposés en un total de 26 sous-facteurs ou critères mesurables, notamment outre les aspects de l'ISO, des facteurs tels que la sécurité, la véracité, l'accessibilité, l'apprenabilité, l'universalité et l'utilité. Chaque facteur correspond à une facette spécifique de l'utilisabilité qui a été identifiée dans les normes ou les modèles existants. Le modèle QUIM avec toutes les mesures a été publié dans le travail de (Padda, 2003). Comme exemples, nous présentons dans le Tableau 1.4 quelques exemples de ces mesures proposées qui correspondent à différents critères.

Tableau 1. 4 : Quelques exemples de mesures proposées dans le modèle QUIM

Critère de qualité	Mesures
Compréhensibilité (*Understandability*)	• Fréquence d'utilisation de l'aide et de la documentation *(Frequency of using help and documentation)* = Nombre de fois où les fonctions d'aide et de documentation sont utilisées (ISO/IEC 9126-2, 2003) • Le temps passé en utilisant de l'aide et de la documentation *(Time spent using help and documentation)* (ISO/IEC 9126-2, 2003) • Densité globale (*Overall density*) (Tullis, 1984)
Action minimale *(Minimal action)*	• Temps pour compléter une tâche (*Time to complete a task*) (Dix et al., 2004) • Nombre des actions utilisées pour réaliser une tâche (*Number of commands used to perform a task*) (Dix et al., 2004)
Capacité de fonctionnement *(operability)*	• Nombre de frames (*Number of frames*) (Donayee, 2001) • Capacité d'entrer-annuler (*Input- undo ability*) = Nombre de fonctions d'entrer-annuler/ Nombre total de fonctions disponibles sur l'interface (ISO/IEC 9126-2, 2003)
Guidage de l'utilisateur *(User guidance)*	• Taux d'annulation *(Rate of cancel)* (Donayee, 2001) • Taux de messages d'erreur *(Rate of error messages)* = Nombre de messages d'erreur affiché/ Nombre de problèmes rencontrés (Padda, 2003)

Récemment, de nombreuses autres mesures portant sur de différents critères de qualité ont été proposées par les standards de SQUARE (tels que ISO/IEC 25021, 2012 et ISO/IEC 25022, 2012).
Comme exemples, nous citons dans le Tableau 1.5 quelques mesures définies par la norme ISO/IEC 25022 portant sur quelques critères et sous-critères de qualité.

Tableau 1. 5 : Quelques exemples de mesures définies par la norme ISO/IEC 25022 (2012)

Critère de qualité		**Mesures**
Efficacité *(Effectiveness)*		• Taux de réalisation correcte (d'efficacité) d'une tâche (*Task effectiveness*) = 1- ∑ la valeur proportionnelle de chaque composante manquante ou incorrecte d'une tâche • Taux d'achèvement des tâches *(Task completion)* = Nombre de tâches achevées/ Nombre total de tâches • Fréquence d'error (*Error frequency*) = Nombre d'erreurs commises par l'utilisateur/ nombre de tâches
Efficience (*Efficiency*)		• Temps d'exécution d'une tâche (*Task time*) • Taux d'efficience d'une tâche (*Task efficiency*) = Taux de réalisation correcte d'une tâche / Temps d'exécution d'une tâche
Flexibilité (*Flexibility)*		• Flexibilité de contexte d'utilisation (*Flexible context of use*) = Nombre de contextes supplémentaires où le produit serait utilisable/ Nombre total de contextes supplémentaires dans lesquelles le produit peut être utilisé.
Satisfaction (*Satisfaction*)	Utilité (*Usefulness*)	• Échelle de satisfaction (*Satisafction scale*) = Questionnaire produisant des échelles psychométriques/ Moyenne de la population • Satisfaction du questionnaire (*Satisfaction questionnaire*) = ∑ (Réponse à une question i) / nombre de réponses
	Plaisir (*Pleasure*)	• Échelle de plaisir (*Pleasure scale*) = Questionnaire produisant des échelles psychométriques/ Moyenne de la population
	Confiance (*Trust*)	• Échelle de confiance (*Trust scale*) = Questionnaire produisant des échelles psychométriques/ Moyenne de la population
	Confort (*Comfort*)	• Échelle de confiance (*Comfort scale*) = Questionnaire produisant des échelles psychométriques/ Moyenne de la population

D'autres travaux de recherche peuvent également être mentionnés, à titre d'exemples nous citons : (Tullis, 1984 ; Dix et al., 2004 ; Bevan, 1997 ; Donyaee et al., 2001 ; Vanderdonckt, 2003 ; Abran et al., 2006 ; Gabillon et al, 2013 ; Zen et Vanderdonckt, 2014 ; Albert et Tullis, 2013). Le Tableau 1.6 présente quelques exemples des mesures d'évaluation des interfaces utilisateur proposées dans la littérature.

Tableau 1. 6 : Quelques exemples de mesures proposées dans la littérature

Références	**Mesures de qualité**
(Tullis, 1984)	Densité globale (*Overall density*) Densité locale (*Local density*)
(Gabillon et al., 2013)	Nombre de labels (*Number of labels*) Nombre des images (*Number of pictures*) Taille d'image de fond (*Size of background image*)
(Zen et Vanderdonckt,2014 ; Vanderdonckt, 2003 ; Buanga, 2011)	Simplicité (*simplicity*)= 3/(n_{vap} + n_{hap}+ n) avec n_{vap} et n_{hap} le nombre de points d'alignement vertical et horizontal et n est le nombre d'objets sur du cadre
	Alignement (*Alignement*)= 4n - (n_{hap}*n_{vap}/n) / 4n – 4 avec n_{vap} et n_{hap} le nombre de points d'alignement vertical et horizontal et n est le nombre d'objets sur du cadre

Tableau 1.6 (suite) : Quelques exemples de mesures proposées dans la littérature

Références	**Mesures de qualité**
(Zen et Vanderdonckt,2014 ; Vanderdonckt, 2003 ; Buanga, 2011)	Regroupement(*Grouping*) = (0.85f)* UMspace + (0.15f)*AL avec UMspace = 1– (aLayout – sumOfAreas / aFrame – sumOfAreas) et AL = min (1 , (4*Ni - (n_{Hap}*n_{Vap} / Ni)) / (4*Ni - 4)) aLayout est l'aire de la mise en page ; sumOfAreas est la somme des surfaces des objets ; aFrame est la surface du cadre ; Ni est le nombre d'objets de type i ; N_{HAP} et n_{Vap} sont respectivement, le nombre de points d'alignement horizontal et le nombre de points d'alignement vertical
(Albert et Tullis, 2013)	Temps d'exécution d'une tâche (*Time on Task*) Taux d'achèvement d'une tâche (*Task completion Rate*)

1.5 L'évaluation ergonomique des IHM des systèmes interactifs

L'évaluation ergonomique des systèmes interactifs vise à mesurer la qualité ergonomique des interfaces utilisateur afin d'améliorer leurs qualités. Cette évaluation permet de détecter les aspects qui peuvent engendrer des incohérences ergonomiques. Elle consiste à vérifier la conformité d'une IHM par rapport à un ensemble de règles ergonomiques formalisées provenant de guides et de recommandations ergonomiques.

Selon Vanderdonckt, une règle ergonomique constitue un principe de conception et/ou d'évaluation à observer en vue d'obtenir et/ou de garantir une interface Homme-Machine ergonomique (Vanderdonckt, 1999). Ces règles ergonomiques sont souvent générales, présentées en langage naturel et dépendent du contexte d'utilisation (Grammenos et al., 2000).

Dans la littérature, différentes règles, recommandations et critères ergonomiques pour l'évaluation de l'utilisabilité des interfaces utilisateur ont été également proposés. Ces règles peuvent correspondre à des guides de style[3] (Mariage, 2005), des algorithmes[4] de conception ergonomique, des normes (comme ISO 9241, ISO/IEC 9126), ou encore des guides, des critères ergonomiques et des recommandations tels que « le Guide Ergonomique des Interfaces Homme-Machine » (Vanderdonckt, 1994), « les Critères Ergonomiques pour l'Évaluation d'Interfaces Utilisateurs » (Bastien et Scapin, 1993), les règles ergonomiques pour le web (Rukshan et Baravalle, 2011),« les Critères Ergonomiques pour les Interactions Homme-Environnements Virtuels» proposés dans (Bach et Scapin, 2005).

La série de standards ISO 9241, par exemple, a fourni un ensemble des recommandations destinées à différents types d'interactions (telles que les interactions textuelles, tactiles, organiques, et haptiques) et des systèmes informatiques ; à titre d'exemples, nous citons : l'accessibilité du logiciel (ISO 9241-171, 2008), les interfaces utilisateur World Wide Web (ISO/IEC 9241-151,2008) et les interactions tactiles et haptiques (ISO/IEC 9241-920, 2009).

En outre, Vanderdonckt (1994) a proposé 3.700 règles pour évaluer des aspects ergonomiques pour les interfaces utilisateur en fonction de huit critères ergonomiques : la compatibilité, la cohérence, la charge de travail, l'adaptabilité, le contrôle de dialogue, la représentativité, le guidage et la gestion d'erreur.
Scapin et Bastien (1997) ont proposé un ensemble des critères, sous-critères et recommandations pour le développement de l'interface utilisateur qui ont été largement utilisés (Filho et al., 2015 ; Castro et al., 2015 ; Tardieu et al., 2015) pour différents types d'interfaces tels que les interfaces utilisateur

[3] Un guide de style est un outil permettant de définir la présentation d'une IHM au sein d'une même plateforme ou d'une même ligne de produits (Mariage, 2005). Nous citons à titre d'exemple le guide proposé par Microsoft nommé « Windows User Experience Interaction Guidelines » (Microsoft, 2011).
[4] Un algorithme décrit un ensemble des instructions à suivre pour atteindre une bonne qualité ergonomique des IHM. A titre d'exemple : « Design Adviser » (Faraday, 2000).

graphique (GUI) comme WIMP, les interfaces 3D, les interfaces utilisateur Web (WUI). Le Tableau 1.7 présente l'ensemble des critères et sous-critères proposés par Bastien et Scapin (1993).

Tableau 1. 7: Critères et sous-critères de guidage proposés par (Bastien et Scapin, 1993)

Critères		**Sous-critères**	**Descriptions**
Guidage	L'ensemble des moyens mis en œuvre pour conseiller, orienter, informer, et conduire l'utilisateur lors de ses interactions avec l'ordinateur (messages, alarmes, labels, etc.), y compris dans ses aspects lexicaux.	*Incitation*	Recouvre les moyens mis en œuvre pour amener les utilisateurs à effectuer des actions spécifiques, qu'il s'agisse d'entrée de données ou autre.
		Groupement /Distinction entre Items	Concerne l'organisation visuelle des Items d'information les uns par rapport aux autres
		Feedback Immédiat	Concerne les réponses de l'ordinateur consécutives aux actions des utilisateurs
		Lisibilité	Concerne les caractéristiques lexicales de présentation des informations sur l'écran pouvant entraver ou faciliter la lecture de ces informations
Charge de Travail	L'ensemble des éléments de l'interface qui ont un rôle dans la réduction de la charge perceptive ou mnésique des utilisateurs et dans l'augmentation de l'efficacité du dialogue	*Brièveté (Concision et Actions Minimales)*	Concerne la charge de travail au niveau perceptif et mnésique à la fois pour les éléments individuels d'entrée ou de sortie et les séquences d'entrées
		Densité Informationnelle	Concerne la charge de travail du point de vue perceptif et mnésique, pour des ensembles d'éléments et non pour des items
Contrôle Explicite	La prise en compte par le système des actions explicites des utilisateurs et le contrôle qu'ont les utilisateurs sur le traitement de leurs actions	*Actions Explicites*	Concerne la relation pouvant exister entre le fonctionnement de l'application et les actions des utilisateurs
		Contrôle Utilisateur	Le fait que l'utilisateur doit toujours avoir la main, pouvoir contrôler le déroulement des traitements informatiques en cours.
Adaptabilité	La capacité à réagir selon le contexte, et selon les besoins et préférences des utilisateurs	*Flexibilité*	Concerne les moyens mis à la disposition des utilisateurs pour personnaliser l'interface afin de rendre compte de leurs stratégies ou habitudes de travail et des exigences de la tâche
		Prise en Compte de l'Expérience de l'Utilisateur	Concerne les moyens mis en œuvre pour respecter le niveau d'expérience de l'utilisateur
Gestion des Erreurs	Tous les moyens permettant d'une part d'éviter ou de réduire les erreurs, et d'autre part de les corriger lorsqu'elles surviennent	*Protection Contre les Erreurs*	Concerne les moyens mis en place pour détecter et prévenir les erreurs d'entrées de données ou de commandes ou les actions aux conséquences néfastes
		Qualité des Messages d'Erreurs	Concerne la pertinence, la facilité de lecture et l'exactitude de l'information donnée aux utilisateurs sur la nature des erreurs commises et sur les actions à entreprendre pour les corriger.
		Correction des Erreurs	Concerne les moyens mis à la disposition des utilisateurs pour leur permettre de corriger leurs erreurs
Homogénéité/Cohérence	Se réfère à la façon avec laquelle les choix de conception de l'interface (codes, dénominations, formats, procédures, etc.) sont conservés pour des contextes identiques, et sont différents pour des contextes différents		
Signifiance des Codes et Dénominations	Concerne l'adéquation entre l'objet ou l'information affichée ou entrée, et son référent		
Compatibilité	Se réfère à l'accord pouvant exister entre les caractéristiques des utilisateurs et des tâches, d'une part, et l'organisation des sorties, des entrées et du dialogue d'une application donnée, d'autre part.		

Quelques recommandations portant sur le critère de guidage et leurs sous-critères proposés par Bastien et Scapin (1993) sont illustrés dans le Tableau 1.8.

Tableau 1. 8 : Exemples des recommandations ergonomiques relatives aux sous-critères de guidage proposés par (Bastien et Scapin, 1993)

Critère	Sous-critère	Exemples de recommandations
Guidage	Incitation	- Afficher les unités de mesure des données à saisir - Pour chaque champ de données, fournir un label - Donner un titre à chaque fenêtre
	Groupement/Distinction entre Items	- Organiser les items en listes hiérarchiques - Établir une distinction visuelle entre les labels et les champs d'entrée.
	Feedback Immédiat	- Suite à l'interruption d'un traitement par l'utilisateur, le système devrait fournir un message indiquant le retour à l'état précédent. - Toujours faire apparaître sur l'écran les entrées effectuées par les utilisateurs, sauf pour les entrées confidentielles.
	Lisibilité	- Les titres doivent être centrés - Les labels doivent être en majuscule - Les lignes de texte continu doivent être d'au moins 50 caractères

En effet, nous distinguons de nombreux autres travaux tels que les huit règles d'or « Golden Rules » proposées par Shneiderman (Shneiderman, 1998) portant sur le design des interfaces, les 247 règles ergonomiques de Travis portant sur le web «Web Usability Guidelines » (Travis, 2014), ou encore les 113 règles proposées par Nielsen « Top Ten Guidelines for Homepage Usability » (Nielsen et Tahir, 2001).

1.6 Synthèse et conclusion

L'évaluation de l'utilisabilité des IHM présente une des disciplines les plus riches en termes de concepts et approches proposés étant donné son importance pour la garantie d'une bonne qualité des systèmes interactifs.

Dans ce premier chapitre, nous avons introduit un aperçu général sur les notions clés de l'évaluation d'utilisabilité des interfaces utilisateur. Nous avons énuméré différents concepts sur lesquels s'appuie l'évaluation, outre leurs définitions de base, (les types de données, les types d'évaluation, le processus d'évaluation, les mesures qui peuvent être choisies, les critères ergonomiques). Nous nous sommes intéressés également à présenter un certain nombre de méthodes et outils d'évaluation qui s'adressent à ce sujet. Bien que ces méthodes convergent toutes vers un objectif commun, portant sur « *l'évaluation de l'utilisabilité des IHM dans le but de détecter les problèmes des interfaces* », chacune possède sa spécificité, contribue à un but spécifique et adopte un principe bien déterminé.

En effet, comme nous avons perçu précédemment, certaines méthodes et outils se basent sur l'intervention des experts, d'autres portent également sur l'implication des utilisateurs pour le recueil de leurs jugements. De plus, une variété de données peut être collectée (telles que, les données relatives aux caractéristiques des éléments graphiques des interfaces utilisateur, les données subjectives portant sur les perceptions des utilisateurs, les données correspondant aux interactions des utilisateurs avec le système, etc.). En outre, l'exploitation des mesures d'utilisabilité existantes définies par les standards et les modèles d'évaluation consolidés s'avère important pour promouvoir la fiabilité des résultats d'évaluation. Dans la littérature, divers outils d'évaluation d'utilisabilité des systèmes interactifs ont été mis en œuvre. Dans le chapitre suivant, nous allons aborder un état de l'art de ces outils.

Chapitre 2 : Etat de l'art sur les outils d'évaluation de l'utilisabilité des IHM

""A wide range of usability evaluation techniques have been proposed, and a subset of these are currently in common use… Each technique has its own requirements, and generally different techniques uncover different usability problems"
(MELODY Y.IVORY & MARTI A. HEARST, 2001)

Introduction

L'évaluation de l'utilisabilité est constituée de méthodologies permettant la mesure des aspects de l'utilisabilité des interfaces utilisateur d'un système interactif et l'identification des problèmes spécifiques (Nielsen, 1993 ; Dix et al., 1998 ; Ivory et Hearst, 2001). Différentes méthodes et outils d'évaluation d'utilisabilité ont été proposés. Ils visent à améliorer la facilité d'utilisation des systèmes interactifs en cherchant à identifier les problèmes rencontrés lors de leur utilisation.

Lors du premier chapitre, nous avons introduit un aperçu général sur les notions clés de l'évaluation d'utilisabilité des IHM des systèmes interactifs (utilisabilité, mesures, processus, méthodes et outils). Dans ce chapitre, nous allons présenter un panorama détaillé sur un ensemble d'outils et des approches d'évaluation existantes assurant des évaluations subjective et objective. Cette présentation de l'état de l'art va introduire, en plus, une étude d'analyse détaillée sur les outils de questionnaires basée sur des critères d'utilisabilité définis par les standards. Ce chapitre est structuré suivant trois sections. Dans la première section, nous allons présenter un état de l'art sur les outils d'utilisabilité des IHM basés sur une évaluation subjective *ou* objective. Cette section sera suivie d'une étude détaillée sur les outils de

questionnaires. La deuxième section présentera les différentes approches basées sur une combinaison des évaluations subjective et objective. Son objectif est de positionner ces travaux par rapport à notre approche qui sera présentée dans le troisième chapitre. La dernière section, comportera des synthèses menées sur les travaux présentés et une conclusion.

2.1 Etat de l'art sur les outils d'évaluation de l'utilisabilité des IHM basés sur une évaluation subjective ou objective

Au fil des années, une grande variété d'outils d'évaluation de l'utilisabilité des interfaces utilisateur ont été proposés. Certains outils s'appuient sur une évaluation subjective dans le but d'identifier les perceptions et les appréciations des utilisateurs concernant l'utilisabilité des interfaces utilisateur (à titre d'exemples, les questionnaires CSUQ (Lewis, 1995), DEEP (Yang et al., 2012)). D'autres outils considèrent les évaluations objectives indépendamment des opinions personnelles des utilisateurs. Ces derniers visent généralement à détecter les problèmes d'utilisabilité concernant des différents aspects d'évaluation tels que : l'aspect ergonomique de l'interface utilisateur (nous citons comme exemples, WAEX (Centeno et al., 2006, 2007) et AWebHUT (Rukshan et Baravalle, 2011)), ou l'aspect d'interaction entre l'utilisateur et le système interactif à évaluer, à titre d'exemples : MESIA (Trabelsi et al, 2006) et EISEval (Tran et al., 2013).

Dans ce chapitre, nous allons présenter un panorama sur ces outils. Ils seront présentés en deux catégories. La première contiendra les outils exploitant une évaluation subjective. La seconde décrit brièvement un ensemble d'outils d'évaluation assurant une évaluation objective (en utilisant les règles ergonomiques ou en capturant des données d'interaction entre l'utilisateur et le système interactif à évaluer).

2.1.1 Les outils assurant une évaluation subjective

L'utilisabilité des systèmes interactifs peut être inspectée d'une façon subjective à travers un ensemble de méthodes et outils d'évaluation. Dans la littérature, plusieurs méthodes et outils d'évaluation subjective des interfaces utilisateur existent. A titre d'exemples, nous citons : les interviews (Olsen, 2002 ; Kuusela et Paul, 2000), les observations (Holzinger, 2004 ; Bentley et al., 1992), le focus groups (Nielsen, 1997), les enquêtes contextuelles (Beyer et Holtzblatt, 1998), les questionnaires (Kirakowski et Corbett, 1993 ; Sauro, 2011b ; Lewis et al., 2013 ; Lewis et al., 2015a), etc.

Dans cette section, nous nous intéressons aux outils de type questionnaire. Ceux-ci permettent de recueillir les problèmes relatifs à la satisfaction des utilisateurs (Nielsen, 1993 ; Yang et al., 2012). Généralement lors d'une évaluation subjective, différents questionnaires d'utilisabilité standardisés sont actuellement disponibles (Sauro et Lewis, 2012). Ces outils ont été conçus afin de fournir une mesure plus fiable d'utilisabilité (Hornbæk, 2006 ; Hornbæk et Law, 2007 ; Sauro et Lewis, 2009 ; Lewis, 2014 ; Sauro, 2015 ; Lewis et Hardzinski, 2015 ; Lewis et al., 2015a). Ils permettent ainsi de contribuer à l'identification des défauts d'utilisabilité portant sur les interfaces utilisateur pour apporter des améliorations (Hamborg et al., 2004 ; Sauro et Lewis, 2012).

Dans ce qui suit, nous exposerons premièrement un état de l'art sur ces outils, leurs caractéristiques, les classifications proposées dans la littérature et les principales études comparatives effectuées à ce propos. Deuxièmement, nous introduirons une étude d'analyse détaillée sur ces outils basée sur des critères d'utilisabilité définis par les standards.

2.1.1.1 Etat de l'art sur les outils des questionnaires d'utilisabilité standardisés

- **Présentation des outils des questionnaires d'utilisabilité standardisés**

Dans cette section, nous allons présenter une synthèse globale sur les questionnaires d'utilisabilité standardisés existants. Ces questionnaires ont été largement utilisés et validés pour l'évaluation et la mesure de l'utilisabilité perçue des interfaces utilisateur. En se basant sur les principales bibliothèques numériques (ACM, IEEE Xplore, Science Directe, Elsevier, Springer Link), nous avons distingué 24 questionnaires standardisés portant sur l'évaluation d'utilisabilité des interfaces utilisateur (Assila et al., 2016a), comme suit :

- Questionnaire for User Interface Satisfaction (QUIS) (Chin et al., 1988)
- Technology Acceptance Model questionnaire (TAM) (Davis, 1989)
- After-Scenario Questionnaire (ASQ) (Lewis, 1990b)
- Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 1990a, 2002)
- Software Usability Measurement Inventory (SUMI)(Kirakowski et Corbett, 1993)
- Single Ease Question (SEQ) (Tedesco et Tullis, 2006 ; Sauro et Dumas, 2009)
- Computer System Usability Questionnaire (CSUQ) (Lewis, 1995)
- System Usability Scale (SUS) (Brooke, 1996)
- Purdue Usability Testing Questionnaire (PUTQ) (Lin et al., 1997)
- Website Analysis Measurement Inventory (WAMMI) (Kirakowski et Cierlik, 1998)
- Usefulness, Satisfaction and Ease of use (USE)(Lund, 1998, 2001)
- Expectation Ratings (ER) (Albert et Dixon, 2003)
- WEBsite Usability Evaluation tool (WEBUSE)(Chiew et Salim, 2003)
- Usability Magnitude Estimation (UME)(McGee, 2003, 2004)
- Mobile Phone Usability Questionnaire (MPUQ)(Ryu et Smith-Jackson, 2005)
- Website Evaluation Questionnaire (WEQ) (Elling et al., 2007)
- Subjective Mental Effort Question (SMEQ) (Zijlstra & van Doorn, 1985 ; Sauro et Dumas, 2009)
- Usability Metric for User Experience (UMUX) (Finstad, 2010)
- Standardized Universal Percentile Rank Questionnaire (SUPR-Q) (Sauro, 2011b, 2015)
- DEsign-oriented Evaluation of Perceived usability (DEEP) (Yang et al., 2012)
- Turkish Computer System Usability Questionnaire (T-CSUQ) (Erdinç et Lewis, 2013)
- Usability Metric for User Experience- LITE (UMUX-LITE) (Lewis et al., 2013)
- Speech User Interface Service Quality questionnaire (SUISQ) (Polkosky, 2008 ; Lewis et Hardzinski, 2015)
- Alternate Usability (AltUsability) (Lewis et al., 2015a)

En commençant par le premier questionnaire (QUIS) qui a été apparu à la fin des années 1980, nous résumons un ensemble des caractéristiques de questionnaires dans un tableau récapitulatif (Tableau 2.1). Il représente une synthèse portant sur les points suivants :

- **Date de création du questionnaire**
- **Degré de fiabilité globale du questionnaire :** indique le degré de fiabilité du questionnaire standardisé à travers un coefficient nommé Cronbach alpha[5]
- **Interface utilisateur (IU)/ système visé :** spécifie le type d'interface ou le système que le questionnaire peut l'évaluer, comme exemples : l'interface WIMP (Windows, Icons, Menus, Pointer), WUI (Web User Interface), l'interface utilisateur des applications mobiles, etc.

[5]Cronbach alpha: un élément fondamental d'évaluation psychométrique proposé par (Nunnally, 1978). C'est une mesure de la cohérence interne (fiabilité) que peut varier de zéro (absolument pas fiable) à un (parfaitement fiable). En outre, pour des besoins de recherche ou d'évaluation dans lequel le score final d'un questionnaire est égal à la moyenne des notes de l'échelle, la fiabilité minimale acceptable typique égale à 0,7 (Nunnally, 1978).

- **Nombre d'éléments :** présente le nombre total des éléments du questionnaire.
- **Type/Forme des éléments :** présente la forme applicable aux éléments du questionnaire (une question ou une phrase)
- **Output :** précise le type de résultats fournis par le questionnaire et indique la méthode d'analyse adoptée pour analyser les données des questionnaires relatives aux réponses des utilisateurs.
- **Échelle adoptée :** identifie le type d'échelle utilisée telles que l'échelle de Likert (Likert, 1932 ; Carifio et Rocco, 2007), l'échelle sémantique différentielle (Dalton et al., 2008), etc. Plus de détails sur les échelles de questionnaires sont présentés dans l'Annexe A (Tableau A.1).

Les outils présentés dans le Tableau 2.1 sont organisés chronologiquement par date de création.
A partir de ce tableau, quelques conclusions peuvent être établies. En ce qui concerne le type de l'interface évaluée, nous constatons que 71% (17/24) des questionnaires peuvent être appliqués à l'évaluation de tous les types d'interfaces (telles que, les interfaces WIMP, Web et Mobile). Sept questionnaires soutiennent l'évaluation des interfaces spécifiques, parmi eux cinq sont dédiés à l'évaluation de la qualité des applications Web ; un seul (SUISQ) se focalise sur l'évaluation des applications interactives de réponse vocale et un seul porte sur l'évaluation des applications mobiles (MPUQ).

En ce qui concerne le degré de fiabilité, tous les questionnaires ont indiqué de bons niveaux impliquant des scores de Cronbach alpha variant entre 0,80 et 0,97. La majorité des questionnaires proposés (17 outils parmi 24) disposent d'un niveau de fiabilité très élevé variant entre 0,9 et 0.97. Tandis que trois outils ont un niveau inférieur à 0.90 et quatre autres (PUTQ, USE, ER et UME) sont validés sans indiquer leurs valeurs de Cronbach alpha. Le questionnaire WEQ possède le niveau le plus élevé égal à 0,97. Toutefois, ce questionnaire est dédié uniquement à l'évaluation des interfaces web. Le deuxième niveau de fiabilité le plus élevé concerne MPUQ (0,96). Celui-ci représente également l'unique questionnaire spécifique aux applications Mobile, néanmoins il comporte un nombre élevé des questions égales à 72. En effet, nous notons que les deux questionnaires ASQ et CSUQ ont également des niveaux élevés de fiabilité touchant respectivement les valeurs 0.96 et 0.95. Ces questionnaires sont caractérisés par un nombre réduit de questions par rapport aux autres questionnaires, égal respectivement à 3 et 19 questions. De plus, les questionnaires QUIS et SUMI (dont les dernières versions datent de 2011) ont indiqué des niveaux de fiabilité élevés se montant respectivement à 0,94 et 0,92. Nous constatons aussi que le questionnaire DEEP dispose d'un haut niveau de fiabilité égal à 0.954. Bien qu'il présente l'un des plus récents questionnaires dédiés aux applications web, il peut être considéré comme un des concurrents des questionnaires Web les plus connus tels que WAMMI et SUPR-Q.

En ce qui concerne les résultats des questionnaires, différentes présentations des résultats ont été proposées (telles que la forme graphique, le nombre, les tableurs, les fichiers csv ...). En outre, il existe plusieurs façons permettant de déterminer le degré de satisfaction qui dépendent des échelles des questionnaires adoptées comme la méthode de calcul de moyenne utilisée en général par les questionnaires exploitant l'échelle de Likert, le programme d'analyse SUMISCO concernant le questionnaire SUMI.

En ce qui concerne ces échelles, il est aussi intéressant de constater que l'échelle de Likert présente la méthode la plus courante qui caractérise les questionnaires. Cette échelle a été adoptée par 80% des questionnaires qui comprennent une variété en nombre d'éléments des réponses (par exemple, de 1 à 7 où 1= Fortement en accord et 7=Fortement en désaccord, voir Annexe A (Tableau A.1)). Tandis que les 20% restants des questionnaires sont axés sur d'autres types d'échelles, tels que l'échelle dichotomique (c'est le cas du questionnaire SUMI) et l'échelle sémantique (par exemple, le questionnaire QUIS).

Tableau 2. 1 : Synthèse globale sur les caractéristiques des questionnaires standardisés existants (Assila et al., 2016a)

Questionnaire	Degré de fiabilité globale	IU/ système(s) visé(s)	Nombre d'éléments	Type des éléments	Outputs	Échelle adoptée
QUIS 1988→2011	0.94	Produits et logiciels	27 : V5	Phrase	Les résultats sont présentés avec un format facile qui peut être importé dans les programmes et les tableurs statistiques.	Echelle sémantique de 10 points numériques
TAM 1989	0.94/0.98	Produits et logiciels	12	Phrase	Aucune information	Echelle sémantique de 7 points nominales
ASQ 1990→1995	0.96	Logiciels	3	Phrase	Les résultats sont calculés en utilisant le score moyen entre les sept points de l'échelle.	Echelle de Likert de 7 points
PSSUQ 1990→2002	0.94	Logiciels et systèmes d'information	18 : V1 19 : V2 16 : V3	Phrase	Les résultats sont calculés en utilisant le score moyen entre les sept points de l'échelle.	Echelle de Likert de 7 points
SUMI 1993→2011	0.92	Applications logicielles	50 (V4.0)	Phrase	Les résultats sont calculés en utilisant le programme SUMISCO qui génère une sortie de fichier csv	Echelle dichotomique de 3 points
CSUQ 1995→2002	0.95	Logiciels et systèmes d'information	18 : V1 19 : V2 16 : V3	Phrase	Les résultats sont calculés en utilisant le score moyen entre les sept points de l'échelle	Echelle de Likert de 7 points
SUS 1996	0.92	Logiciels	10	Phrase	Pour obtenir la valeur globale du SUS ; additionner le score de chaque item et le multiplier par 2,5. Les scores sont compris entre 0 et 100	Echelle de Likert de 5 points
PUTQ 1997	Non publié	Systèmes d'information	100	Question	Aucune information	Echelle de Likert de 7 points
WAMMI 1998→2000	0.90	IU web	20	Phrase	Les résultats sont présentés sous forme graphique	Echelle de Likert de 5 points
USE 2001	Non publié	Produits et logiciels	30	Phrase	Aucune information	Echelle de Likert de 7 points
ER 2003	Non publié	Logiciels	2	Phrase	Les résultats sont présentés avec un graphique qui indique un diagramme de dispersion des scores retournés	Echelle de Likert de 7 points
WEBUSE 2003	>0.8	Tout type de sites Web	24	Phrase	Rapport de synthèse indiquant l'aspect de l'utilisabilité, le niveau pour chaque critère et le score moyen.	Echelle de Likert de 5 points nominaux
UME 2003→2004	Non publié	Produits et logiciels	1	Question	Les résultats sont calculés en utilisant une formule mathématique liée à l'UME.	Echelle spécifique UME
MPUQ 2005	0.96	Applications de téléphonie mobile	72	Question	Les résultats sont fondés sur l'analyse hiérarchique des procédés y compris dans un des modèles développés de prise de décision.	Echelle de Likert de 7 points
SEQ 2006	>0.94	Logiciels	1	Phrase	Les résultats sont calculés en utilisant le score moyen entre les cinq/sept points de l'échelle.	Echelle de Likert de 5 points
WEQ 2007	0.97	Sites web des organisations gouvernementales	32	Phrase	Les résultats sont présentés dans un rapport qui contient les évaluations de l'analyse des commentaires des utilisateurs à propos de leurs scores.	Echelle de Likert de 5 points
SMEQ 2009	>0.94	Logiciels	1	Question	Aucune information	Échelle spécifique graduée de 0 à 150
UMUX 2010	0.94	Logiciels	4	Phrase	Score global = [(∑(Item1 + Item2+Item 3+Item4) /24)] *100	Echelle de Likert de 7 points
SUPR-Q 2011	0.94	Interfaces web	13	Question Phrase	Les résultats indiquent une comparaison entre les scores retournés et les scores des autres sites et fournissent des classements relatifs exprimés en pourcentages.	Echelle de Likert de 5/11 points
DEEP 2012	0.954	IU web	19	Phrase	Aucune information	Echelle de Likert de 5 points nominaux
T-CSUQ 2013	0.85	Logiciels	13	Phrase	Les résultats sont calculés en utilisant le score moyen entre les sept points de l'échelle.	Echelle de Likert de 7 points
UMUX-LITE 2013	0.82/ 0.83	Logiciels	2	Phrase	Score global = [(Item1 + Item2)-2] * (100/ 12). Les scores sont compris entre 0 et 100.	Echelle de Likert de 7 points
SUISQ 2008 **SUISQ-R** **SUISQ-MR** 2015	0.93 0.88 -	Applications interactives de réponse vocale	25 : V 1 14 : V2 9 : V3	Phrase	Les résultats sont calculés en utilisant le score moyen	Echelle de Likert de 5 points nominaux
AltUsability 2015	0.9	Logiciels	7	Phrase	Score global=(∑(Item1, Item2, Item3, Item4, Item5, Item6, Item7) -7)*(100/42) ; Les scores sont compris entre 0 et 100.	Echelle de Likert de 7 points

{V : Version}

Divers questionnaires ont été proposés dans la littérature. Ils sont focalisés davantage sur l'évaluation d'une variété d'aspects de qualité, autre que l'utilisabilité, utilisés dans différents contextes d'évaluation. Parmi ceux-ci, nous mentionnons AttrackDiff (Hassenzahl et al, 2003), un instrument qui assure l'évaluation de nombreux aspects relatifs à l'expérience des utilisateurs tel que l'attractivité d'un produit à travers la technique de paires de mots. Ensuite, nous indiquons Service User eXperience (Väänänen et Segerståhl, 2009), un autre questionnaire qui sert à évaluer les capacités des services Web modernes pour promouvoir et soutenir l'expérience des utilisateurs. De plus, McNamara et Kirakowski (2011) ont proposé un questionnaire, nommé Consumer Products Questionnaire, permettant de mesurer la satisfaction des utilisateurs vis-à-vis des produits de consommation électroniques. Plus récemment, Lewis et Mayes (2014) ont proposé un questionnaire standardisé nommé Emotional Metric Outcomes dédié à l'évaluation des résultats affectifs de l'interaction du client, que ce soit avec des services numériques ou humains. Néanmoins, ce questionnaire concerne un contexte plus particulier à propos des études de l'utilisabilité non modérée en utilisant un grand échantillon. De travaux récents ont recommandé l'utilisation de ce questionnaire dans les recherches de l'expérience d'utilisateur au tant qu'une mesure qui peut complémenter les questionnaires d'utilisabilité standardisés existants (Lewis et Mayes, 2014 ; Lewis et al., 2015b ; Lewis, 2014) (pour plus de détails, voir (Lewis et Mayes, 2014)).

Face à la diversité des questionnaires, différents autres travaux se sont intéressés à proposer des classifications selon différentes catégories. D'autres travaux se sont focalisés sur l'élaboration des études comparatives entre ces questionnaires. Les sections suivantes seront consacrées à l'introduction de quelques classifications récemment proposées et à présenter les études comparatives les plus abordées et référencées dans la littérature.

- **Classifications existantes des questionnaires d'utilisabilité standardisés**

Il existe de différentes classifications des questionnaires d'utilisabilité standardisés. Bien que certains travaux ont présenté les questionnaires d'une manière générale (Ryu, 2009 ; Sauro, 2015), d'autres ont fourni diverses classifications (Sauro et Lewis, 2012 ; Chiew et Salim 2003 ; Elling et al., 2007 ; Yang et al., 2012).

Récemment, Sauro et Lewis (2012) ont proposé trois catégories de questionnaires. La première regroupe les questionnaires Poststudy dédiés à assurer une évaluation à la fin d'une étude, en particulier après avoir terminé une série de scénarios de test. La seconde concerne les questionnaires post-task qui fournissent une évaluation plus contextuelle. Ils sont utilisés immédiatement à la fin de chaque tâche ou un scénario dans une étude d'utilisabilité. La dernière catégorie comprend des questionnaires spécifiques dédiés à l'évaluation des applications Web. Nous citons comme exemples WAMMI (Kirakowski et Cierlik, 1998) et SUPR-Q (Sauro, 2011b). Parmi les questionnaires poststudy les plus connus, nous distinguons QUIS, SUMI, PSSUQ et SUS (Sauro et Lewis, 2012). Concernant les questionnaires post-task, nous citons ASQ, SEQ, SMEQ, ER et UME (Sauro et Lewis, 2012).

Yang et al. (2012) ont proposé une autre classification. Elle répertorie les questionnaires d'utilisabilité autour de trois catégories qui dépendent du type de système à évaluer. La première concerne les questionnaires universels applicables à l'évaluation de tous les types de produits électroniques. Nous citons comme exemples USE, CSUQ, TAM, QUIS, SUS et PUTQ. Quant au deux autres catégories, elles portent sur l'évaluation des applications web (par exemple, WAMMI) et mobiles (comme MPUQ).

En tenant compte du Tableau 2.1 de synthèse des questionnaires, nous proposons de classifier les 24 outils présentés suivant les deux classifications comme le montre le Tableau 2.2.

Tableau 2. 2 : Classifications de questionnaires

Questionnaire	Questionnaires basés sur la classification de Yang et al. (2012)			Questionnaires basés sur la classification de Sauro et Lewis (2012)		
	Universels	**Pour les sites web**	**Pour les mobiles**	**Poststudy**	**Post-task**	**Pour les sites web**
QUIS	X			X		
TAM	X			X		
ASQ	X				X	
PSSUQ	X			X		
SUMI	X			X		
CSUQ	X			X		
SUS	X			X		
PUTQ	X			X		
WAMMI		X				X
USE	X			X		
ER	X				X	
WEBUSE		X				X
UME	X				X	
MPUQ			X	X		
WEQ		X				X
SMEQ	X				X	
SEQ	X				X	
UMUX	X			X		
SUPR-Q		X				X
DEEP		X				X
T-CSUQ	X			X		
UMUX-LITE	X			X		
SUISQ/ SUISQ-R				X		
AltUsability	X			X		
Nombre total	*17 (de 24)*	*5 (de 24)*	*1 (de 24)*	*14 (de 24)*	*5 (de 24)*	*5 (de 24)*

Dans ce qui suit, nous allons adopter le terme « Questionnaire spécifique » pour désigner les questionnaires spécifiques pour les applications mobiles, Web ou d'autres applications spécifiques (comme le cas du questionnaire SUISQ qui concerne les applications interactives de réponse vocale) présentés dans le Tableau 2.2.

- **Comparaison des questionnaires d'utilisabilité standardisés**

En examinant la littérature, nous notons l'existence d'un nombre réduit d'études qui ont procédé aux comparaisons directes des divers questionnaires standardisés d'utilisabilité (Sauro et Lewis, 2012). Ces études ont porté seulement sur neuf questionnaires parmi les 24 présentés, qui sont : SUS, QUIS, CSUQ, UMUX, UMUX-Lite, AltUsability, SEQ, UME et SMEQ.

Différents travaux (Tullis et Stetson, 2004 ; Borsci et al., 2015 ; Lewis et al., 2015a) ont accordé plus d'attention à la comparaison de SUS avec les autres questionnaires standardisés (QUIS, UMUX et UMUX-LITE) ainsi l'étude de la corrélation entre eux. Cela peut être justifié par le fait que SUS présente un standard de l'industrie connu sous la dénomination ―Quick and Dirty". Il est fréquemment utilisé dans les études d'utilisabilité et a été référencé dans plus de 600 publications (Sauro et Lewis, 2012 ; Lewis, 2014 ; Sauro, 2011a, 2015 ; Lewis et al., 2015a). Néanmoins, ce questionnaire sert principalement à assurer une évaluation globale et rapide d'utilisabilité, plutôt que la découverte de problèmes d'utilisabilité (Brooke, 1996).

Dans une autre étude conduite par Tullis et Stetson (2004) pour l'évaluation de l'utilisabilité des sites web, SUS a été comparé avec quatre questionnaires (QUIS, CSUQ, Words[6] et Ours[7]). Les résultats

[6]Words (adapté de Microsoft's Product Reaction Cards) : est un questionnaire basé sur les 118 mots utilisés par Microsoft sur leurs cartes de réaction de produit. L'utilisateur doit choisir les mots qui décrivent le mieux leur interaction avec le site Web. Chaque mot est présenté par une case à cocher (Benedek et Miner, 2002).

[7]Ours : est un ancien questionnaire d'évaluation de sites web personnalisé qui a été proposé par Tullis et Stetson. Il est composé de neuf éléments qui se basent sur une échelle de sept points allant de «Fortement en désaccord » à « Fortement en accord » (Tullis et Stetson, 2004).

d'analyse rapportés de cette étude ont montré que le questionnaire SUS a été le plus rapide à converger sur la bonne conclusion avec des résultats fiables à travers toutes les tailles adoptées d'échantillons. En outre, deux travaux récents (Borsci et al., 2015 ; Lewis et al., 2015a) se sont focalisés sur l'étude de la correspondance et la corrélation entre SUS et d'autres questionnaires. La première a comparé SUS avec UMUX et UMUX-LITE et la seconde par rapport à UMUX-LITE et AltUsability. Par conséquent, les résultats fournis par ces deux études ont rapporté une forte corrélation et des correspondances élevées entre ces questionnaires (pour plus de détails voir (Tullis et Stetson, 2004 ; Borsci et al., 2015 ; Lewis et al., 2015a)).

En outre, deux autres études significatives de comparaisons directes des questionnaires ont été menées par Tedesco et Tullis (2006) et Sauro et Dumas (2009). Ces travaux s'articulent autour de cinq questionnaires standardisés classés selon Sauro et Lewis (2012) dans la catégorie des questionnaires post-task. Dans ces études, les auteurs se sont concentrés sur la mesure de la sensibilité[8] entre les questionnaires afin de déterminer celui le plus sensible.
Dans la première étude, Tedesco et Tullis (2006) ont effectué une comparaison sur cinq questionnaires (SEQ-V1, V2-SEQ, ASQ, ER et SEQ-V3). Par conséquent, l'analyse psychométrique a montré de bons résultats au niveau de tous les questionnaires en utilisant des échantillons de larges tailles. Néanmoins, il a prouvé que SEQ-V1 a été le plus sensible en utilisant des échantillons de plus petites tailles. Dans la deuxième étude, Sauro et Dumas (2009) ont comparé SEQ avec d'autres questionnaires (SMEQ et UME). Par conséquent, en utilisant des échantillons de petites tailles (<5), les analyses ont indiqué que les questionnaires étaient insensibles avec un peu de différence entre eux. Pour les échantillons de tailles supérieur à cinq, les résultats ont révélé que SMEQ dispose de meilleur pourcentage significatif **—t**ests ", mais il reste insuffisant (Sauro et Dumas, 2009 ; Sauro et Lewis, 2012). Toutefois, ces deux études recommandent, également, l'utilisation des deux questionnaires SEQ et SMEQ (Tedesco et Tullis, 2006 ; Sauro et Dumas, 2009).

- **Les aspects de qualité existants des questionnaires d'utilisabilité standardisés**

En analysant les questionnaires d'utilisabilité standardisés, nous avons identifié qu'ils déclarent explicitement différents aspects de qualité. Dans certains questionnaires, ces aspects se réfèrent aux critères de qualité (comme, la satisfaction, l'utilisabilité, l'efficience, l'efficacité, etc.). D'autres correspondent à des caractéristiques de l'interface utilisateur, citons à titre d'exemples (les facteurs d'écran, les liens, la mise en page, etc.).
Nous synthétisons chaque aspect des questionnaires dans l'Annexe A (Tableaux A.2 et A.3). Ces tableaux représentent l'ensemble des aspects identifiés dans les 24 questionnaires standardisés.

Nous avons trouvé 30 aspects de qualité existants qui concernent les questionnaires universels. Les aspects de l'utilité du système, l'apprentissage, l'utilisabilité, la facilité globale de l'achèvement de la tâche et le système global présentent les critères les plus fréquents adoptés par les questionnaires universels. Ces critères concernent des aspects généraux de qualité. A titre d'exemple, le questionnaire CSUQ qui couvre quatre critères généraux : le système global, l'utilité du système, la qualité de l'information et la qualité de l'interface.
En outre, certains questionnaires ont mis l'accent sur la mesure des aspects plus spécifiques, y compris PUTQ, SUMI, AltUsability. En effet, PUTQ couvre huit critères (la compatibilité, la cohérence, la flexibilité, l'action minimale, la charge minimale de la mémoire, la limitation de la perception et le guidage de l'utilisateur). Néanmoins avec 100 éléments, il présente le questionnaire le plus long par rapport aux autres trouvés. SUMI est le deuxième plus long questionnaire avec 50 éléments. Il couvre cinq critères d'utilisabilité (l'apprentissage, l'efficience, l'influence, la serviabilité et le contrôle). Par

[8]Sensibilité : capacité d'un questionnaire à indiquer la différence significative entre les systèmes.

conséquent, le nombre élevé des éléments peut influencer le rendement des utilisateurs lors du test d'utilisabilité (Yang et al., 2012). En outre, AltUsability (Lewis et al., 2015a) est un questionnaire récent qui porte sur des aspects plus spécifiques d'utilisabilité (la facilité de navigation, la capacité de recherche, la familiarité, le besoin, et l'attractivité). De plus, nous constatons qu'ER, SMEQ et SEQ sont les plus courts questionnaires (avec seulement un élément) qui couvrent un seul aspect général portant sur la facilité globale de l'achèvement de la tâche.

Concernant les questionnaires spécifiques d'utilisabilité, ils couvrent 38 aspects de qualité. Cependant, nous avons trouvé que la totalité d'entre eux ne sont pas différents. En effet, certains aspects diffèrent en termes de terminologies utilisées, tel est le cas du critère de l'apprentissage (WAMMI) et la facilité d'apprentissage (MPUQ)). En outre, nous avons constaté que certains autres portent sur des aspects généraux de qualité, citons par exemple SUPR-Q qui couvre quatre aspects (y compris la facilité d'utilisation et l'apparence). D'autres questionnaires s'adressent à des aspects plus spécifiques, comme par exemple le critère de navigation (relatif au questionnaire DEEP) qui a été décrit par le questionnaire WEQ en fonction de cinq sous-critères (la convivialité, la structure, les hyperliens, la vitesse et l'option de recherche). Certains autres questionnaires ont combiné des sous-critères dans un seul critère, tels que WEBUSE (contenu, organisation et lisibilité), DEEP (structure et architecture de l'information) et MPUQ (contrôle et efficacité ; la facilité d'apprentissage et utilisation).

A la fin de cette section, nous pouvons conclure que malgré l'existence de plusieurs aspects de qualité adressés par les questionnaires, la majorité d'entre eux considèrent des aspects généraux. De ce fait, il parait intéressant de rendre ces aspects uniformes par rapport aux critères de qualité traditionnels définis par les normes et les guides les plus connus. Cela va permettre de distinguer les différents critères d'utilisabilité pouvant couvrir chaque questionnaire afin d'aider les évaluateurs de l'utilisabilité pour l'identification des problèmes des interfaces utilisateur.

Nous allons présenter dans la section suivante une analyse des 24 questionnaires présentés selon des critères de qualité connus, définis par les standards.

2.1.1.2 Etude d'analyse des questionnaires selon des critères standards d'utilisabilité

Pour effectuer notre analyse, nous décidons de prendre en compte deux ensembles de critères d'utilisabilité largement utilisés et référencés dans la littérature : ceux proposés par la norme ISO 9241-11(1998) (efficacité, efficience et satisfaction) et les critères ergonomiques proposés par Scapin et Bastien (1997) (guidage, charge de travail, contrôle explicite, adaptabilité, gestion des erreurs, cohérence, signifiance des codes et dénominations et compatibilité). Pour compléter ces critères ergonomiques, nous avons trouvé intéressant d'évaluer la qualité des informations sur la base des critères d'utilisabilité définis par la norme ISO/WD 9241-112 (2013) (détectabilité, discriminabilité, adéquation, cohérence, compréhensibilité). Ce standard définit les principes de conception ergonomiques pour les systèmes interactifs liés à la présentation des informations. Ces derniers sont utiles pour la conception et l'évaluation de tous les types de interfaces utilisateur (ISO 9241-112, 2013). Sur la base de ces critères, nous avons effectué une analyse de tous les questionnaires dans le but d'identifier clairement les critères d'utilisabilité de trois groupes qui sont considérés ou non par les questionnaires standardisés. Environ 475 éléments de questionnaires ont été analysés selon les listes sélectionnées de critères. Nous présentons dans ce qui suit, les résultats obtenus d'analyse des questionnaires (universels et spécifiques).

Pour chaque questionnaire, nous avons suivi une analyse détaillée sur chaque élément selon les listes de critères choisis. Nous nous sommes basés principalement sur le sens de l'élément afin d'identifier à quel critère d'utilisabilité, il peut être plutôt lié. Par exemple, l'élément suivant *"I can effectively complete my work using this system"* (extrait du questionnaire CSUQ) est davantage lié au critère d'efficacité.

Nous avons adopté ce processus pour analyser tous les éléments de questionnaires. Par conséquent, nous synthétisons les résultats de l'analyse effectuée pour les questionnaires spécifiques ceux dans le Tableau 2.3 et ceux qui sont universels dans le Tableau 2.4. Toutefois, nous avons exclu de cette analyse quatre questionnaires universels (SEQ, SMEQ, ER, UME), du fait qu'ils contiennent un seul élément portant sur le critère de la facilité globale de l'exécution des tâches. De plus, nous avons exclu le questionnaire spécifique SUISQ, vu la spécificité de ses critères qui s'adressent davantage aux applications interactives de réponse vocale.
Dans le Tableau 2.5, nous avons introduit des exemples[9] sur les éléments des questionnaires universels et spécifiques relatifs à quelques critères d'utilisabilité considérés (les listes des exemples complètes sont présentées dans les Tableaux A.4 et A.5 dans l'annexe A). Nous sommes conscients que certains critères sont généralement liés entre eux et que certains éléments peuvent être liés à plusieurs critères. A titre d'exemple, nous avons attribué l'élément suivant *"The organization of information on the system screens is clear"* (extrait de CSUQ et PSSUQ) aux deux critères de discriminabilité et guidage (voir Annexe A, Tableau A.4).

Comme le montre les tableaux 2.3 et 2.4, nous avons identifié différents critères d'utilisabilité couverts par les questionnaires universels et spécifiques. A partir du Tableau 2.4, nous notons que la majorité des questionnaires standardisés universels couvrent plus de 6 sur 15 critères d'utilisabilité. Par exemple, nous avons constaté que les cinq aspects de qualité de SUMI couvrent tous les critères d'utilisabilité mentionnés. Concernant les questionnaires spécifiques, nous avons constaté que la plupart d'entre eux (WAMMI, WEBUSE, WEQ et DEEP) couvrent plus de 7 sur 15 critères d'utilisabilité en plus de leurs aspects actuels de qualité. En outre, nous notons que le seul questionnaire portant sur l'évaluation des interfaces utilisateur mobiles (MPUQ) couvre tous les critères d'utilisabilité considérés.

Tableau 2. 3 : Les résultats d'analyse des questionnaires spécifiques d'utilisabilité

Critères d'utilisabilité		WAMMI	WEBUSE	WEQ	SUPR-Q	DEEP	MPUQ
Critères de l'ISO 9421-11	Efficacité		X				X
	Efficience	X	X	X	X	X	X
	Satisfaction	X			X		X
Critères de l'ISO/WD 9241-112	Détectabilité	X	X	X		X	X
	Discriminabilité	X	X	X	X	X	X
	Adéquation	X	X	X	X	X	X
	Cohérence		X			X	X
	Compréhensibilité	X	X	X		X	X
Critères de Scapin et Bastien	Guidage	X	X	X	X	X	X
	Charge de travail	X	X	X	X	X	X
	Contrôle explicite		X			X	X
	Adaptabilité	X	X				X
	Gestion des erreurs						X
	Cohérence		X				X
	Signifiance des codes	X	X	X		X	X
	Compatibilité						X

[9]Pour éviter tout problème qui peut être engendré par la traduction des éléments des questionnaires standardisés, nous avons décidé de les présenter dans leur langue originale (Anglais).

Tableau 2. 4 : Les résultats d'analyse des questionnaires universels d'utilisabilité

Critères d'utilisabilité		QUIS	TAM	PSSUQ	CSUQ	Alt-Usability	T-CSUQ	SUMI	SUS	PUTQ	USE	UMUX	UMUX-Lite	ASQ
Critères de l'ISO 9421-11	Efficacité	X		X	X		X	X			X	X	X	X
	Efficience		X	X	X	X	X	X	X			X		X
	Satisfaction	X		X	X		X	X	X		X	X		X
Critères de l'ISO/WD 9241-112	Détectabilité	X	X					X		X				
	Discriminabilité	X		X	X		X	X		X				
	Adéquation (Appropriateness)							X	X	X			X	
	Cohérence	X						X	X	X	X			
	Compréhensibilité	X	X					X		X				
Critères de Scapin et Bastien	Guidage	X		X	X		X	X		X	X			X
	Charge de travail			X	X		X	X		X	X			
	Contrôle explicite					X		X		X				
	Adaptabilité	X	X			X		X		X	X			
	Gestion des erreurs	X		X	X		X	X		X	X			
	Cohérence	X						X	X	X	X			
	Signifiance des codes	X						X		X				
	Compatibilité							X		X				

Cette analyse nous a fourni les critères d'utilisabilité les plus abordés par les questionnaires standardisés. Pour les questionnaires universels, nous notons que les trois critères de la norme ISO 9241-11 (efficacité, efficience et satisfaction) sont les plus considérés (y compris chacun des neuf questionnaires) (voir Tableau 2.4). Nous constatons aussi que l'analyse a montré que 8 questionnaires universels parmi 13 ont utilisé le critère du guidage. Pour les questionnaires spécifiques, nous observons que tous les questionnaires couvrent les cinq critères suivants : l'efficience, la discriminabilité, l'adéquation, la charge de travail et le guidage.

Tableau 2. 5 : Exemples des éléments de questionnaires d'utilisabilité en fonction des quelques critères standards

Critères d'utilisabilité	Items	Questionnaire
Efficacité (ISO 9241-11, 1998)	Tasks can be performed in a straight-forward manner I can effectively complete my work using this system Tasks can be performed in a straight forward manner using this software I can use it successfully every time [This system's] capabilities meet my requirements (Finstad, 2010 ; Kerzazi et Lavallée, 2011) Overall, I am satisfied with the ease of completing the tasks in this scenario (Lewis, 1995)	QUIS CSUQ/PSSUQ/T-CSUQ SUMI USE UMUX/UMUX-Lite ASQ
Efficience (ISO 9241-11, 1998)	I am able to complete my work quickly using this system I found the system very cumbersome to use This software responds too slowly to inputs I have to spend too much time correcting things with [this system] (Kerzazi et Lavallée, 2011) Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario (Lewis, 1995) Using [this product] in my job would enable me to accomplish tasks more quickly This system helps me to do my job more efficiently I am able to find what I need quickly on this website	CSUQ/PSSUQ/T-CSUQ SUS SUMI UMUX ASQ TAM AltUsability SUPR-Q
Detectabilité (ISO/WD 9241-112, 2013)	Characters on the computer screen (hard to read,…, easy to read) Either the amount or quality of the help information varies across the system Are selected data highlighted? My interaction with [this product] would be clear and understandable Reading content at this website is easy	QUIS SUMI PUTQ TAM WEBUSE
Compréhensibilité (ISO/WD 9241-112, 2013)	My interaction with [this product] would be clear and understandable I can understand and act on the information provided by this software Do the commands have distinctive meanings? Messages on screen which prompt user for input (confusing,...,clear) I find the information in this website easy to understand The content (including text, pictures, audios, and videos etc.) was easy to understand Is the interface with this product clear and understandable?	TAM SUMI PUTQ QUIS WEQ DEEP MPUQ
Guidage (Scapin et Bastien, 1997)	Computer keeps you informed about what it is doing (never always) Help messages on the screen (unhelpful,…, helpful) The information (such as online help, on-screen messages and other documentation) provided with this system is clear The organization of the menus seems quite logical Is the guidance information always available? Is HELP provided? I can easily know where I am at this website I always know where I am on this website The website has a clean and simple presentation This website helped me find what I was looking for Is the backlighting feature for the keyboard and screen helpful?	QUIS QUIS CSUQ/PSSUQ/T-CSUQ SUMI USE PUTQ WEBUSE WEQ SUPR-Q DEEP MPUQ
Charge de travail (Scapin et Bastien, 1997)	It is easy to find the information I needed There is never enough information on the screen when it's needed There are too many steps required to get something to work It requires the fewest steps possible to accomplish what I want to do with it Is the screen density reasonable? The information on this website is valuable It was easy to find the information I needed on the website Are data items kept short?	CSUQ/PSSUQ/T-CSUQ SUMI SUMI USE PUTQ SUPR-Q DEEP MPUQ

2.1.2 Les outils assurant une évaluation objective

Dans la littérature, il existe une grande variété des outils d'évaluation de l'utilisabilité des interfaces utilisateur qui s'appuient sur une évaluation objective indépendamment des perceptions des utilisateurs. A titre d'exemples, nous citons : ERGOVAL (Farenc et al., 2001) ; Destine (Mariage et al., 2004) ; AWebHUT (Rukshan et Baravalle, 2011) ; ErgoCoIn (Morandini et al., 2011) ; Morae (TechSmith, 2012) et Usabilla (Usabilla, 2012).

L'Environnement intégré d'aide à l'évaluation des systèmes interactifs adopté dans le cadre de notre thèse comporte deux outils d'évaluation objective (le mouchard électronique EISEval et l'inspecteur des règles ergonomiques). Dans cette section, nous présentons des synthèses sur quelques outils de recueil des données d'évaluation sur l'interaction entre le système interactif à évaluer et l'utilisateur et quelques outils utilisant les règles ergonomiques proposés dans la littérature.

2.1.2.1 Les outils utilisant les règles ergonomiques

Comme mentionné dans le Chapitre 1 (Section 1.5), l'évaluation ergonomique des systèmes interactifs porte sur la mesure de la qualité des leurs interfaces utilisateur dans le but d'améliorer leurs qualités. Nombreux outils d'évaluation portant sur l'inspection de la conformité des IHM en exploitant les connaissances et les règles ergonomiques ont été proposés, tels que : SYNOP (Kolski, 1989), Sherlock (Grammenos et al., 2000), le système de gestion de règles (Parush, 2000), ERGOVAL (Farenc et al., 2001), Web Tango (Ivory et Hearst, 2001), DESTINE (Mariage et al, 2004), l'outil d'évaluation ergonomique (Charfi et Ezzedine, 2011), AWebHUT (Rukshan et Baravalle, 2011) et l'inspecteur des règles ergonomiques (Charfi et al., 2014, 2015).

Ces outils se focalisent sur l'évaluation des interfaces utilisateur de types WIMP et Web dont seulement deux outils applicables aussi à l'évaluation des applications de type Mobile (le système de gestion de règles (Parush, 2000) et l'inspecteur des règles ergonomiques (Charfi et al., 2014)). L'usage de la majorité de ces outils requiert principalement l'intervention des évaluateurs. Certains outils n'exigent aucunes connaissances pré-requises de l'évaluateur (tels que AWebHUT et Web Tango). Des outils requièrent un niveau basique (par exemples, l'outil d'évaluation ergonomique et Ergoval), et d'autres requièrent des bonnes connaissances en ergonomie (tel est le cas de Sherlock). De plus, quelques outils (Sherlock et l'inspecteur des règles ergonomiques) impliquent aussi l'intervention des experts pour l'aide à la configuration des règles ergonomiques.

En effet, ces règles sont stockées en général dans de bases de données (appelées aussi base des règles, base des connaissances) et structurées selon différents attributs (propriétés) (tels que la clef, le titre, la classe de règles, type d'inspection). Au niveau de la flexibilité des règles ergonomiques, nous distinguons une variété selon les outils. Certains (comme, Sherlock et le système de gestion des règles, Destine, l'outil d'évaluation ergonomique et l'inspecteur des règles ergonomiques) fournissent différentes fonctionnalités pour gérer les règles telles que le choix, l'ajout, la modification et la suppression. Tandis que les autres (Ergoval, WebTango, et AWebHUT) ne fournissent aucune fonctionnalité.

Au niveau des sorties fournies par ces outils, nous constatons que tous permettent d'analyser d'une façon automatisée les résultats et les présenter dans des rapports d'évaluation contenant les violations d'utilisabilité détectées sur l'interface évaluée. Par contre, seulement deux outils sur huit (Destine et l'inspecteur des règles ergonomiques) suggèrent des recommandations et des corrections automatiques dans le but d'améliorer la qualité de l'interface utilisateur.

Tous les détails sur ces outils sont présentés dans l'annexe A, Tableau A.6.

2.1.2.2 Les outils de recueil des données d'interaction entre l'utilisateur et le système interactif à évaluer

Dans cette section, nous rapportons une synthèse sur quelques outils permettant d'assurer une évaluation objective portant sur la capture des interactions des utilisateurs avec le système interactif à évaluer. Ces outils visent en général à capturer différents types des données d'interaction en temps réel, sauvegarder ces informations et spécifier une méthode d'analyse de ces données pour inspecter la qualité des systèmes interactifs. Différentes méthodes d'analyse ont été proposées, telles que l'analyse manuelle, l'analyse semi-automatique ou l'analyse automatique (Bastien et Scapin, 2002 ; Ezzedine et Abed, 1997 ; Ezzedine et Trabelsi, 2005 ; Mariage, 2005). Dans la littérature, nous distinguons une variété d'outils (sous forme de logiciel ou de service web) portant sur l'évaluation de différents types des systèmes interactifs : WIMP (comme MESIA (Trabelsi et al, 2006 ; Trabelsi et Ezzedine, 2013) et EISEval (Tran et al., 2013)), Web (comme les outils WebQuilt (Hong et al., 2001 ; Waterson, 2002) et Web RemUSINE (Paganelli et Paternò, 2002 ; 2003)) et Mobile (tel est le cas de Multi Device RemUSINE (Paterno et al., 2007)). Tous les détails sur ces outils sont présentés dans l'annexe A, Tableau A.7.

Au niveau des intervenants, ces outils nécessitent l'intervention d'un évaluateur pour gérer l'évaluation. En outre, ils se basent tous sur la participation des utilisateurs pour capturer leurs données d'interactions réelles. La capture de ces données se fait d'une manière automatique. Nous notons une diversité au niveau des données capturées. En effet, tous les outils permettent de capturer les événements physiques (de bas niveau) (par exemples, les événements du clavier et les actions de la souris). Certains outils (tels que Web RemUsine (Paganelli et Paternò, 2002 ; 2003), Mutli Device RemUsine (Paterno et al., 2007), et EISEval (Tran et al., 2013)) capturent en plus les évènements d'interface utilisateur et quatre outils sur cinq (WebQuilt, Web RemUsine, Mutli Device RemUsine, et EISEval) capturent aussi les évènements de niveau abstrait d'interaction (tel que les actions relatives à un champ de texte). Des nombreuses techniques de capture sont utilisées. Chaque technique dépend du processus d'évaluation adopté. L'utilisation de l'outil QC/ Replay[10] (log file) est considérée par la famille des outils « Web RemUsine et Mutli Device RemUsine » (Paganelli et Paternò, 2002 ; 2003 ; Paterno et al., 2007). Les deux outils MESIA et EISEval se focalisent sur l'injection de code dans le code source. Quant à l'outil WebQuilt, il considère le proxy pour capturer les données d'interactions.
En outre, tous ces outils offrent une phase automatique d'analyse. Les résultats d'analyse pour la majorité de ces outils sont affichés sous formes textuelle et/ou graphique et s'adressent à une diversité des problèmes d'utilisabilité portant par exemples sur les tâches exécutées, les pages visitées, etc. Toutefois, nous constatons qu'aucun de ces outils ne fournit une phase de critique automatisée des résultats. De ce fait, l'évaluateur doit intervenir pour interpréter les résultats et proposer des améliorations.

Dans la littérature, nous notons que l'intérêt actuel des travaux de recherche s'articule autour de l'usage de différentes méthodes et/ou outils d'évaluation assurant à la fois des évaluations subjective et objective sur l'utilisabilité des interfaces utilisateur.

Au niveau de la deuxième section, nous allons présenter un ensemble de ces travaux.

[10] **QC/Relay** est un outil utilisé par les interfaces graphiques GUI. Il permet de capturer, enregistrer, lire et relire les actions exécutées par les utilisateurs. Il utilise le langage TCL (Tool Command Language).
http://sqa.fyicenter.com/FAQ/Testing-Tools/Software_Test_Tools_QC_REPLAY.html

2.2 Les approches basées sur une combinaison des évaluations subjective et objective

Nous avons vu qu'il existe de nombreux outils d'évaluation d'utilisabilité. Chacun peut assurer une évaluation subjective ou objective. Au cours de cette décennie, nous distinguons que l'utilisation de différentes méthodes et/ ou outils d'évaluation d'utilisabilité subjective et objective à la fois est largement recommandée pour avoir des meilleurs résultats (Tran et al., 2008 ; Al-Wabil et Al-Khalifa, 2009 ; Lewis, 2014 ; Charfi et al., 2015). Cela offre une évaluation plus précise et complète.

Quelques travaux ont été effectués dans ce contexte. Dans ce qui suit, nous présentons un ensemble de ces approches portant sur la combinaison des évaluations subjective et objective, suivi par un tableau récapitulatif de synthèse sur ces approches étudiées et une discussion. De ce fait, nous traiterons dans chaque approche les points suivants :

- **Interface utilisateur visée :** spécifie le type d'interface utilisateur à évaluer, citons par exemples : les interfaces WIMP, WUI, Mobile.
- **Aspects et ou critères de qualité à évaluer :** concerne les aspects et/ou les critères de qualité à inspecter, tels que : la navigation, la qualité d'information, etc.
- **Nature des évaluations combinées :** identifie la nature des évaluations subjective et objective combinées telles que la combinaison des méthodes, la combinaison des outils, la combinaison des données, etc.
- **Principe de combinaison :** concerne le principe adopté pour assurer la combinaison.
- **Problèmes détectés :** indique les problèmes de qualité pris en considération lors de l'approche d'évaluation.
- **Résultats fournis à la fin de l'évaluation :** concerne le rapport d'évaluation et précise la manière (graphique/textuelle) ou/et le format de présentation des résultats, par exemples pdf, html, etc.
- **Avantages et inconvénients de l'approche :** spécifie les principaux atouts et faiblesses de l'approche étudiée.

2.2.1 Présentation des approches étudiées

Les approches étudiées seront présentées dans l'ordre croissant de leurs apparitions.

2.2.1.1 WebUSE : approche d'évaluation d'utilisabilité des interfaces Web

WebUSE est une approche d'évaluation d'utilisabilité des interfaces utilisateur destinée aux systèmes interactifs de type Web (Nikov et al, 2003). Elle porte sur l'intégration des données d'évaluation subjective et objective d'utilisabilité dans le but d'inspecter les problèmes de l'interface utilisateur. Les données objectives sont collectées depuis le site web à évaluer, à titre d'exemples, citons la vitesse d'accès à l'internet, le nombre d'interactions, etc. Elles correspondent à quatre critères comme le montre la Figure 2.1 (background d'utilisateur, environnement, interactions et paramètres physiologiques). Les données subjectives concernent les avis et les opinions des utilisateurs. Elles correspondent comme exemples à la détectabilité de la structure de site, la satisfaction, la lisibilité des informations, etc. Ces données sont extraites d'une façon subjective depuis des questionnaires d'utilisabilité. Elles sont regroupées selon chacun des critères suivants (qualité de navigation, qualité de conception d'interface, qualité d'information, qualité globale du système et attitudes d'utilisateur). En outre, cette approche s'appuie sur un modèle hybride qui applique un algorithme nommé QuickFBP[11] pour agréger les données subjectives et objectives relatives aux différents critères dans un

[11] L'algorithme QuickFBP (Nikov et Stoeva, 2001) est un algorithme de propagation de logique floue. Il offre une accélération de calcul importante pour améliorer les réseaux multicouches. De plus, il fournit une accélération considérable de la convergence de l'algorithme FBP basée sur un calcul modifié de la fonction nette.

indice unique d'utilisabilité. Ce modèle combine les principes de la logique floue avec les réseaux de neurones. L'algorithme de la combinaison a été implémenté dans un outil logiciel MS EXCEL afin de calculer l'indice d'utilisabilité. Les données sont mesurées dans des échelles de types différents comme l'échelle d'intervalle (pour les données objectives) et l'échelle nominale ou ordinale (pour les données subjectives). Par la suite, tous les résultats sont convertis sur la base d'une échelle d'intervalle [1 ; 5], où 1 est le pire et 5 est le meilleur résultat. De plus, chaque ensemble de données pour tous les critères d'utilisabilité a été estimé par un coefficient de concordance[12].
A la fin de l'évaluation, l'approche WebUSE fournit de différents résultats qui peuvent être présentés sous forme graphique. Ceux-ci correspondent aux : (i) indices relatifs aux résultats subjective et objective comme le montre la Figure 2.2, (ii) à l'indice de chaque critère évalué et (iii) l'indice d'agrégation de deux évaluations subjective et objective (voir Figure 2.3). Ce dernier est comparé avec les valeurs calculées par l'approche conventionnelle moyenne pondérée et les valeurs des évaluations des experts.

Néanmoins, l'usage de cette approche est restreint aux sites web. Nous notons l'absence de la corrélation entre les critères subjectifs et objectifs adoptés. En outre, cette approche ne fournit pas d'indications concernant la façon de collecter les données d'évaluation. De plus, nous distinguons un manque de détail concernant le questionnaire adopté lors de l'évaluation subjective ainsi l'algorithme QuickFBP utilisé pour l'agrégation des données d'évaluation. Celle-ci, à son tour, peut entrainer la perte de données (subjective ou objective) qui peut affecter l'exactitude des problèmes d'utilisabilité.

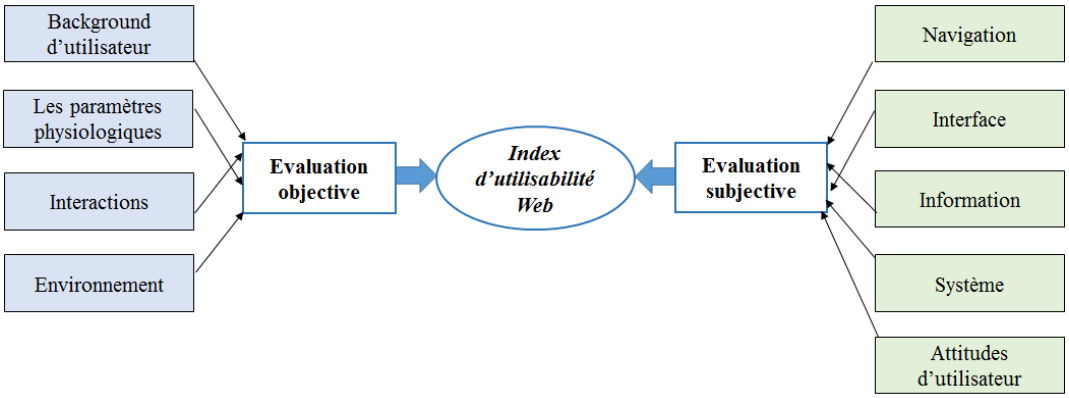


Figure 2. 1 : Modèle hiérarchique de WebUSE (inspiré de (Nikov et al., 2003))

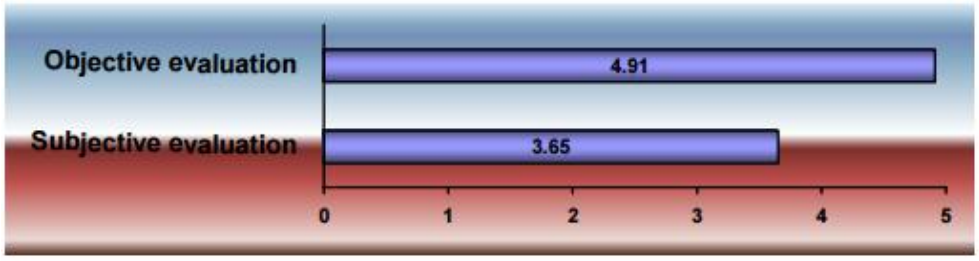


Figure 2. 2 : Exemples des résultats d'évaluations objective et subjective fournis par WebUSE (Nikov et al., 2003)

[12] Coefficient de concordance : test statistique d'accord entre deux ou plusieurs jugements. Il permet de chiffrer l'intensité ou la qualité de l'accord réel entre des jugements qualitatifs appariés. http://kappa.chez-alice.fr/Kappa_2juges_Def.htm

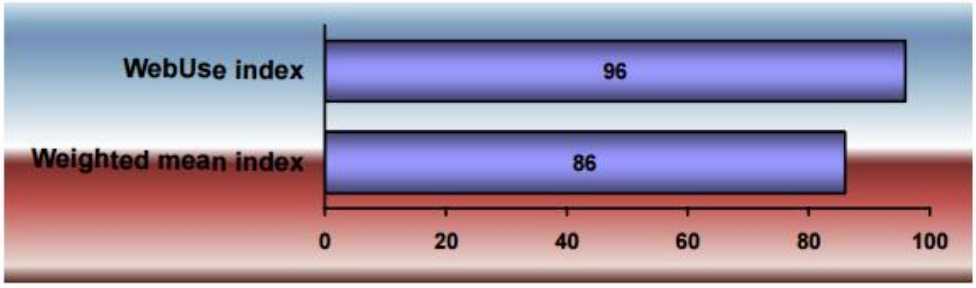


Figure 2. 3 : Exemple de résultat final d'évaluation fourni par WebUSE (Nikov et al., 2003)

2.2.1.2 Méthode de normalisation des mesures d'utilisabilité en un score unique (SUM)

Cette méthode d'évaluation d'utilisabilité des systèmes interactifs est proposée par (Sauro et Kindlund, 2005). Elle vise à assurer une évaluation sommative d'un produit logiciel ainsi que de comparer des différentes tâches, études ou produits. En outre, cette méthode permet de simplifier les trois aspects (efficacité, efficience et satisfaction) d'utilisabilité définis par les normes ISO 9241-11 (1998) et ANSI/NCITS 354-2001 (2001) en une seule mesure sommative et standardisée nommée SUM (Summated Usability Metric) comme le montre la Figure 2.4. Celle-ci s'appuie sur l'agrégation de quatre mesures d'évaluation, qui correspondent aux trois critères d'utilisabilité, dont trois mesures sont objectives et une subjective. Ces mesures sont : le temps de tâche, le nombre d'erreurs, l'achèvement de la tâche et la satisfaction moyenne des utilisateurs. Ces quatre mesures sont combinées en utilisant des méthodes de normalisations adaptées de Six Sigma[13]. L'agrégation porte sur le calcul de la moyenne des mesures normalisées lors de chaque tâche en utilisant la technique de l'Analyse en Composantes Principales[14]. L'utilisation de cette technique comporte divers avantages tels que la construction d'un modèle d'utilisabilité qui minimise l'erreur aléatoire pour chaque mesure, la suppression des données redondantes à partir des variables qui se chevauchent (par exemple, si l'information contenue dans les erreurs peut être représentée par le temps ou la satisfaction).

Deux avantages majeurs supportent l'utilisation de cette méthode. D'abord, ce score fournit une variable continue qui peut être exploitée dans l'analyse de régression et les tests d'hypothèses. De la même manière les mesures existantes peuvent être utilisées pour inspecter l'utilisabilité. Deuxièmement, une seule mesure basée sur les limites de spécification logique donne une idée de la façon avec laquelle une tâche ou un produit peut être utilisable sans avoir à référencer des données historiques. Toutefois, cette méthode possède des limites. En effet, l''utilisation de ce score dépend des données brutes utilisées dans cette étude et ne peut pas être comparé à d'autres valeurs des composantes provenant d'autres ensembles de données. En outre, l'agrégation des mesures subjective et objective en un score unique ne peut pas être bénéfique pour la détection des problèmes d'utilisabilité (Sauro, 2005). Selon Bradner et Dawe (2012), d'autres limites ont été relevées par rapport à l'équation d'agrégation de SUM. Celles-ci portent sur le manque d'orientation autour de décisions importantes (telles que la façon de choisir les niveaux cibles pour chaque mesure), ainsi les interprétations lors de la déclaration des résultats qui requièrent généralement de longues études pour avoir des conclusions.

[13] Six Sigma est un ensemble de techniques et d'outils d'amélioration des processus. Il a été introduit par l'ingénieur Bill Smith et utilisé dans de nombreux secteurs industriels. Cette méthode vise à améliorer la qualité de sortie d'un processus par l'identification et l'élimination des causes de défauts (les erreurs) et de minimiser la variabilité dans les processus de fabrication et d'affaires (Tennant, 2001).

[14] Analyse en Composantes Principales (ACP) est une technique statistique qui transforme linéairement un ensemble original de variables dans un ensemble plus restreint de variables non corrélées (Jolliffe, 2002). Son objectif est de réduire la dimension de l'ensemble de données d'origine (Sauro et Kindlund, 2005).

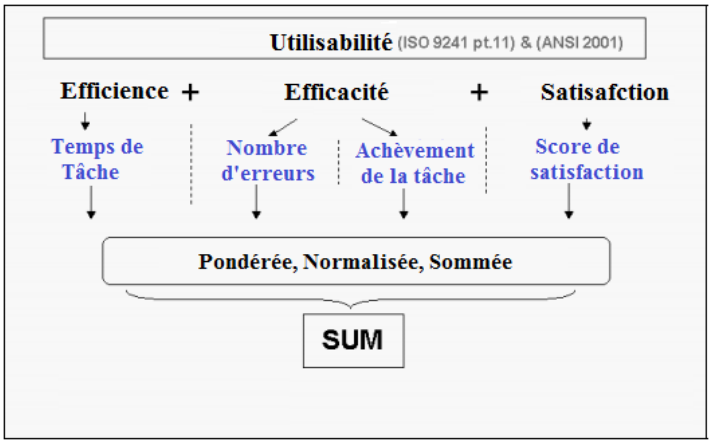


Figure 2. 4 : Modèle de la méthode d'évaluation d'utilisabilité SUM (Sauro et Kindlund, 2005)

2.2.1.3 Approche d'intégration de méthodes d'évaluation d'utilisabilité en utilisant des profils de problèmes d'utilisabilité

Al-Wabil et Al Khalifa (Al-Wabil and Al Khalifa, 2009) ont élaboré une approche permettant d'assurer l'évaluation d'utilisabilité des applications Web en combinant des différentes méthodes d'évaluation afin de détecter les problèmes d'utilisabilité. La combinaison est assurée en faisant correspondre les capacités et les limites des méthodes avec une classification de problèmes d'utilisabilité proposés par Chattratichart et Lindgaard (2008). Cette classification concerne sept segments de profils de problèmes mentionnés dans la Figure 2.5. Elle est définie comme étant un modèle de classification permettant de guider à la sélection de méthodes dans des projets d'évaluation d'utilisabilité. Par exemple, les problèmes liés à la visibilité des éléments de l'interface sont classés comme des problèmes liés aux graphiques et aux objets d'interface utilisateur.

Dans cette approche, les problèmes d'utilisabilité ont été assignés à cinq méthodes et outils spécifiques d'utilisabilité. La première concerne la méthode d'inspections portant sur les heuristiques d'évaluation en se basant sur celles proposées par Nielsen (1993) pour évaluer la conformité d'un portail web avec les normes d'utilisabilité et les guides de conception. La deuxième considère les tests d'utilisabilité avec les utilisateurs finaux pour mesurer l'utilisabilité d'une façon subjective via le questionnaire SUS (Brooke, 1996). En effet, ces deux méthodes ont été appliquées afin d'identifier les problèmes d'utilisabilité selon la classification adoptée. Quant aux trois autres méthodes, elles ont été utilisées pour compléter les deux méthodes mentionnées dans le but de confirmer les problèmes d'utilisabilité détectés. L'outil « Eye Tracking » permet de détecter les problèmes de visibilité d'éléments d'interfaces utilisateur. La « Card Storing » a porté sur la détection des problèmes liés à la désorientation (Al-Wabil and Al Khalifa, 2009). Ceux-ci sont liés aux utilisateurs qui ne sont pas capables de comprendre la structure d'une page web à évaluer, ou ils ont confondu avec l'étiquetage des sections de page et l'organisation de l'information au sein de la page. La dernière méthode concerne le « *Focus Groups* » permettant d'évaluer la satisfaction des utilisateurs (Al-Wabil and Al Khalifa, 2009).

Cette approche a été appliquée pour évaluer un portail web afin de montrer leur applicabilité et utilisation. A la fin de l'évaluation, elle se contente de donner un aperçu des analyses et des rapports des résultats d'utilisabilité des différentes méthodes d'évaluation. Toutefois, vu que les résultats fournis par chaque méthode d'évaluation sont présentés séparément, cette approche n'introduit aucun aspect d'intégration effective. Elle se focalise plutôt sur la façon avec laquelle les méthodes d'évaluation de l'utilisabilité peuvent se compléter mutuellement.

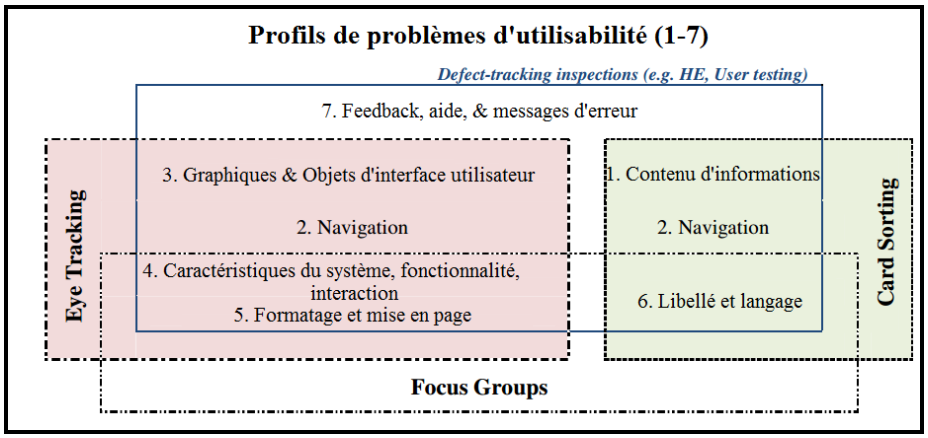


Figure 2. 5 : Approche d'intégration de méthodes d'évaluation d'utilisabilité en utilisant des profils de problèmes d'utilisabilité (Al-Wabil et Al Khalifa, 2009)

2.2.1.4 Evaluation d'utilisabilité des systèmes logiciels en utilisant des mesures subjectives et objectives

Cette méthode est proposée par (Kerzazi et Lavallée, 2011) dans le but d'évaluer l'utilisabilité des systèmes interactifs de type WIMP. Elle s'appuie sur l'utilisation d'un ensemble des mesures subjectives et objectives. Cette méthode est définie lors d'une étude expérimentale qui porte sur la comparaison de l'utilisabilité de deux systèmes logiciels selon les trois critères définis par la norme ISO 9241-11 (l'efficacité, l'efficience et la satisfaction). Ces critères ont été d'abord évalués à travers un questionnaire qui adopte une échelle de Likert et d'autre part en utilisant des mesures objectives. L'efficacité a été mesurée d'une façon objective en validant que les critères de modélisation requises ont été remplis. L'efficience a été mesurée par le temps requis pour réaliser complétement la tâche avec chaque système. La satisfaction a été mesurée d'une façon objective par le nombre de problèmes signalés.

Lors de l'évaluation, cette méthode comporte deux phases. La première porte sur la comparaison de ces systèmes selon les trois critères en utilisant des mesures subjectives basées sur des questions d'utilisabilité existantes. La deuxième phase porte sur l'utilisation de mesures objectives relatives aux trois critères afin d'assurer la comparaison entre les deux systèmes. A la fin de l'évaluation, les résultats de comparaison de deux systèmes sont rapportés. Ils comportent les résultats des mesures subjectives et objectives correspondant aux deux systèmes. Toutefois, aucune information concernant les méthodes ou les outils d'évaluation adoptés n'est indiquée.

Par conséquent, les résultats de l'étude comparative ont conclu que les évaluations subjectives et objectives peuvent se compléter mutuellement. De plus, cette approche a recommandé la combinaison de données subjectives et objectives pour fournir une vue plus complète d'utilisabilité.

2.2.1.5 Approche d'aide à la prise de décision par la combinaison des méthodes d'évaluation subjectives et objectives

Il s'agit d'une approche d'évaluation pour l'aide à la prise de décision sur l'utilisabilité et l'utilité des systèmes interactifs de type WIMP (Hentati et al., 2013). Elle vise à définir des scores uniques d'utilisabilité et d'utilité à travers des données recueillies par différentes méthodes et outils d'évaluation en exploitant les concepts de *logique floue*. Elle s'appuie sur l'utilisation de questionnaire, du mouchard électronique et la technique d'inspection de la qualité ergonomique. Cette approche est structurée autour de cinq modules. Le premier concerne la capture de données d'évaluation. Le second module porte sur l'association des données à un ensemble des critères de qualité portant sur les deux aspects de l'utilisabilité et l'utilité. Le troisième module concerne la quantification de ces deux aspects. Il exploite les concepts de logique floue qui dépendent d'un

ensemble des paramètres à fixer par l'évaluateur (les bases de règles, la fuzzification et la diffuzzification). Le module quatre se base sur le calcul du degré de confiance pour chaque méthode adoptée. En outre, des calculs statistiques, concernant les données des utilisateurs, des tâches, et les poids de pondération qui correspondent à chaque critère, seront déterminés par le module quatre. Les résultats obtenus pour chaque critère provenant des différentes méthodes sont agrégés entre eux. Par la suite, ils sont présentés selon un format approprié basé sur Matlab via le module cinq. Pour chacun des aspects à évaluer, un ensemble des formules sont proposées permettant de déterminer les correspondances entre les critères relatifs à l'aspect de qualité et les méthodes d'évaluation. Chaque formule est déterminée sur la base de la moyenne des scores de données agrégées. A son tour, chaque score prend en considération les sous-critères et les données agrégées ensemble en se basant sur leurs poids relatifs.
A la fin de l'évaluation, les résultats obtenus de chaque critère sont projetés en appliquant les règles de fuzzification adoptées afin de déterminer la signification de chaque aspect qui peut être faible, moyenne, bonne, très bonne. Néanmoins, aucune interprétation des résultats n'est fournie.
Par ailleurs, l'utilisation de cette approche s'avère assez critique. Son principal inconvénient réside dans le fait qu'elle se base sur l'utilisation des poids pondérés pour les critères et les méthodes utilisées. Généralement, la sélection de ces poids reste un inconvénient majeur portant sur les difficultés trouvées pour déterminer l'importance relative entre les critères. De plus, l'agrégation des données provenant d'évaluations subjective et objective en un unique score est non bénéfique pour l'identification des problèmes d'utilisabilité, le fait que celui-ci peut ignorer certains résultats qui peuvent affecter l'exactitude des conclusions fournies.

2.2.1.6 Framework RITA pour l'évaluation des interfaces utilisateur

RITA l'acronyme de « useR Interface evaluaTion frAmework » est une approche générique dédiée à l'évaluation de l'utilisabilité des systèmes interactifs de différents types des interfaces utilisateur (WIMP, Web, Mobile) (Charfi et al., 2015). Elle a été proposée lors d'un travail antérieur et préliminaire établi dans le même cadre de développement de l'environnement intégré d'aide à l'évaluation des systèmes interactifs. Cette approche est constituée des applications logicielles structurées selon une architecture modulaire permettant de garantir une utilisation flexible et configurable. En outre, elle s'appuie sur trois outils d'évaluation. Le premier est le mouchard électronique (Tran, 2009). Il permet de recueillir des informations sur les actions des utilisateurs avec le système à évaluer en temps réel pour les analyser et aider l'évaluateur à identifier les problèmes de l'utilisation de l'interface utilisateur. Le second est un inspecteur de la qualité ergonomique. Il porte sur l'évaluation ergonomique de la conformité des interfaces utilisateur avec un ensemble des règles ergonomiques (spécifiées au préalable par l'évaluateur) afin de détecter les incohérences ergonomiques (Charfi, 2013). Le troisième outil est un questionnaire permettant de collecter les appréciations des utilisateurs à propos des différents aspects d'utilisabilité de l'interface utilisateur.
En outre, RITA comprend quatre modules comme l'illustre la Figure 2.6. *Le premier concerne le gestionnaire de règles ergonomiques*. Il consiste à définir et identifier les règles ergonomiques à considérer lors de l'évaluation d'un système interactif. Une règle se caractérise par un ensemble des attributs (identifiant, libellé, description, erreur engendrée lors du non-respect de la règle, recommandation, technique d'évaluation, critère (sous-critère) ergonomique, contexte, corpus et référence de la règle). Les règles ergonomiques sont spécifiées en associant des valeurs de l'attribut d'un (des) composant(s) de l'interface à des valeurs spécifiées par l'évaluateur. Celles-ci vont permettre la spécification d'un modèle de référence sous forme d'un fichier XML.
Le deuxième module consiste à capturer l'ensemble des données d'évaluation à travers trois outils : un outil de génération de questionnaire, un parseur de données et un mouchard électronique. Les données de questionnaire sont capturées via un questionnaire d'utilisabilité spécifié par l'évaluateur pendant la

phase d'évaluation. Quant au parseur de données, il permet de dégager les valeurs des attributs des composants graphiques qui sont sauvegardées sous forme de fichier XML. Le mouchard électronique (Tran, 2009) permet, à son tour, de capturer l'ensemble des évènements relatifs aux interactions des utilisateurs avec le système interactif à évaluer. Ceux-ci sont sauvegardés dans une base de données de type MS Access.

Le troisième module permet d'analyser les différentes données provenant du mouchard, du questionnaire et la description des interfaces utilisateur. L'évaluation ergonomique est effectuée via un inspecteur des règles ergonomiques, son objectif est de vérifier la conformité de la valeur (ou des valeurs) des attributs de composants graphiques (spécifiés dans des fichiers XML) par rapport aux valeurs spécifiées par l'évaluateur (le modèle de référence). Les comparaisons entre ces deux entrées permettent de générer en sortie les incohérences ergonomiques qui peuvent être sauvegardées sous forme Pdf et Xml.

Le quatrième module consiste à générer le rapport final des résultats de l'évaluation basée sur les trois outils. Celui-ci peut être présenté selon trois formats : txt, Pdf et Html. Ce rapport comprend les problèmes d'utilisabilité détectés et inspectés par les différents outils. Il fournit aussi quelques recommandations concernant la qualité ergonomique pour l'amélioration de l'interface utilisateur.

Cette approche a permis de synthétiser l'ensemble des résultats capturés de différents outils dans un seul rapport d'évaluation. Toutefois, nous avons noté des difficultés qui s'imposent aux évaluateurs lors du traitement des différents résultats obtenus afin de corriger les problèmes d'utilisabilité. Etant donné de l'hétérogénéité des résultats et les différents aspects d'évaluation traités, l'évaluateur se trouve obligé d'interpréter une grande quantité des résultats qui peuvent être parfois contradictoires. Par exemple, lors de l'évaluation du critère de lisibilité, l'inspecteur indique la conformité des éléments de l'interface aux règles ergonomiques, cependant les résultats du questionnaire indique l'insatisfaction des utilisateurs. De ce fait, nous attestons que la combinaison des évaluations subjective et objective avec une présentation séparée de leurs résultats est insuffisante. Par conséquent, l'intégration de résultats d'évaluation subjective et objective d'une façon cohérente et complémentaire sera sans doute plus bénéfique dans le processus d'évaluation.

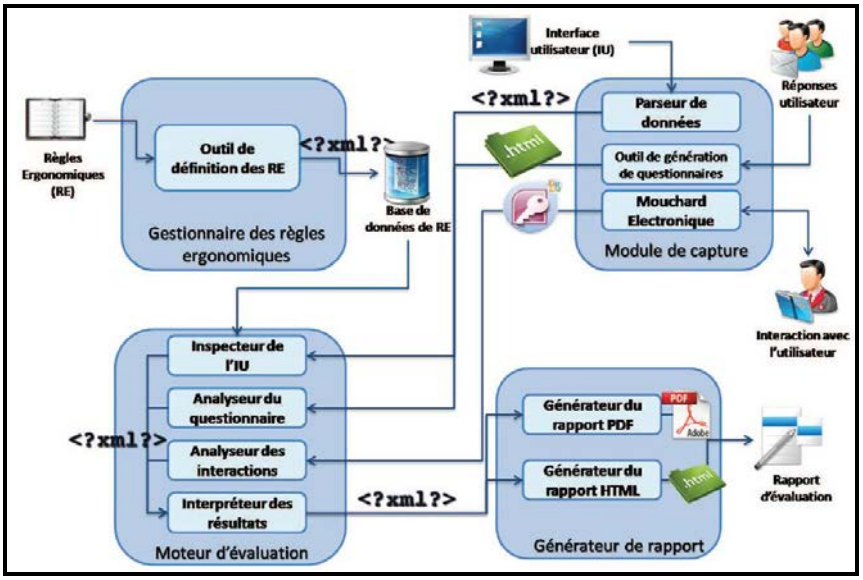


Figure 2. 6 : Architecture de Framework RITA (Charfi et al., 2015)

2.2.2 Récapitulatif et Discussion

Après avoir présenté les approches étudiées visant à assurer des combinaisons des évaluations subjective et objective. Nous présentons dans ce qui suit, un tableau récapitulatif (Tableau 2.6) représentant une synthèse des approches en se référant à la majorité des aspects énoncés.

Tableau 2. 6 : Tableau récapitulatif des approches présentées

Approches étudiées	Interface utilisateur visée	Aspects et/ou critères de qualité à évaluer	Nature des évaluations combinées	Principe de combinaison	Problèmes détectés	Résultats fournis
WebUSE (Nikov et al., 2003)	Web	Différents aspects tels que : background d'utilisateur, qualité d'information, la lisibilité des informations, etc.	Intègre des données d'évaluation subjective et objective d'utilisabilité	Agrégation des données subjectives et objectives dans un indice unique d'utilisabilité basé sur un modèle hybride qui applique un algorithme spécifique.	Problèmes généraux sur les interfaces utilisateur	sous forme textuelle, Excel et des graphiques à barre
SUM (Sauro et Kindlund, 2005)	Logiciels	Efficacité, efficience et satisfaction (ISO 9241-11, 1998)	Intègre des mesures d'utilisabilité standardisées	Agrégation des mesures en une seule mesure sommative et standardisée en utilisant la technique de l'Analyse en Composantes Principales.	Ne précise pas les problèmes d'utilisabilité	sous forme textuelle
Approche de (Al-Wabil et Al Khalifa, 2009)	Web	Applique les Heuristiques de Nielsen	Combine des méthodes subjective et objective d'utilisabilité	Combinaison des méthodes en faisant correspondre leurs capacités et limites avec une classification de problèmes d'utilisabilité. Pas d'intégration effective entre les données.	Problèmes d'utilisabilité proposés par Chattratichart et Lindgaard (Chattratichart et Lindgaard, 2008)	sous forme textuelle et des graphiques à barre
Approche de (Kerzazi et Lavalée, 2011)	WIMP	Efficacité, efficience et satisfaction (ISO 9241-11, 1998)	Combine des mesures subjective et objective	Pas d'intégration effective entre les mesures. La combinaison est assurée via les critères à évaluer d'une façon indirecte.	Problèmes d'utilisabilité portant sur les trois critères à évaluer	sous forme textuelle et des graphiques à barre
Approche de (Hentati et al., 2013)	WIMP	Utilisabilité : les huit critères ergonomiques de Vanderdonckt (1994) Utilité : le taux d'erreur, la facilité d'exécution	Intègre différentes données d'évaluation extraites des outils d'évaluation subjective et objective	L'agrégation des données est assurée via les concepts de la logique floue en utilisant la moyenne entre les scores.	Aucune information n'est fournie sur les problèmes détectés. Aucune interprétation des scores n'est fournie.	sous forme graphique selon un format spécifique basé sur Matlab
RITA (Charfi et al., 2015)	WIMP, Web et Mobile	Critères et sous-critères de Bastien et Scapin (1993)	Combine des outils d'évaluation subjective et objective sous forme des modules logiciels	Les outils sont combinés à travers une architecture modulaire.	Ils correspondent aux problèmes extraits de chaque outil, tels que les problèmes ergonomiques des interfaces utilisateur, les problèmes lors de l'exécution des tâches des utilisateurs.	sous forme des fichiers : PDF, html, txt

A travers les approches présentées et le Tableau récapitulatif 2.6, nous pouvons déduire ce qui suit :

- Toutes ces approches s'appuient sur les évaluations subjective et objective. La plupart d'entre elles ont adopté les questionnaires pour assurer l'évaluation subjective. Néanmoins, nous notons une diversité au niveau des méthodes et/ou outils utilisés qui permettent d'assurer l'évaluation objective. Certains visent à capturer des données portant sur les composants graphiques des interfaces utilisateur, d'autres se focalisent sur les données provenant d'interactions avec le système interactif à évaluer (comme par exemple, le Mouchard électronique).
- Au niveau des interfaces visées, certains travaux sont dédiés à l'évaluation des applications WIMP (Kerzazi et Lavalée, 2011 ; Hentati et al., 2013). D'autres concernent les applications Web (Nikov et al., 2003 ; Al-Wabil et Al khalifa, 2009). Tandis que l'approche la plus récente (Charfi et al., 2015) a été adressée à l'évaluation des interfaces WIMP, web et Mobiles et celle proposée par Sauro et Kindlund (2005) a été destinée à l'évaluation des logiciels d'une façon générale.
- Différents critères d'utilisabilité ont été considérés. A l'exception de l'approche WebUSE, les auteurs des autres travaux se basent sur des critères connus définis par les standards et très référenciés dans la littérature, nous citons par exemples les critères ISO 9241-11 (1998) (considérés par l'approche de Kerzazi et Lavalée et SUM), les heuristiques de Nielsen optés par l'approche d'Al-Wabil et Al Khalifa et les critères ergonomiques de Bastien et Scapin (1993) considérés par l'approche RITA. Cela montre l'importance de se référer aux standards pour assurer l'évaluation de l'utilisabilité des interfaces utilisateur.
- Concernant la nature des évaluations combinées, nous distinguons trois catégories. La première adoptée par la plupart des approches (WebUSE, SUM, l'approche de Kerzazi et Lavalée et l'approche de Hentati et al.) concerne l'intégration des données, mesures ou résultats subjective et objective. La seconde catégorie concerne la combinaison des différentes méthodes d'évaluation subjective et objective adoptée par l'approche d'Al-Wabil et Al Khalifa. La troisième catégorie porte sur la combinaison des différents outils dans un seul.
- A propos des principes utilisés dans les combinaisons dans ces travaux, nous distinguons deux catégories. La première concerne les approches SUM, WebUSE et l'approche Hentati et al. portant sur l'intégration effective des données, mesures ou résultats de différentes évaluations. Ces approches se basent sur l'agrégation en un seul score des données en utilisant différentes techniques telles que la logique floue, l'analyse en composantes principales. Bien que ces approches assurent une intégration effective des évaluations subjective et objective, elles ne précisent pas les problèmes d'utilisabilité qui peuvent être corrigés. En outre, à l'exception de WebUSE, ces approches ne fournissent aucune aide pour l'interprétation de leurs scores globaux. Par ailleurs, beaucoup d'études devraient être effectuées pour soutenir l'interprétation du score général.

  La deuxième catégorie concerne les approches proposées par Al-Wabil et Al Khalifa (2009), Kerzazi et Lavalée (2011) et Charfi et al. (2015). Elles sont destinées en général à déterminer comment différentes évaluations peuvent être complémentaires. Ces approches ont comporté une diversité à propos de la nature des évaluations combinées (méthode, outils, ou mesures). Elles ont également recommandé la combinaison des évaluations subjective et objective à la fois afin d'améliorer les résultats d'évaluation. Différentes techniques de combinaison significatives ont été proposées à travers ces travaux, telles que la mesure des critères d'utilisabilité de deux façons : subjective et objective. Bien que ces approches assurent la spécification des problèmes d'utilisabilité, aucune intégration effective des évaluations subjective et objective n'a été fournie ni des interprétations des résultats.

2.3 Synthèse et conclusion

Dans ce chapitre nous avons présenté une revue de la littérature sur les outils d'évaluation d'utilisabilité des systèmes interactifs. Nous avons vu qu'il existe nombreux outils d'évaluation des interfaces utilisateur assurant des évaluations subjective et/ou objective.

Pour l'évaluation subjective, nous nous sommes focalisés sur l'utilisation des outils des questionnaires d'utilisabilité standardisés. L'utilisation de ces questionnaires a été largement recommandée dans la littérature (Lewis, 2014 ; Hornbæk et Law de 2007 ; Lewis et Hardzinski, 2015) vue leurs validations et fiabilités offertes par la normalisation.

En outre, l'usage des questionnaires peut dépendre des aspects de qualité adoptés. La domination des aspects généraux dans la majorité des questionnaires largement utilisés dans la littérature (tels que SUMI, SUS, CSUQ) nous a conduit à élaborer une étude d'analyse sur tous les questionnaires étudiés en se référant aux critères d'utilisabilité proposés par les standards. Cela nous a fourni des informations supplémentaires sur quelques critères implicitement considérés par les questionnaires standardisés existants.

Concernant l'évaluation objective, des nombreux outils ont été présentés. Dans un premier temps, nous nous sommes focalisés sur ceux utilisant les règles et les connaissances ergonomiques. Dans un second temps, nous avons abordé ceux qui se basent sur la capture des données d'interaction entre l'utilisateur et le système interactif à évaluer.

Tous ces outils présentés visent en général à atteindre l'objectif de détection des problèmes d'utilisabilité en vue d'améliorer la qualité des interfaces utilisateur des systèmes interactifs. Par ailleurs, nous avons distingué que chacun d'entre eux a son propre principe et possède un ensemble des caractéristiques qui le différencie par rapport aux autres.

Ce chapitre a révélé aussi l'intérêt des approches qui s'appuient sur la combinaison des évaluations objective et subjective. Nous avons étudié et analysé chaque approche. Une synthèse sur ces approches a été présentée en se référant à l'ensemble des critères agissant sur l'objectif de la combinaison. A partir de cette synthèse, nous pouvons déduire ce qui suit :

- Tous ces travaux ont assuré la combinaison des évaluations subjective et objective.
- Les travaux qui ont assuré l'intégration effective des évaluations subjective et objective ont été basés sur l'agrégation des données (mesures, résultats) en un score unique. Bien que cette façon s'avère importante, l'agrégation en un score unique généralement ne permet pas la détection et l'identification avec précision des problèmes d'utilisabilité (Sauro, 2005). De plus, elle peut ignorer certains aspects (subjectifs et / ou objectifs) qui peuvent être pertinents lors de l'exécution de l'évaluation.
- La majorité de ces travaux ne fournissent aucun support pour assurer l'interprétation de leurs résultats, ni des suggestions de recommandations pour corriger les problèmes détectés d'utilisabilité.

Dans ce cadre, nous allons proposer dans le chapitre suivant une nouvelle contribution d'évaluation de l'utilisabilité des systèmes interactifs. Cette contribution sera introduite dans *l'Environnement intégré d'aide à l'évaluation des systèmes interactifs* adopté dans le cadre de notre thèse. Elle vise principalement à intégrer effectivement les résultats provenant des différents outils d'évaluations subjective et objective adoptés dans l'environnement d'une façon *cohérente et complémentaire,* en vue de détecter les problèmes d'utilisabilité des interfaces utilisateur et avoir une bonne mesure d'utilisabilité.

Chapitre 3 : Approche d'intégration des résultats d'évaluation d'utilisabilité subjective et objective

"Usability evaluation methods uncover different types of usability problems; therefore, it is often recommended for evaluators to use multiple assessment methods"
(JAKOB NIELSEN, 1993)

Introduction

Après avoir présenté dans le chapitre précédent une revue de la littérature sur les outils et les approches d'évaluation d'utilisabilité des interfaces utilisateur, nous pouvons affirmer que l'intégration des méthodes, outils ou données d'évaluation est une solution susceptible d'aboutir à des meilleurs résultats. En effet, la considération des évaluations subjective et objective à la fois s'avère importante pour avoir une évaluation complète. En tenant compte que les approches étudiées n'ont pas contribué à l'intégration effective de ces évaluations d'une façon cohérente et complémentaire dans le but de détecter des problèmes d'utilisabilité, nous visons dans ce chapitre à répondre à cet intérêt. Pour atteindre cet objectif, nous avons proposé une approche d'évaluation subjective et objective en spécifiant des mesures sur la base de la norme ISO/IEC 15939.

Dans la première section, nous commencerons par introduire la motivation et la présentation globale de l'approche. Dans la deuxième section, nous décrirons d'une manière détaillée les étapes à suivre dans l'approche. Celles-ci permettent la spécification d'un indicateur d'utilisabilité en appliquant le « *Modèle de mesure d'information* » proposé par la norme ISO/IEC 15939. Par la suite, nous présenterons dans la troisième section l'ensemble des modèles de mesure proposés. Et finalement, nous achèverons le chapitre par une conclusion.

3.1 Motivation et présentation générale de l'approche

3.1.1 Motivation

Nous rappelons que notre objectif principal est inclu dans le cadre de l'accomplissement et la mise en œuvre de l'environnement intégré d'aide à l'évaluation des systèmes interactifs de supervision du trafic proposé par Ezzedine (Ezzedine, 2002). Notre motivation majeure consiste à proposer une approche pour l'intégration effective des résultats hétérogènes d'évaluation subjective et objective provenant des trois outils d'évaluation, comme le montre la Figure 3.1.

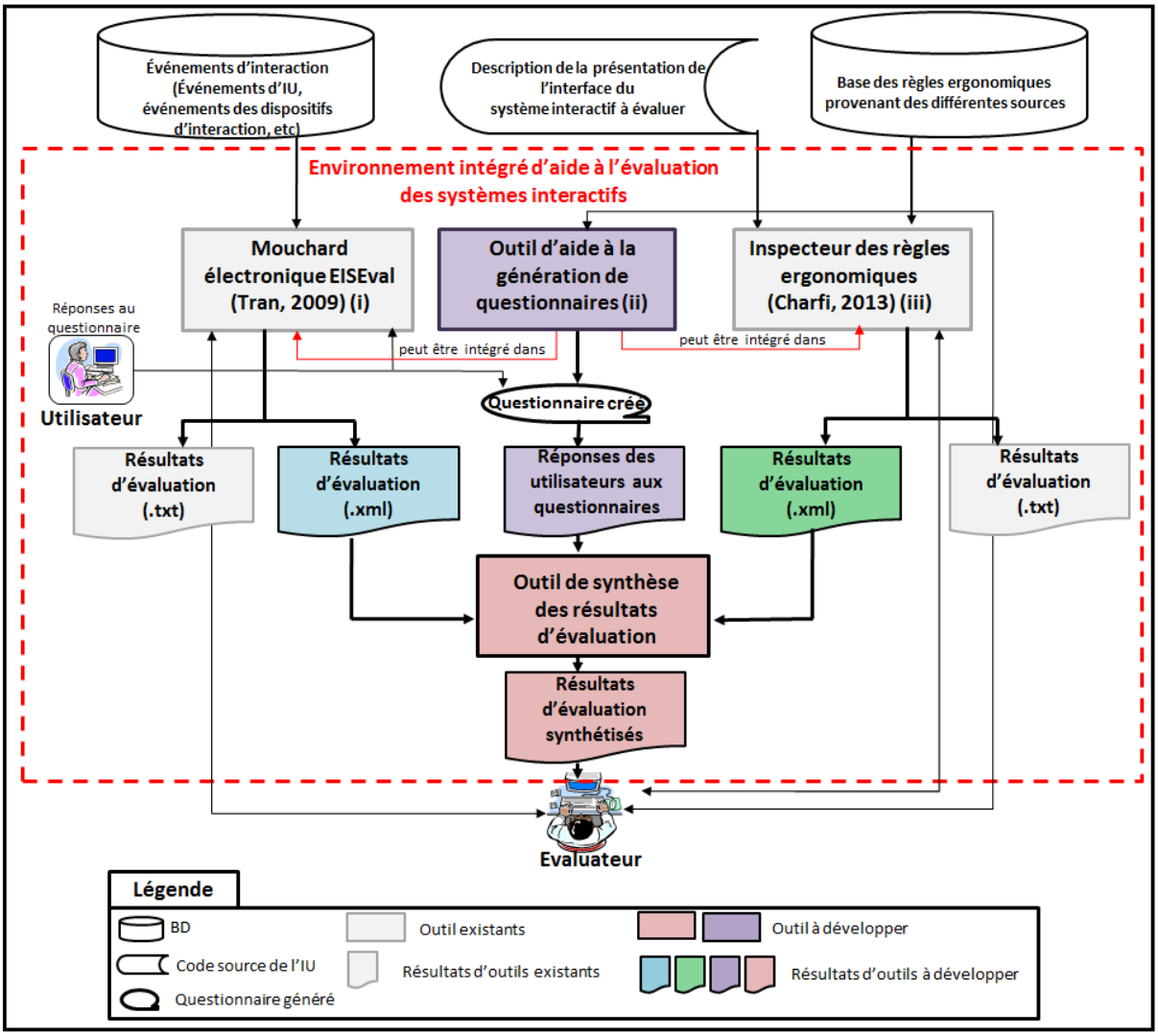


Figure 3. 1 : Environnement intégré d'aide à l'évaluation des systèmes interactifs (Ezzedine, 2002)

Les deux premiers outils (i) et (iii) ont été développés dans des recherches précédentes :

- *Un mouchard électronique* (nommé ***E**nvironment for **I**nteractive **S**ystem **Eval**uation*) (Tran et al., 2013) qui a été proposé par Tran pendant sa thèse (Tran, 2009). Il assure une évaluation

objective portant sur la capture et l'analyse des actions des utilisateurs et leurs interactions avec le système interactif à évaluer en temps réel. Particulièrement, cet outil est dédié à l'évaluation des systèmes interactifs à base d'agents appliqués dans la supervision du trafic. En outre, il porte sur l'analyse de l'activité des utilisateurs utilisant des interfaces et tous les événements générés par les dispositifs d'interaction. Par conséquent, un ensemble des données peut être déterminé (voir Figure 3.2 (a)).

- *Un inspecteur des règles ergonomiques* (Charfi et al., 2014) qui a été proposé et réalisé par Charfi pendant sa thèse (Charfi, 2013). Cet outil constitue un sous-module de la Framework RITA (Cf. Chapitre 2, Section 2.2.1.6, Figure 2.6). Il assure une évaluation objective sur la qualité ergonomique des interfaces utilisateur des systèmes interactifs. Il se base sur un ensemble des règles ergonomiques afin de détecter les incohérences ergonomiques dans les interfaces utilisateur portant sur certains critères d'utilisabilité (tels que la lisibilité, la densité d'information, etc.). En outre, cet outil permet également de gérer et générer des recommandations ergonomiques. Par conséquent, il capture différentes données concernant les attributs des composants graphiques des interfaces utilisateur (voir Figure 3.2 (b)).

Le troisième outil assure une évaluation subjective basée sur la capture des appréciations des utilisateurs vis-à-vis du système interactif à évaluer. Il s'agit de l'*outil d'aide à la génération de questionnaires* (Figure 3.1) développé au cours de cette thèse. Une première version a été proposée dans (Assila et al., 2013a, 2013b) et une version plus étendue est incluse dans l'outil de synthèse des résultats qui sera présentée dans le Chapitre 4. Son but principal est d'acquérir les opinions des utilisateurs concernant des critères spécifiques d'utilisabilité. Par conséquent, un ensemble des données peut être généré (voir Figure 3.2 (c)).

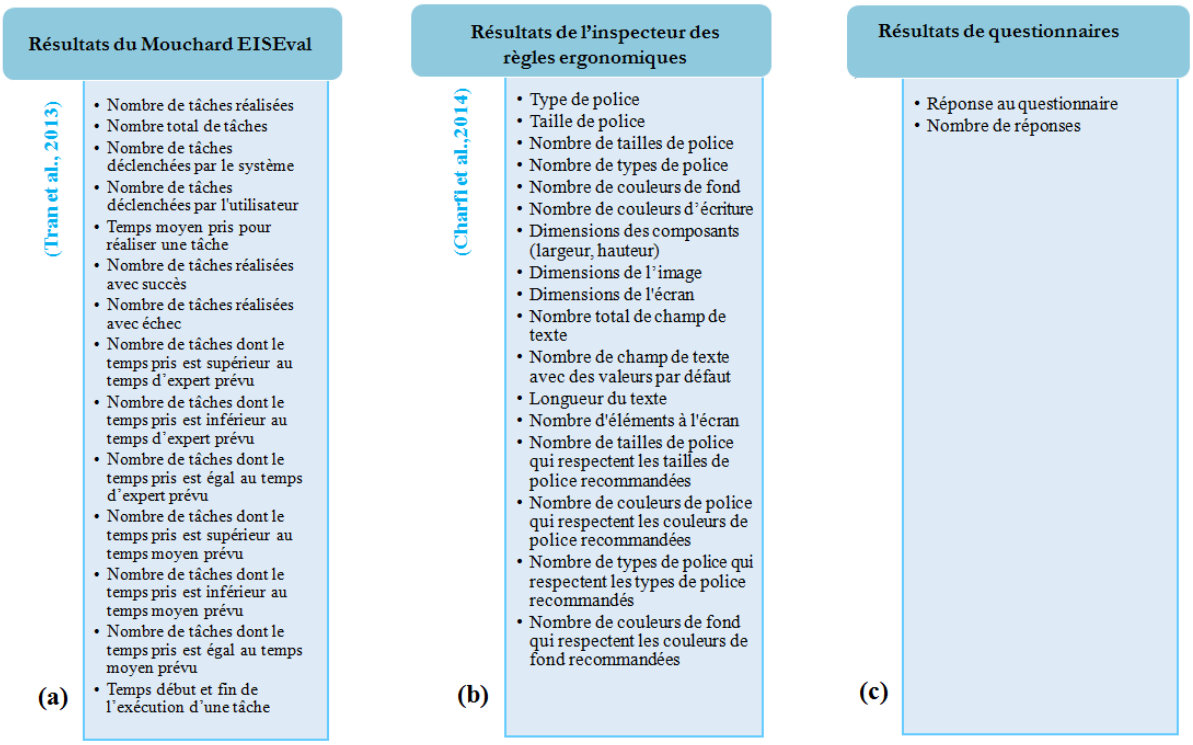


Figure 3. 2 : Les données capturées des différents outils de l'environnement intégré d'aide à l'évaluation des systèmes interactifs

Au sein de cet environnement, une évaluation est effectuée en suivant trois phases principales. La première porte sur l'évaluation statique du système à évaluer par l'évaluateur en utilisant l'inspecteur des règles ergonomiques. Dans la deuxième phase, un scénario d'exécution des tâches est planifié par

l'évaluateur. L'utilisateur doit donc exécuter ce scénario. Celui-ci spécifie un problème ou une situation où l'utilisateur doit utiliser le système. Il peut comprendre une ou plusieurs tâches qui correspondent au fonctionnement du système à évaluer. Une fois que l'utilisateur commence l'utilisation du système, le mouchard électronique EISEval est également lancé. Dans la troisième phase, l'utilisateur répond à un questionnaire d'utilisabilité préparé par l'évaluateur. À la fin de la session d'évaluation, un ensemble de données extraites à partir des trois outils est généré.

Par ailleurs, les travaux antérieurs réalisés sur cet environnement concernant le développement d'outils mentionnés ont permis à chaque outil de générer ses propres résultats d'évaluation d'une façon séparée pour être analyser et interpréter par l'évaluateur (voir Figure 3.2).

En général, la compréhension et l'interprétation d'un ensemble hétérogène de données présentent des tâches complexes. De ce fait, notre objectif principal est donc d'intégrer ces données hétérogènes d'une manière uniforme et cohérente afin de mieux aider l'évaluateur à la prise de décision lors de l'évaluation d'un système interactif donné. Par conséquent, nous avons proposé une approche permettant d'intégrer les différents résultats. Celle-ci sera mise en œuvre dans l'environnement via l'outil de synthèse des résultats d'évaluation (*cf. Chapitre 4*) comme l'illustre la Figure 3.1.

Dans la sous-section suivante, nous introduisons une présentation générale de l'approche proposée.

3.1.2 Présentation générale de l'approche proposée

Pour permettre l'intégration de l'ensemble des données d'évaluation d'utilisabilité extraites de trois outils dans notre environnement, nous proposons d'utiliser « *le modèle de mesure d'information* (Measurement Information Model) » défini par la norme ISO / IEC 15939 (voir *Chapitre 1, Section 1.1.3*) pour les raisons suivantes :

a) C'est une norme internationale applicable et adoptée dans l'industrie (Staron et al., 2009) ;

b) Elle fournit un processus complet de construction de mesure (le Modèle de mesure d'information) allant depuis la spécification de ses attributs jusqu'à l'établissement des indicateurs qui répondent aux exigences des intervenants et de leurs besoins d'information ; et,

c) Elle offre la possibilité de définir les procédures de collecte des mesures, l'analyse et l'interprétation des résultats en utilisant des critères de décision prédéfinis pour appuyer la prise de décision.

Pour définir les modèles de mesure, nous avons d'abord identifié comment associer les données provenant de différents outils d'évaluations subjective (les réponses des questionnaires d'utilisabilité) et objective (les données provenant du mouchard électronique EISEval et l'inspecteur des règles ergonomiques).

Étant donné que toutes ces données visent à évaluer l'utilisabilité, nous avons conclu que la meilleure façon de les intégrer est de chercher à quel(s) critère(s) d'utilisabilité chaque donnée pourrait répondre. Nous avons eu recours à trois principaux critères d'utilisabilité de la norme ISO 9241-11 (1998) (efficacité[15], efficience[16] et satisfaction[17]) et les critères ergonomiques proposés par Bastien et Scapin (1993) qui sont largement utilisés et bien structurés (voir *Chapitre 1, Section 1.5, Tableau 1.7*).

En se basant sur les critères identifiés pour chacune des données, nous avons défini, par conséquent, un Mapping entre les données (d'évaluation subjective et objective) relatives aux mêmes critères.

[15] L'efficacité est la précision et le degré d'achèvement selon lesquels l'utilisateur atteint des objectifs spécifiés (ISO 9241-11, 1998)
[16] L'efficience est le rapport entre les ressources dépensées et la précision et le degré d'achèvement selon lesquels l'utilisateur atteint des objectifs spécifiés (ISO 9241-11, 1998)
[17] La satisfaction est l'absence d'inconfort, et attitudes positives dans l'utilisation du produit (ISO 9241-11, 1998)

Dans ce qui suit, nous spécifions ce mapping et la façon dont nous l'utilisons pour soutenir la définition de nos indicateurs en utilisant la structure du modèle de mesure d'information.

3.2 La construction des indicateurs d'utilisabilité des interfaces utilisateur basés sur la norme ISO/IEC 15939

3.2.1 Mapping des données basé sur les critères d'utilisabilité

L'évaluation subjective dans l'environnement est effectuée par l'utilisation de questionnaires standardisés d'utilisabilité. En se basant sur les 24 questionnaires existants dans la littérature (mentionnés dans *le Chapitre 2, Section 2.1.1),* nous avons sélectionné cinq questionnaires (QUIS (Chin et al., 1988), PSSUQ (Lewis et al., 1990a), SUMI (Kirakowski et Corbett, 1993), CSUQ (Lewis, 1995) et USE (Lund, 1998, 2001)), dont les quatre premiers ayant les plus hauts degrés de fiabilité globale (supérieure ou égale à 0,92) et considérés comme universels selon la classification de Yang et al. (2012) (voir *Chapitre 2, Tableau 2.1*). Cela signifie qu'ils sont applicables à l'évaluation de tous les types d'interfaces utilisateur et produits logiciels.

Toutefois, ces questionnaires portent plutôt sur l'évaluation des aspects généraux de qualité dans le but de fournir des scores globaux de satisfaction sur l'utilisabilité perçue du système évalué (Hartson et Pyla, 2012 ; Sauro, 2015).

Compte tenu de notre cible d'intégration, il était essentiel d'effectuer une analyse plus précise sur les éléments de ces questionnaires par rapport à chaque critère d'utilisabilité considéré. Cette analyse vise à identifier les éléments de questionnaires portant davantage sur les critères d'utilisabilité adoptés. Il est important de noter que lors de l'extraction de ces éléments à partir des questionnaires, chacun a une fiabilité indépendante du degré global du questionnaire dans lequel il est introduit. Celui-ci a été utilisé uniquement pour l'identification des meilleurs questionnaires pour notre analyse.

De ce fait, nous avons considéré les résultats de l'étude d'analyse d'éléments de questionnaires que nous avons élaborés lors du Chapitre 2 (*Section 2.1.1.2, Tableau 2.5*).

Sur la base de l'analyse des questionnaires et l'étude des données extraites des deux autres outils d'évaluation objective (le mouchard électronique EISEval et l'inspecteur des règles ergonomiques), nous avons défini notre *Mapping* en considérant chaque critère. Ce *Mapping* consiste à associer d'une façon complémentaire et cohérente les données des évaluations subjective et objective aux mêmes critères d'utilisabilité comme l'illustre la Figure 3.3.

Nous sommes conscients que les critères d'utilisabilité sont généralement liés et que certaines données sont importantes pour plusieurs critères. Dans ce cas, nous pouvons définir plusieurs Mappings. En outre, certains critères sont constitués de sous-critères, tel est le cas du critère de Charge de travail qui comporte deux sous-critères : la densité informationnelle et la brièveté (Scapin et Bastien, 1997). Pour cette raison, nous avons considéré en plus dans notre *Mapping* les sous-critères.

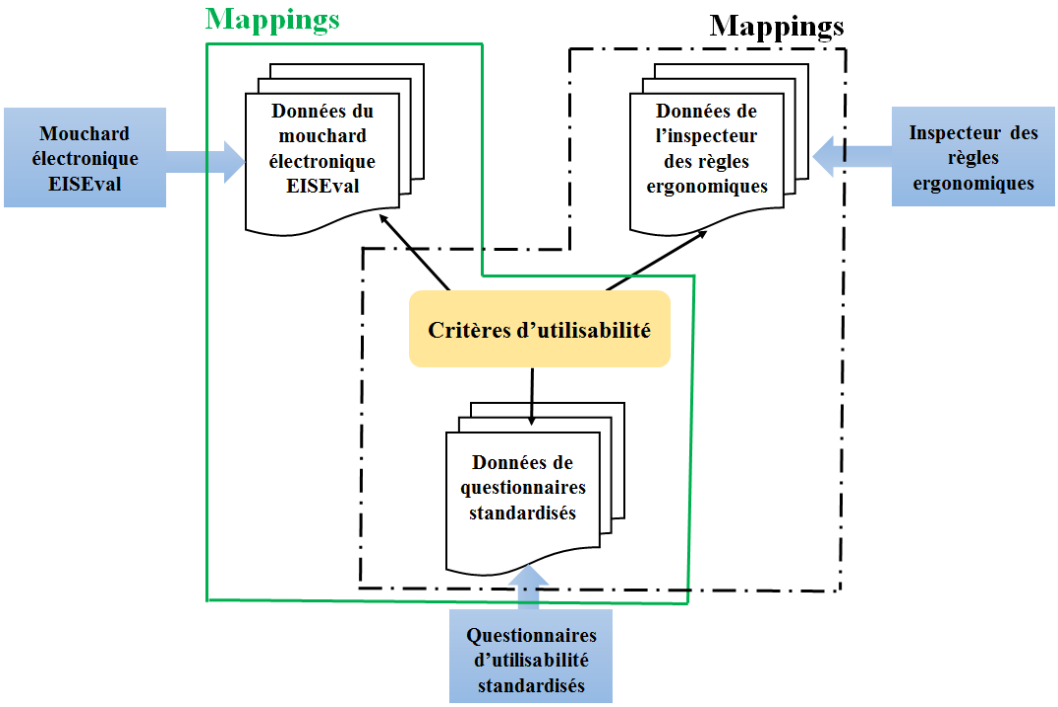


Figure 3. 3 : Le modèle proposé de *Mapping* des données

Le Tableau 3.1 présente des exemples de résultats des *Mappings* entre les éléments de questionnaires adoptés et les données de l'inspecteur des règles ergonomiques, ainsi que du mouchard électronique EISEval.

Tableau 3. 1 : Exemples de résultats des *Mappings* par rapport à quelques critères d'utilisabilité

Critère d'utilisabilité	**Données mappées**	
	Provenant de questionnaires (CSUQ ; SUMI ; PSSUQ ; CSUQ ; USE)	***Provenant de deux outils (1-EISEval, 2-inspecteur ergonomique)***
Efficacité *(ISO 9241-11, 1998)*	—J'ai complété mon travail correctement en utilisant ce système" (*I can effectively complete my work using this system*: **CSUQ***)	Nombre total de tâches **(1)** Nombre de tâches réalisées **(1)** Nombre de tâches réalisées avec succès **(1)** Nombre de tâches réalisées avec échec **(1)**
Charge de Travail : Densité informationnelle *(Scapin et Bastien, 1997)*	—Il est facile de trouver l'information dont j'avais besoin" (*It is easy to find the information I needed*: **CSUQ***)	Dimensions des composants (largeur, hauteur) **(2)** Dimensions de l'image **(2)** Dimensions de l'écran **(2)** Nombre d'éléments à l'écran **(2)**
Efficience *(ISO 9241-11, 1998)*	—J'ai été en mesure de compléter rapidement ma tâche avec ce système" (*I am able to complete my work quickly using this system*: **CSUQ***)	Le temps de réalisation d'une tâche **(1)** Nombre de tâches dont le temps pris est supérieur au temps d'expert prévu **(1)** Nombre total de tâches **(1)** Nombre de tâches réalisées **(1)** Nombre de tâches réalisées avec succès **(1)** Nombre de tâches réalisées avec échec **(1)**
Charge de Travail : Brièveté : Action Minimale *(Scapin et Bastien, 1997)*	—Il faut passer par trop d'étapes pour faire marcher quelque chose " (*There are too many steps required to get something to work*: **SUMI****)	Nombre de tâches réalisées **(1)** Nombre d'évènements /actions pour réaliser une tâche **(1)** Nombre total de tâches **(1)**

*Echelle de CSUQ : 7 points allant de «1 : fortement en accord» jusqu' à «7 : fortement en désaccord»

**Echelle de SUMI : 3 points (d'accord, ne sait pas et pas d'accord)

Tableau 3.1 (suite) : Exemples de résultats des *Mappings* par rapport à quelques critères d'utilisabilité

Critère d'utilisabilité	Données mappées	
	Provenant de questionnaires (CSUQ ; SUMI ; PSSUQ ; CSUQ ; USE)	***Provenant de deux outils (1-EISEval, 2-inspecteur ergonomique)***
Guidage : Lisibilité *(Scapin et Bastien, 1997)*	"La lecture d'informations sur l'écran" (*Reading characters on the screen*: **QUIS*****)	Nombre de tailles de police **(2)** Nombre de types de police **(2)** Nombre de couleurs de fond **(2)** Nombre de couleurs **(2)** Nombre de tailles de police qui respectent les tailles de police recommandées **(2)** Nombre de types de police qui respectent les types de police recommandés **(2)** Nombre de couleurs de fond qui respectent les couleurs de fond recommandées **(2)** Nombre de couleurs de police qui respectent les couleurs de police recommandées **(2)**
Guidage : Feedback immédiat *(Scapin et Bastien, 1997)*	"La rétroaction du système" (*System Feedback*: **QUIS*****)	Le nombre des tâches déclenchées par le système **(1)** Le nombre des tâches déclenchées par l'utilisateur **(1)**
Cohérence *(Scapin et Bastien, 1997)*	"Je ne remarque aucune incohérence lorsque je l'utilise" (*I don't notice any inconsistencies as I use it*: **USE******)	Type de police **(2)** Taille de police **(2)** Couleur de police **(2)** Couleur de fond **(2)**
Adaptabilité : Flexibilité *(Scapin et Bastien, 1997)*	"Il est flexible" (*It is flexible*: **USE******)	Bouton « Modifier » **(2)** Nombre de champ de texte avec des valeurs par défaut **(2)**

**Echelle de SUMI : 3 points (d'accord, ne sait pas et pas d'accord)
***Echelle de QUIS : 10 points numériques allant de 0 à 9.
****Echelle de USE : 7 points allant de «1 : Désaccord» jusqu'à «7 : Accord»

En résumé, ce *Mapping* a constitué une phase préliminaire également primordiale permettant d'établir une correspondance entre les données d'évaluation subjective et objective. La spécification de cette correspondance sera exploitée pour définir des indicateurs d'utilisabilité permettant d'assurer l'intégration des données (Assila et al., 2014a). Nous détaillerons les étapes à suivre lors de la création d'un indicateur d'utilisabilité dans la section suivante.

3.2.2 Les étapes de spécification d'un indicateur d'utilisabilité

Rappelons que lors du Chapitre 1 (*Section 1.1.3*), nous avons présenté les notions de base d'un ensemble des concepts portant sur la définition des mesures nommées «*indicateurs* ». Ceux-ci ont été spécifiés à travers le modèle de mesure d'information défini par la norme ISO/IEC 15939 (2007) (*cf. Chapitre1*).

Compte tenu de notre objectif d'intégration des résultats d'évaluations subjective et objective, nous proposons de définir des indicateurs d'utilisabilité en nous appuyant sur les concepts de mesures définis par la norme ISO/IEC 15939 (2007).
De ce fait, nous nous intéressons dans cette section à définir la manière avec laquelle nous pouvons spécifier ces indicateurs.

Comme l'illustre la Figure 3.4, la spécification d'un indicateur d'utilisabilité s'effectue en trois étapes principales :

- *Etape 1 :* Identification de besoin d'information.
- *Etape 2 :* Spécification des mesures de base et dérivées.
- *Etape 3 :* Spécification de l'indicateur.

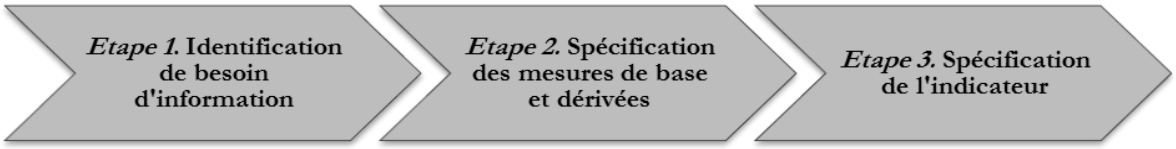


Figure 3. 4 : Les étapes de spécification d'un indicateur d'utilisabilité

Dans ce qui suit, nous décrivons en détail chaque étape.

3.2.2.1 Etape 1 : L'identification de besoin d'information

L'identification de besoin d'information consiste à indiquer l'objectif principal en utilisant un indicateur. En général, les besoins d'information doivent être définis en tenant compte des entités à évaluer. Dans notre cas, les entités présentent l'interface utilisateur à évaluer et le système interactif lui-même. Notre objectif concerne notamment l'évaluation des interfaces utilisateur d'un système interactif en fonction de chaque critère d'utilisabilité considéré (les critères ISO 9241-11 et ceux de Scapin et Bastien), (voir Tableau 3.2).

Tableau 3. 2 : Les critères d'utilisabilité de ISO 9241-11 (1998) et de (Scapin et Bastien, 1997)

	ISO 9241-11	***Scapin et Bastien***
Critère d'utilisabilité	Efficacité Efficience Satisfaction	Guidage (Incitation, Groupement/Distinction entre Items, Feedback Immédiat, Lisibilité) Charge de Travail (Brièveté, Densité Informationnelle) Contrôle Explicite (Actions Explicites et Contrôle Utilisateur) Adaptabilité (Flexibilité et Prise en Compte de l'Expérience de l'Utilisateur) Gestion des Erreurs (Protection Contre les Erreurs, Qualité des Messages d'Erreurs, Correction des Erreurs) Homogénéité/Cohérence Signifiance des Codes et Dénominations Compatibilité
Nombre des critères/ sous-critères	3	16
Nombre total	19	

Comme l'illustre le Tableau 3.2, nous disposons de 19 critères potentiels pour lesquelles nous pouvons définir au minimum 19 besoins d'information relatifs à 19 indicateurs d'utilisabilité (si on considère pour chaque critère au minimum un besoin). Afin de concevoir et valider la faisabilité de notre approche de définition des indicateurs d'utilisabilité, nous avons décidé de définir les besoins d'information d'une façon évolutive en considérant un premier groupe de critères, comme l'illustre la Figure 3.5.

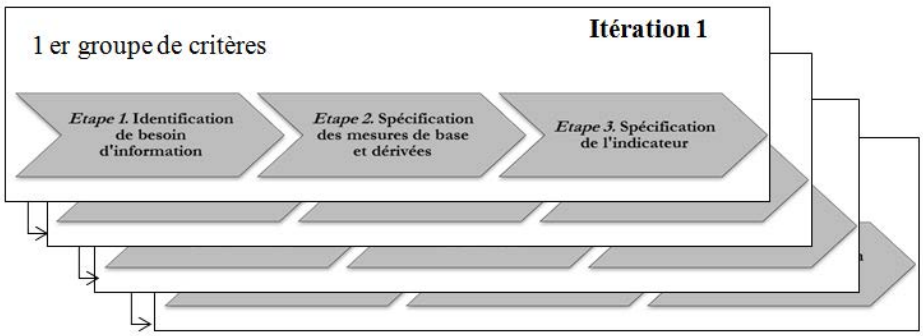


Figure 3. 5 : Définition des besoins d'information d'une façon évolutive

Pour choisir les critères inclus dans la première itération, nous nous sommes basés sur ceux qui expoitent les données existantes pouvant être extraites à travers les outils existants du mouchard électronique et l'inspecteur des règles ergonomiques (Figure 3.2). De ce fait, dans un premier lieu, nous avons considéré pour les critères ISO 9241-11 (1998) l'efficacité et l'efficience pour les raisons suivantes :

- Ces deux critères sont essentiels pour l'évaluation de l'utilisabilité du système interactif.
- La possibilité de les évaluer d'une façon subjective à travers les questionnaires d'utilisabilité standardisés et d'une façon objective indépendamment des avis des utilisateurs en utilisant le mouchard électronique EISEval.

Par contre, nous avons éliminé le critère de satisfaction puisque les données recueillies lors de l'évaluation de la satisfaction sur l'utilisabilité d'un système interactif dépendent uniquement des avis et perceptions des utilisateurs en utilisant un outil de questionnaire. Pour cela, il n'est pas possible de l'évaluer d'une façon objective en utilisant le mouchard électronique ni l'inspecteur des règles ergonomiques.

Dans un second lieu, concernant les critères ergonomiques de Scapin et Bastien, nous étions plus précis en définissant les indicateurs pour les sous-critères.

Dans ce premier groupe, nous avons considéré d'une part, la densité informationnelle et les actions minimales qui portent sur le critère de charge de travail. D'autre part, la lisibilité et le feedback immédiat qui concernent le critère de guidage. Le choix de ces sous-critères dépend de :

- La disponibilité des données de base dans les outils existants qui permettent de déterminer les mesures d'utilisabilité.
- La possibilité de mesurer ces critères, à la fois, en utilisant les données génèrées par les outils d'évaluation objective (le mouchard électronique EISEval et l'inspecteur des règles ergonomiques) ainsi que les questionnaires d'utilisabilité standardisés.
- L'existence de certaines mesures d'utilisabilité très connues et utilisées qui peuvent être exploitées ou adaptées lors de l'évaluation de ces critères.
- L'existence des règles et recommandations ergonomiques qui peuvent aider à l'interprétation de ces mesures.

Pour les autres critères ergonomiques, nous distinguons, d'une part, que pour certains (tels que la flexibilité et la prise en compte de l'expérience de l'utilisateur), l'utilisation des outils existants de l'environnement intégré est insuffisante. En effet, l'évaluation de l'adaptabilité des interfaces utilisateur mesure leurs capacités à réagir selon le contexte, et selon les besoins et préférences des utilisateurs. En considérant les données de base des outils existants, la mesure de ce critère en utilisant le mouchard électronique ou l'inspecteur des règles ergonomiques n'est pas possible. Cela requiert, en plus, l'application des méthodes d'évaluation appropriées et l'implémentation des outils supplémentaires pour mesurer la personnalisation des interfaces utilisateur, générer les données de

base ainsi que leurs mesures d'utilisabilité associées. Ces critères vont composer un deuxième groupe à traiter dans nos futurs travaux.
D'autre part , pour certains autres (tels que l'incitation, la qualité des messages d'erreurs, la correction des erreurs, la cohérence, le groupement/distinction entre Items, la concision, la signifiance des codes, la compatibilité), nous sommes conscients qu'ils dépendent de la sémantique des informations des interfaces utilisateur à évaluer. Cela exigera un effort additionnel par l'implication d'un expert en ergonomie pour prendre en compte l'aspect sémantique des informations. Cette catégorie peut constituer un troisième groupe à considérer dans nos perspectives.

Le Tableau 3.3 représente l'ensemble des besoins d'information spécifiés par rapport à chaque critère d'utilisabilité choisi.

Tableau 3. 3 : Les besoins d'information spécifiés pour les critères/ sous-critères d'utilisabilité choisis

Critères d'utilisabilité choisis	Besoin d'information
Efficacité	Évaluer l'efficacité de réalisation des tâches pendant l'exécution du système.
Efficience	Évaluer l'efficience de réalisation des tâches pendant l'exécution du système.
Densité informationnelle	Évaluer la densité globale des informations d'une interface utilisateur d'un système
Actions minimales	Évaluer la brièveté par rapport aux actions minimales lors de la réalisation d'une tâche d'un système interactif.
Lisibilité	Évaluer la lisibilité d'une interface utilisateur par rapport à la conformité des caractéristiques textuelles des informations présentées.
	Évaluer la lisibilité d'une interface utilisateur par rapport à la conformité des couleurs utilisés.
Feedback immédiat	Évaluer le feedback immédiat aux actions des utilisateurs sur les interfaces utilisateur du système.

3.2.2.2 Etape 2 : La spécification des mesures de base et dérivées

Après avoir identifié le besoin d'information relatif à un indicateur, il est important de bien spécifier les mesures (de base ou/et dérivées) pouvant être adoptées lors de la spécification de l'indicateur en prenant en compte son objectif visé. En suivant la même structure du modèle de mesure d'information défini par la norme ISO/IEC 15939, nous avons commencé d'abord par l'identification des entités et leurs attributs les plus adéquats pouvant répondre à chaque besoin d'information.

En effet, concernant les entités nous avons choisi, comme mentionné précédemment, le système interactif et l'interface utilisateur à évaluer. Pour le système interactif, nous nous sommes intéressés à toutes les informations, relatives à l'interaction entre l'utilisateur et le système, portant sur les tâches exécutées. Ces informations peuvent être capturées à travers le mouchard électronique EISEval (voir Tableau 3.1). Pour l'interface utilisateur finale, nous avons considéré toutes les données portant sur les composants graphiques de l'interface utilisateur (citons comme exemples, les labels, les images, les icônes, les couleurs, les polices, etc.) qui sont capturées à partir de l'inspecteur des règles ergonomiques. En plus, nous avons tenu compte également des résultats des perceptions des utilisateurs recueillis par l'outil de questionnaire.

Ensuite, nous avons décidé d'exploiter les mesures d'utilisabilité déjà définies dans la littérature tout en respectant nos besoins d'information et les données que nous avions dans chaque outil. Le Tableau 3.4 représente l'ensemble des mesures d'utilisabilité considérées d'évaluation objective portant sur chaque critère/ sous-critère choisi.

Tableau 3. 4 : Les mesures utilisées pour l'évaluation des critères/ sous-critères d'utilisabilité choisis

Critères d'utilisabilité choisis	Mesures d'utilisabilité considérées
Efficacité	Nombre de tâches achevées (*Number of tasks completed*) : (ISO/IEC 25022,2012 ; ISO/IEC 9126-4, 2004)
	Nombre total de tâches (To*tal number of tasks*) (ISO/IEC 25022,2012)
	Taux de réalisation correcte (d'efficacité) d'une tâche (*Task effectiveness*) = 1- ∑ la valeur proportionnelle de chaque composante manquante ou incorrecte d'une tâche (ISO/IEC 25022, 2012)
	Taux moyen de réalisation correcte des tâches (*Mean goal achievement*) = ∑ (Taux de réalisation correcte d'une tâche) / Nombre total des tâches (ISO/IEC 25022, 2012)
Efficience	Temps d'exécution d'une tâche (*Task time*) (ISO/IEC 25022,2012 ; Padda, 2003)
	Temps total des tâches (*Total Task time*) =∑ Temps d'exécution d'une tâche (ISO/IEC 25022,2012)
	Taux d'efficience d'une tâche (*Task efficiency*) = Taux de réalisation correcte d'une tâche / Temps d'exécution d'une tâche (ISO/IEC 25022,2012 ; Bevan et Macleod, 1994 ; ISO/IEC 9126-4,2004 ; Padda, 2003)
	Taux moyen d'efficience des tâches (*Goal achievement efficiency*) =∑ Taux d'efficience d'une tâche/ Nombre total des tâches = ∑ (Taux de réalisation correcte d'une tâche) / Temps total des tâches = Taux moyen de réalisation correcte des tâches/ Temps moyen des tâches (ISO/IEC 25022,2012)
Densité informationnelle	Densité globale (*Overall density*) = Espace utilisé sur une interface / Espace total d'une interface (Gabillon et al., 2013) = Nombre de pixels éclairés / Nombre total de pixels disponibles (Tullis, 1983 ; Vanderdonckt, 2003)
Actions minimales	Nombre des actions utilisées pour réaliser une tâche (*Number of commands used to perform a task)* (Dix et al., 2004 ; Padda, 2003)
	Nombre d'actions réellement nécessaire (*Number of actions actually needed*) (ISO/IEC 25022, 2012)
Lisibilité	Nombre de tailles de police (*Number of font sizes*) (Padda, 2003 ; Seffah et al., 2006)
	Nombre de types de police (*Number of font types*) (Padda, 2003 ; Seffah et al., 2006)
	Nombre de couleurs de police (*Number of Font color*) (Padda, 2003 ; Seffah et al., 2006)
	Nombre de couleurs de fond (Number of Background color) (Padda, 2003 ; Seffah et al., 2006)
Feedback immédiat	Nombre de fonctions de Feedback (*Number of feedback functions*) (Padda, 2003)

Pour les mesures d'évaluation subjective, nous avons considéré les deux mesures suivantes :

- La réponse à une question définie par (ISO/IEC 25022, 2012).
- Le pourcentage des utilisateurs ayant des réponses positives à la question défini par (Albert et Tullis, 2013). Cette mesure est égale au *Nombre des utilisateurs qui ont donné des réponses positives* divisé par le *Nombre total des utilisateurs*.

Comme nous ne disposons pas de toutes les mesures nécessaires qui répondent aux besoins d'information, nous avons défini quelques nouvelles mesures de base et dérivées en l'adaptant à partir des mesures existantes. Comme l'illustre le Tableau 3.5, ces mesures concernent les trois sous-critères suivants : actions minimales, lisibilité et feedback immédiat. Plus de détails sur ces mesures sont présentés dans la Section 3.3 (Tableaux 3.12, 3.13, 3.14 et 3.15).

En outre, nous avons révélé que certaines autres données devraient également être collectées à partir des interfaces utilisateur afin de répondre à tous les besoins d'information. Pour générer ces données nous avons étendu les deux outils d'inspecteur des règles ergonomiques et du mouchard électronique EISEval (*Cf. Chapitre 4, Section 4.2.et Section 4.3*).

Sur la base des mesures d'utilisabilité spécifiées correspondant aux besoins d'information, nous spécifions par la suite les indicateurs appropriés.

Tableau 3. 5 : Les mesures d'utilisabilité adaptées sur la base des mesures existantes

Sous-critères d'utilisabilité	Mesures de base	Mesures dérivées
Actions minimales	- Nombre d'utilisateurs réalisant les actions minimales requises d'une tâche	- Pourcentage des participants réalisant les actions requises d'une tâche
Lisibilité	- Nombre de types de police qui respectent les types recommandés - Nombre de tailles de police qui respectent les tailles recommandées - Nombre de couleurs d'écriture qui respectent les couleurs d'écriture recommandées - Nombre de couleurs de fond qui respectent les couleurs de fond recommandées	- Taux de conformité des types de police recommandés d'une interface - Taux de conformité des tailles de police recommandées d'une interface - Taux de conformité des couleurs d'écriture recommandées - Taux de conformité des couleurs de fond recommandées
Feedback immédiat	- Nombre total des actions d'un utilisateur lors de la réalisation des tâches - Nombre des actions d'un utilisateur ayant au moins un feedback	- Feedback aux actions d'un utilisateur

3.2.2.3 Etape 3 : La spécification d'un indicateur d'utilisabilité

La définition d'un indicateur qui répond à un besoin d'information requiert la spécification d'un modèle d'analyse. Ceci est défini sur la base d'un algorithme qui combine les mesures concernées caractérisant l'indicateur par les critères de décision associés (ISO/IEC 15939, 2007). Pour cela, la définition d'un indicateur dépend principalement de la spécification de deux composantes qui sont les suivantes : les mesures concernées par le modèle d'analyse de l'indicateur et leurs critères de décision associés.

Dans une première phase, nous nous intéressons à déterminer les mesures appropriées du modèle d'analyse qui correspond à un indicateur d'utilisabilité. Étant donné que les indicateurs à spécifier visent principalement à assurer l'intégration des résultats d'évaluation subjective et objective, l'identification des mesures appropriées à un indicateur a été basée sur les *Mappings* définis auparavant. En prenant en compte chaque besoin d'information, nous nous focalisons sur les Mappings des critères d'utilisabilité concernés. Par conséquent, nous avons sélectionné les mesures d'utilisabilité qui correspondent mieux aux données associées à ces *Mappings.* De ce fait, nous avons instancié le modèle d'analyse d'un indicateur, d'une façon générale, *comme étant une combinaison des mesures portant sur l'évaluation subjective (les questionnaires) et celles de l'évaluation objective (Mouchard électronique et inspecteur des règles ergonomiques).*

Afin de mieux illustrer les principaux concepts utilisés pour spécifier un indicateur, nous introduisons un exemple d'indicateur hypothétique qui porte sur le besoin d'information suivant : « évaluer le critère d'utilisabilité X des interfaces utilisateur d'un système interactif donné». Cet indicateur vise à évaluer le critère d'utilisabilité X sur la base des perceptions des utilisateurs et l'évaluation objective dans le but de vérifier l'existence ou l'absence des problèmes d'utilisabilité liés à ce critère.

Plus précisément, le modèle d'analyse correspondant à cet exemple d'indicateur peut être instancié comme étant une paire de mesures (M1, M2). Nous supposons que M1 est une mesure portant sur les

perceptions des utilisateurs (relative au critère évalué X) extraite d'un questionnaire utilisant une échelle ordinale comme celle proposée par CSUQ (de 1 Fortement en accord jusqu' à 7 Fortement en désaccord). Aussi, nous supposons que M2 est une mesure permettant d'assurer une évaluation objective extraite de l'un d'autres outils (le mouchard électronique EISEval ou l'inspecteur des règles ergonomiques) utilisant une échelle de ratio allant de 0 à 1 (ou de 0% à 100%).

Dans une seconde phase, nous nous intéressons à définir une méthode permettant de soutenir l'interprétation des résultats de mesures caractérisant l'indicateur. En adoptant la norme ISO/IEC 15939, l'interprétation des résultats d'un indicateur est construite sur la base des *critères de décision*. Rappelons que selon la norme ISO/IEC 15939, ces critères constituent des seuils, cibles, ou modèles permettant l'aide à l'interprétation des résultats de la mesure (ISO/IEC 15939, 2007).
Pour cela et en vue d'établir ces critères, nous avons défini un ensemble des règles simples qui tiennent compte de chaque modèle d'analyse choisi. Une règle consiste à interpréter les valeurs liées aux mesures combinées de l'indicateur sur la base des seuils bien déterminés pour chaque mesure. A leur tour, ces derniers permettent de vérifier si les valeurs des mesures ont de bons résultats.

Le Tableau 3.6 illustre un exemple des critères de décision portant sur l'exemple d'indicateur hypothétique mentionné précédemment (notant que les mesures adoptées, leurs valeurs et leurs seuils sont proposés uniquement à titre d'exemple illustratif et générique, nous allons présenter des exemples spécifiques dans la section suivante).

Tableau 3. 6 : Exemple des critères de décision associés à l'indicateur hypothétique

		M2	
		[0 ; 0.5 [	[**0.5** ; 1]
M1	{1, 2, **3**}	Mauvais niveau de qualité : Il y a des problèmes	**Bon niveau de qualité :** **Aucun problème**
	{4, 5, 6, 7}	Très mauvais niveau de qualité : Il y a des problèmes	Mauvais niveau de qualité : Il y a des problèmes

Comme exemples, la valeur 3 présente le seuil relatif à la mesure M1 jusqu'auquel les valeurs de M1 sont considérées bonnes et la valeur 0.5 présente le seuil relatif à M2 à partir duquel les valeurs de M2 sont considérées bonnes.

Comme le montre le Tableau 3.6, les critères de décision de l'exemple de l'indicateur (énoncé comme une paire de (M1, M2)) incluent quatre règles. A titre d'exemple, les résultats de l'indicateur indiquent un « bon niveau de qualité » si la mesure M2 a une valeur comprise entre l'intervalle [0,5 ; 1] **et** la mesure M1 a une valeur appartenant à l'ensemble de {1, 2, 3}.

En général, la spécification des règles de critères de décision d'un indicateur d'utilisabilité dépend nécessairement des valeurs de mesures associées et leurs interprétations générales (qui existent déjà dans la littérature), ainsi que du seuil choisi pour chaque mesure.

En outre, la sélection de ces seuils est cruciale pour supporter l'interprétation de l'indicateur. Toutefois, il s'avère très difficile et délicat de décider du choix des valeurs appropriées de ces seuils.

Bien que les mesures existantes se caractérisent généralement par des seuils généraux prédéfinis (par exemple « le plus proche de 1 est mieux » (ISO/IEC 25022, 2012)), l'intervention d'un expert est généralement essentielle pour interpréter les résultats (Oliveira et al., 2012).

Compte tenu que notre approche est destinée à l'évaluation des systèmes interactifs dédiés à la supervision du trafic du transport terrestre, nous avons décidé d'impliquer des experts d'évaluation de qualité des systèmes de supervision du trafic. Cet expert doit avoir une expérience dans l'évaluation de l'utilisabilité des interfaces utilisateur de la même technologie (dans notre cas, les applications WIMP) et, si c'est possible, dans le même type de domaine d'application (par exemple dans notre cas, les

systèmes en temps réel). Pour définir les seuils, l'expert doit tenir compte les informations du contexte d'utilisation du système à évaluer.

Selon la norme ISO 9241-11 (1998), le contexte d'utilisation est constitué des « utilisateurs, tâches et équipement (logiciels et matériels), ainsi que les environnements physiques et sociaux dans lesquels un produit est utilisé ».

Plus particulièrement, nous avons considéré un contexte d'utilisation bien déterminé dans lequel se déroule l'évaluation d'un système interactif donné de type WIMP.

Concernant les *utilisateurs*, nous avons considéré leurs âges, sexe et leurs expériences portant à la fois sur le domaine de la régulation du trafic du transport terrestre ainsi que la technologie informatique.
A propos des *tâches*, un scénario bien déterminé doit être défini visant à atteindre les objectifs d'évaluation concernés.
Pour les *équipements* (les technologies matérielles et logicielles à adopter dans l'évaluation), ils concernent notamment des ordinateurs personnels et une application WIMP (un **S**ystème d'**A**ide à l'**I**nformation Voyageurs : SAI).
Pour *l'environnement physique*, il porte sur les trois types d'environnement d'évaluation proposés par Dumas et Fox (Dumas et Fox, 2009) qui sont définis comme suit :

- Environnement en laboratoire : l'évaluation se déroule dans un laboratoire de recherche dans lequel les participants sont invités pour exécuter un test d'utilisabilité, même s'il ne s'agit pas concrètement de laboratoire typique d'utilisabilité ;
- Environnement à distance : lorsque l'administrateur de l'évaluation et les participants sont dans des endroits différents ; et
- Environnement sur place : dans le cas où l'évaluation se déroule dans l'environnement réel dans lequel les utilisateurs finaux travaillent.

De ce fait, l'expert définit les seuils des mesures sur la base de sa propre expérience dans les évaluations précédentes en tenant compte des différents aspects du contexte d'utilisation.

Pour les mesures portant sur l'évaluation objective, nous nous sommes basés sur les recommandations et les guides ergonomiques définies dans la littérature, appropriées aux applications WIMP afin d'aider les experts lors du choix du seuil. La technique de définition d'un seuil a été utilisée comme étant un moyen utile pour l'analyse des données telles que « le temps d'exécution d'une tâche » (Albert et Tullis, 2013). Ce seuil permet d'indiquer les résultats acceptables, qui peuvent être soit au-dessus ou en dessous de celui-ci.

En adoptant ce principe, nous décidons de fixer un seul seuil par mesure et de découper les résultats d'une mesure selon deux catégories. La première inclut les résultats que l'expert juge acceptables (ou bons) et la seconde concerne les résultats qu'il juge non acceptables (ou mauvais). Par conséquent, nous avons découpé les valeurs d'une mesure en deux intervalles : Acceptable et Non acceptable (dont l'expert devrait définir le seuil).

D'une façon similaire, le même principe a été appliqué pour les mesures portant sur l'évaluation subjective. Bien que pour les critères d'efficacité, d'efficience et de feedback immédiat nous avons considéré des mesures subjectives portant également sur les réponses directes des participants, nous avons adopté le même principe. De ce fait, l'expert définit un seuil unique en fonction de l'échelle du questionnaire adopté. Cette valeur va dépendre des points d'échelle utilisée ainsi que la tonalité de la question. Prenant l'exemple de l'échelle du questionnaire CSUQ comme l'illustre le Tableau 3.3 (7 points allant de «1 : fortement en accord» jusqu' à «7 : fortement en désaccord»), si l'expert fixe

comme seuil la valeur 3, donc les réponses allant de 1 à 3 représentent l'*Accord* des participants ; et celles allant de points 4 jusqu'à 7 représentent le *Désaccord* des participants.

Le Tableau 3.7 représente le même exemple de l'indicateur hypothétique illustré dans le Tableau 3.6 en spécifiant les intervalles appropriés.

Tableau 3. 7 : Exemple des critères de décision en spécifiant les intervalles associés à l'indicateur hypothétique

		M2	
		[0 ; 0.5 [: **Non acceptable**	[**0.5** ; 1] : **Acceptable**
M1	{1, 2, **3**} **Accord**	Mauvais niveau de qualité : Il y a des problèmes dans la mesure M2	**Bon niveau de qualité : Aucun problème**
	{4, 5, 6, 7} **Désaccord**	Très mauvais niveau de qualité : Il y a des problèmes dans les deux mesures	Mauvais niveau de qualité : Il y a des problèmes dans la mesure M1

Après avoir introduit le principe de définition des critères de décision d'un indicateur, nous avons jugé nécessaire de le représenter graphiquement dans le but d'assurer efficacement et concrètement l'intégration entre les mesures combinées. Ces représentations visent à offrir une compréhension simple et rapide des résultats de l'indicateur ainsi que de leurs interprétations sur la base de leurs critères de décision.

Pour cela, nous avons décidé d'adopter des graphiques appropriés selon les types de mesures concernées caractérisant chaque indicateur. Toutefois, malgré leurs relations complémentaires et cohérentes, ces mesures se diffèrent en termes de types de leurs résultats qui peuvent se baser sur différents types d'échelles (comme exemples, ordinale, et ratio) et données (discrètes ou continues).

Sur la base de la littérature de statistiques descriptives qui établissent des relations entre les variables hétérogènes (Tufte, 2002 ; Hongquan, 2004 ; Duquia et al., 2014) et tout en respectant les différents types de mesures concernées, nous décidons d'utiliser principalement les graphiques du nuage de points (2D, 3D). Ces graphiques vont illustrer chaque indicateur proposé. De plus, d'autres graphiques (tels que le diagramme à barres et le camembert) ont été adoptés lors de l'analyse détaillée des résultats des indicateurs.

La Figure 3.6 illustre un exemple graphique de l'indicateur hypothétique présenté précédemment. Le graphique montre l'intégration effective de deux mesures M1 (d'évaluation subjective) et M2 (d'évaluation objective) en un unique résultat (x). Il schématise en plus les règles définies des critères de décision pour l'aide à l'interprétation des résultats de l'indicateur. Ces dernières indiquent les trois niveaux de qualité présentés dans le Tableau 3.7 compte tenu des seuils établis par un expert (Seuil Y = 3 et Seuil X = 0,5). Par exemple, lorsqu'un résultat se situe entre les deux intervalles de *Désaccord et Non acceptable*, cela indique *un Très mauvais niveau* de qualité pour l'interface utilisateur (Figure 3.4).

Généralement, pour chaque indicateur nous identifions trois niveaux de qualité, comme suit :

- Très mauvais niveau de qualité : lorsque toutes les mesures de l'indicateur ont des mauvais résultats ;
- Mauvais niveau de qualité : lorsque au moins une des mesures a un mauvais résultat ; et
- Bon niveau de qualité : lorsque toutes les mesures de l'indicateur ont des bons résultats.

En outre, chaque niveau de qualité a été complémenté par certaines recommandations incluses dans les critères de décision en vue d'améliorer la qualité de l'interface utilisateur. Par ailleurs, nous avons

introduit des graphiques supplémentaires pour élaborer des analyses détaillées des résultats qui peuvent être utilisés dans le cas de détection des problèmes afin d'avoir une bonne évaluation.

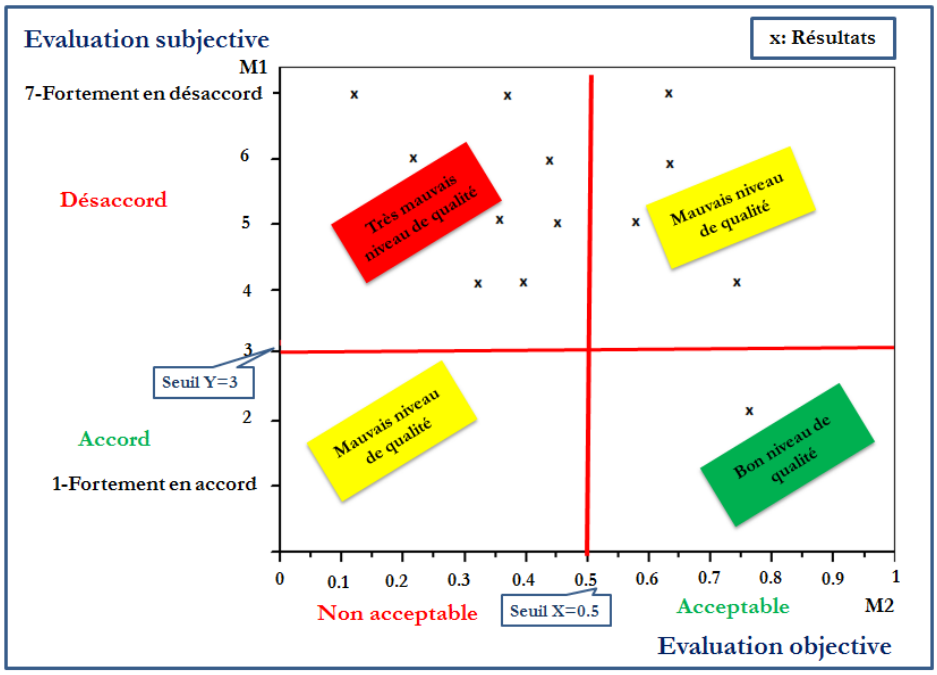


Figure 3. 6 : Exemple générale de présentation graphique d'un indicateur avec leurs critères de décision

En résumé de cette section, nous avons suivi les trois étapes spécifiées auparavant pour définir un ensemble d'indicateurs portant sur les différents critères d'utilisabilité choisis. Le Tableau 3.8 liste ces indicateurs.

Tableau 3. 8 : La liste des indicateurs proposés portant sur les critères/ sous-critères d'utilisabilité choisis

Critères/sous-critères d'utilisabilité choisis	Indicateurs d'utilisabilité
Efficacité	Efficacité de réalisation des tâches
Efficience	Efficience de réalisation des tâches
Densité informationnelle	Densité informationnelle d'une/des interface(s) utilisateur
Actions minimales	Actions minimales des tâches réalisées
Lisibilité	Conformité des caractéristiques textuelles d'une/des interface(s) utilisateur
	Conformité des couleurs utilisées d'une/des interface(s) utilisateur
Feedback immédiat	Feedback immédiat aux actions des utilisateurs sur les interfaces utilisateur

Dans la section suivante, nous présentons les spécifications détaillées de ces indicateurs proposés.

3.3 La spécification des modèles de mesure d'information des indicateurs d'utilisabilité des IHM basés sur la norme ISO/IEC 15939

En suivant les étapes de définition d'un indicateur d'utilisabilité indiquées précédemment, nous allons présenter dans cette section des spécifications détaillées des modèles de mesure d'information portant sur l'ensemble d'indicateurs d'utilisabilité proposés.

Cette section comporte deux parties. La première introduit les modèles de mesure d'information de deux indicateurs portant sur l'évaluation des critères ISO 9241-11 (efficacité et efficience). La seconde partie spécifie les indicateurs d'utilisabilité proposés concernant l'évaluation de l'ensemble des critères ergonomiques de Scapin et Bastien considérés.

3.3.1 Les modèles de mesure d'information proposés pour l'évaluation des critères ISO 9241-11 (Efficacité et Efficience)

Dans cette section, nous présentons les spécifications des principaux concepts des modèles de mesure d'information proposés pour définir deux indicateurs d'utilisabilité correspondent aux critères ISO 9241-11 d'efficacité (voir Tableau 3.9) et d'efficience (voir Tableau 3.10).

La spécification de ces deux indicateurs a été définie en considérant qu'une évaluation d'un système interactif donné se base sur la définition d'un scénario des tâches visant à résoudre un problème spécifique. Cela requiert l'intervention des participants pour exécuter l'ensemble des tâches spécifiées en utilisant le système. Compte tenu de *Mappings* définis dans le Tableau 3.1, l'évaluation de l'efficacité de toutes les tâches réalisées (par les participants) considère les mesures (subjective et objective) présentées dans le Tableau 3.4. De ce fait, nous avons instancié l'indicateur d'efficacité de réalisation des tâches comme étant une paire de mesures : *Réponse à la question* et *Taux moyen de réalisation correcte des tâches d'un utilisateur* (Assila et al., 2014b, 2014c, 2016c).
De la même façon, en vue d'évaluer l'efficience des toutes les tâches réalisées, nous avons adopté la mesure de «*Taux moyen d'efficience des tâches (Goal achievement efficiency)*» (définie par la norme ISO/IEC 25022, 2012). Cette mesure est obtenue en divisant la mesure de «*Taux moyen de réalisation correcte des tâches* » par «*le temps moyen des tâches*» (ISO/IEC 25022, 2012). Par conséquent, l'indicateur d'efficience de réalisation des tâches est instancié comme une paire de mesures de *Réponse à la question* et *Taux moyen d'efficience des tâches d'un utilisateur* (Assila et al., 2014b ; Assila et al., 2014c ; Assila et al., 2016c).

Ces deux indicateurs sont présentés sous forme d'un graphique du nuage de points. Les résultats d'un utilisateur sont schématisés sur ce graphique en couplant les deux mesures (subjective et objective) associées à l'indicateur. Généralement, ces deux types de mesures sont sensiblement corrélées (Sauro et Lewis, 2012 ; Sauro et Kindlund, 2005 ; Sauro et Lewis, 2009). Toutefois, nous avons préféré d'assurer la corrélation pour interpréter les résultats de ces deux indicateurs et fournir des recommandations. De ce fait, nous avons décidé de procéder à l'interprétation et l'analyse uniquement dans le cas où l'ensemble des résultats des utilisateurs fourni un coefficient de corrélation $r \geq 0{,}50$ (c.-à-d $r^2 \geq 0{,}25$) sur la base de (Cohen, 1988).

Concernant la spécification de seuils de ces indicateurs, l'expert doit considérer notamment l'ensemble des tâches qui composent le scénario spécifié ainsi que le contexte d'utilisation adopté de l'évaluation.

Par conséquent, ces indicateurs indiquent les niveaux de qualité et les problèmes potentiels relatifs à l'exécution de l'ensemble des tâches. En cas de détection des problèmes, il est important de considérer l'ensemble des recommandations appropriées définies dans chaque indicateur pour rectifier les problèmes détectés. Comme recommandation, l'évaluateur peut analyser les résultats détaillés par tâche.

Tableau 3. 9 : Spécification du modèle de mesure d'information de l'indicateur d'efficacité de réalisation des tâches

Besoin d'information	Évaluer l'efficacité de réalisation des tâches pendant l'exécution du système
Indicateur	Efficacité de réalisation des tâches
Modèle d'analyse	Le modèle est énoncé comme une paire de mesures (**Réponse à la question** et **Taux moyen de réalisation correcte des tâches d'un utilisateur**) représenté par un nuage de points. Tout d'abord, vérifier la corrélation entre les deux mesures. Si le coefficient de corrélation (r) $\geq 0{,}5$ (i.e, $r^2 \geq 0{,}25$), cela indique que les *réponses des utilisateurs* et leurs *Taux moyen de réalisation correcte des tâches* sont bien corrélés (Cohen, 1988). Par conséquent, l'interprétation des résultats des utilisateurs peut être procédée. Dans le cas contraire, cela signifie que les résultats de l'évaluation ne peuvent pas être interprétés avec cet indicateur. Il est recommandé d'identifier et de corriger les problèmes potentiels et de planifier de nouvelles sessions d'évaluation.

Tableau 3.9 (suite) : Spécification du modèle de mesure d'information de l'indicateur d'efficacité de réalisation des tâches

Critères de décision

Interprétation de chaque mesure

- Pour la mesure du *Taux moyen de réalisation correcte des tâches* : une valeur proportionnelle plus élevée est préférable (valeur maximale = 100%) (ISO/IEC 25022, 2012).
  0≤ *Taux moyen de réalisation correcte des tâches*≤100%.

  Les résultats sont divisés en deux groupes (intervalles) en fonction du seuil X spécifié par l'expert :
  - *Non acceptable* : Pour les valeurs ϵ [0% ; X[
  - *Acceptable* : Pour les valeurs ϵ [X ; 100%]

- La mesure de *Réponse à la question* est interprétée en fonction de l'échelle du questionnaire adoptée (par exemple, Likert) et la tonalité de la question.
  Exemple d'une question portant sur l'efficacité :
  « J'ai complété mon travail correctement en utilisant ce système" (*I can effectively complete my work using this system*") (une question extraite du questionnaire CSUQ).
  Cette question se base sur une échelle de Likert de 7 points allant de 1 : Fortement en accord jusqu'à 7 : Fortement en désaccord dont sa tonalité est positive.
  Par conséquent, la plus petite valeur est mieux (vu que la valeur 1 est Fortement en accord).

  Deux groupes sont définis en fonction du seuil Y spécifié par l'expert :
  - *Accord* : Pour les valeurs ϵ {1 à Y}
  - *Désaccord* : Pour les valeurs ϵ à {la valeur suivante après Y à 7}.

  Remarque : L'expert précise les seuils X et Y en tenant en compte du contexte d'utilisation de l'évaluation.

Critères de décision de l'indicateur

		Taux moyen de réalisation correcte des tâches	
		[0% ; X[: *Non acceptable*	[X ; 100%] : *Acceptable*
Réponse à la question	{1 à Y} *Accord*	Mauvais niveau d'efficacité : Il y a des problèmes dans la mesure « Taux moyen de réalisation correcte des tâches» **Recommandations :** - D'autres tests sont requis en se basant sur le contexte d'utilisation adéquat. - Analyser les résultats détaillés*.	Bon niveau d'efficacité : Pas de problèmes.
	{la valeur suivant Y à 7} *Désaccord*	Très mauvais niveau d'efficacité : Il existe des problèmes dans les deux mesures. **Recommandations :** - Probablement, il est nécessaire de reconcevoir les interfaces utilisateur. - Analyser les résultats détaillés*. - Evaluer d'autres critères pour fixer mieux les problèmes d'utilisabilité.	Mauvais niveau d'efficacité : Il existe des problèmes au niveau de satisfaction des utilisateurs. **Recommandations :** - Réviser les besoins des utilisateurs et du système. - Analyser les résultats détaillés*. - Evaluer d'autres critères pour fixer mieux les problèmes d'utilisabilité.

*Les résultats détaillés par rapport à chaque tâche sont représentés dans un graphique à barres. Ce dernier présente les valeurs de la mesure de *Taux moyen de réalisation correcte d'une tâche* (par rapport à tous les utilisateurs) pour chaque tâche (Albert et Tullis, 2013, Chapitre 2). Par conséquent, ce graphique permet l'aide à la détection des tâches qui ont des valeurs faibles et priorise celles qui doivent être améliorées en premier.

Mesures dérivées

Taux moyen de réalisation correcte des tâches (appelé *Mean goal achievement* en se basant sur des études de cas présentées dans la norme ISO /IEC 25022 (2012)) : mesure la moyenne de la proportion des objectifs de tâches correctement réalisées d'un utilisateur.

Fonctions de mesures

Taux moyen de réalisation correcte des tâches = ∑ (Taux de réalisation correcte d'une tâche)/ Nombre total des tâches

Mesures de base

1. Taux de réalisation correcte d'une tâche (appelé *Task effectiveness*) (ISO/IEC 25022, 2012 ; ISO/IEC 9126-4, 2004).
2. Nombre total de tâches (ISO/IEC 9126-4,2004 ; ISO/IEC 25022,2012)
3. **Réponse à la question** (ISO/IEC 25022, 2012)

Méthodes de mesures

1. Taux de réalisation correcte d'une tâche = (1-∑valeur proportionnelle de chaque composante manquante ou incorrecte lors de la réalisation de la tâche) *100 =la valeur proportionnelle de l'achèvement correct de la tâche
2. Compter le nombre total de tâches
3. Demander aux utilisateurs leurs perceptions sur l'efficacité de réalisation de leurs tâches

Remarque : Toutes les mesures sont déterminées de façon automatique en utilisant la version étendue du mouchard électronique EISEval et l'outil de synthèse des résultats (*Cf. Chapitre 4*).

Tableau 3. 10 : Spécification du modèle de mesure d'information de l'indicateur d'efficience de réalisation des tâches

Besoin d'information

Évaluer l'efficience de réalisation des tâches pendant l'exécution du système

Indicateur

Efficience de réalisation des tâches

Modèle d'analyse

Le modèle est énoncé comme une paire de mesures (**Réponse à la question** et **Taux moyen d'efficience des tâches d'un utilisateur**) représenté par un nuage de points.
Tout d'abord, vérifier la corrélation entre les deux mesures. Si le coefficient de corrélation (r) ≥ 0,5 (i.e, $r^2 \geq 0,25$), cela indique que les *réponses des utilisateurs* et leurs *Taux moyen de réalisation correcte des tâches* sont bien corrélés (Cohen, 1988). Par conséquent, l'interprétation des résultats des utilisateurs peut être procédée. Dans le cas contraire, cela signifie que les résultats de l'évaluation ne peuvent pas être interprétés avec cet indicateur. Il est recommandé d'identifier et de corriger les problèmes potentiels et de planifier de nouvelles sessions d'évaluation.

Critères de décision

Interprétation de chaque mesure

- Pour la mesure de *Taux moyen d'efficience des tâches* : Une valeur élevée indique une forte proportion de succès d'utilisateurs dans une petite quantité de temps (ISO / IEC 25022, 2012).
  0≤ *Taux moyen d'efficience des tâches* ≤N, où N est la valeur maximale fixée par l'expert.

  Les résultats sont divisés en deux groupes (intervalles) en fonction du seuil X spécifié par l'expert :
  - *Non acceptable* : Pour les valeurs ϵ [0 ; X[
  - *Acceptable* : Pour les valeurs ϵ [X ; N]

  Comme recommandation, un expert devrait exécuter les tâches et calculer une valeur théorique en se basant sur ses propres résultats de la mesure de l'efficience réalisation des tâches (Trabelsi et Ezzedine, 2013). Selon Albert et Tullis (2013), une manière de déterminer cette valeur théorique est d'assurer le suivi du temps d'exécution des tâches. Celui-ci peut être doublé ou triplé pour calculer la valeur limite acceptable d'efficience des tâches (Albert et Tullis, 2013).

- La mesure de *Réponse à la question* est interprétée en fonction de l'échelle du questionnaire adopté (par exemple, Likert) et la tonalité de la question.
  Exemple d'une question portant sur l'efficience :
  —J'ai été en mesure de compléter rapidement ma tâche avec ce système" (I am able to complete my work quickly using this system) (une question extraite du questionnaire CSUQ).
  Cette question se base sur une échelle de Likert de 7 points allant de 1 : Fortement en accord jusqu'à 7 : Fortement en désaccord dont sa tonalité est positive.
  Par conséquent, la plus petite valeur est mieux (vu que la valeur 1 est Fortement en accord).

  Deux groupes sont définis en fonction du seuil Y spécifié par l'expert :
  - *Accord* : Pour les valeurs ϵ {1 à Y}
  - *Désaccord* : Pour les valeurs ϵ à {la valeur suivante après Y à 7}.

  Remarque : L'expert précise les seuils X et Y en tenant en compte du contexte d'utilisation de l'évaluation.

Critères de décision de l'indicateur

		Taux moyen d'efficience des tâches	
		[0 ; X[: *Non acceptable*	[X ; N] : *Acceptable*
Réponse à la question	{1 à Y} *Accord*	Mauvais niveau d'efficience : Il y a des problèmes dans la mesure « Taux moyen d'efficience des tâches » **Recommandations :** - D'autres tests sont requis en se basant sur le contexte d'utilisation adéquat. - Analyser les résultats détaillés*.	Bon niveau d'efficience : Pas de problèmes.
	{la valeur suivant Y à 7} *Désaccord*	Très mauvais niveau d'efficience : Il existe des problèmes dans les deux mesures. **Recommandations :** - Probablement, il est nécessaire de reconcevoir les interfaces utilisateur. - Analyser les résultats détaillés*. - Evaluer d'autres critères pour fixer mieux les problèmes d'utilisabilité.	Mauvais niveau d'efficience : Il existe des problèmes de satisfaction des utilisateurs. **Recommandations :** - Réviser les besoins des utilisateurs et du système. - Analyser les résultats détaillés*. - Evaluer d'autres critères pour fixer mieux les problèmes d'utilisabilité.

* Les résultats détaillés par rapport à chaque tâche sont représentés dans un graphique à barres. Ce dernier présente les valeurs de la mesure de *Taux moyen d'efficience d'une tâche* (de tous les utilisateurs) pour chaque tâche (Albert et Tullis, 2013, chapitre 2) par rapport aux valeurs de la mesure *Taux moyen/théorique d'efficience d'une tâche* défini par l'expert. Par conséquent, ce graphique permet l'aide à la détection des tâches qui ont des valeurs faibles et priorise celles qui doivent être améliorées en premier.

Tableau 3.10 (suite) : Spécification du modèle de mesure d'information de l'indicateur d'efficience de réalisation des tâches

Mesures dérivées	**Taux moyen d'efficience des tâches** (appelé *Goal achievement efficiency* en se basant sur des études de cas présentées dans la norme ISO/IEC 25022 (2012)) : mesure la moyenne de la proportion des objectifs de tâches correctement réalisées d'un utilisateur par rapport à chaque unité de temps.
Fonctions de mesures	Taux moyen d'efficience des tâches= Taux moyen de réalisation correcte des tâches/ Temps moyen de réalisation de toutes les tâches par un utilisateur = ∑ (Taux de réalisation correcte d'une tâche) / Temps total des tâches.
Mesures de base	1. Taux de réalisation correcte d'une tâche (appelé *Task effectiveness*) (ISO/IEC 25022, 2012 ; ISO/IEC 9126-4, 2004). 2. Temps total des tâches (appelé *Total Task time* en se basant sur des études de cas présentées dans la norme ISO/IEC 25022 (2012)) 3. **Réponse à la question** (ISO/IEC 25022, 2012)
Méthodes de mesures	1. Taux de réalisation correcte d'une tâche = 1-∑valeur proportionnelle de chaque composante manquante ou incorrecte lors de la réalisation de la tâche =la valeur proportionnelle de l'achèvement correct de la tâche. 2. Temps total de tâches = ∑ Durée d'une tâche, Où la *Durée d'une tâche* est définie par (ISO/IEC 25022, 2012) 3. Demander aux utilisateurs leurs perceptions sur l'efficience de réalisation de leurs tâches. **Remarque :** Toutes les mesures sont déterminées de façon automatique en utilisant la version étendue du mouchard électronique EISEval et l'outil de synthèse des résultats (*Cf. Chapitre 4*).

3.3.2 Les modèles de mesure d'information proposés pour l'évaluation d'un ensemble des critères ergonomiques de Scapin et Bastien

Cette section introduit des autres indicateurs d'utilisabilité proposés relatifs aux sous-critères ergonomiques considérés de Scapin et Bastien. Les indicateurs qui portent sur la densité informationnelle et la lisibilité concernent notamment les interfaces utilisateur du système interactif à évaluer, en vue de déterminer la qualité ergonomique de chaque interface concernée par rapport à chaque critère. Tandis que l'indicateur qui concerne le sous-critère d'actions minimales porte plutôt sur l'évaluation des actions minimales de chaque tâche effectuée par les utilisateurs. Ainsi, l'indicateur qui concerne le feedback immédiat porte sur l'évaluation des rétroactions du système suite aux actions des utilisateurs lors de la réalisation de leurs tâches.

La spécification de ces indicateurs a été déterminée en considérant principalement les différents *Mappings* définis dans le Tableau 3.1 sur la base des mesures spécifiées dans les Tableaux 3.4 et 3.5. Chaque indicateur est instancié comme étant une paire ou triplet des mesures d'évaluation subjective et objective.

En effet, pour les mesures d'évaluation subjective, nous avons adopté la mesure de « Réponse à une question » ainsi que celle du « *Pourcentage des utilisateurs ayant des réponses positives à la question»* (Albert et Tullis, 2013). Celle-ci indique le pourcentage de satisfaction des utilisateurs par rapport à la question posée relative à une interface (IU) ou à une tâche afin de mieux identifier les IU /tâches qui ont des problèmes ergonomiques. Afin de calculer cette mesure, nous avons proposé l'utilisation de scores top box (e.g, top un, deux ou trois) avant de fixer les seuils. Le score Top-Box est une technique généralement utilisée lors de l'analyse et la synthèse des données à partir des échelles d'évaluation (Albert et Tullis, 2013, p.128). Les top boxes sont l'extrémité(s) positives de l'échelle, celles qui indiquent les réponses positives par rapport à l'énoncé de la question (Tullis et Albert, 2008, p.128). En général, les scores Top-2-Box sont utilisés plus souvent avec les échelles les plus grandes, telles que les échelles de 7 ou 9 points. L'expert décide s'il/elle veut travailler avec le top-box, top-2-box ou top-3-box lors de l'analyse des résultats de l'indicateur. Par conséquent, le score top-box calculé est donc schématisé dans le graphique de l'indicateur. Par la suite, l'expert se charge de définir un seuil unique pour l'aide à l'interprétation des résultats obtenus. Pour les mesures concernant l'évaluation objective, nous nous sommes basés notamment sur celles proposées dans la littérature pour adapter certaines nouvelles mesures. Celles-ci, nous allons les présenter dans les sous-sections suivantes.

Enfin, pour mieux aider l'expert lors du choix des seuils de l'indicateur, nous avons inclus un ensemble des recommandations ergonomiques existant en prenant en compte le type des interfaces graphiques concernées (les IU statiques de type WIMP), en plus du contexte d'utilisation adopté de l'évaluation. En outre, lors d'une évaluation, toutes les interfaces d'un système interactif donné sont évaluées sur la base d'un même seuil. A la fin de l'évaluation, chaque indicateur indique le niveau de qualité et les problèmes potentiels relatifs à chaque interface. En cas de détection des problèmes, l'indicateur suggère des recommandations appropriées pour améliorer la qualité de l'interface utilisateur évaluée.
Dans ce contexte, nous présentons dans ce qui suit, les spécifications complètes des indicateurs proposés pour l'évaluation de la densité informationnelle, la lisibilité, les actions minimales et le feedback immédiat.

3.3.2.1 L'évaluation de la densité informationnelle d'une interface utilisateur

La densité informationnelle est l'un des sous-critères de la charge de travail proposés par Scapin et Bastien (1997). Elle concerne la charge de travail du point de vue perceptif et mnésique, pour des ensembles d'éléments et non pour des items (Bastien et Scapin, 1993). Le tableau 3.11 fournit une spécification du modèle de mesure proposé pour l'évaluation de ce critère. La mesure subjective concerne le "*Pourcentage des utilisateurs ayant des réponses positives à la question*". Pour la mesure objective, nous avons choisi la "*densité globale*" en tant qu'une mesure de la densité informationnelle largement référencée dans la littérature (Tullis, 1983, 1997 ; Vanderdonckt et Gillo, 1994 ; Vanderdonckt, 2003 ; Padda, 2003 ; Gabillon et al, 2013 ; Galitz, 2007). Dans les interfaces utilisateur graphiques, la densité est calculée en divisant le nombre de pixels éclairés par le nombre total de pixels disponibles (Vanderdonckt et Gillo, 1994 ; Vanderdonckt, 2003). Gabillon et al. (2013) ont calculé la densité globale en divisant l'espace utilisé d'une interface sur l'espace total d'une interface.

Pour la définition de seuil, étant donné que nous nous sommes focalisés dans notre travail sur l'interface utilisateur statique et non pas les IUs qui sont dynamiquement adaptées, l'expert devrait tenir compte des recommandations proposées dans la littérature pour évaluer la densité globale de l'interface utilisateur graphique (Tullis, 1983 ; Vanderdonckt et Gillo, 1994 ; Galitz, 2007) (voir Tableau 3.11). Par conséquent, l'indicateur indique pour chaque interface si elle a des problèmes de densité et précise quelques recommandations pour améliorer leur qualité (Assila et al., 2016b, 2016c).

Tableau 3. 11 : Spécification du modèle de mesure d'information de l'indicateur de la densité informationnelle d'une/des interface(s) utilisateur

Besoin d'information	Évaluer la densité globale des informations d'une interface utilisateur d'un système
Indicateur	Densité informationnelle d'une/des interface(s) utilisateur
Modèle d'analyse	Le modèle est énoncé comme une paire de mesures (**Pourcentage des utilisateurs ayant des réponses positives** et **la densité globale d'une interface**) représenté par un nuage de points.
Critères de décision	**Interprétation de chaque mesure** • Pour la mesure de *densité globale d'une interface* : 0%<*Densité globale*≤100% Recommandations pour le choix du seuil appliqué à une interface graphique : - La densité globale ne doit pas dépasser 25% (Tullis, 1983 ; Vanderdonckt et Gillo, 1994). - Restreindre la densité de l'écran à pas plus de 30% (Galitz, 2007). Les résultats sont divisés en deux groupes (intervalles) en fonction du seuil X spécifié par l'expert : - *Acceptable* : Pour les valeurs ϵ]0% ; X] - *Non acceptable* : Pour les valeurs ϵ]X ;100%] • Pour la mesure de *Pourcentage des utilisateurs ayant des réponses positives de la question:* une valeur proportionnelle plus élevée est mieux (valeur maximale = 100%) 0≤ *Pourcentage des utilisateurs ayant des réponses positives à la question par interface* ≤100% Les réponses positives d'une question dépendent des meilleures réponses en utilisant les scores Top-Boxes (Top-Box, Top-2-Box, ou Top-3-Box). Ces dernières sont interprétées en fonction de l'échelle du questionnaire adopté (par exemple, Likert) et la tonalité de la question.

Tableau 3.11 (suite) : Spécification du modèle de mesure d'information de l'indicateur de la densité informationnelle d'une/des interface(s) utilisateur

Critères de décision

Exemple d'une question portant sur la densité informationnelle :

—Il est facile de trouver l'information dont j'avais besoin" (It is easy to find the information I needed) (une question extraite du questionnaire CSUQ)

Cette question se base sur une échelle de Likert de 7 points allant de 1 : Fortement en accord jusqu'à le point 7 : Fortement en désaccord dont sa tonalité est positive.
Par conséquent, la plus petite valeur est mieux (vu que la valeur 1 est Fortement en accord).

Deux intervalles sont définis en fonction du seuil Y spécifié par l'expert :
- *Acceptable* : Pour les valeurs ϵ [Y ; 100%]
- *Non acceptable* : Pour les valeurs ϵ [0% ; Y[

Remarque : L'expert précise les seuils X et Y en tenant en compte du contexte d'utilisation de l'évaluation.

Critères de décision de l'indicateur

		Densité globale d'une interface	
		]0% ; X] : *Acceptable*	]X ; 100%] : *Non acceptable*
Pourcentage des utilisateurs ayant des réponses positives à la question par interface	[Y ; 100%] Acceptable	Bon niveau de densité de l'interface : Pas de problèmes	Mauvais niveau de densité : Il y a des problèmes de densité globale de l'interface **Recommandations :** - Des révisions sont nécessaires pour minimiser la densité élevée de l'interface. - Analyser les résultats*
	[0% ; Y[Non acceptable	Mauvais niveau de densité : Il y a des problèmes de satisfaction des utilisateurs. **Recommandations :** - Évaluer la possibilité de réorganiser les éléments représentés sur l'écran pour faciliter la localisation de l'information par l'utilisateur. - Évaluer d'autres critères. - Analyser les résultats**	Très mauvais niveau de densité : Il y a des problèmes sur les deux mesures. **Recommandations :** - Évaluer la possibilité de reconcevoir l'interface pour minimiser sa densité et faciliter la localisation de l'information par l'utilisateur - Analyser les résultats***

*Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus haut niveau de densité par rapport aux autres.
** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus bas niveau de satisfaction par rapport aux autres.
*** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus haut niveau de densité et le plus bas niveau de satisfaction.

Mesures dérivées

Densité globale d'une interface (Tullis, 1983 ; Vanderdonckt et Gillo, 1994 ; Padda, 2003 ; Gabillon et al., 2013 ; Galitz, 2007) : mesure la valeur proportionnelle de l'affichage global utilisé pour présenter les informations d'une interface.
Pourcentage des utilisateurs ayant des réponses positives à la question par interface (Albert et Tullis, 2013) : mesure la valeur proportionnelle des participants qui ont donné des réponses positives à la question. Cette mesure est déterminée sur la base de *scores Top Boxes* sélectionnés par l'expert (Top-Box, Top-2-Box, ou Top-3-Box).

Fonctions de mesures

Densité globale d'une interface = Espace utilisé sur une interface / Espace total d'une interface

Pourcentage des utilisateurs ayant des réponses positives à la question= Nombre des utilisateurs qui ont donné des réponses positives /Nombre total des utilisateurs

Mesures de base

1. Espace utilisé sur une interface (*le nombre de pixels de tous les objets de l'interface*)
2. Espace total d'une interface (*le nombre total de pixels disponibles)*
3. Nombre des utilisateurs qui ont donné des réponses positives
4. Nombre total des utilisateurs (ISO/IEC 25022, 2012)

Méthodes de mesures

1. Calculer le nombre de pixels de tous les objets de l'interface.
2. Calculer le nombre total de pixels de l'interface
3. Compter le nombre de utilisateurs qui ont choisi la/les réponse(s) positive(s).
4. Compter le nombre total des interrogés/utilisateurs.

Remarque : Toutes les mesures sont déterminées de façon automatique en utilisant la version étendue de l'inspecteur des règles ergonomiques et l'outil de synthèse des résultats (*Cf. Chapitre 4*).

3.3.2.2 L'évaluation de la lisibilité d'une interface utilisateur

La lisibilité est un sous-critère du guidage proposé par Scapin et Bastien (1997). Elle concerne les caractéristiques lexicales de présentation des informations sur l'écran pouvant entraver ou faciliter la

lecture de ces informations (Bastien et Scapin, 1993). Par exemple, le couleur de fond, la taille, le type, le couleur de police de l'information sont liées au critère de la lisibilité (Bastien et Scapin, 1993).

Le Tableau 3.12 fournit une spécification d'un modèle de mesure proposé pour l'évaluation de ce critère en fonction de la conformité des caractéristiques textuelles des informations présentées. L'objectif principal de cet indicateur est de vérifier si les caractéristiques textuelles relatives au type de police et la taille de la police des informations présentées sont conformes aux règles ergonomiques choisies par l'évaluateur et si les utilisateurs sont satisfaits. Il considère un triplet de mesures représentées sous forme du nuage de points de trois dimensions (voir le tableau 3.12). La mesure de la conformité peut être déterminée de façon automatique en utilisant l'inspecteur des règles ergonomiques.

Dans cet indicateur, nous avons adapté deux mesures d'évaluation objective (le *Taux de conformité des tailles de police recommandées* et le *Taux de conformité des types de police recommandés)* en se basant sur les mesures ergonomiques existantes (voir *Tableau 3.12*).

Tableau 3. 12 : Spécification du modèle d'information de mesure de l'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur

Besoin d'information	Évaluer la lisibilité d'une interface utilisateur par rapport à la conformité des caractéristiques textuelles des informations présentées
Indicateur	Conformité des caractéristiques textuelles d'une/des interface(s) utilisateur
Modèle d'analyse	Le modèle est énoncé comme un triplet de mesures (**Pourcentage des utilisateurs ayant des réponses positives** et **Taux de conformité des types de police recommandés** et **Taux de conformité des tailles de police recommandées**) représenté par un nuage de points en 3D.
Critères de décision	**Interprétation de chaque mesure** • Pour la mesure de *Taux de conformité des types de police recommandés d'une interface utilisateur* : une valeur proportionnelle plus élevée indique que les types de police de l'interface utilisateur sont conformes aux types de police recommandés par l'évaluateur (plus grand est mieux) (valeur maximale = 100%). 0≤*Taux de conformité des types de police*≤100%. Les résultats sont divisés en deux groupes (intervalles) en fonction du seuil X1 spécifié par l'expert : - *Non acceptable* : Pour les valeurs ϵ [0% ; X1[- *Acceptable :* Pour les valeurs ϵ [X1 ; 100%] • Pour la mesure de *Taux de conformité des tailles de police recommandées d'une interface utilisateur* : une valeur proportionnelle plus élevée indique que les tailles de police de l'interface utilisateur sont conformes aux tailles recommandées par l'évaluateur (plus grand est mieux) (valeur maximale = 100%). 0≤*Taux de conformité des tailles de police*≤100%. Les résultats sont divisés en deux groupes (intervalles) en fonction du seuil X2 spécifié par l'expert : - *Non acceptable* : Pour les valeurs ϵ [0% ; X2[- *Acceptable* : Pour les valeurs ϵ [X2 ; 100%] • Pour la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question :* une valeur proportionnelle plus élevée est mieux (valeur maximale = 100%) 0≤ *Pourcentage des utilisateurs ayant des réponses positives à la question par interface*≤100% Les réponses positives d'une question dépendent des meilleures réponses adoptées en utilisant les scores Top-Boxes (Top-Box, Top-2-Box, ou Top-3-Box). Ces dernières sont interprétées en fonction de l'échelle du questionnaire (par exemple, Likert) et la tonalité de la question. *Exemple d'une question portant sur la lisibilité :* ―La lecture d'informations sur l'écran" (Reading characters on the screen) (une question extraite du questionnaire QUIS) Cette question se base sur une échelle de Likert de 10 points allant de 0 : Difficile jusqu'à le point 9 : Facile dont sa tonalité est positive. Par conséquent, la plus grande valeur est mieux (vu que la valeur 9 est Facile). Deux intervalles sont définis en fonction du seuil Y spécifié par l'expert : - *Acceptable :* Pour les valeurs ϵ [Y ; 100%] - *Non acceptable* : Pour les valeurs ϵ [0% ; Y[**Remarque :** L'expert précise les seuils X 1, X2, et Y en tenant en compte du contexte d'utilisation de l'évaluation.

Tableau 3.12 (suite) : Spécification du modèle d'information de mesure de l'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur

Critères de décision

Critères de décision de l'indicateur

		Taux de conformité des types de police recommandés d'une interface			
		[X1 ; 100%] : *Acceptable*		[0% ; X1[: *Non acceptable*	
		Taux de conformité des tailles de police recommandées d'une interface		**Taux de conformité des tailles de police recommandées d'une interface**	
		[X2 ; 100%] : *Acceptable*	[0 ; X2[: *Non acceptable*	[X2 ; 100%] : *Acceptable*	[0 ; X2[: *Non acceptable*
Pourcentage des utilisateurs ayant des réponses positives	[Y ; 100%] *Acceptable*	Bon niveau de lisibilité de l'interface : Pas de problèmes	Mauvais niveau de lisibilité : Il existe des problèmes sur les tailles de police de l'interface **Recommandations** Certaines révisions sont nécessaires pour répondre aux règles ergonomiques de tailles de police. -Analyser les résultats*	Mauvais niveau de lisibilité : Il existe des problèmes sur les types de police de l'interface **Recommandations** Certaines révisions sont nécessaires pour répondre aux règles ergonomiques de types de police. - Analyser les résultats***	Mauvais niveau de lisibilité : Il existe des problèmes sur les types et les tailles de police de l'interface. **Recommandations** Des révisions sont nécessaires pour répondre aux règles ergonomiques. -Analyser les résultats*****
	[0% ; Y[*Non acceptable*	Mauvais niveau de lisibilité : Il existe des problèmes sur la satisfaction des utilisateurs **Recommandations** - Réviser les besoins des utilisateurs. - Évaluer d'autres critères. - Analyser les résultats**.	Mauvais niveau de lisibilité : Il existe des problèmes sur la satisfaction des utilisateurs et les tailles de police de l'interface. **Recommandations** -Des révisions sont nécessaires pour répondre aux règles ergonomiques de tailles de police. -Réviser les besoins des utilisateurs. - Évaluer d'autres critères. - Analyser les résultats****	Mauvais niveau de lisibilité : Il existe des problèmes sur la satisfaction des utilisateurs et les types de police utilisés de l'interface. **Recommandations** -Des révisions sont nécessaires pour répondre aux règles ergonomiques de types de police. -Réviser les besoins des utilisateurs. -Évaluer d'autres critères - Analyser les résultats******	Très mauvais niveau de lisibilité : Il existe des problèmes au niveau de trois mesures **Recommandations** - Révisions totales sont nécessaires pour améliorer les caractéristiques textuelles en respectant les recommandations ergonomiques. -Analyser les résultats*******

* Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus bas niveau de Taux de conformité des tailles de police recommandées.

** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus bas niveau de satisfaction

*** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus bas niveau de Taux de conformité des types de police recommandés.

**** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a les plus bas niveaux de satisfaction et du Taux de conformité des tailles de police recommandées.

***** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a les plus bas niveaux du Taux de conformité des tailles de police recommandées et du Taux de conformité des types de police recommandés.

****** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a les plus bas niveaux de satisfaction et du Taux de conformité des types de police recommandés.

******* Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a les plus bas niveaux des trois mesures.

Mesures dérivées

Taux de conformité des types de police recommandés d'une interface : mesure la valeur proportionnelle de la conformité des types de police recommandés d'une interface.
Taux de conformité des tailles de police recommandées d'une interface : mesure la valeur proportionnelle de la conformité des tailles de police recommandées d'une interface.
Pourcentage des utilisateurs ayant des réponses positives de la question par interface (Albert et Tullis, 2013) : mesure la valeur proportionnelle des participants qui ont donné des réponses positives de la question. Cette mesure est déterminée sur la base de scores *Top Boxes* sélectionnés par l'expert (Top-Box, Top-2-Box, ou Top-3-Box).

Tableau 3.12 (suite) : Spécification du modèle d'information de mesure de l'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur

Fonctions de mesures	Taux de conformité des types de police recommandés = Nombre de types de police qui respectent les types de police recommandés / Nombre des types de police d'une interface Taux de conformité des tailles de police recommandées = Nombre de tailles de police qui respectent les tailles de police recommandées / Nombre de tailles de police d'une interface. Pourcentage des utilisateurs ayant des réponses positives à la question= Nombre des utilisateurs qui ont donné des réponses positives /Nombre total des utilisateurs
Mesures de base	1. Nombre de types de police qui respectent les types recommandés 2. Nombre de types de police d'une interface (Padda, 2003 ; Seffah et al., 2006) 3. Nombre de tailles de police qui respectent les tailles recommandées 4. Nombre de tailles de police d'une interface (Padda, 2003 ; Seffah et al., 2006) 5. Nombre des utilisateurs qui ont donné des réponses positives 6. Nombre total des utilisateurs (ISO/IEC 25022, 2012)
Méthodes de mesures	1. Compter le nombre de types de police qui respectent les types de police recommandés 2. Compter le nombre de types de police d'une interface 3. Compter le nombre de tailles de police qui respectent les tailles de police recommandées 4. Compter le nombre de tailles de police d'une interface 5. Compter le nombre de utilisateurs qui ont choisi la/les réponse(s) positive(s). 6. Compter le nombre total des interrogés/utilisateurs. **Remarque :** Toutes les mesures sont déterminées d'une façon automatique en utilisant l'inspecteur des règles ergonomiques (version étendue) et l'outil de synthèse des résultats (*Cf. Chapitre 4*).

De plus, nous avons proposé un deuxième indicateur pour l'évaluation de la lisibilité en fonction de la conformité des couleurs utilisées des interfaces utilisateur. En suivant le même principe de l'indicateur précédant, celui-ci considère, également, un triplet de mesures représentées sous forme du nuage de points de trois dimensions. Le Tableau 3.13 fournit la spécification du modèle de mesure proposé. Dans cet indicateur, nous avons adapté deux mesures d'évaluation objective (le *Taux de conformité des couleurs d'écriture recommandées* et le *Taux de conformité des couleurs de fond recommandées)* en se basant sur les mesures ergonomiques existantes (voir *Tableau 3.13*).

Tableau 3. 13 : Spécification du modèle de mesure d'information de l'indicateur de conformité des couleurs des interfaces utilisateur

Besoin d'information	Évaluer la lisibilité d'une interface utilisateur par rapport à la conformité des couleurs utilisés
Indicateur	Conformité des couleurs utilisées d'une/des interface(s) utilisateur
Modèle d'analyse	Le modèle est énoncé comme un triplet de mesures (**Pourcentage des utilisateurs ayant des réponses positives** et **Taux de conformité des couleurs d'écriture recommandées** et **Taux de conformité des couleurs de fond recommandées**) représenté par un nuage de points en 3D.
Critères de décision	**<u>Interprétation de chaque mesure</u>** • Pour la mesure de *Taux de conformité des couleurs d'écriture recommandées d'une interface utilisateur* : une valeur proportionnelle plus élevée indique que les couleurs d'écriture de l'interface utilisateur sont conformes aux couleurs recommandées par l'évaluateur (plus grand est mieux) (valeur maximale = 100%). $0 \leq$ *Taux de conformité des couleurs d'écriture recommandées* $\leq 100\%$. Les résultats sont divisés en deux groupes (intervalles) en fonction du seuil X1 spécifié par l'expert : - *Non acceptable* : Pour les valeurs ϵ [0 ; X1[- *Acceptable* : Pour les valeurs ϵ [X1 ; 100%] • Pour la mesure de *Taux de conformité des couleurs de fond recommandées d'une interface utilisateur* : une valeur proportionnelle plus élevée indique que les couleurs de fond de l'interface utilisateur sont conformes aux couleurs recommandées par l'évaluateur (plus grand est mieux) (valeur maximale = 100%). $0 \leq$ *Taux de conformité des couleurs de fond recommandées* $\leq 100\%$. Les résultats sont divisés en deux groupes (intervalles) en fonction du seuil X2 spécifié par l'expert : - *Non acceptable* : Pour les valeurs ϵ [0 ; X2[- *Acceptable* : Pour les valeurs ϵ [X2 ; 100%] • Pour la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question :* une valeur proportionnelle plus élevée est mieux (valeur maximale = 100%) $0 \leq$ *Pourcentage des utilisateurs ayant des réponses positives à la question par interface* $\leq 100\%$ Les réponses positives d'une question dépendent des meilleures réponses en utilisant les scores Top-Boxes (Top-Box, Top-2-Box, ou Top-3-Box). Ces dernières sont interprétées en fonction de l'échelle du questionnaire (par exemple, Likert) et la tonalité de la question.

Tableau 3.13 (suite) : Spécification du modèle de mesure d'information de l'indicateur de conformité des couleurs des interfaces utilisateur

Critères de décision

Exemple d'une question portant sur la lisibilité :

—La lecture d'informations sur l'écran" (Reading characters on the screen) (une question extraite du questionnaire QUIS)

Cette question se base sur une échelle de Likert de 10 points allant de 0 : Difficile jusqu'au point 9 : Facile, dont sa tonalité est positive. Par conséquent, la plus grande valeur est mieux (vu que la valeur 9 est Facile).

Deux intervalles sont définis en fonction du seuil Y spécifié par l'expert :
- *Acceptable* : Pour les valeurs ϵ [Y ; 100%]
- *Non acceptable* : Pour les valeurs ϵ [0 ; Y[

Remarque : L'expert précise les seuils X 1, X2, et Y en tenant en compte du contexte d'utilisation de l'évaluation.

Critères de décision de l'indicateur

		Taux de conformité des couleurs d'écriture recommandées d'une interface			
		[X1 ; 100%] : *Acceptable*		[0 ; X1[: *Non acceptable*	
		Taux de conformité des couleurs de fond recommandées d'une interface		**Taux de conformité des couleurs de fond recommandées d'une interface**	
		[X2 ; 100%] : *Acceptable*	[0 ; X2[: *Non acceptable*	[X2 ; 100%] : *Acceptable*	[0 ; X2[: *Non acceptable*
Pourcentage des utilisateurs ayant des réponses positives	[Y ; 100%] *Acceptable*	Bon niveau de lisibilité de l'interface : Pas de problèmes	Mauvais niveau de lisibilité : Il existe des problèmes sur les couleurs de fond de l'interface **Recommandations** -Certaines révisions sont nécessaires pour répondre aux règles ergonomiques de couleurs de fond. - Analyser les résultats*	Mauvais niveau de lisibilité : Il existe des problèmes sur les couleurs d'écriture de l'interface **Recommandations** -Certaines révisions sont nécessaires pour répondre aux règles ergonomiques de couleurs d'écriture. - Analyser les résultats***	Mauvais niveau de lisibilité : Il existe des problèmes sur les couleurs d'écriture et de fond de l'interface. **Recommandations** -Des révisions sont nécessaires pour répondre aux règles ergonomiques. - Analyser les résultats*****
	[0 ; Y[*Non acceptable*	Mauvais niveau de lisibilité : Il existe des problèmes sur la satisfaction des utilisateurs **Recommandations** -Réviser les besoins des utilisateurs. Évaluer d'autres critères. - Analyser les résultats**	Mauvais niveau de lisibilité : Il existe des problèmes sur la satisfaction des utilisateurs et les couleurs de fond de l'interface. **Recommandations** -Des révisions sont nécessaires pour répondre aux règles ergonomiques de couleurs de fond. -Réviser les besoins des utilisateurs. -Évaluer d'autres critères. Analyser les résultats****	Mauvais niveau de lisibilité : Il existe des problèmes sur la satisfaction des utilisateurs et les couleurs d'écriture utilisées de l'interface. **Recommandations** -Des révisions sont nécessaires pour répondre aux règles ergonomiques de couleurs d'écriture. -Réviser les besoins des utilisateurs. - Évaluer d'autres critères. Analyser les résultats******	Très mauvais niveau de lisibilité : Il existe des problèmes dans les trois mesures **Recommandations** - Des révisions totales sont nécessaires pour améliorer les caractéristiques textuelles en respectant les recommandations ergonomiques. - Analyser les résultats*******

***** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus bas niveau de Taux de conformité des couleurs de fond recommandées.

** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus bas niveau de satisfaction

*** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a le plus bas niveau de Taux de conformité des couleurs d'écriture recommandées.

**** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a les plus bas niveaux de satisfaction et du Taux de conformité des couleurs de fond recommandées.

***** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a les plus bas niveaux du Taux de conformité des couleurs de fond recommandées et du Taux de conformité des couleurs d'écriture recommandées.

****** Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a les plus bas niveaux de satisfaction et du Taux de conformité des couleurs d'écriture recommandées.

******* Dans le cas de plusieurs IUs, commencez à améliorer l'interface utilisateur qui a les plus bas niveaux des trois mesures.

Tableau 3.13 (suite) : Spécification du modèle de mesure d'information de l'indicateur de conformité des couleurs des interfaces utilisateur

Mesures dérivées	**Taux de conformité des couleurs d'écriture recommandées :** mesure la valeur proportionnelle de la conformité des couleurs d'écriture recommandées d'une interface. **Taux de conformité des couleurs de fond recommandées :** mesure la valeur proportionnelle de la conformité des couleurs de fond recommandées d'une interface. **Pourcentage des utilisateurs ayant des réponses positives à la question par interface** (Albert et Tullis, 2013) : mesure la valeur proportionnelle des participants qui ont donné des réponses positives à la question. Cette mesure est déterminée sur la base de scores *Top Boxes* sélectionnés par l'expert (Top-Box, Top-2-Box, ou Top-3-Box).
Fonctions de mesures	Taux de conformité des couleurs d'écriture recommandées = Nombre de couleurs d'écriture qui respectent les couleurs recommandées / Nombre de couleurs d'écriture d'une interface. Taux de conformité des couleurs de fond recommandées = Nombre de couleurs de fond qui respectent les couleurs recommandées / Nombre de couleurs de fond d'une interface. Pourcentage des utilisateurs ayant des réponses positives à la question= Nombre des utilisateurs qui ont donné des réponses positives /Nombre total des utilisateurs.
Mesures de base	1. Nombre de couleurs d'écriture qui respectent les couleurs d'écriture recommandées 2. Nombre de couleurs d'écriture d'une interface (Padda, 2003 ; Seffah et al., 2006) 3. Nombre de couleurs de fond qui respectent les couleurs de fond recommandées 4. Nombre de couleurs de fond d'une interface (Padda, 2003 ; Seffah et al., 2006) 5. Nombre des utilisateurs qui ont donné des réponses positives 6. Nombre total des utilisateurs (ISO/IEC 25022, 2012)
Méthodes de mesures	1. Compter le nombre de couleurs d'écriture qui respectent les couleurs recommandées 2. Compter le nombre de couleurs d'écriture d'une interface 3. Compter le nombre de couleurs de fond qui respectent les couleurs de fond recommandées 4. Compter le nombre de couleurs de fond d'une interface 5. Compter le nombre d'utilisateurs qui ont choisi la/les réponse(s) positive(s). 6. Compter le nombre total des interrogés/utilisateurs. **Remarque :** Toutes les mesures sont déterminées de façon automatique en utilisant l'inspecteur des règles ergonomiques (version étendue) et l'outil de synthèse des résultats (*Cf. Chapitre 4*).

3.3.2.3 L'évaluation des actions minimales des tâches réalisées

Les actions minimales est un sous-critère de la charge de travail proposé par Scapin et Bastien (1997). Il concerne la charge de travail quant aux actions nécessaires à l'atteinte d'un but et à l'accomplissement d'une tâche (Bastien et Scapin, 1993). Pour évaluer ce critère, nous avons proposé un indicateur qui considère une paire de mesures représentées sous forme du nuage de points. Son objectif principal est de vérifier la brièveté par rapport aux actions requises effectuées par les utilisateurs lors de la réalisation de leurs tâches avec le système à évaluer. Nous avons commencé par la définition d'une mesure objective qu'on a nommé «*Pourcentage des participants réalisant les actions requises d'une tâche »*. Généralement, une tâche peut être effectuée selon une ou différentes façons/chemins en suivant un ensemble d'actions. Avant de commencer les tests d'utilisabilité par les utilisateurs, un utilisateur expert doit exécuter la tâche à travers le système à évaluer. Par conséquent, le mouchard électronique capture puis analyse toutes les actions effectuées. Par la suite, il génère un réseau de Pétri qui correspond aux chemins possibles effectués par l'utilisateur expert lors de l'exécution de la tâche. Celui-ci peut être généré sous forme d'un fichier XML qui est utilisé, ensuite par la version étendue du mouchard électronique pour déterminer le plus court chemin effectué par l'utilisateur expert. Ce chemin est considéré comme étant un chemin de référence pour déterminer si un utilisateur quelconque a réalisé les mêmes actions minimales effectuées par l'utilisateur expert lors de la réalisation de la tâche. En suivant le même principe pour tous les utilisateurs impliqués dans l'évaluation, notre outil réalisé SOUSEval (*Cf. Chapitre 4*) permet de déterminer le nombre des utilisateurs réalisant les mêmes actions minimales requises de la tâche que celles effectuées par l'utilisateur expert (chemin de référence). Et par conséquent, calculer leur pourcentage.

Concernant la mesure subjective, elle porte sur le "*Pourcentage des utilisateurs ayant des réponses positives à la question* ». Le Tableau 3.14 fournit une spécification du modèle de mesure de l'indicateur proposé pour l'évaluation de ce critère.

Tableau 3. 14 : Spécification du modèle de mesure d'information de l'indicateur d'actions minimales des tâches réalisées

Besoin d'information

Évaluer la brièveté par rapport aux actions minimales lors de la réalisation d'une procédure ou une tâche d'un système interactif

Indicateur

Actions minimales des tâches réalisées.

Modèle d'analyse

Le modèle est énoncé comme une paire de mesures (**Pourcentage des utilisateurs ayant des réponses positives** et **Pourcentage des participants réalisant les actions requises d'une tâche)** représenté par un nuage de points.

Critères de décision

Interprétation de chaque mesure

- Pour la mesure de *Pourcentage des participants réalisant les actions requises d'une tâche :* Une valeur élevée indique qu'une forte proportion des utilisateurs a réalisé les actions minimales requises de la tâche que celles effectuées par l'utilisateur expert.
  0≤*Pourcentage des participants réalisant les actions requises d'une tâche*≤100%
  Les résultats sont divisés en deux groupes (intervalles) en fonction du seuil X spécifié par l'expert :
  - *Non acceptable* : Pour les valeurs ϵ [0% ; X[
  - *Acceptable* : Pour les valeurs ϵ [X ; 100%]

- Pour la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question :* une valeur proportionnelle plus élevée est mieux (valeur maximale = 100%)
  0≤ *Pourcentage des utilisateurs ayant des réponses positives à la question par interface*≤100%
  Les réponses positives d'une question dépendent des meilleures réponses adoptées en utilisant les scores Top-Boxes (Top-Box, Top-2-Box, ou Top-3-Box). Ces dernières sont interprétées en fonction de l'échelle du questionnaire (par exemple, Likert) et la tonalité de la question.
  Exemple d'une question portant sur les actions minimales :
  ―Il faut passer par trop d'étapes pour faire marcher quelque chose " (There are too many steps required to get something to work) (une question extraite du questionnaire SUMI)
  Cette question se base sur une échelle dichotomique comportant trois options (d'accord, ne sait pas, pas d'accord) dont sa tonalité est négative. Par conséquent, la réponse positive dans ce cas est "Désaccord". De ce fait, on n'applique pas la technique des scores top boxes.

  Deux intervalles sont définis en fonction du seuil Y spécifié par l'expert :
  - *Acceptable :* Pour les valeurs ϵ [Y ; 100%]
  - *Non acceptable* : Pour les valeurs ϵ [0% ; Y[

Remarque : L'expert précise les seuils X et Y en tenant en compte du contexte d'utilisation de l'évaluation

Critères de décision de l'indicateur

		Pourcentage des participants réalisant les actions requises d'une tâche	
		[0% ; X[: *Non acceptable*	[X ; 100%] : *Acceptable*
Pourcentage des utilisateurs ayant des réponses positives	[Y ; 100%] *Acceptable*	Mauvais niveau de qualité de tâche : Il y a des problèmes des actions minimales pour réaliser la tâche. **Recommandations :** - Des révisions sont nécessaires pour améliorer les résultats. - Analyser les résultats*	Bon niveau de qualité : Pas de problèmes
	[0% ; Y[*Non acceptable*	Très mauvais niveau de qualité de la tâche : Il y a des problèmes sur les deux mesures. **Recommandations :** - Reconcevoir les interfaces utilisateur du système évalué en minimisant les efforts des utilisateurs lors de l'exécution des tâches. - Analyser les résultats**	Mauvais niveau de qualité : Il y a des problèmes de satisfaction des utilisateurs. **Recommandations :** - Des révisions sont nécessaires pour satisfaire les besoins des utilisateurs. Réduire le nombre d'actions minimales requises pour la tâche. - Analyser les résultats***

*Dans le cas de plusieurs tâches, commencez par la modélisation de la tâche qui a un *Pourcentage des participants réalisant les actions requises* le plus petit par rapport aux autres.

** Dans le cas de plusieurs tâches, commencez par la modélisation de la tâche qui a un le plus bas niveau de satisfaction par rapport aux autres.

*** Dans le cas de plusieurs tâches, commencez par la modélisation de la tâche qui a les plus bas niveaux de deux mesures.

Tableau 3.14 (suite) : Spécification du modèle de mesure d'information de l'indicateur d'actions minimales des tâches réalisées

Mesures dérivées	**Nombre des utilisateurs réalisant les actions minimales requises de la tâche :** cette mesure détermine le nombre des utilisateurs réalisant les mêmes actions minimales requises de la tâche (c.-à-d., le plus court chemin défini à partir de l'exécution du système par l'utilisateur expert) **Pourcentage des participants réalisant les actions requises d'une tâche :** Pour chaque tâche, cette mesure détermine le pourcentage des utilisateurs réalisant les mêmes actions minimales requises de la tâche (le plus court chemin) que celles effectuées par l'utilisateur expert. **Pourcentage des utilisateurs ayant des réponses positives à la question par interface** (Albert et Tullis, 2013) : mesure la valeur proportionnelle des participants qui ont donné des réponses positives à la question.
Fonctions de mesures	Nombre des utilisateurs réalisant les actions requises de la tâche : le calcul de cette mesure est effectué via l'outil de synthèse des résultats en utilisant un algorithme permettant, de comparer pour une tâche donnée, pour chaque utilisateur, si l'ensemble des actions qu'il a réalisé sont les mêmes que celles identifiées dans le plus court chemin (référence). Par conséquent, il détermine le nombre des utilisateurs réalisant les actions requises. Pourcentage des participants réalisant les actions requises d'une tâche =Nombre d'utilisateurs réalisant les actions minimales requises de la tâche / Nombre total des utilisateurs. Pourcentage des utilisateurs ayant des réponses positives à la question= Nombre des utilisateurs qui ont donné des réponses positives /Nombre total des utilisateurs.
Mesures de base	1. Nombre des actions utilisées pour réaliser une tâche (Dix et al., 2004 ; Padda, 2003) 2. Nombre des utilisateurs qui ont donné des réponses positives 3. Nombre total des utilisateurs (ISO/IEC 25022, 2012)
Méthodes de mesures	1. Déterminer via la version étendue du mouchard électronique EISEval. 2. Compter le nombre de utilisateurs qui ont choisi la/les réponse(s) positive(s). 3. Compter le nombre total des utilisateurs **Remarque :** Toutes les mesures sont déterminées de façon automatique en utilisant la version étendue du mouchard électronique EISEval et l'outil de synthèse des résultats (*Cf. Chapitre 4).*

3.3.2.4 L'évaluation du feedback immédiat aux actions des utilisateurs sur les IHM à évaluer

Le feedback immédiat est un sous-critère de guidage proposé par Scapin et Bastien (1997). Il concerne les réponses de l'ordinateur consécutives aux actions des utilisateurs (Bastien et Scapin, 1993). L'indicateur proposé du Feedback immédiat aux actions des utilisateurs vise principalement à évaluer la capacité du système à réagir à chacune des actions de l'utilisateur en lui donnant un retour ainsi que leur satisfaction par rapport au feedback reçu.

Il se compose d'une paire de mesures représentées sous forme du nuage de points. La mesure subjective concerne la *Réponse de l'utilisateur à la question*. Pour l'évaluation objective, nous avons défini une mesure qu'on a appelée « *Feedback aux actions d'un utilisateur »*. Cette mesure consiste à déterminer la valeur proportionnelle de rétroactions du système par rapport aux actions effectuées par un utilisateur lors de la réalisation de leur tâches. Celle-ci permet de vérifier l'absence ou la présence de Feedback par rapport à chacune des actions de l'utilisateur.

Pour déterminer cette mesure, nous nous sommes basés sur un ensemble des événements pouvant être capturés par le mouchard électronique EISEval lors de l'exécution des tâches par l'utilisateur. Nous distinguons, d'une part, les évènements interfaces utilisateur (EVIU) généralement déclenchés par l'utilisateur et qui peuvent se produire sur l'interface utilisateur (EVIU) (par exemples, *ouverture d'une fenêtre d'une interface*, *click sur un bouton Annuler*) (Tran, 2009). D'autre part, les événements déclenchés par le système (appelés services) qui se produisent suite à un événement(s) déclencheur(s). Ces services sont considérés comme des actions externes, (par exemple, *le clic sur un bouton Annuler* d'une fenêtre peut déclencher *la fermeture de cette fenêtre*) (Tran, 2009) ou comme étant des actions internes. Par exemple, dans le cas d'un système interactif de supervision du transport public terrestre, lorsqu'une perturbation se passe sur le trafic, le service d'avertissement de perturbation d'un agent de ce système va déclencher une action interne concernant *l'affichage d'une fenêtre d'avertissement* destinée à avertir le régulateur du trafic et lui demander de traiter cette perturbation (Tran, 2009). Un autre type de service peut être déclenché par des actions internes permettant d'assurer la

communication entre les agents application et contrôle du système sans apparaître sur l'interface utilisateur.

Etant donné que selon Scapin et Bastien (1997), le système doit réagir à chacune des actions de l'utilisateur, nous avons proposé dans un premier lieu de calculer le nombre des actions d'un utilisateur ayant au moins une rétroaction. Celui-ci est déterminé via la version étendue du mouchard électronique en vérifiant si chaque événement (action) déclenché par l'utilisateur, lors de la réalisation de leur tâche, produit une action externe sur l'interface utilisateur. Par conséquent, nous avons défini la mesure du *Feedback aux actions d'un utilisateur* en divisant le *Nombre d'actions d'un utilisateur ayant au moins un feedback* par la mesure du *Nombre total d'actions d'un utilisateur lors de la réalisation des tâches*. Ces deux mesures sont définies en l'adaptant de la mesure du *Nombre des actions utilisées pour réaliser une tâche (*Dix et al., 2004 ; Padda, 2003).

Notons que dans cet indicateur, nous n'avons pas évalué la rapidité du feedback (le temps de réponse du système) d'une façon objective. Toutefois, nous avons considéré cet aspect d'une façon subjective. Dans le Tableau 3.15, nous présentons la spécification du modèle de mesure de l'indicateur proposé.

Tableau 3. 15 : Spécification du modèle de mesure d'information de l'indicateur de feedback immédiat aux actions des utilisateurs

<table>
<tr><td>Besoin d'information</td><td colspan="3">Évaluer le feedback immédiat aux actions des utilisateurs sur les interfaces utilisateur</td></tr>
<tr><td>Indicateur</td><td colspan="3">Feedback immédiat aux actions des utilisateurs</td></tr>
<tr><td>Modèle d'analyse</td><td colspan="3">Le modèle est énoncé comme une paire de mesures (Réponse à la question et Feedback aux actions d'un utilisateur) représenté par un nuage de points.</td></tr>
<tr><td>Critères de décision</td><td colspan="3"><u>Interprétation de chaque mesure</u>
• Pour la mesure du Feedback aux actions d'un utilisateur : elle vérifie la présence ou l'absence de Feedback par rapport à chaque action de l'utilisateur.
Une valeur proportionnelle égale à 100% indique que le système a réagi à chacune des actions de l'utilisateur.
En général, 0%≤ Feedback aux actions d'un utilisateur ≤100%.
Les résultats sont divisés en deux groupes (intervalles) :
- Acceptable : Pour les valeurs = 100%
- Non acceptable : Pour les valeurs < 100%
• La mesure de Réponse à la question est interprétée en fonction de l'échelle du questionnaire adoptée (par exemple, Likert) et la tonalité de la question.
Exemple d'une question portant sur le feedback immédiat :
"La rétroaction du système" (System feedback) (une question extraite du questionnaire QUIS)
Cette question se base sur une échelle de Likert de 10 points allant de 0 : Faible jusqu'à le point 9 : Bonne, dont sa tonalité est positive.
Par conséquent, la plus grande valeur est mieux (vu que la valeur 9 est bonne).
Deux groupes sont définis en fonction du seuil Y spécifié par l'expert :
- Désaccord : Pour les valeurs ϵ {1 à Y}
- Accord : Pour les valeurs ϵ à {la valeur suivante après Y à 9}.
Remarque : L'expert précise le seuil Y en tenant en compte du contexte d'utilisation de l'évaluation.
<u>Critères de décision de l'indicateur</u>
<table>
<tr><td colspan="2"></td><td>Feedback aux actions d'un utilisateur</td><td></td></tr>
<tr><td colspan="2"></td><td><100% : Non acceptable</td><td>=100% : Acceptable</td></tr>
<tr><td rowspan="2">Réponse à la question</td><td>{la valeur suivant Y à 9}

Accord</td><td>Mauvais niveau de feedback immédiat :
Il y a des problèmes sur les feedbacks aux actions de l'utilisateur.
Recommandations :
- Des révisions sont nécessaires pour améliorer les résultats.
- Analyse détaillée des résultats*</td><td>Bon niveau de feedback immédiat :
Pas de problèmes</td></tr>
<tr><td>{1 à Y}
Désaccord</td><td>Très mauvais niveau de feedback immédiat
Il y a des problèmes sur les deux mesures.
Recommandations :
- Reconcevoir les interfaces utilisateur du système évalué en s'assurant des feedbacks sur les actions des utilisateurs.
- Analyser les résultats détaillés*</td><td>Mauvais niveau de feedback immédiat
Il y a des problèmes de satisfaction des utilisateurs.
Recommandations :
- Des révisions sont nécessaires pour satisfaire les besoins des utilisateurs.
- Améliorer la qualité de feedback.
- Analyser les résultats détaillés*</td></tr>
</table></td></tr>
</table>

Tableau 3.15 (suite) : Spécification du modèle de mesure d'information de l'indicateur de feedback immédiat aux actions des utilisateurs

Critères de décision	*L'analyse détaillée des résultats est représentée dans un graphique à barres. Elle fournit la mesure du *Feedback immédiat programmé* par rapport à chaque interface. Celle-ci est capturée à travers la version étendue de l'inspecteur des règles ergonomiques et elle est égale au *Nombre de fonctions de Feedback dans une interface* divisé par le *Nombre de fonctions disponibles à l'utilisateur dans une interface* (Padda, 2003). Par conséquent, ce graphique permet l'aide à la détection des interfaces qui ont des valeurs faibles et priorise celles qui doivent être améliorées en premier.
Mesures dérivées	**Feedback aux actions d'un utilisateur :** mesure la valeur proportionnelle de rétroactions du système par rapport aux actions d'un utilisateur. Celle-ci permet de vérifier l'absence ou la présence de Feedback par rapport à chacune des actions d'utilisateur. Cette mesure est adaptée des mesures définies par (Dix et al., 2004 ; Padda, 2003).
Fonctions de mesures	Feedback aux actions d'un utilisateur = Nombre d'actions ayant au moins un feedback pour un utilisateur / Nombre d'actions d'un utilisateur lors de la réalisation des tâches.
Mesures de base	1. Nombre d'actions d'un utilisateur ayant au moins un feedback (adaptée de (Dix et al., 2004 ; Padda, 2003)) 2. Nombre total d'actions d'un utilisateur lors de la réalisation des tâches (adaptée de (Dix et al., 2004 ; Padda, 2003)) 3. **Réponse à la question** (ISO/IEC 25022, 2012).
Méthodes de mesures	1. Cette mesure est déterminée de façon automatique via la version étendue du mouchard électronique EISEval. 2. Cette mesure est déterminée de façon automatique via l'outil de synthèse des résultats. 3. Demander aux utilisateurs leurs perceptions sur la qualité de feedback lors de la réalisation de leurs tâches. **Remarque :** Toutes les mesures sont déterminées d'une façon automatique en utilisant le mouchard électronique EISEval et l'outil de synthèse des résultats (*Cf. Chapitre 4*).

3.4 Conclusion

La construction complète de l'environnement intégré d'aide à l'évaluation des systèmes interactifs a requis d'abord la définition d'une approche permettant d'assurer l'intégration des résultats hétérogènes issus de différents outils. Lors de ce chapitre, nous avons présenté notre approche d'intégration des résultats d'évaluation subjective et objective d'une façon complémentaire et cohérente. Nous avons commencé par la proposition d'un modèle de *Mapping* dans le but d'établir la correspondance entre les données d'évaluation subjective et objective à des critères d'utilisabilité (par exemple, pour le critère d'efficience, nous avons associé l'opinion de l'utilisateur sur sa perception de la rapidité de la réalisation correcte d'une tâche par rapport au temps d'exécution collecté par le mouchard électronique). Ceci corrobore l'idée avancée par Farenc et al. (1995) qui porte sur la mise en correspondance des facteurs d'évaluation (dans leur cas des critères ergonomiques) aux critères, plus génériques, d'utilisabilité et d'utilité. Par la suite, en vue d'assurer l'intégration effective entre elles, nous nous sommes basés sur les différents concepts définis par la norme ISO/IEC 15939 pour définir un ensemble d'indicateurs d'utilisabilité. Nous avons introduit les étapes nécessaires pour spécifier ces indicateurs. En plus d'assurer l'intégration, ceux-ci servent notamment à détecter des problèmes précis d'utilisabilité, interpréter les résultats fournis et suggérer des recommandations pour améliorer la qualité des systèmes interactifs.

Afin d'achever le développement de l'environnement, nous aborderons, dans le chapitre suivant, la mise en œuvre de l'approche proposée par la présentation de l'outil de synthèse des résultats d'évaluation ainsi les développements réalisés.

Chapitre 4 : SOUSEval[18] : un outil pour l'intégration d'évaluations objective et subjective de l'utilisabilité des IHM

"… Automation is therefore a promising way to augment existing approaches"
(MELODY Y. IVORY & MARTI A. HEARST, 2001)

Introduction

Au cours du chapitre précédent, nous avons présenté une approche d'évaluation de l'utilisabilité des interfaces utilisateur des systèmes interactifs. Elle assure notamment l'intégration effective des résultats d'évaluation subjective et objective via l'utilisation des indicateurs d'utilisabilité proposés sur la base de la norme ISO/IEC 15939. Nous avons détaillé les étapes nécessaires à suivre pour la spécification d'un ensemble des indicateurs portant sur l'évaluation des différents critères de l'utilisabilité de l'IHM d'un système interactif.

Ce chapitre présente l'ensemble des développements réalisés pour mettre en pratique l'environnement intégré d'aide à l'évaluation des systèmes interactifs proposé par (Ezzedine, 2002). Il contribue à la mise en œuvre des indicateurs d'utilisabilité par le développement de l'outil de synthèse des résultats d'évaluation qui va inclure également un outil d'aide à la génération de questionnaires. Cet outil est

[18] SOUSEval : Subjective Objective USability Evaluation

baptisé **SOUSEval** : **S**ubjective **O**bjective **US**ability **Eval**uation. Ce chapitre est structuré suivant six sections. Nous commencerons par la présentation de l'environnement intégré ainsi que le positionnement de notre travail réalisé. Ensuite, nous décrirons, dans les quatre sections suivantes l'ensemble des développements effectués permettant de mettre en exécution l'environnement intégré. Finalement, nous conclurons ce chapitre par une synthèse portant sur le travail mené.

4.1 Environnement intégré d'aide à l'évaluation des systèmes interactifs

Comme l'illustre la Figure 4.1, nous rappelons que l'environnement intégré d'aide à l'évaluation des systèmes interactifs est composé de quatre outils d'évaluation dont deux existent et deux qui manquent. En effet, le premier outil existant concerne un mouchard électronique EISEval (Tran, 2009) qui porte sur l'évaluation de l'interaction entre les utilisateurs et le système interactif à évaluer. Le second outil existant présente un inspecteur des règles ergonomiques (Charfi, 2013) visant à assurer une évaluation ergonomique de l'utilisabilité des interfaces utilisateur.

Concernant les outils qui manquent, le premier présente un outil d'aide à la génération de questionnaires dont l'objectif est d'assurer une évaluation subjective basée sur la prise en compte des opinions et perceptions des utilisateurs vis-à-vis du système interactif à évaluer. Tandis que le second outil concerne la synthèse des résultats d'évaluation qui vise à intégrer les résultats de trois outils d'évaluation mentionnés afin de fournir une bonne mesure de l'utilisabilité d'un système interactif donné.

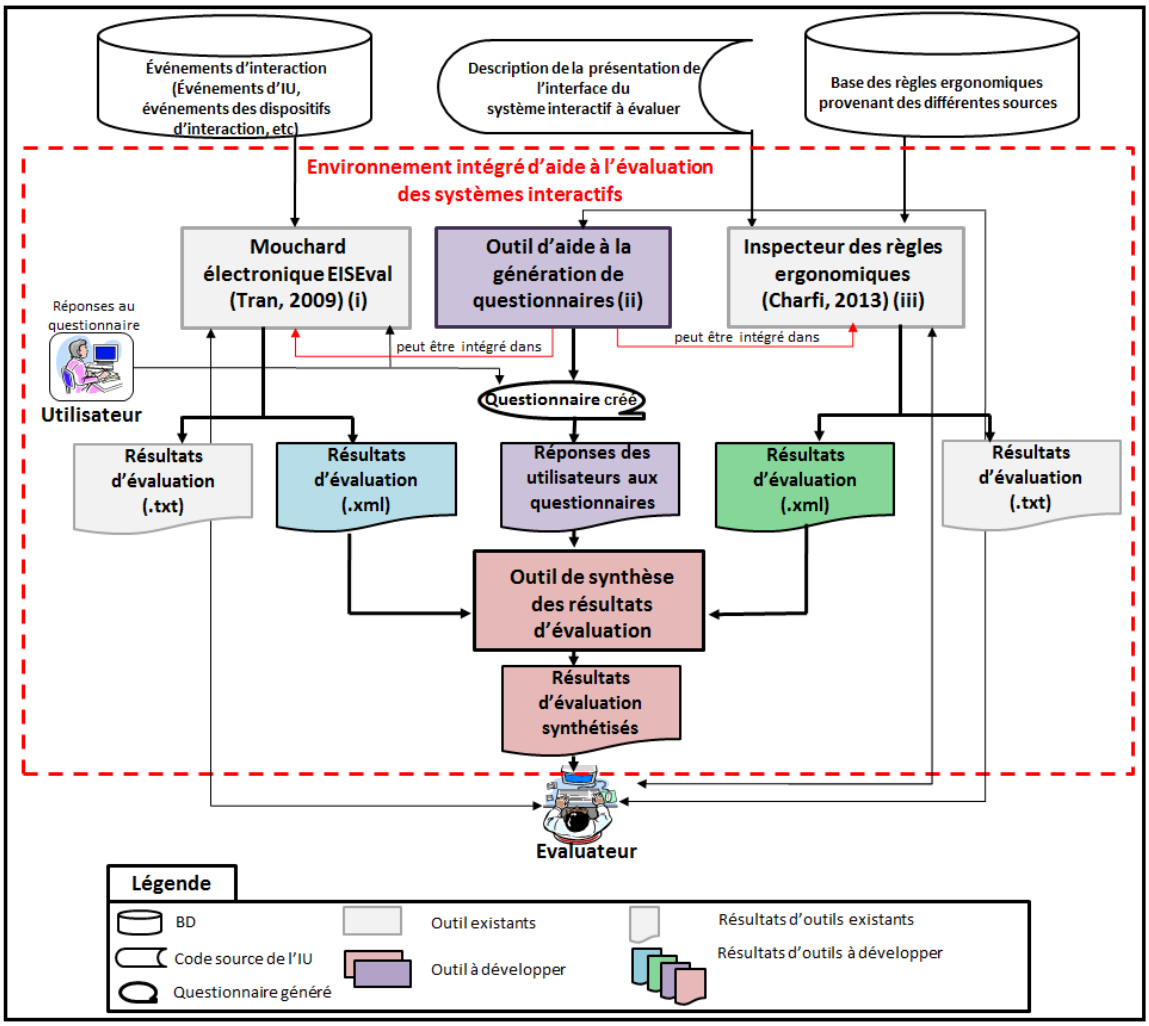


Figure 4. 1 : Environnement intégré d'aide à l'évaluation des systèmes interactifs

Nous rappelons que notre objectif principal dans le cadre de cette thèse consiste à achever le développement complet de l'environnement intégré à travers la réalisation d'outils manquants. Afin de mettre en œuvre d'une façon complète l'environnement intégré, le développement de l'outil de synthèse des résultats d'évaluation « SOUSEval » est réalisé à travers la mise en place d'indicateurs d'utilisabilité proposés lors du Chapitre 3 (voir Section 3.3). Ceux-ci contribuent à assurer l'intégration effective et complémentaire des résultats d'évaluation objective provenant de deux outils existants (le mouchard électronique et l'inspecteur des règles ergonomiques) avec ceux obtenus de l'évaluation subjective qui sont obtenus grâce à l'outil d'aide à la génération de questionnaires.

Compte tenu de notre objectif d'intégration des résultats de trois outils mentionnés précédemment, nous avons décidé de mettre en place l'outil d'aide à la génération de questionnaires directement dans l'outil de synthèse des résultats d'évaluation « SOUSEval ». De cette façon, celui-ci assure une double fonction. La première comme étant un outil d'évaluation subjective d'aide à la génération de questionnaires permettant de créer, gérer, générer et analyser des questionnaires d'utilisabilité. La seconde fonction porte sur l'évaluation intégrée de l'utilisabilité des systèmes interactifs sur la base de l'intégration des résultats de trois outils d'évaluation subjective et objective définis dans l'environnement.

Comme nous l'avons mentionné dans le Chapitre 3 (Section 3.1), le Tableau 4.1 introduit l'ensemble des mesures d'utilisabilité qui sont générées à partir de deux outils : Mouchard électronique EISEval et Inspecteur des règles ergonomiques.

Tableau 4. 1 : Les mesures d'utilisabilité capturées des outils d'évaluation existants

	Outils existants	
	Mouchard électronique EISEval *(Tran, 2009)*	**Inspecteur des règles ergonomiques** *(Charfi, 2013)*
Mesures de base capturées	Nombre de tâches réalisées Nombre total de tâches Nombre de tâches déclenchées par le système Nombre de tâches déclenchées par l'utilisateur Temps moyen pris pour réaliser une tâche Nombre de tâches réalisées avec succès Nombre de tâches réalisées avec échec Nombre de tâches dont le temps pris est supérieur au temps d'expert prévu Nombre de tâches dont le temps pris est inférieur au temps d'expert prévu Nombre de tâches dont le temps pris est égal au temps d'expert prévu Nombre de tâches dont le temps pris est supérieur au temps moyen prévu Nombre de tâches dont le temps pris est inférieur au temps moyen prévu Nombre de tâches dont le temps pris est égal au temps moyen prévu Temps début et fin de l'exécution d'une tâche	Nombre de tailles de police Nombre de types de police Nombre de couleurs de fond Nombre de couleurs d'écriture Dimensions des composants (largeur, hauteur) Dimensions de l'image Dimensions de l'écran Nombre total de champ de texte Nombre de champ de texte avec des valeurs par défaut Longueur du texte Nombre d'éléments à l'écran Nombre de tailles de police qui respectent les tailles de police recommandées Nombre de couleurs de police qui respectent les couleurs de police recommandées Nombre de types de police qui respectent les types de police recommandés Nombre de couleurs de fond qui respectent les couleurs de fond recommandées

Comme présenté dans le Tableau 4.1 et en considérant l'ensemble des mesures d'utilisabilité associées à tous les indicateurs proposés (*Cf. Chapitre 3*), nous avons distingué que les outils existants ne permettent pas de capturer toutes les mesures nécessaires. A cette fin, nous avons décidé d'abord d'étendre la version actuelle du mouchard électronique EISEval ainsi celle de l'inspecteur des règles ergonomiques comme l'illustre la Figure 4.2.

Dans ces deux versions étendues d'outils, nous avons ajouté la capture de l'ensemble de mesures manquantes requises lors du développement des indicateurs d'utilisabilité. De plus nous avons fourni la possibilité de générer les résultats de ces mesures sous forme des fichiers (*.XML*) pour rendre possible leurs exploitations via l'outil de synthèse des résultats d'évaluation.

En outre, nous avons implémenté la collecte de certaines mesures d'utilisabilité d'une façon directe dans notre outil réalisé SOUSEval. Celles-ci sont dérivées à partir des mesures extraites des versions étendues d'outils existants.

Le tableau 4.2 représente l'ensemble des mesures d'utilisabilité mises en place dans les outils existants et les nouvelles implémentations (les versions étendues d'outils existants ainsi que notre outil SOUSEval réalisé), réparties sur l'ensemble des critères considérés décrivant les indicateurs.

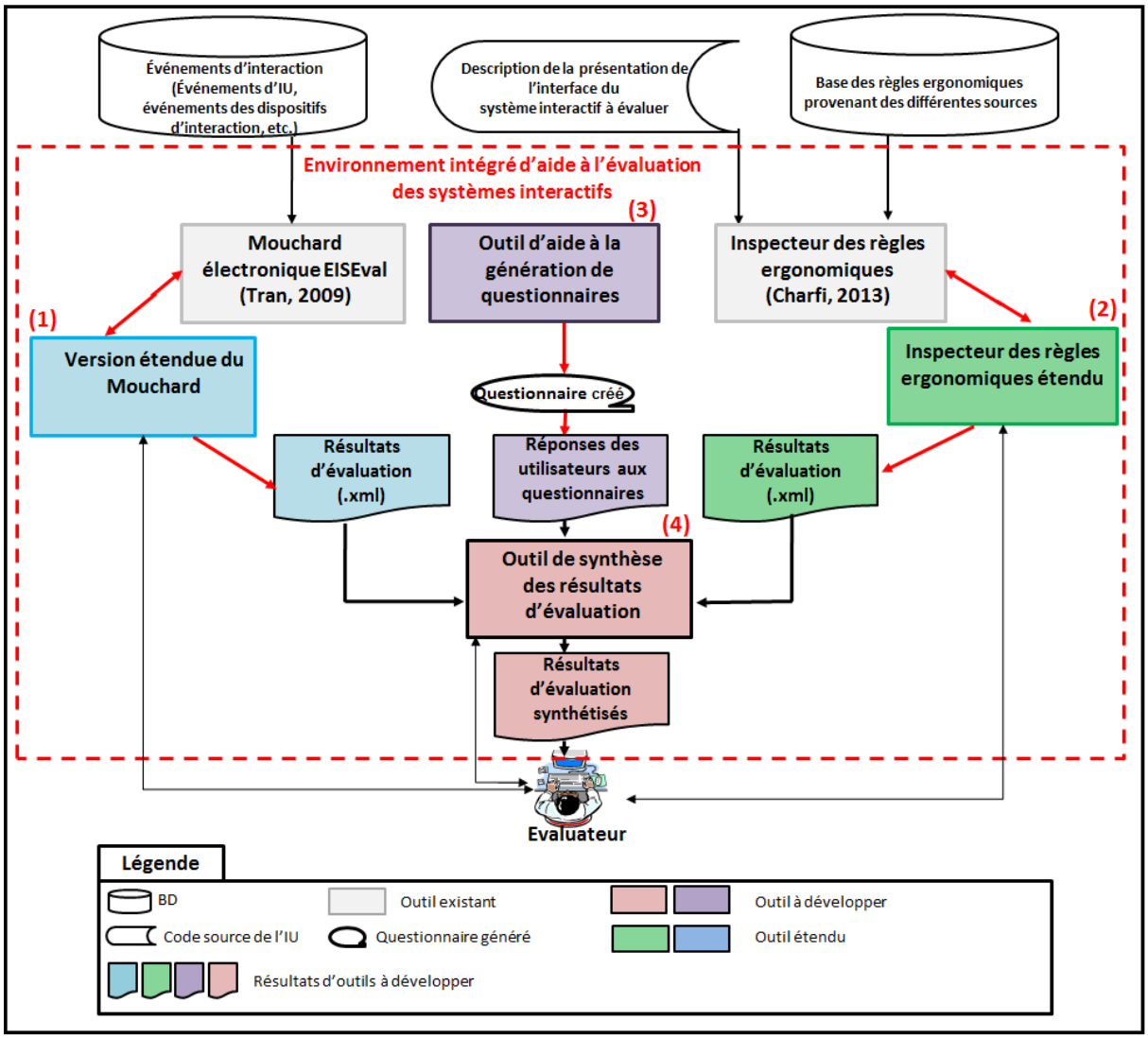


Figure 4. 2 : Environnement intégré d'aide à l'évaluation des systèmes interactifs basé sur des versions étendues d'outils existants

Tableau 4. 2 : Les mesures d'utilisabilité mises en place dans les outils existants et les nouvelles implémentations

Critères d'utilisabilité	Mesures d'utilisabilité mises en place	Outils existants		Nouvelles implémentations
		Mouchard électronique EISEval	Inspecteur des règles ergonomiques	
Efficacité	Nombre de tâches réalisées	X		
	Nombre total de tâches	X		
	Taux de réalisation correcte d'une tâche			Version étendue du mouchard électronique
	Taux moyen de réalisation correcte des tâches			Outil SOUSEval
Efficience	Heure du début de l'exécution d'une tâche	X		
	Heure de fin de l'exécution d'une tâche	X		
	Temps d'exécution d'une tâche			Version étendue du mouchard électronique
	Temps total d'exécution des tâches	X		
	Temps moyen d'exécution des tâches			Outil SOUSEval
	Taux d'efficience d'une tâche			Outil SOUSEval
	Taux moyen d'efficience des tâches			Outil SOUSEval
Charge de travail *Densité informationnelle*	Densité globale d'une interface			Version étendue de l'inspecteur des règles ergonomiques
Charge de travail *Actions minimales*	Nombre des actions pour réaliser une tâche			Version étendue du mouchard électronique
	Nombre d'utilisateurs réalisant les actions minimales requises d'une tâche			Outil SOUSEval
	Nombre total des utilisateurs			Outil SOUSEval
	Pourcentage des participants réalisant les actions requises d'une tâche			Outil SOUSEval
Guidage *Feedback immédiat*	Nombre de fonctions de Feedback dans une interface			Version étendue de l'inspecteur des règles ergonomiques
	Nombre des actions d'un utilisateur ayant au moins un feedback			Version étendue du mouchard électronique
	Nombre total des actions d'un utilisateur lors de la rélisation des tâches			Outil SOUSEval
	Nombre de fonctions disponibles à l'utilisateur dans une interface			Version étendue de l'inspecteur des règles ergonomiques
	Feedback aux actions d'un utilisateur			Outil SOUSEval

Tableau 4.2 (suite) : Les mesures d'utilisabilité mises en place dans les outils existants et les nouvelles implémentations

Critères d'utilisabilité	Mesures d'utilisabilité mises en place	Outils existants		Nouvelles implémentations
		Mouchard électronique EISEval	Inspecteur des règles ergonomiques	
Guidage : *Lisibilité*	Nombre de tailles de police qui respectent les tailles de police recommandées		X	
	Nombre de couleurs de police qui respectent les couleurs de police recommandées		X	
	Nombre de types de police qui respectent les types de police recommandés		X	
	Nombre de couleurs de fond qui respectent les couleurs de fond recommandées		X	
	Nombre de tailles de police		X	
	Nombre de types de police		X	
	Nombre de couleurs de fond		X	
	Nombre de couleurs d'écriture		X	
	Taux de conformité des couleurs d'écriture recommandées			Outil SOUSEval
	Taux de conformité des couleurs de fond recommandées			Outil SOUSEval
	Taux de conformité des types de police recommandés			Outil SOUSEval
	Taux de conformité des tailles de police recommandées			Outil SOUSEval

En plus de ces mesures représentées dans le Tableau 4.2, notre outil SOUSEval permet de générer un ensemble des mesures d'utilisabilité portant sur l'évaluation subjective de l'utilisabilité des systèmes interactifs en utilisant les questionnaires. Nous allons les présenter lors de deux Section 4.4 et 4.5.

Dans ce qui suit, nous présenterons l'ensemble des développements réalisés et mis en œuvre dans l'environnement intégré.

4.2 La version étendue du Mouchard électronique EISEval

Le Mouchard électronique EISEval présente le premier outil d'évaluation qui a été réalisé par (Tran, 2009) dans le cadre du développement de l'environnement intégré. Celui-ci permet d'assurer une évaluation objective des systèmes interactifs dédiés à la supervision du trafic terrestre (Tramway et Bus) en impliquant des utilisateurs en temps réel. En effet, la spécificité de ce mouchard consiste à capturer et analyser toutes les actions effectuées par les utilisateurs ainsi celles produites par le système à évaluer lors de l'exécution d'un ensemble des tâches planifiées par un évaluateur.

Ce mouchard comporte sept modules qui communiquent entre eux, définis comme suit :

- *Module 1* - Capture des événements aux niveaux base et intermédiaire
- *Module 2* - Association des événements au niveau intermédiaire (services, EVIUs) aux tâches correspondantes
- *Module 3* - Analyse des données capturées et des tâches
- *Modules 4 et 5* - Génération et confrontation des réseaux de Pétri (RdPs)
- *Module 6* - Critique de système et proposition au concepteur des suggestions d'amélioration
- *Module 7* - Configuration de l'environnement pour évaluer différents systèmes interactifs

Pour plus de détails sur ces modules, le lecteur peut se référer aux travaux de (Tran, 2009 ; Tran et al., 2008).

Actuellement, ce mouchard génère l'ensemble des résultats, mentionnés dans le Tableau 4.1, portant sur les tâches effectuées après une phase d'analyse des tâches effectuée par l'évaluateur. Il représente ainsi les résultats sous forme graphique et fournit la possibilité de les enregistrer sous forme textuelle (.txt).

De plus, le mouchard EISEval permet à travers ses deux modules 4 et 5 de générer les RdPs (Réseau de Pétri) pour chaque tâche effectuée et de l'enregistrer sous forme de fichier (.pnml/.xml) qui contient notamment la description détaillée sur la réalisation de la tâche d'un utilisateur.

Vu notre besoin d'extraire quelques autres mesures d'utilisabilité (mentionnées dans le Tableau 4.2), nous avons étendu la version du mouchard par la mise en œuvre d'un module supplémentaire (Module 8) comme l'illustre la Figure 4.3.

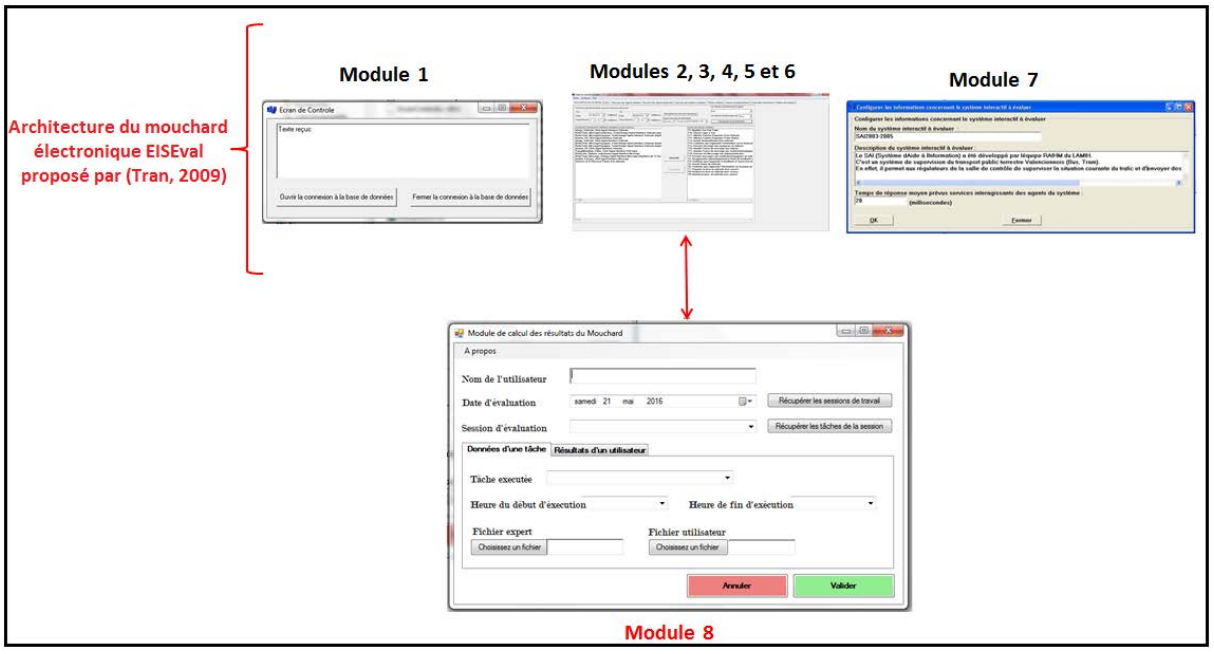


Figure 4. 3 : Architecture du mouchard électronique EISEval étendu

La Figure 4.4 illustre un aperçu de l'interface du module supplémentaire réalisé. Le module ajouté permet de récupérer la session de travail relative à un utilisateur donné à partir de la base de données du mouchard. Par la suite, il récupère l'ensemble des tâches réalisées ainsi que l'ensemble des données relatives à chaque tâche (l'heure du début d'exécution, l'heure de fin d'exécution). Ensuite, l'évaluateur doit charger les deux fichiers contenant les descriptions détaillées sur la réalisation de la

tâche, générées à partir de deux modules 4 et 5 du mouchard. Le premier fichier est celui de l'expert et le second est celui de l'utilisateur.

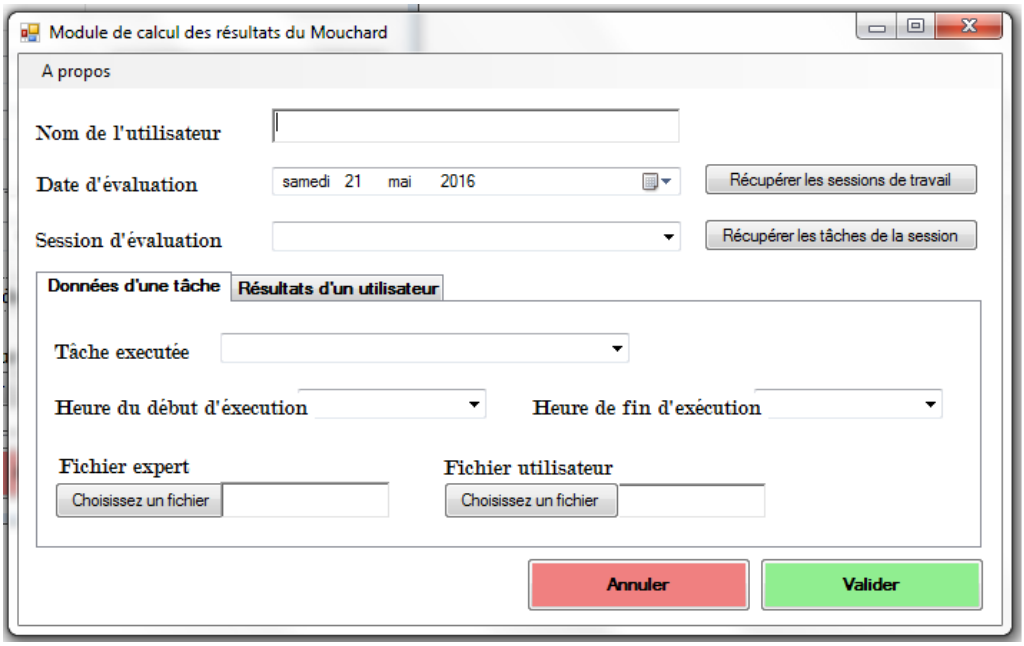


Figure 4. 4 : Aperçu de l'interface du module supplémentaire ajouté dans le mouchard électronique EISEval

A la fin de l'évaluation, ce module génère les mesures suivantes : le taux de réalisation correcte de la tâche, le temps d'exécution d'une tâche (en seconde), le nombre des actions pour réaliser une tâche et le nombre des actions ayant au moins un feedback. Par conséquent, il permet de générer ces résultats sous format d'un fichier XML comme l'illustre la Figure 4.5.

```xml
<?xml version="1.0" encoding="utf-8"?>
<Tâche>
   <Taux_de_réalisation_correcte>65</Taux_de_réalisation_correcte>
   <Temps_exécution>170.716</Temps_exécution>
   <Nombre_des_actions>5</Nombre_des_actions>
   <Nombre_des_actions_feedback>5</Nombre_des_actions_feedback>
</Tâche>
```

Figure 4. 5 : Exemple des résultats du fichier XML fourni par le module 8

4.3 La version étendue de l'inspecteur des règles ergonomiques

L'inspecteur des règles ergonomiques constitue le deuxième outil d'évaluation qui a été conçu par (Charfi, 2013) dans le cadre du développement de l'environnement intégré. Il permet d'assurer une évaluation objective de la qualité ergonomique des systèmes interactifs de type WIMP afin de détecter les problèmes ergonomiques des interfaces utilisateur.

En outre, cet outil vise principalement à inspecter les incohérences ergonomiques des interfaces utilisateur à travers la validation de leurs conformités par rapport à un ensemble des règles ergonomiques. Celles-ci sont spécifiées par un évaluateur et constituent un modèle de référence permettant d'assurer une bonne qualité de l'interface utilisateur.

L'inspecteur des règles ergonomiques dispose d'un parseur de code source des interfaces utilisateur du système interactif à évaluer. Il permet de capturer les valeurs des attributs des composants graphiques

ainsi que l'ensemble des mesures d'utilisabilité mentionnées dans le Tableau 4.1. Certaines mesures (le nombre de tailles de police qui respectent les tailles de police recommandées, le nombre de couleurs de police qui respectent les couleurs de police recommandées, le nombre de types de police qui respectent les types de police recommandés et le nombre de couleurs de fond qui respectent les couleurs de fond recommandées) sont déterminées sur la base du modèle de référence spécifié par un évaluateur.

Toutefois, nous avons distingué que la version actuelle de l'inspecteur des règles ergonomiques ne permet pas d'inspecter la totalité de mesures d'utilisabilité relatives aux critères ergonomiques considérés dans notre travail (voir Tableau 4.2). A cette fin, nous avons étendu cette version au niveau du code source du parseur adopté afin de fournir la possibilité de capturer l'ensemble des mesures d'utilisabilité suivantes : la densité globale d'une interface, le nombre de fonctions de feedback dans une interface et le nombre de fonctions disponibles à l'utilisateur dans une interface.

L'une de puissances de la version étendue de l'inspecteur des règles ergonomiques est sa capacité à communiquer d'une façon implicite avec notre outil d'évaluation réalisé SOUSEval. La communication est effectuée lors de chargement du code source du système interactif à évaluer via le sous-module de l'évaluation ergonomique mis en place dans l'outil SOUSEval (plus de détails sur ce module seront présentés dans la Section 4.5.2). De plus, nous avons lié la version étendue de l'inspecteur des règles ergonomiques avec la base de données de notre outil réalisé SOUSEval. Cette liaison va permettre l'enregistrement du modèle de référence spécifié par l'évaluateur ainsi que la génération des résultats fournis des mesures capturées directement dans notre outil SOUSEval.

4.4 L'outil SOUSEval

SOUSEval constitue le cœur de l'environnement intégré. Il présente une application web développée sous l'outil logiciel *Netbeans* en utilisant le langage de programmation *PHP 5*.
Lors de la conception de notre outil, nous avons adopté le design pattern MVC (Modèle, Vue, Contrôleur) afin de réduire la complexité du code et l'organiser (voir Figure 4.6). Ce pattern permet de séparer la modélisation du domaine, la présentation et les actions fondées sur les entrées de l'utilisateur, selon trois classes distinctes (Burbeck, 1992):

- *Le modèle :* Il gère le comportement et les données du domaine d'application.
- *La vue :* elle gère l'affichage d'informations.
- *Le contrôleur* : Il interprète les actions de l'utilisateur et prend en charge la gestion des événements de synchronisation pour mettre à jour la vue ou le modèle et les synchroniser.

En outre, l'outil SOUSEval permet de réaliser des évaluations subjectives de l'utilisabilité des interfaces utilisateur des systèmes interactifs comme étant un outil d'aide à la génération de questionnaires pour l'évaluation subjective. D'autre part, notre outil permet d'effectuer des évaluations intégrées subjective et objective.
Lors de la mise en œuvre de notre outil SOUSEval, nous nous sommes basés sur le processus d'évaluation proposé par la norme ISO/IEC 25040 (ISO/IEC 25040, 2011). Comme le montre le Tableau 4.3, ce processus est structuré autour de cinq étapes : la spécification des exigences de l'évaluation, la spécification de l'évaluation, la conception de l'évaluation, l'exécution et la conclusion de l'évaluation (voir *Chapitre 1, Section 1.2.1*).

En se basant sur celui-là, notre processus s'étend sur trois étapes essentielles. La première concerne la préparation de l'évaluation dans laquelle nous avons synthétisé les trois premières étapes de la norme ISO/IEC 25040. Dans cette étape, nous avons répondu aux questions suivantes, comme le montre le Tableau 4.3 (Pourquoi évaluer ? Qu'est ce qu'on va évaluer? Quelles sont les méthodes et les critères

à adopter? Qui va procéder à l'évaluation ? et Quand évaluer ?). La deuxième étape consiste à l'exécution de l'évaluation. Quant à la troisième, elle concerne la conclusion de l'évaluation qui porte sur l'analyse des résultats obtenus. Pour bien mener une évaluation, l'évaluateur doit préparer un plan d'évaluation à décrire dans un cahier de charge avant de procéder à l'évaluation.

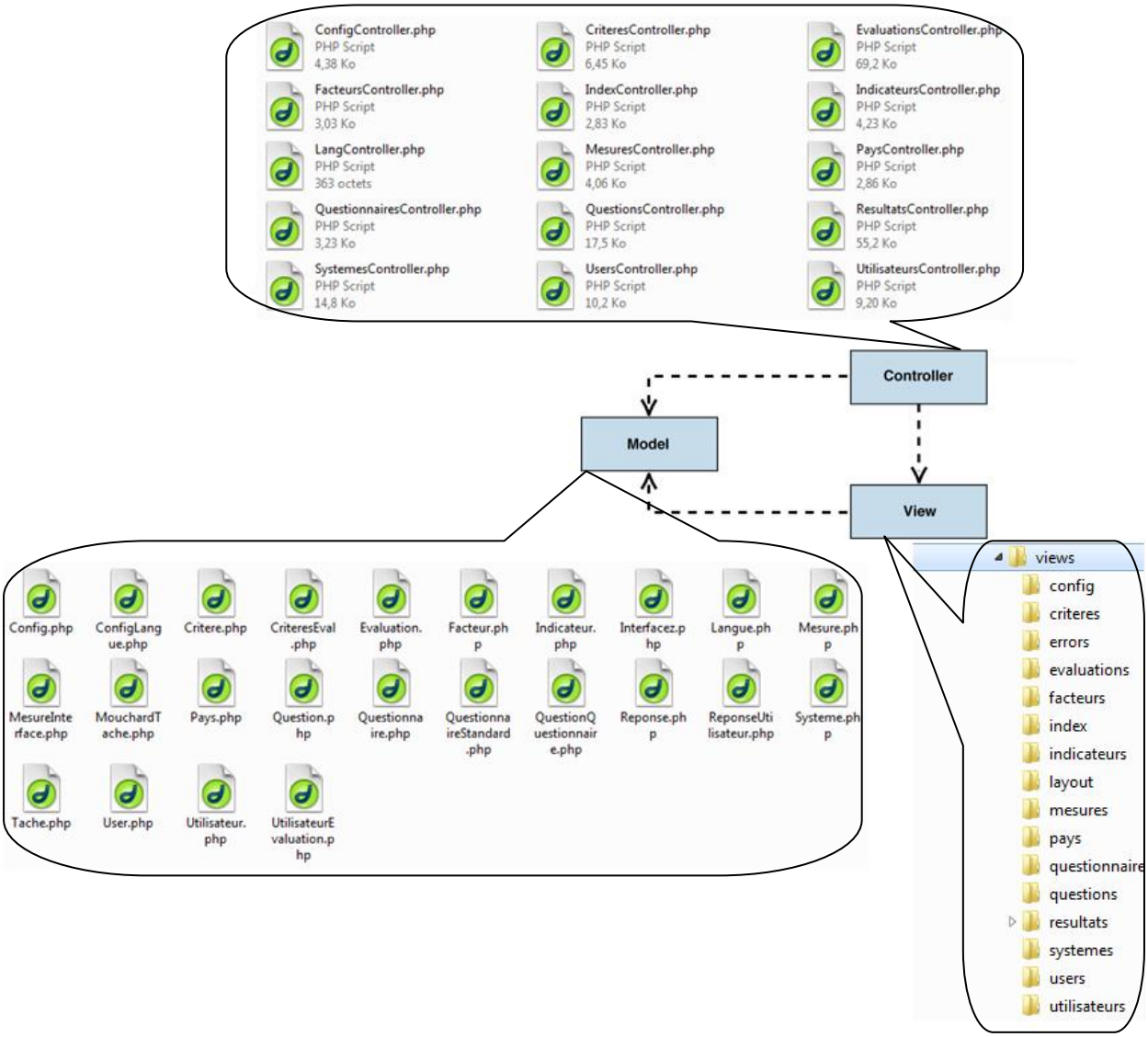


Figure 4. 6 : La structure du design pattern MVC réalisé

Tableau 4. 3 : Le processus de l'évaluation adopté en se basant sur la norme ISO/IEC 25040

Processus de la Norme ISO/IEC 25040	**Notre processus adopté**
La spécification des exigences d'évaluation ○ Pourquoi évaluer ? ○ Qu'est ce qu'on va évaluer ○ Qui va procéder à l'évaluation?	*La préparation de l'évaluation* ○ Description de l'objectif de l'évaluation ○ Les critères d'utilisabilité ○ Les utilisateurs
La spécification de l'évaluation ○ Quelles sont les méthodes à adopter?	○ Les questions relatives aux critères à évaluer
La conception de l'évaluation • Quand évaluer?	○ Les rendez-vous des utilisateurs pour effectuer les évaluations
L'exécution de l'évaluation	*L'exécution de l'évaluation*
La conclusion de l'évaluation	*La conclusion de l'évaluation*

Etant donné que notre outil permet d'assurer des évaluations subjectives et intégrées, nous avons proposé un processus à adopter pour chacune d'évaluation que nous allons les présenter dans ce qui suit.

4.4.1 L'outil SOUSEval comme étant un outil d'aide à la génération des questionnaires

Notre première étape lors de la mise en œuvre de notre outil SOUSEval consiste à implémenter un composant qui assure l'évaluation subjective de l'utilisabilité des systèmes interactifs. Ce composant est défini comme étant un outil d'aide à la génération des questionnaires.
Comme l'illustre la Figure 4.7, une évaluation subjective est constituée de trois modules :

- Le module de préparation qui porte sur la création de l'évaluation subjective
- Le module de l'évaluation qui permet d'assurer l'exécution de l'évaluation
- Le module d'analyse des résultats de l'évaluation qui sert à fournir les résultats obtenus.

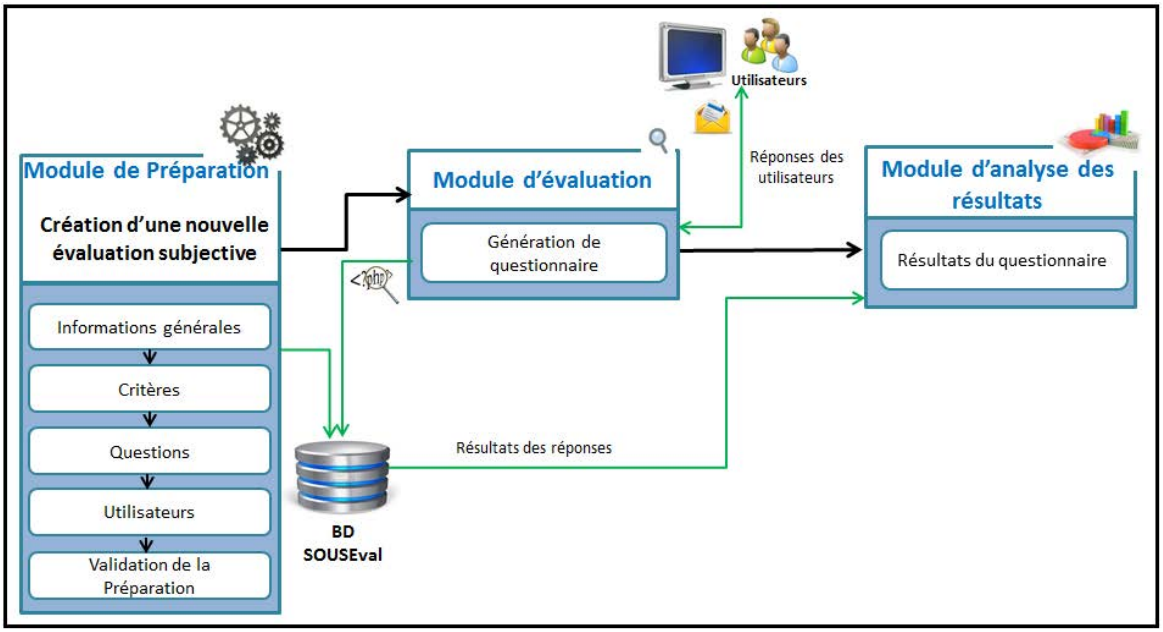


Figure 4. 7 : L'architecture du composant de l'évaluation subjective mis en place dans l'outil SOUSEval

Dans ce qui suit, nous décrivons chaque module.

4.4.1.1 Le module de préparation d'une évaluation subjective

Comme l'illustre la Figure 4.9, la préparation d'une évaluation subjective porte sur la création d'une nouvelle évaluation qui comporte quatre étapes.

La première consiste à la spécification des informations générales sur l'évaluation en indiquant le libellé, la description et la date de création de l'évaluation.

La deuxième étape concerne la spécification des critères d'utilisabilité à considérer lors de l'évaluation. Ceux-ci sont sélectionnés à partir de l'ensemble des critères configurés dans l'outil SOUSEval qui doivent être ajoutés par l'évaluateur avant le démarrage de l'évaluation.

La troisième étape porte sur la spécification de questions pour la construction du questionnaire d'évaluation. Celui-ci vise d'une façon directe à détecter les problèmes de l'utilisabilité d'un système interactif portant sur chaque critère sélectionné lors de la phase précédente.
Etant donné que dans notre approche (*Cf. Chapitre 3*) nous avons adopté les questionnaires standardisés de l'utilisabilité existants, notre outil permet d'assurer la configuration de ces

questionnaires et leurs questions associées. En effet, la création d'une question standard nécessite l'identification de l'énoncé, le critère/sous-critère associé, le nom de son questionnaire et les réponses possibles, comme l'illustre la Figure 4.8. Pour cela, lors de la spécification des questions, notre outil SOUSEval génère la liste des questions appropriées à chaque critère sélectionné en tenant compte de toutes leurs informations associées.

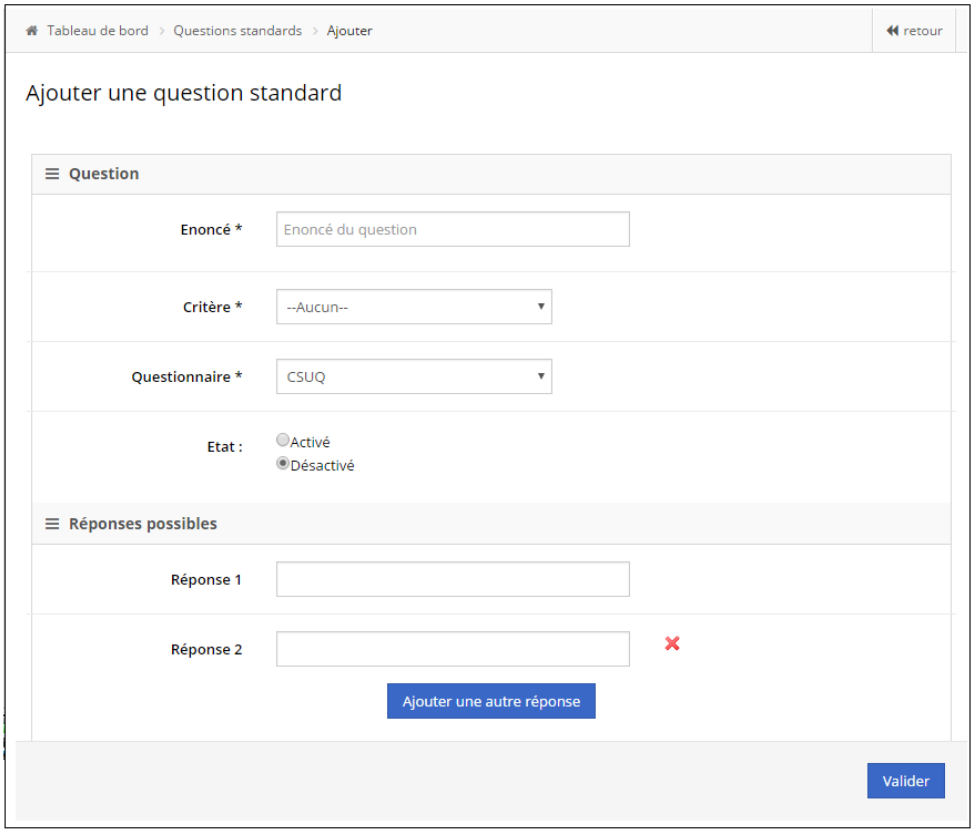


Figure 4. 8 : Aperçu de l'interface d'ajout d'une question

Finalement, après avoir spécifié l'ensemble des questions portant sur les critères de qualité sélectionnés, la quatrième étape consiste en l'identification des utilisateurs à impliquer lors de l'évaluation à partir de la liste des utilisateurs déjà saisie comme l'illustre la Figure 4.9.

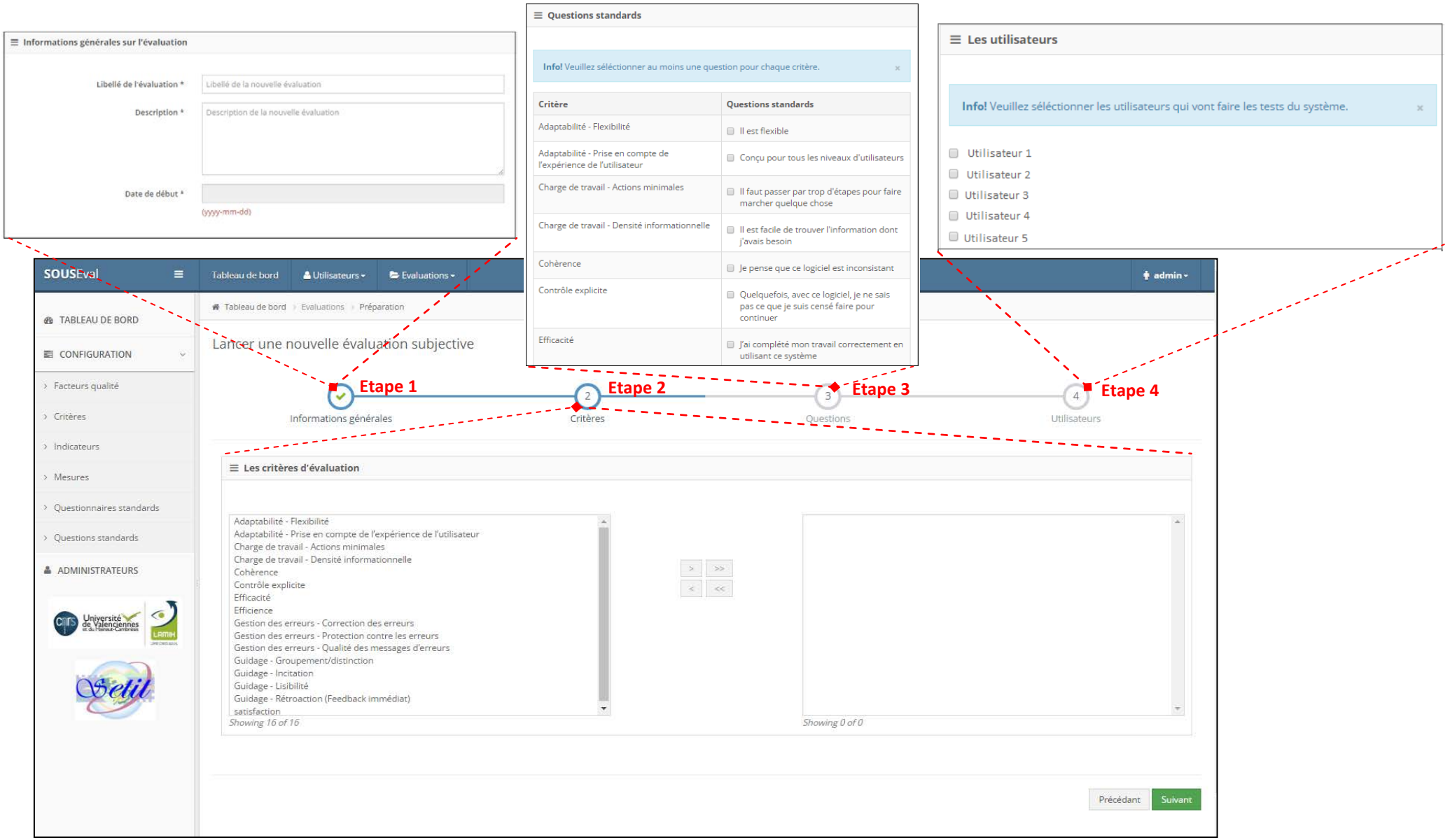


Figure 4. 9 : Aperçus des interfaces du module de préparation d'une évaluation subjective

La spécification de quatre étapes précédentes permet d'effectuer 75 % de la phase de préparation. Afin de conclure cette phase, notre outil assure la planification des évaluations pour les utilisateurs impliqués à travers le composant «*Gérer les utilisateurs*» comme l'illustre la Figure 4.10. Celui-ci permet en plus d'ajouter des nouveaux utilisateurs. En effet, la planification d'une évaluation pour chaque utilisateur s'effectue en spécifiant la date, l'heure de l'évaluation et son expertise par rapport au système à évaluer. Trois choix sur l'expertise sont possibles (expert, moyen, novice). Par conséquent, tous les rendez-vous des utilisateurs seront sauvegardés dans un calendrier défini dans l'outil SOUSEval.

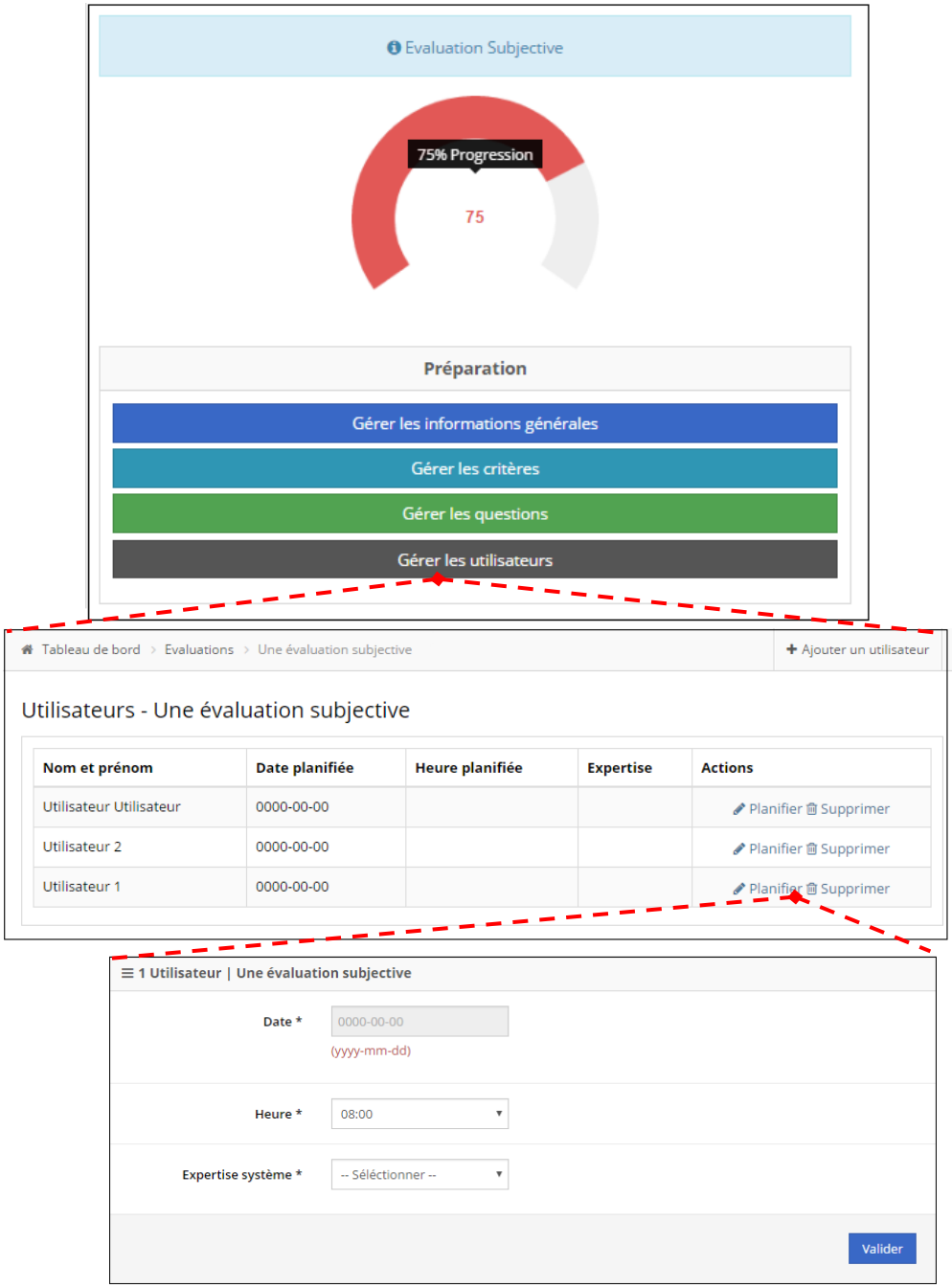


Figure 4. 10 : Aperçus des interfaces de planification d'une évaluation subjective

En outre, l'outil SOUSEval offre une flexibilité permettant d'apporter des modifications au niveau des informations générales de l'évaluation, les critères de qualité spécifiés et leurs questions associées

ainsi que les informations sur les rendez-vous des utilisateurs, avant de valider définitivement la phase de préparation.

Comme l'illustre la Figure 4.11, la gestion des critères d'utilisabilité consiste à ajouter un (des) critère(s) à la liste de critères sélectionnées à partir de la liste de critères disponible ou le(s) supprimer. Notons que la suppression d'un critère d'évaluation engendrera automatiquement la suppression de toutes les questions associées.

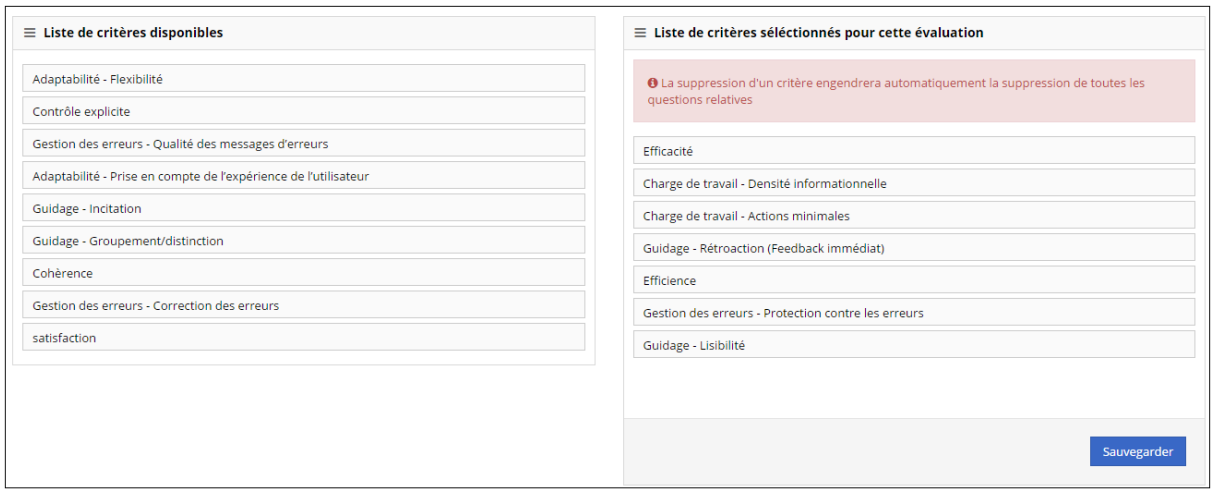


Figure 4. 11 : Aperçu de l'interface du sous-module de gestion des critères d'une évaluation subjective

Concernant la gestion des questions, ce composant permet de gérer l'ensemble des informations relatives au questionnaire spécifié de l'évaluation telles que le renommage du titre du questionnaire, la suppression d'une question existante et l'ajout d'une nouvelle question. Celle-ci peut être définie d'une façon personnalisée ou extraite de la liste des questions configurées qui correspondent aux questionnaires standardisées, Figure 4.12.

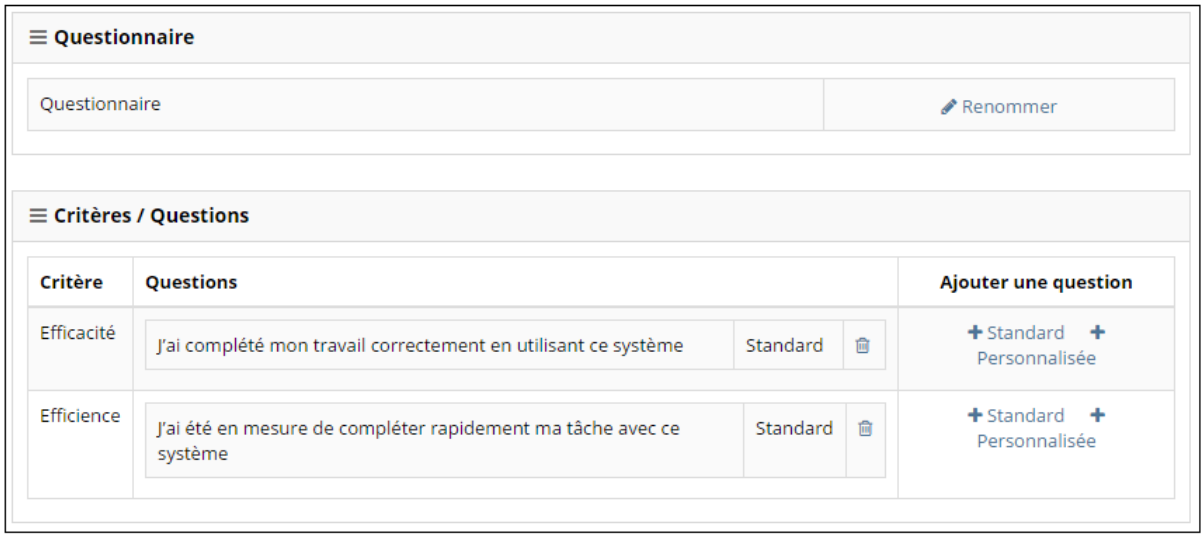


Figure 4. 12 : Aperçu de l'interface du sous-module de gestion des questions

4.4.1.2 Le module d'évaluation

La validation de la phase de préparation de l'évaluation conduit à l'activation du module de l'évaluation et la désactivation de deux sous-modules de gestion des critères et gestion des questions. De sorte qu'aucune modification sur les critères d'évaluation sélectionnés ni les questions spécifiées ne sera pas possible. Comme l'illustre la Figure 4.13, le module de l'évaluation se charge d'exécuter

l'évaluation à travers le lancement du questionnaire d'utilisabilité créé qui va être généré dans les sessions des utilisateurs impliqués dans l'évaluation.

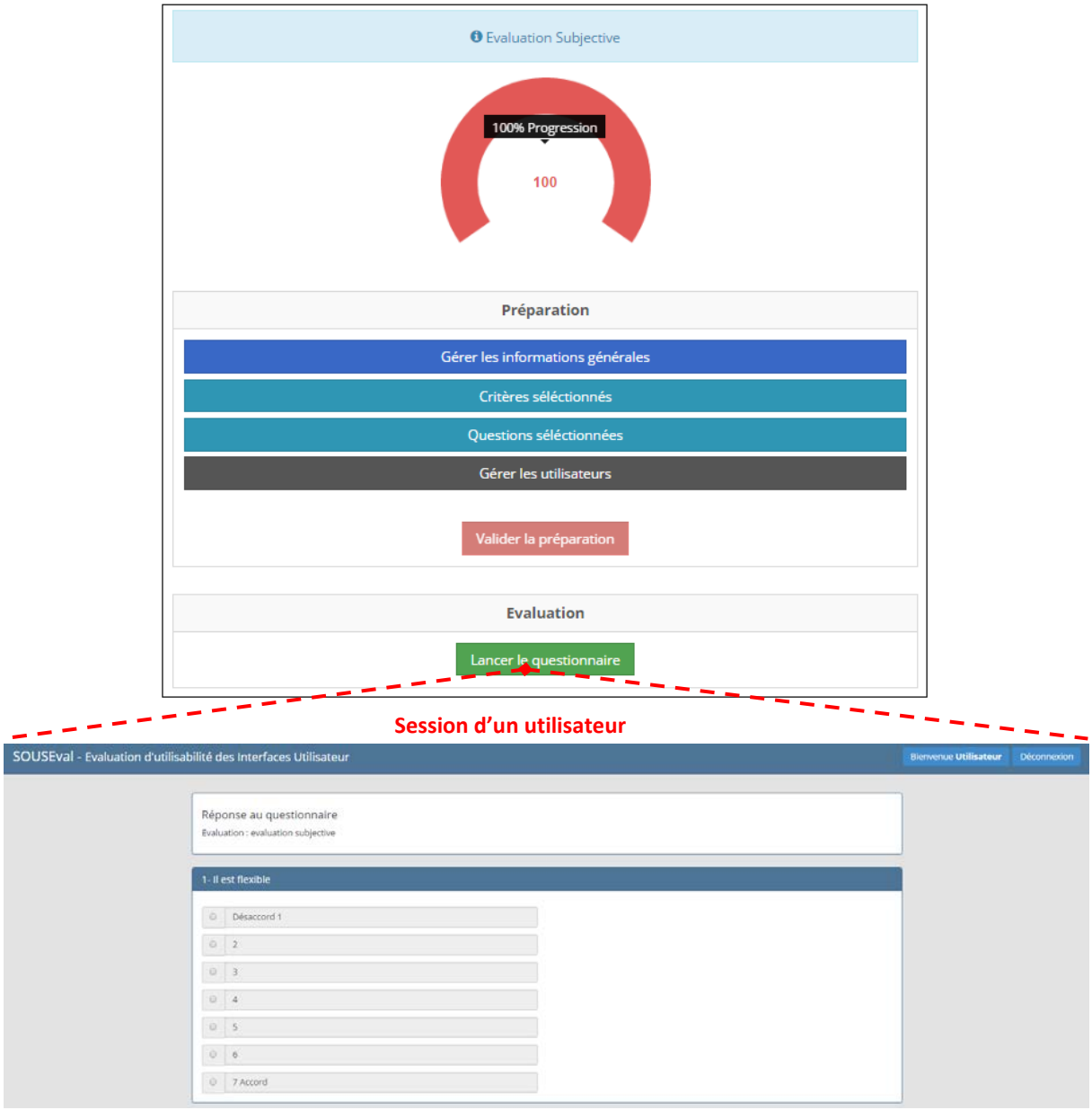


Figure 4. 13 : Aperçus des interfaces du module de l'évaluation subjective

4.4.1.3 Le module d'analyse des résultats d'évaluation

A la fin de l'évaluation, toutes les réponses des utilisateurs sont capturées et analysées dans notre outil puis générées à travers le module d'analyse des résultats. Comme l'illustre la Figure 4.14, ce module se charge de fournir les résultats obtenus des réponses des utilisateurs impliqués dans l'évaluation par rapport à chaque question du questionnaire. Il permet ainsi d'afficher ces résultats sous forme graphique ou textuelle pour aider l'évaluateur lors de l'interprétation des résultats. En outre, il permet de générer les réponses détaillées de chaque utilisateur qui peuvent être exportées vers Excel.

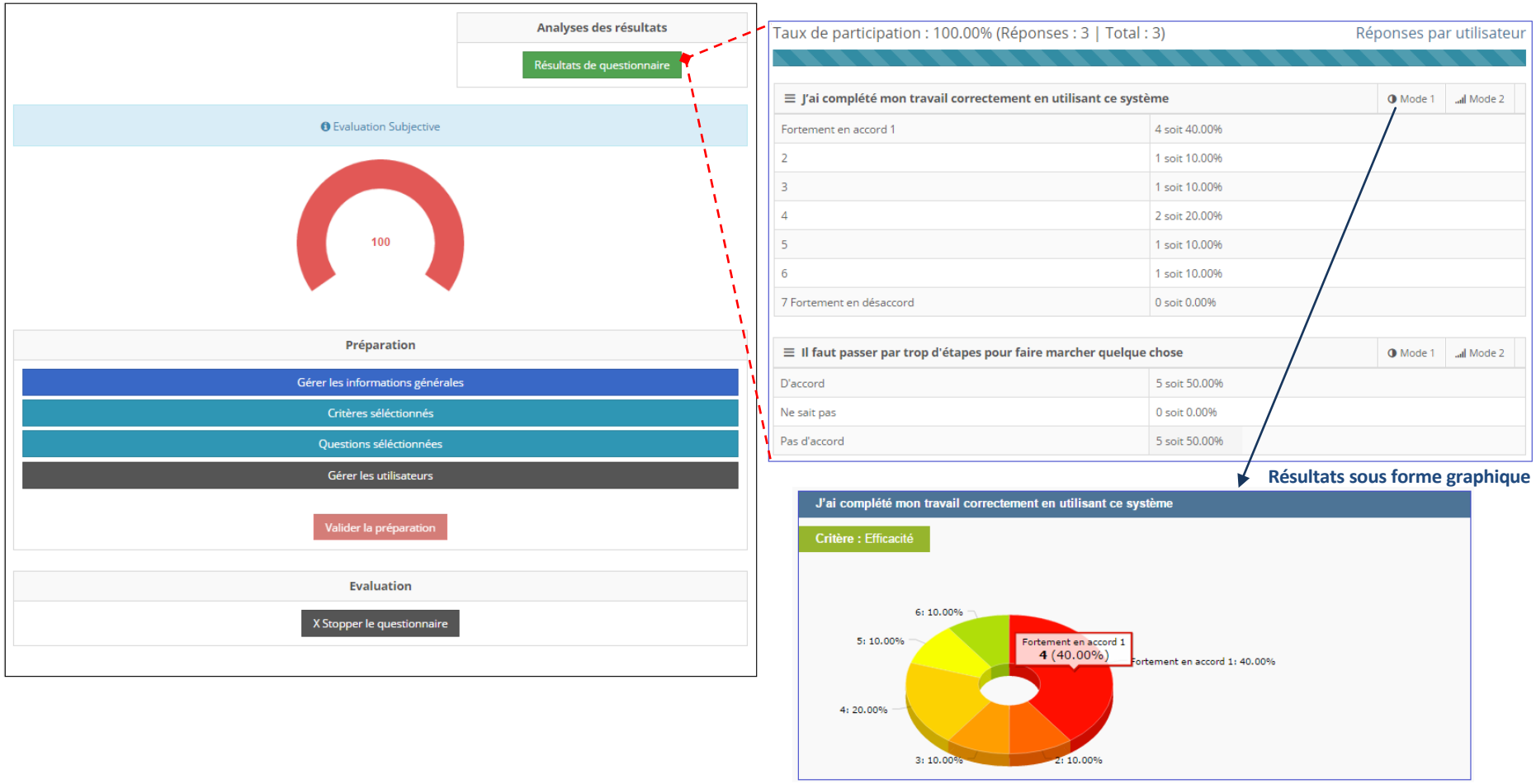


Figure 4. 14 : Aperçus des interfaces du module d'analyse des résultats de l'évaluation subjective

4.4.2 L'outil SOUSEval pour l'intégration des évaluations subjective et objective de l'utilisabilité des interfaces utilisateur

Après avoir réalisé le premier composant de notre outil SOUSEval qui porte sur l'évaluation subjective, nous avons mis en place un deuxième composant. Celui-ci sert à assurer une évaluation intégrée subjective et objective de l'utilisabilité des interfaces utilisateur. Cette évaluation porte sur l'intégration effective des résultats provenant de l'évaluation subjective (celle fondée sur l'utilisation d'un questionnaire d'utilisabilité) et les résultats des évaluations objectives (celles extraites du mouchard électronique EISEval et l'inspecteur des règles ergonomiques).

L'intégration est établie via la mise en place des indicateurs d'utilisabilité proposés lors de Chapitre 3 dans l'outil SOUSEval.

En suivant le même principe adopté lors d'une évaluation subjective, la réalisation d'une évaluation intégrée est basée notamment sur les trois modules suivants :

- Le module de préparation qui concerne la création de l'évaluation intégrée.
- Le module de l'évaluation qui porte sur l'exécution de l'évaluation
- Le module d'analyse des résultats de l'évaluation qui se charge de fournir les résultats synthétisés d'une façon intégrée et assurer leurs interprétations.

La Figure 4.15 représente l'architecture complète de notre outil SOUSEval qui est structurée selon deux composants de l'évaluation subjective et celle intégrée.

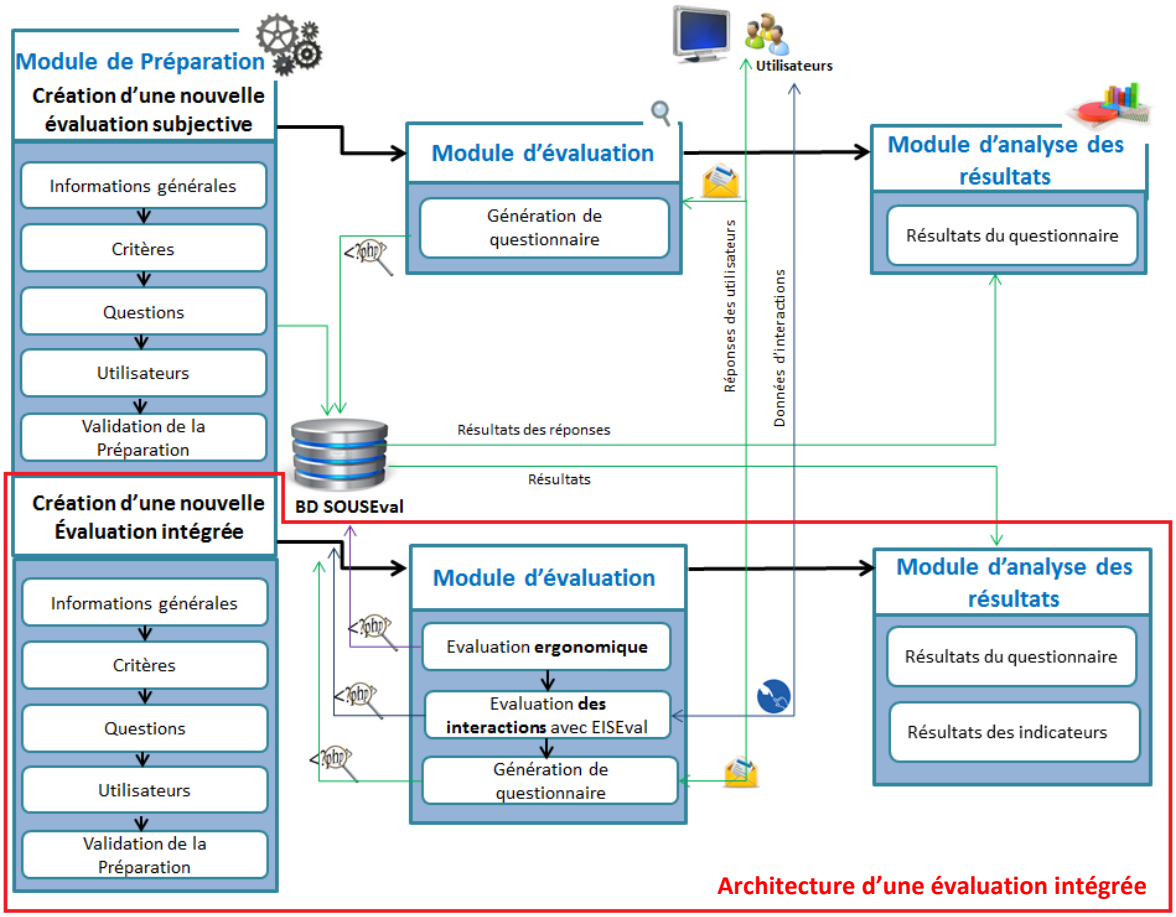


Figure 4. 15 : L'architecture globale de l'outil réalisé SOUSEval

Dans ce qui suit, nous présenterons les spécifications conceptuelles mises en œuvre dans chaque module de l'évaluation intégrée.

4.4.2.1 Le module de préparation de l'évaluation intégrée

En suivant la même structure adoptée lors de la phase de préparation d'une évaluation subjective, la préparation d'une évaluation intégrée requiert également la spécification de quatre éléments de base suivants : les informations générales de l'évaluation, les critères d'utilisabilité à considérer, les questions associées à chaque critère et les utilisateurs impliqués dans l'évaluation. Tout comme dans l'évaluation subjective (*Section 4.4.1*), pour conclure la phase de préparation d'une évaluation intégrée, il est important de planifier les rendez-vous des évaluations pour tous les utilisateurs impliqués dans l'évaluation.

4.4.2.2 Le module d'évaluation

Après avoir conclu la phase de préparation, nous pouvons commencer l'exécution de l'évaluation intégrée à travers le module d'évaluation. Nous rappelons que l'environnement intégré (Section 4.1) vise à établir une évaluation intégrée, via l'exécution des outils du mouchard électronique, l'inspecteur des règles ergonomiques et le questionnaire d'utilisabilité, permettant de synthétiser leurs résultats.

Rappelons aussi que nous avons étendu les deux outils existants du mouchard électronique EISEval et l'inspecteur des règles ergonomiques afin de rendre possible la récupération de l'ensemble des mesures de l'utilisabilité requises dans notre approche ainsi que la récupération de leurs résultats.
De ce fait, une évaluation intégrée s'effectue en trois étapes :

- Une évaluation ergonomique de la qualité des interfaces utilisateur du système à évaluer en utilisant la version étendue de l'inspecteur des règles ergonomiques.
- Une évaluation des interactions des utilisateurs avec le système à évaluer en utilisant la version étendue du mouchard électronique.
- Une évaluation subjective dans le but de capturer les perceptions des utilisateurs sur l'utilisabilité des IHM à évaluer. Celle-ci est effectuée par notre outil SOUSEval.

Etant donné que l'objectif principal de notre outil est la synthèse des résultats hétérogènes extraits des différents outils existants et le questionnaire, il est important que lors de la phase d'exécution de l'évaluation d'assurer la récupération de tous les résultats des outils existants.

Pour atteindre notre objectif, nous avons décidé, via le module d'évaluation, d'établir une communication entre les versions étendues des outils existants et notre outil développé. Cette communication est réalisée via la récupération et l'enregistrement direct des fichiers (.xml) des résultats des outils existants dans la base de données de notre outil à la fin de la phase d'exécution de l'évaluation. Pour l'évaluation subjective, notre outil permet de l'effectuer d'une façon directe en utilisant le questionnaire d'utilisabilité spécifié pendant la phase de préparation de l'évaluation intégrée comme mentionné précédemment.

Dans ce qui suit, nous présenterons chaque étape de l'exécution de l'évaluation ainsi que la façon avec laquelle notre outil permet de récupérer les résultats des outils existants.

4.4.2.2.1 L'évaluation ergonomique

L'évaluation ergonomique présente la première étape à effectuer lors de l'exécution de l'évaluation intégrée. Dans la Figure 4.16, nous présentons le processus adopté lors de la conduite de l'évaluation ergonomique. Dans un premier lieu, l'évaluateur doit spécifier d'abord à travers la version étendue de l'inspecteur des règles ergonomiques un modèle de référence qui sera enregistré dans la base de données de notre outil SOUSEval. Ce modèle de référence spécifie les règles ergonomiques à adopter

lors de l'évaluation concernant les attributs des composants graphiques des interfaces utilisateur. Par conséquent, ce modèle sera exploité lors du calcul de certaines mesures (voir Section 4.3).
Dans un second lieu, l'évaluateur procède au chargement du code source du système interactif à évaluer (à introduire sous forme d'un dossier compressé (.zip), voir Figure 4.17). Ce chargement est établi en faisant appel au code source de la version étendue de l'inspecteur des règles ergonomiques. Celui-ci se charge de parcourir et analyser tous les fichiers de codes source relatifs à chaque interface du système à évaluer (qui sont souvent d'extension. dfm) afin de déterminer toutes les valeurs des mesures adoptées.
A la fin de l'évaluation ergonomique, l'outil SOUSEval génère les résultats obtenus des mesures relatives à chaque interface. Il indique ainsi le libellé de chaque interface en générant uniquement les mesures relatives aux critères sélectionnés dans l'évaluation (voir Figure 4.18).

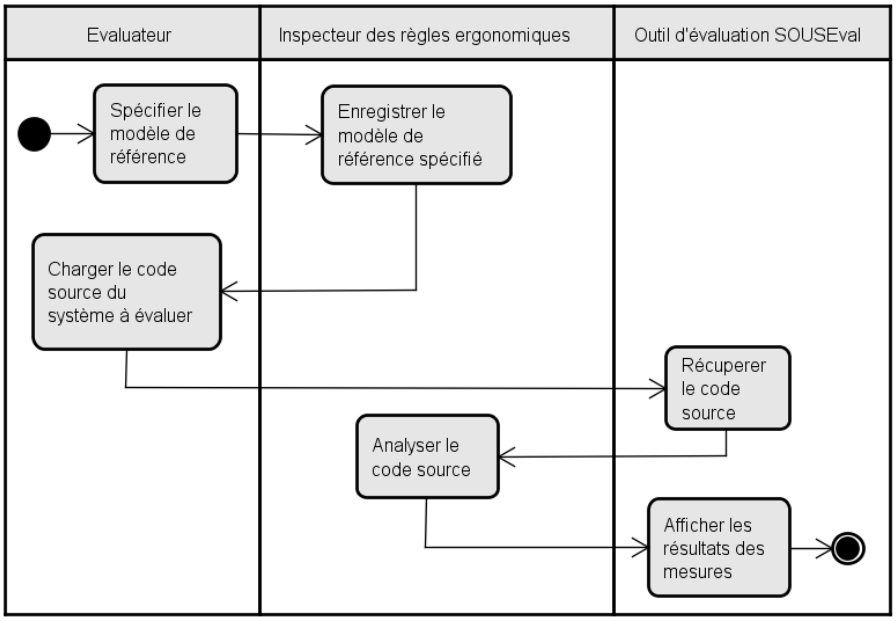


Figure 4. 16 : Le processus adopté lors de la phase de la conduite d'une évaluation ergonomique

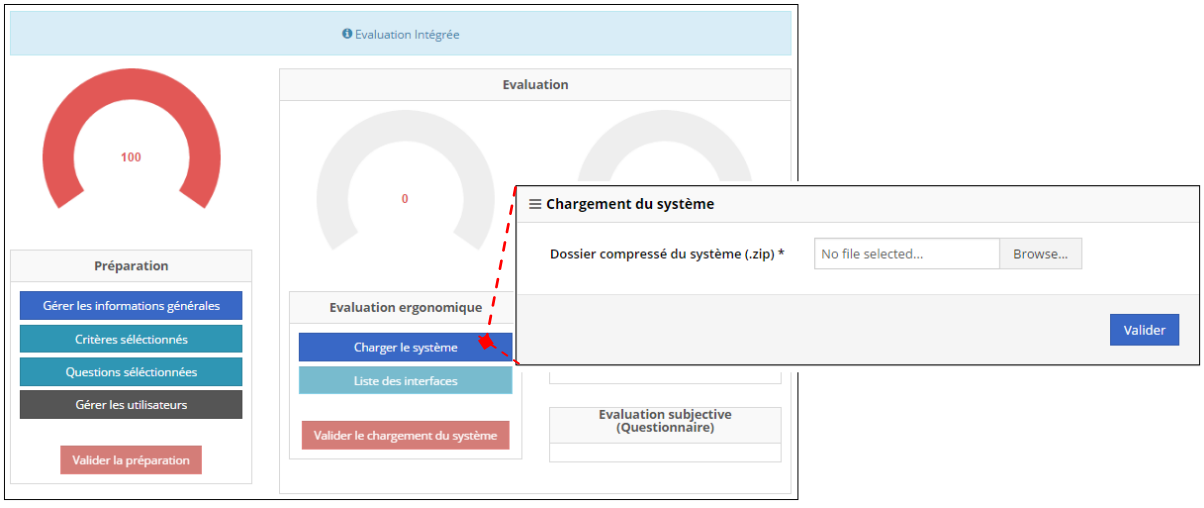


Figure 4. 17 : Aperçu de l'interface de chargement du système de sous-module d'évaluation ergonomique

Cette première étape de l'évaluation intégrée est conclue par la validation finale de chargement du système à évaluer, comme l'illustre la Figure 4.18. Cela signifie qu'aucune possibilité de changement du système chargé n'est autorisée une fois le chargement définitif est effectué.

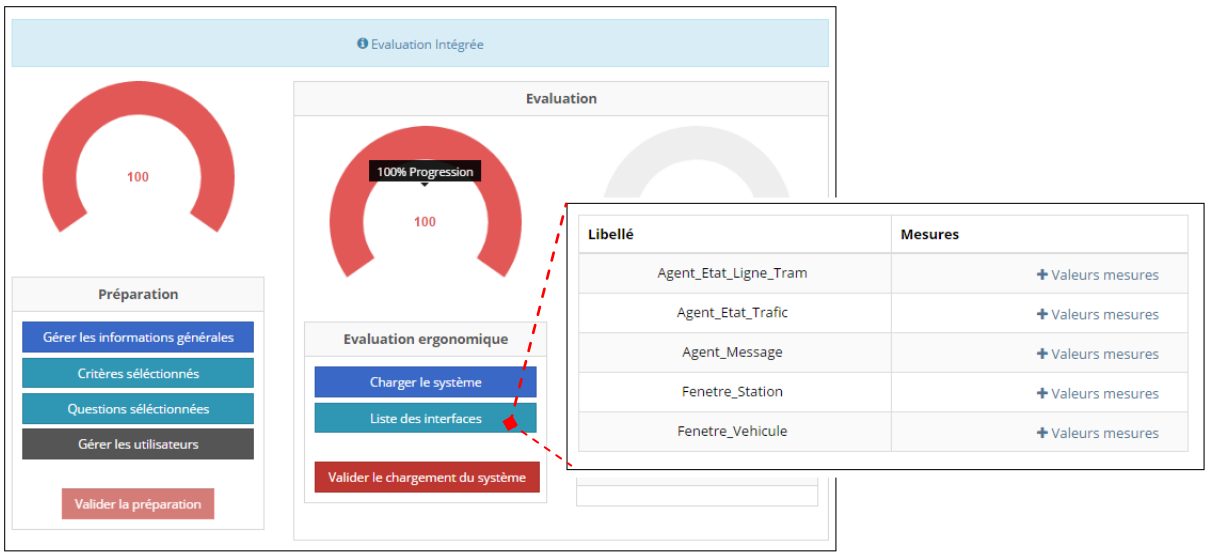


Figure 4. 18 : Aperçu de l'interface de la liste des interfaces générées de sous-module d'évaluation ergonomique

4.4.2.2.2 L'évaluation des interactions des utilisateurs avec le système interactif à évaluer

Passant par la suite à la deuxième étape de l'évaluation intégrée. Elle consiste à effectuer une évaluation basée sur l'implication des utilisateurs afin de capturer leurs interactions avec le système à évaluer en temps réel. L'exécution de cette évaluation est réalisée en utilisant la version étendue du Mouchard électronique EISEval (*Section 4.2*). Lors de cette phase nous visons à capturer l'ensemble des mesures d'utilisabilité portant sur les interactions des utilisateurs avec le système interactif à évaluer (mentionné dans le Tableau 4.2). Ces mesures peuvent être déterminées suite à l'exécution d'un ensemble des tâches effectuées par des utilisateurs impliqués dans l'évaluation. Ces tâches sont souvent relatives à un scénario prédéfini par l'évaluateur qui traite un problème ou une situation dans laquelle l'utilisateur doit utiliser le système.

Avant de commencer la phase de conduite de l'évaluation en utilisant la version étendue du mouchard électronique, notre outil SOUSEval permet de gérer l'ensemble des tâches à exécuter par les utilisateurs en ajoutant chaque tâche via le sous-module d'évaluation des interactions comme l'illustre la Figure 4.19. Cela va servir à déterminer certains résultats des indicateurs qui portent sur l'exécution des tâches (l'indicateur d'efficacité de réalisation des tâches, l'indicateur d'efficience de réalisation des tâches, l'indicateur d'action minimale pour réaliser une tâche et l'indicateur de feedback aux actions des utilisateurs).

Comme l'illustre la Figure 4.19, l'ajout d'une nouvelle tâche est spécifié par un libellé, le fichier de résultats d'un expert extrait à partir du module 8 de la version étendue du mouchard électronique (Voir *Section 4.2*), ainsi que son fichier détaillé de chaque tâche exécutée. Le fichier de résultats permet la récupération de la mesure du nombre d'actions d'expert requises pour réaliser une tâche et le temps de réalisation d'une tâche par un expert. Ceux-ci seront utilisés comme des références pour l'aide à l'interprétation des résultats de l'indicateur d'efficience de réalisation des tâches.

En outre, ces deux fichiers contribuent au calcul de certaines mesures dérivées relatives aux critères d'efficience (le taux d'efficience d'une tâche d'un expert, le taux d'efficience moyen des tâches de l'expert et le taux d'efficience théorique) et le critère d'actions minimales (le nombre d'utilisateurs réalisant les actions requises de la tâche et le pourcentage des participants réalisant les actions requises d'une tâche, voir *Chapitre 3*).

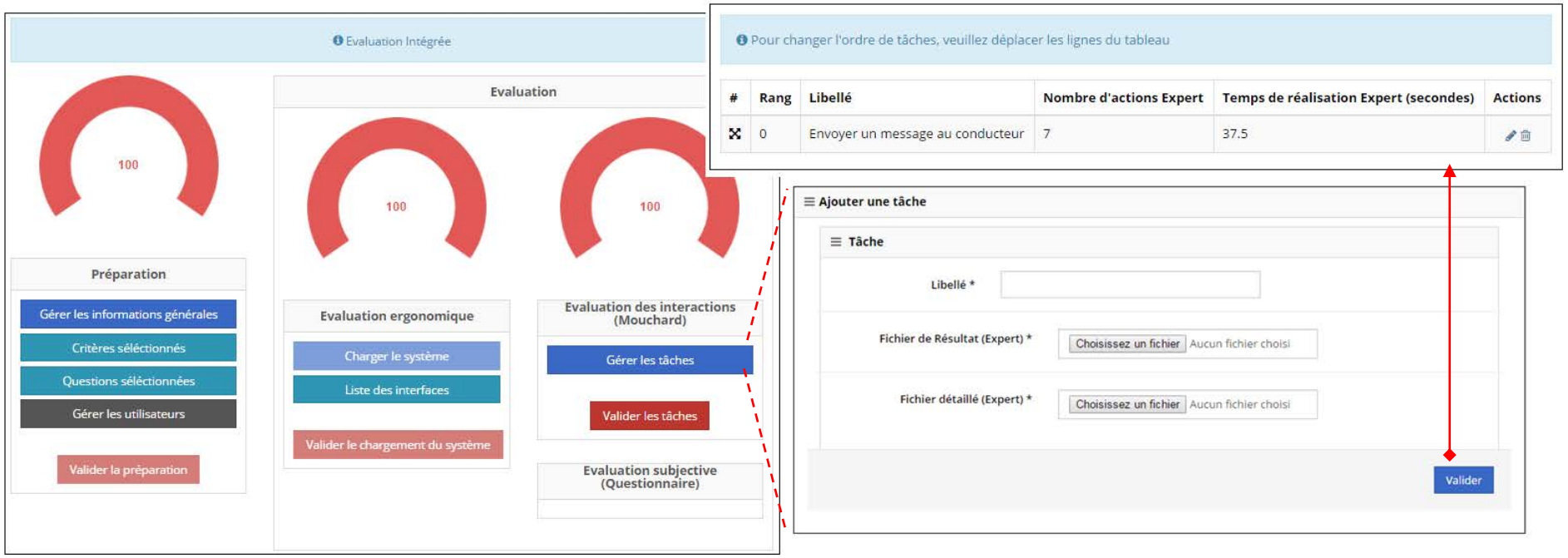


Figure 4. 19 : Aperçus des interfaces du sous-module d'évaluation des interactions

Après la spécification des toutes les tâches à exécuter par les utilisateurs, il s'avère nécessaire de « *Valider les tâches* » pour démarrer la phase de l'exécution de l'évaluation en impliquant les utilisateurs (Figure 4.19).

Cette phase est effectuée en utilisant la version étendue du mouchard électronique EISEval, le questionnaire d'utilisabilité (spécifié pendant la phase de préparation) et un système interactif à évaluer. Pendant sa session d'évaluation, l'utilisateur doit d'abord exécuter l'ensemble des tâches spécifiées par l'évaluateur en utilisant le système interactif à évaluer. Une fois l'utilisateur commence l'utilisation du système, le mouchard électronique EISEval est également lancé pour capturer toutes ses interactions. Après l'exécution de toutes les tâches, l'utilisateur sera dirigé vers sa session en ligne pour répondre au questionnaire spécifié.

4.4.2.2.3 L'évaluation subjective

L'évaluation subjective présente la troisième étape à effectuer lors de la conduite d'une évaluation intégrée en utilisant l'outil SOUSEval. Lors de cette phase, l'évaluateur va lancer le questionnaire déjà spécifié pendant la phase de préparation via le sous-module d'évaluation subjective (voir la Figure 4.20). Par conséquent, ce questionnaire va être généré dans les sessions des utilisateurs impliqués dans l'évaluation. A son tour, l'utilisateur doit répondre au questionnaire. Un exemple de celui-ci est présenté dans la Figure 4.20. Il est constitué d'un ensemble des questions qui dépendent des critères d'évaluation sélectionnés. Certaines présentent des questions générales, d'autres sont relatives aux interfaces et aux tâches exécutées.

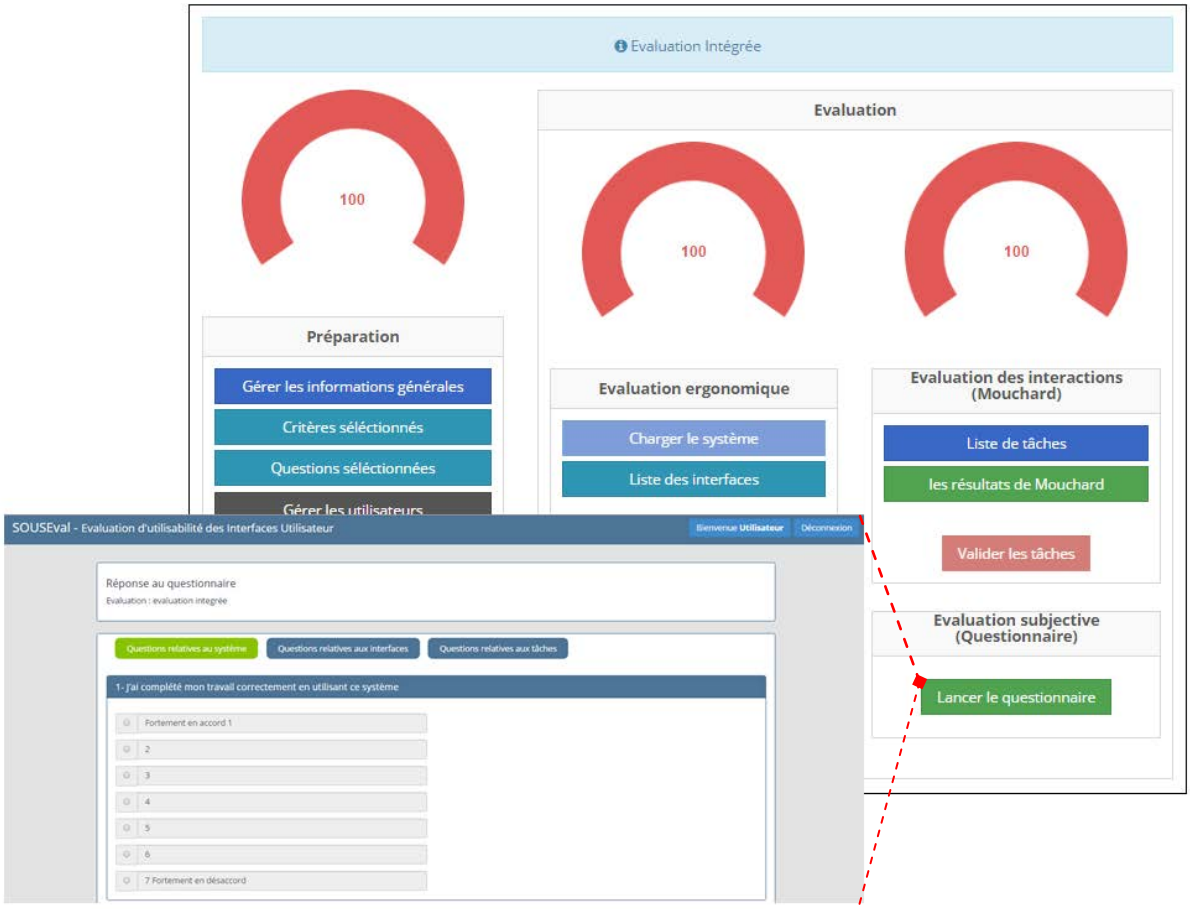


Figure 4. 20 : Aperçu du questionnaire généré à partir du sous-module de l'évaluation subjective

A la fin de l'évaluation, étant donné que notre outil se charge de capturer les réponses des utilisateurs au questionnaire, la dernière étape lors de la conduite de l'évaluation consiste à la récupération des

résultats fournis par la version étendue du mouchard électronique EISEval. Cette étape est effectuée via le sous-module d'évaluation des interactions à travers le composant de Résultats du mouchard (voir Figure 4.21).

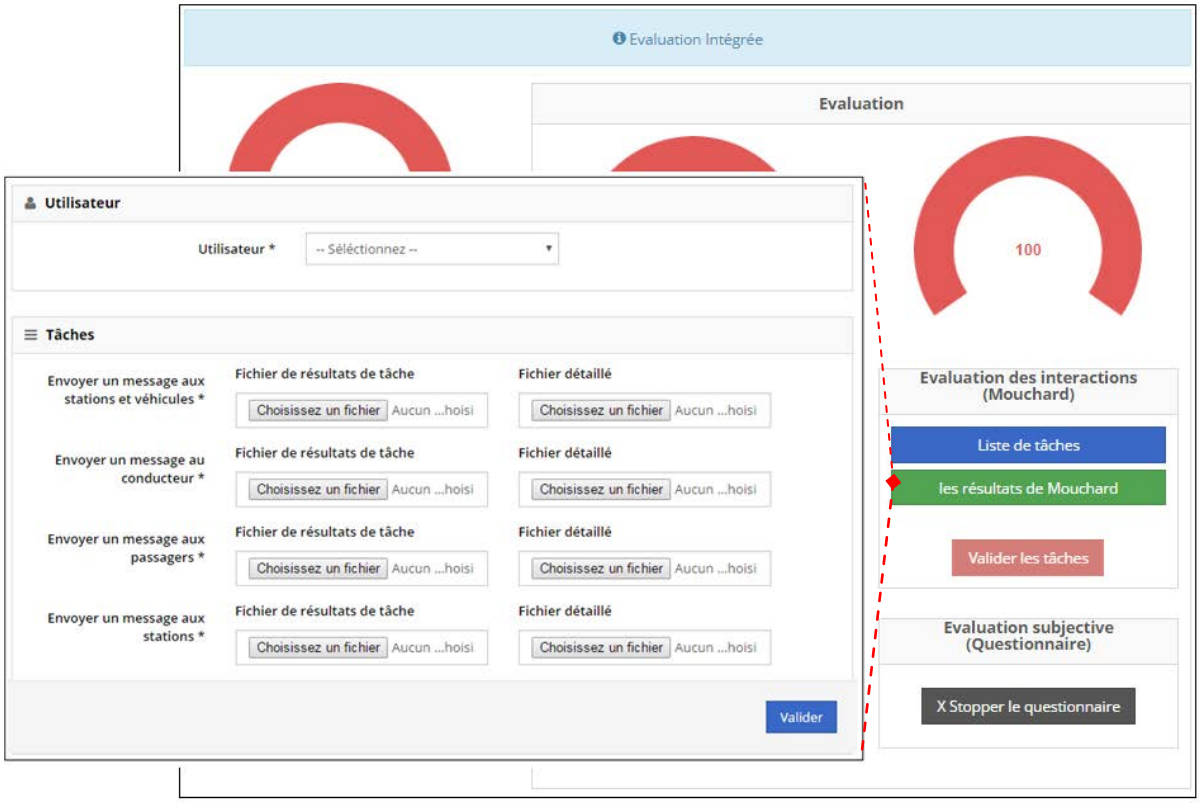


Figure 4. 21 : Aperçu de l'interface du sous-module d'insertion des résultats du Mouchard

Celui-là permet d'insérer les résultats du mouchard pour chaque utilisateur impliqué dans l'évaluation. Comme l'illustre la Figure 4.21, en suivant le même principe appliqué lors de la spécification des tâches à exécuter, l'insertion des résultats de chaque tâche est réalisée par le chargement de deux fichiers de résultats fournis par la version étendue du mouchard électronique. Le premier contient les résultats des mesures d'utilisabilité (mentionnées dans le Tableau 4.2) et le second fichier porte sur la description détaillée de la réalisation de la tâche.

Après avoir capturé tous les résultats du mouchard, notre outil SOUSEval se charge de calculer certaines mesures dérivées portant sur les critères d'efficacité, efficience, le feedback immédiat et actions minimales (mentionnées dans le Tableau 4.2).

4.4.2.3 Le module d'analyse des résultats d'évaluation

Après avoir fini la phase d'exécution de l'évaluation, le module d'analyse des résultats sera lancé. Il permet de synthétiser, générer et analyser tous les résultats des mesures capturées pendant la phase de conduite de l'évaluation.

Ce module comporte deux parties. La première concerne les résultats de questionnaire et la deuxième synthétise d'une façon intégrée via les indicateurs d'utilisabilité tous les résultats obtenus lors de trois étapes d'évaluation effectuées.

- *L'analyse des Résultats du questionnaire*

Généralement, SOUSEval génère les résultats des réponses des utilisateurs impliqués dans l'évaluation immédiatement à la fin de chaque session d'évaluation. Il effectue ainsi une phase d'analyse par

rapport à chaque question du questionnaire qui peut être classifiée selon trois catégories (questions générales, questions relatives aux interfaces et questions relatives aux tâches). Tout comme dans le module d'analyse des résultats d'une évaluation subjective (Section 4.4.3), ce module permet d'afficher ces résultats sous forme graphique ou textuelle pour aider l'évaluateur lors de la phase d'interprétation. En outre, il permet de générer les réponses détaillées de chaque utilisateur qui peuvent être exportées vers Excel.

La Figure 4.22 illustre un aperçu d'un exemple des résultats de questionnaire d'une évaluation intégrée.

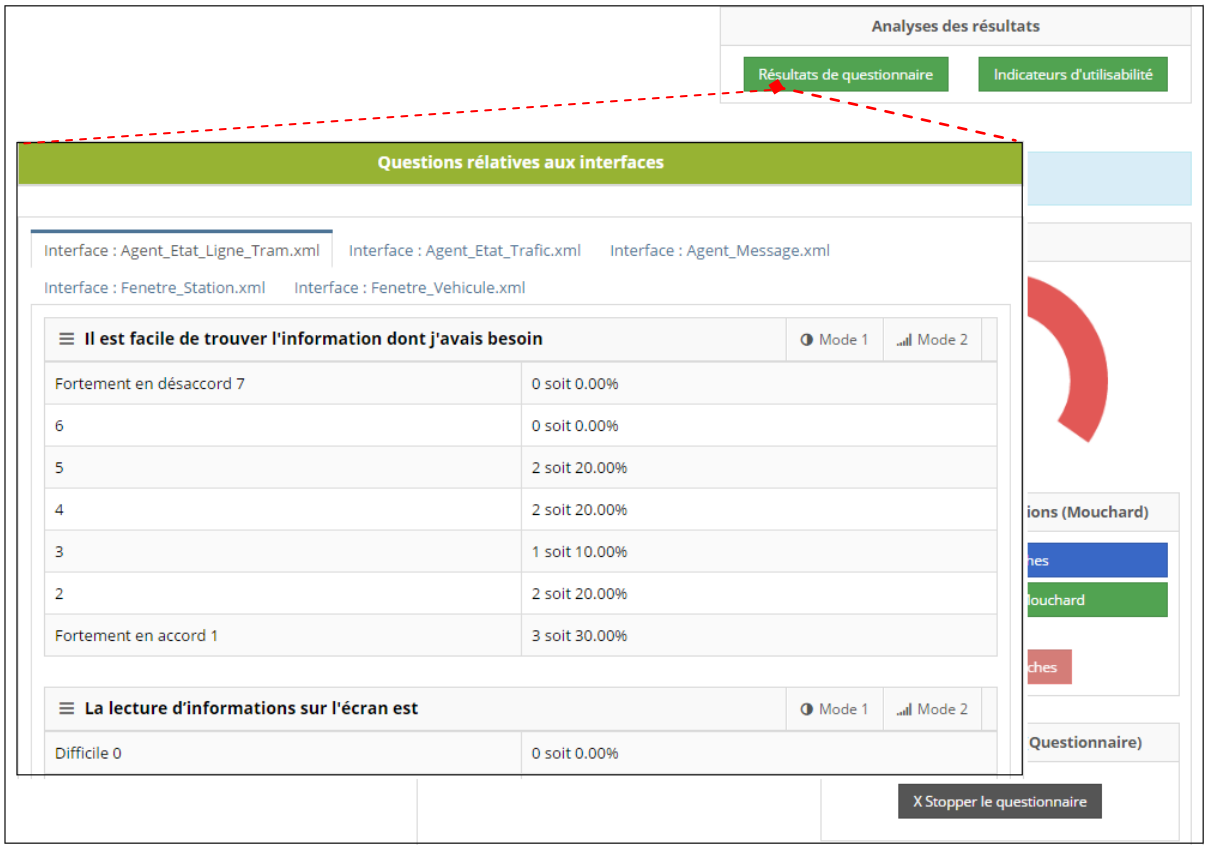


Figure 4. 22 : Aperçu d'un exemple des résultats de questionnaire lors d'une évaluation intégrée

- *L'analyse des résultats de l'évaluation via les indicateurs d'utilisabilité*

La synthèse de résultats d'une évaluation intégrée est assurée en utilisant les indicateurs d'utilisabilité mis en place dans le module d'analyse des résultats. Nous rappelons que ceux-ci servent notamment à l'intégration des résultats du questionnaire d'utilisabilité (évaluation subjective) avec ceux issus du mouchard et de l'inspecteur des règles ergonomiques (évaluation objective) d'une façon cohérente et complémentaire.

Ces indicateurs (spécifiés lors du Chapitre 3) portent sur l'évaluation d'un ensemble de critères d'utilisabilité (l'efficacité, l'efficience, la densité informationnelle, l'action minimale, la lisibilité et le feedback immédiat).

Le sous-module des indicateurs d'utilisabilité se charge de générer les résultats obtenus des indicateurs relatifs aux critères sélectionnés pendant une évaluation.

En effet, chaque indicateur (instancié souvent comme un couple de deux mesures une subjective et une objective) est désigné par un critère d'utilisabilité bien défini. Il est représenté avec un graphique contenant un nuage de points où l'axe d'abscisse concerne la mesure objective et l'axe d'ordonné porte sur la mesure subjective, comme l'illustre la Figure 4.23.

Au niveau du traitement des résultats des indicateurs, ce sous-module permet l'aide à l'interprétation des résultats obtenus via l'ensemble des critères de décision relatifs à chaque indicateur (décrits dans le *Chapitre 3*) mis en place dans l'outil SOUSEval.

En tenant compte qu'un expert en évaluation doit être impliqué pendant l'analyse des résultats pour définir les seuils des mesures d'un indicateur, ce sous-module lui fournit une analyse des résultats obtenus. La définition de ces seuils relatifs aux mesures associées à un indicateur est réalisée via l'implémentation de deux droites dynamiques qui se déplacent tout au long des axes d'abscisse et d'ordonné comme le montre la Figure 4.23.

Par conséquent, une fois les seuils sont fixés, l'outil SOUSEval génère les analyses appropriées qui portent sur le niveau de qualité de chaque interface utilisateur du système interactif évalué (ou tâche réalisée) ainsi que leurs recommandations associées.

Différents indicateurs d'utilisabilité qui sont mis en place dans notre outil seront présentés dans le chapitre suivant concernant les études expérimentales.

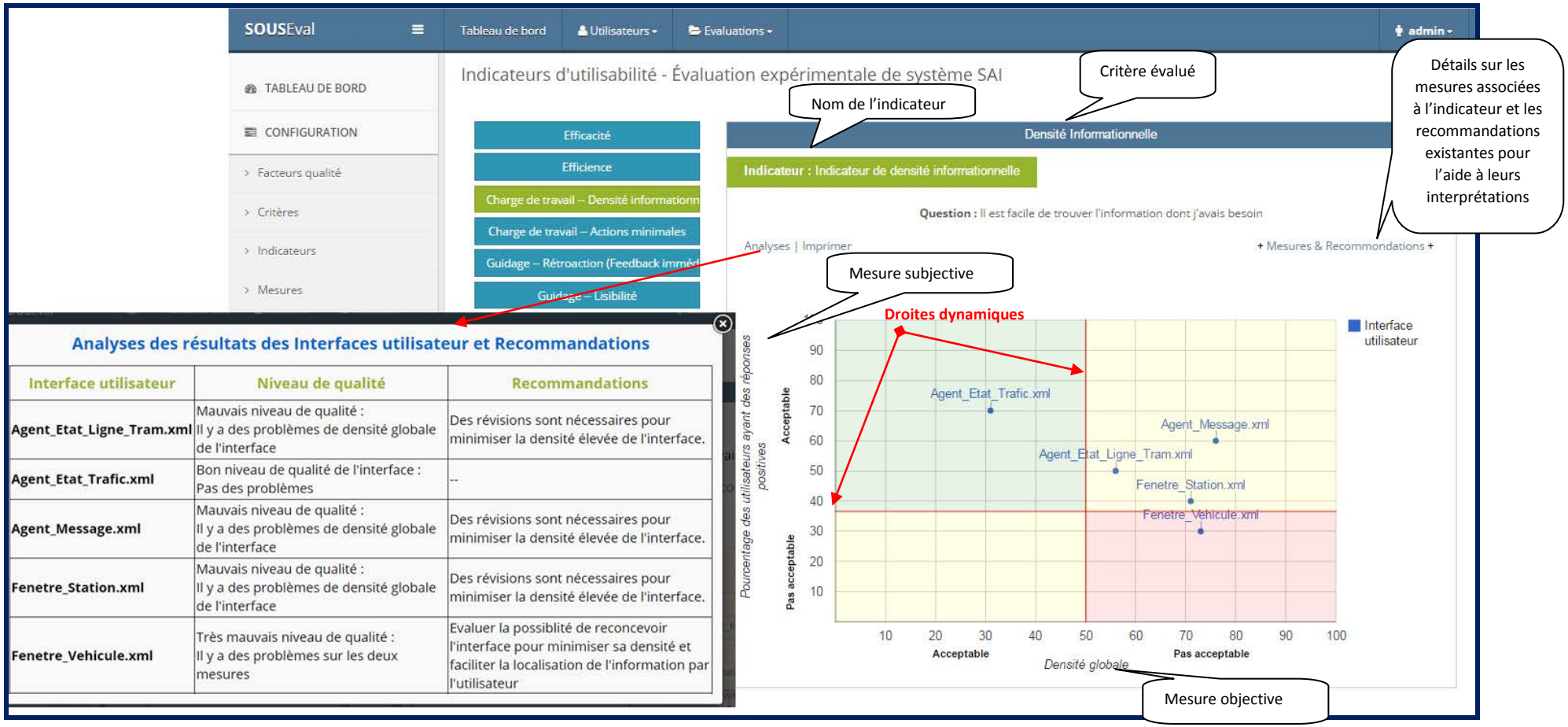


Figure 4. 23 : Aperçu de l'interface d'analyse des résultats en utilisant les indicateurs d'utilisabilité

4.5 Synthèse et conclusion

Dans ce chapitre, nous avons présenté l'ensemble des développements réalisés pour l'accomplissement de la mise en œuvre de l'environnement intégré d'aide à l'évaluation des systèmes interactifs. Nous avons commencé par la présentation des modifications apportées sur les outils existants afin de rendre plus flexible leurs exploitations. Par la suite, nous avons spécifié notre outil d'évaluation réalisé. Celui-ci permet d'effectuer d'évaluations subjective et objective intégrées dans lequel nous avons mis en œuvre notre approche proposée. Il a inclus en plus un module permettant d'assurer des évaluations subjectives, sous la forme d'un outil d'aide à la génération de questionnaires.

A travers ce chapitre, nous avons présenté l'architecture fonctionnelle de l'outil lors de la réalisation d'une évaluation subjective ou intégrée tout en détaillant les différents modules conçus : le module de préparation de l'évaluation, le module d'évaluation et le module d'analyse des résultats de l'évaluation. Celui–ci a contribué lors de l'évaluation intégrée au développement de l'ensemble des indicateurs d'utilisabilité proposés. Il a permis ainsi la mise en œuvre de leurs critères de décisions associées pour garantir une interprétation automatique.

Dans le chapitre suivant, nous nous contenterons de mettre en pratique l'outil réalisé à travers une étude de cas réel dans le but d'illustrer le fonctionnement complet de l'environnement intégré ainsi que l'étude de faisabilité de l'approche proposée.

Chapitre 5 : Evaluation Expérimentale

Introduction

Dans le chapitre précédent, nous avons présenté les développements réalisés qui ont permis la réalisation totale de l'environnement intégré d'aide à l'évaluation des systèmes interactifs. Dans ce chapitre, nous allons nous focaliser sur la mise en pratique de notre environnement et l'étude de faisabilité de notre approche à travers une étude de cas. Celle-ci porte sur l'évaluation de l'utilisabilité d'un système interactif nommé « **S**ystème d'**A**ide à l'**I**nformation Voyageurs ». Le choix de cette application est relatif à la nature spécifique de l'environnement intégré qui est dédié à l'évaluation des systèmes interactifs de supervision du Trafic urbain.

Ce chapitre est structuré autour de trois sections. Dans la première, nous présenterons une étude de cas en exécutant notre outil. Ensuite, dans la deuxième section, nous aborderons une discussion qui porte sur les risques de validité de notre travail mené. Enfin, nous achèverons ce chapitre par une conclusion.

5.1 Etude de Cas : Système d'Aide à l'Information Voyageurs

Dans cette section, nous présenterons une étude de cas concernant l'utilisation de l'environnement intégré, dans lequel nous avons implémenté notre approche proposée via l'outil SOUSEval, pour l'évaluation d'un système de supervision du Trafic urbain. Celui-ci est baptisé SAI : Système d'Aide à l'Information voyageurs. Il présente un système interactif à base d'agents. Il a été conçu dans le cadre d'un projet coopératif du Groupement Régional Nord-Pas-de-Calais pour la Recherche dans les Transports (Bonte et al., 2007). Ce système consiste à aider les régulateurs humains se trouvant dans une salle de contrôle pendant l'exécution de leur tâche de surveillance du réseau de transports en commun.
En plus, le SAI permet de les aider à effectuer leurs tâches de manière optimale dans les situations normales et anormales (Trabelsi et Ezzedine, 2013) grâce à un système d'aide à la décision (SAD), qui a été développé par une autre équipe de recherche du LAGIS, Lille (Bonte et al., 2007). Ce système vise essentiellement à informer les régulateurs sur le positionnement des différents véhicules (bus, tramway) sur le réseau de transport en temps réel. En outre, il leurs permet de communiquer des informations avec les conducteurs ou les passagers de différents véhicules à travers des messages qui sont injectés via des sockets TCP /IP via un système de communication.

La Figure 5.1 illustre un exemple d'envoi de messages effectué lors de notre expérimentation. Elle présente ainsi des aperçus des principales interfaces utilisateur du système d'aide à l'information (interface Agent_Etat_Trafic, interface Agent_Etat_Ligne_Tram, interface Agent_Message, interface Fenetre_Station et interface Fenetre_Vehiclue). Pour plus de détails, le lecteur intéressé peut se référer aux travaux de (Bonte et al., 2007 ; Ezzedine et al., 2008 ; Trabelsi, 2006 ; Tran, 2009).

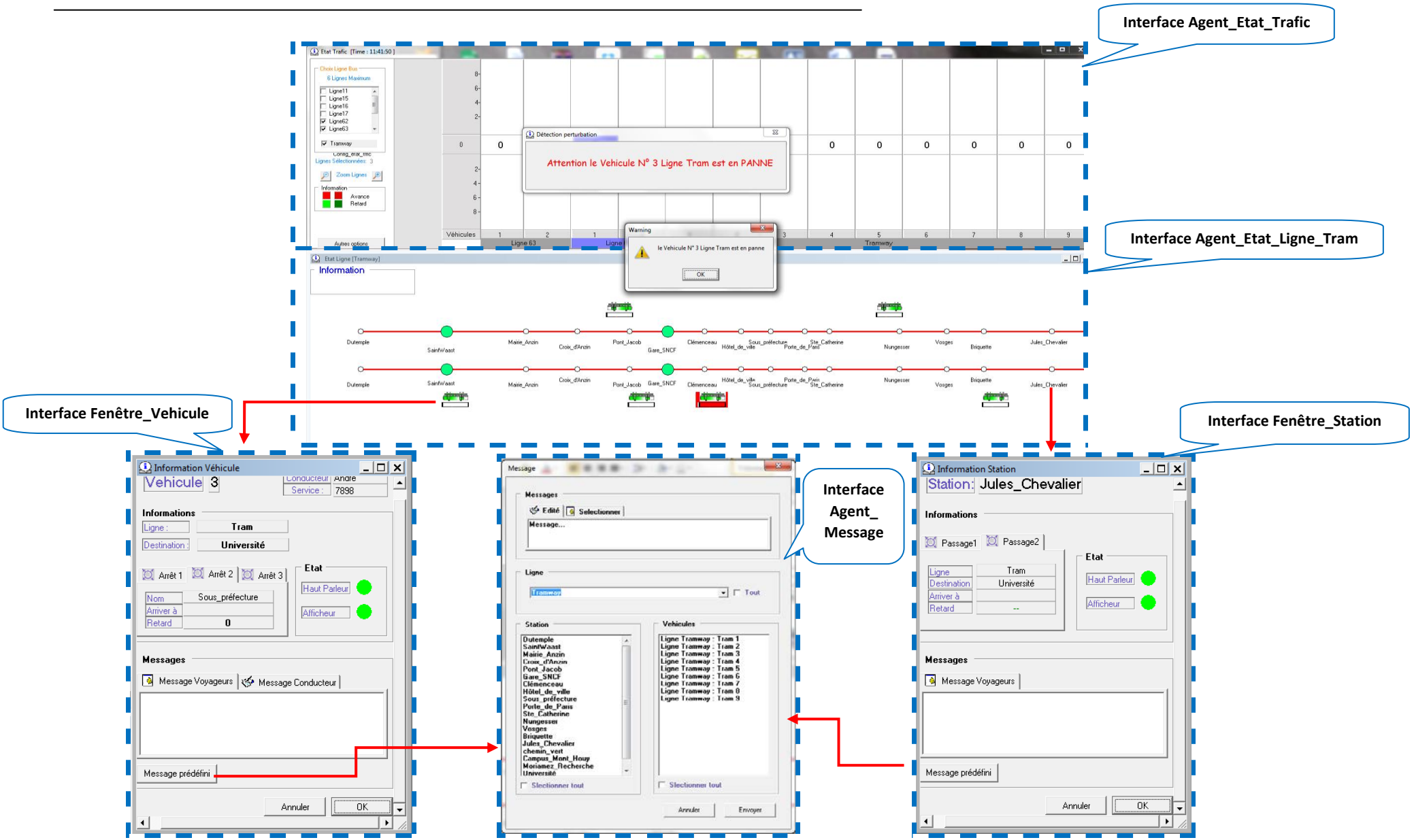


Figure 5. 1 : Aperçus des principales interfaces utilisateur du Système d'Aide à l'Information (SAI)

Pour bien conduire notre expérimentation, nous avons planifié et réalisé notre étude de cas en suivant les principales étapes du processus expérimental d'évaluation d'IHM proposé par Helen C. Purchase (2012). Ces étapes portent sur la spécification de l'objectif de l'étude de cas, la définition de la procédure expérimentale (planification de l'expérimentation) et l'exécution de l'expérimentation. Celle-ci va permettre la collecte et l'analyse des résultats d'expérimentation.

Dans ce qui suit, nous détaillerons chaque étape réalisée du processus expérimental suivi lors de notre étude de cas.

5.1.1 Objectif de l'étude de cas

Lors de cette étude de cas, notre objectif principal consiste à étudier la faisabilité de notre approche proposée des indicateurs d'utilisabilité via la mise en pratique de l'environnement intégré à travers l'exécution de notre outil SOUSEval. En plus, notre objectif est de montrer comment l'environnement permet de détecter des problèmes d'utilisabilité des interfaces utilisateur en intégrant des résultats d'évaluation d'utilisabilité subjective et objective.
De plus, nous visons à partir de cette étude de cas d'illustrer la procédure à suivre lors de l'utilisation de l'environnement intégré pour évaluer un système interactif donné.

5.1.2 Définition de la procédure expérimentale : Planification de l'expérimentation

La spécification d'une procédure expérimentale est une phase primordiale pour mener l'expérimentation avec succès. Elle présente l'étape initiale lors de la conception expérimentale et doit être bien liée au but visé par l'expérimentation (Purchase, 2012). En outre, cette phase consiste à planifier l'expérimentation via la préparation d'un plan de test à appliquer.

Notons que lors de l'exécution d'une évaluation (subjective ou intégrée) en utilisant notre outil SOUSEval, nous nous sommes basés sur l'implication des utilisateurs pour effectuer des tests d'utilisatbilité. Généralement, la réalisation de ces tests n'est pas une tâche facile et exige beaucoup de planification à l'avance. En s'inspirant des travaux de Rubin et Chisnell (2008), l'élaboration d'un plan de test d'utilisabilité s'articule autour de différentes étapes :

- La spécification de l'environnement et l'équipement du test
- La spécification des caractéristiques des participants
- La spécification de la liste des tâches à réaliser
- La spécification de la méthode de conception de test

En se basant sur ces étapes, nous détaillerons dans ce qui suit notre plan de test considéré lors de nos expérimentations.

- **L'environnement et les équipements du test (le contexte d'utilisation)**

La conception de notre étude de cas est déterminée en spécifiant les principales composantes du contexte d'utilisation pour la session d'évaluation à effectuer. Nous avons adopté la définition du contexte d'utilisation définie par la norme ISO 9241-11 (1998) qui comprend quatre composants : "les utilisateurs, les tâches, l'équipement et l'environnement" (ISO 9241-11,1998). Durant nos expérimentations, nous nous sommes basés sur le contexte d'utilisation suivant :

- ✓ *Les utilisateurs* sont :
  - Des étudiants de master et doctorat, ingénieurs et enseignants qui vont jouer le rôle des utilisateurs du SAI. Etant donné que lors de cette étude de cas, notre objectif principal ne s'adresse pas en premier lieu à l'évaluation du Système d'Aide à l'Informations

voyageurs (SAI), mais plutôt il porte sur l'étude de faisabilité de notre approche proposée, nous avons impliqué des utilisateurs qui ne connaissent pas le SAI.

- Un expert en évaluation des systèmes de supervision du trafic urbain dont le rôle est de fixer les seuils des mesures associées aux indicateurs d'utilisabilité.
- Un évaluateur pour conduire l'évaluation et analyser les résultats obtenus.

Notons que pendant notre expérimentation, nous avons agi en tant qu'évaluateur et modérateur de tests des participants au cours de l'exécution des scénarios expérimentaux.

- ✓ *Les tâches* sont définies sur la base d'un scénario à partir duquel les utilisateurs doivent utiliser le système SAI pour résoudre un problème de déraillement d'un Tramway ;
- ✓ *L'équipement* est un PC et l'application WIMP de la supervision du trafic SAI, et ;
- ✓ *L'environnement* est un laboratoire de recherche. Nos tests d'évaluation se sont déroulés dans le laboratoire LAMIH à l'Université de Valenciennes et du Hainaut Cambrésis.

- **Les caractéristiques des participants**

Pendant l'expérimentation, 10 utilisateurs ont participé en tant qu'utilisateurs du système SAI. L'échantillon des utilisateurs est constitué de cinq femmes (50%) et cinq hommes (50%) âgés entre 27 à 50. Ils ont des profils différents, comprenant des étudiants de master, de doctorat, des ingénieurs et des enseignants en sciences informatiques ou d'autres domaines de l'Université de Valenciennes et du Hainaut Cambrésis (France).

Les expériences des utilisateurs en utilisant le système à évaluer ont été capturées via l'outil SOUSEval à l'aide d'une échelle ordinale : novice, moyen et expert. Quant à leur expertise par rapport à l'utilisation de la technologie, elle a été déterminée via l'échelle suivante : faible, moyenne et bonne. Certains participants ont déjà utilisé ce système (SAI) lors d'autres évaluations, ce qui justifie le pourcentage de 30% qui déclare une expérience moyenne avec le système.

La Figure 5.2 présente les informations sur les utilisateurs.

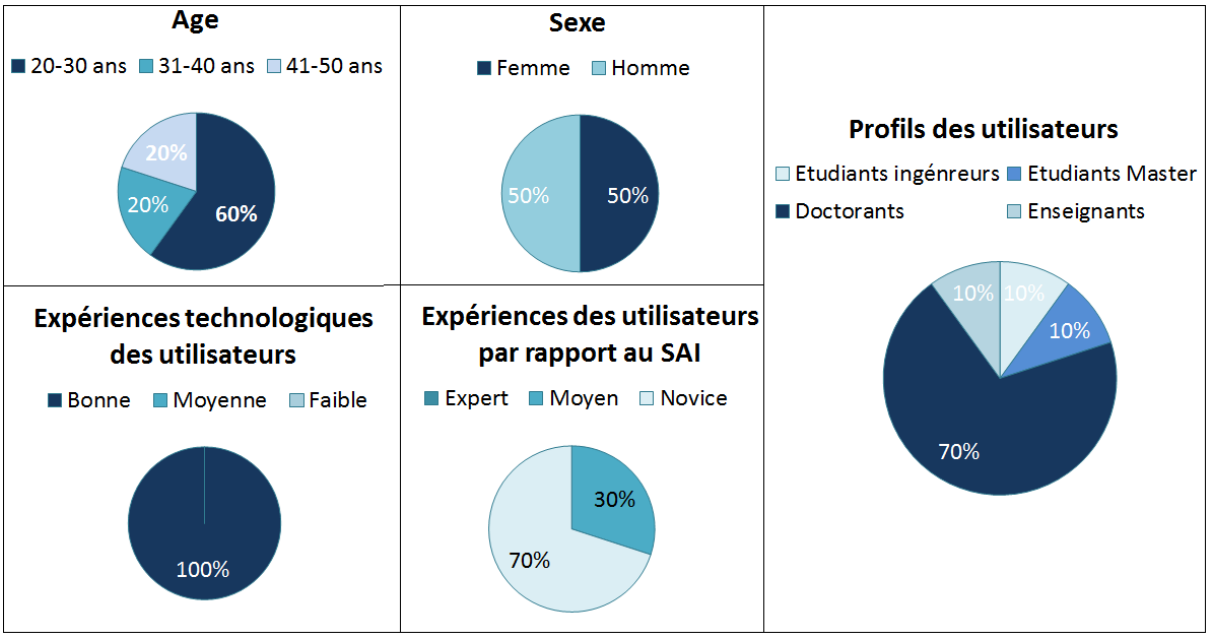


Figure 5. 2 : Les informations sur les utilisateurs

- **La liste des tâches à réaliser pendant la session de test : Scénario d'évaluation**

Les participants ont été invités à agir au tant qu'un régulateur du réseau du trafic urbain pour exécuter un scénario expérimental composé d'un ensemble de tâches. Celui-ci concerne une situation spécifique de l'utilisation du système SAI dans laquelle nous allons traiter le cas de déraillement d'un Tramway qui présente un des incidents techniques les plus constatés sur le réseau.

Ce scénario a été inspiré d'un cas réel qui a été annoncé en ligne par l'équipe UNSA Soléa « Le saviez-vous ? » le 12 Avril 2014 : « *Suite à un incident technique, le Tramway a déraillé et a été découpé en deux. L'incident qui n'a pas fait des victimes n'a pas seulement bloqué la presque totalité du trafic Tram ce matin, mais il a aussi endommagé une rame, et causé des retards sur les lignes de bus qui passent d'habitude à côté du lieu de l'incident* ».

Dans ce cas, les utilisateurs du SAI doivent interagir rapidement pour gérer les trafics, signaler les perturbations, prévenir les conducteurs des Tramways et de bus, les voyageurs dans les stations et dans les Tramways et les bus, les informer des retards inattendus qui peuvent se produire et endommager leurs circulations.

Pour gérer cette situation, nous avons demandé aux participants pendant leurs sessions d'évaluation d'exécuter l'ensemble des tâches présentées dans le Tableau 5.1.

Tableau 5. 1 : Le scénario des tâches de l'étude de cas

Un problème technique est survenu dans le TRAM 3, vous devez :	
Tâche 1 :	Signaler cette perturbation à tous les véhicules en circulation **Tram 7** et **Tram 1** et les stations situées en aval du véhicule incidenté du *Dutemple* à *Clémenceau* : « *Suite à un incident technique, le terminus est **Gare_ SNCF*** »
Tâche 2 :	Informer le conducteur du véhicule endommagé **Tram 3** sur l'intervention du service de dépannage. *« Le service de dépannage arrive dans 10 minutes »*
Tâche 3 :	Envoyer un message aux passagers de la station « **Hôtel de ville** » : *« Trafic interrompu »*
Tâche 4 :	Envoyer un message à toutes les stations, situées après l'incident *d'Hôtel-de_ville* à *Université*, pour prévenir les passagers. *« Suite à un incident technique. Pas du prochain Tramway. Des bus Relais-Tram seront à votre disposition. »*

- **La méthode de conception de test**

En se basant sur (Rubin et Chisnell, 2008), nous décrivons dans cette section la méthode de conception de notre plan de test suivi lors de la conduite des tests d'utilisabilité.

a) <u>*Méthodologie*</u> *:* Ces tests d'utilisabilité visent à capturer un ensemble des mesures d'utilisabilité relatives à un ensemble des critères de qualité pour assurer l'évaluation du système SAI. Nous avons collecté d'une part à travers le mouchard électronique EISEval (version étendue) des données qui portent sur l'interaction entre l'utilisateur et le système SAI pendant leur exécution du scénario d'évaluation. D'autre part, nous avons recueilli leurs avis et perceptions vis-à-vis du système SAI en utilisant un questionnaire d'utilisabilité.

b) <u>*Conception de test expérimental*</u> *:* Toutes les sessions de tests relatives à tous les participants ont été réalisées dans une salle de recherche au laboratoire LAMIH à l'Université de Valenciennes en respectant le même contexte d'utilisation. Chaque participant a exécuté le

scénario de tâches mentionné précédemment. Nous avons invité jusqu'à 10 participants. Chacun d'entre eux est impliqué dans une session de test qui a duré 50 minutes au total.

Comme illustre la Figure 5.3, les 10 premières minutes ont été réservées pour l'introduction à la session de test au participant (accueil, description des objectifs de l'évaluation et présentation globale du système à évaluer (SAI)). Avant de commencer la session, 10 autres minutes ont été accordées pour la phase d'apprentissage et d'essai libre du SAI. La session du test a duré 30 minutes dans lesquelles au maximum 15 minutes sont accordées à l'utilisateur afin d'exécuter leur tâches du scénario de test. Les 15 minutes restantes sont destinées aux réponses au questionnaire d'utilisabilité à la fin de la session du test.

Avant de commencer les vrais tests, nous avons effectué une phase de pré-test (d'essai) avec deux utilisateurs afin de confirmer la planification en termes de temps.

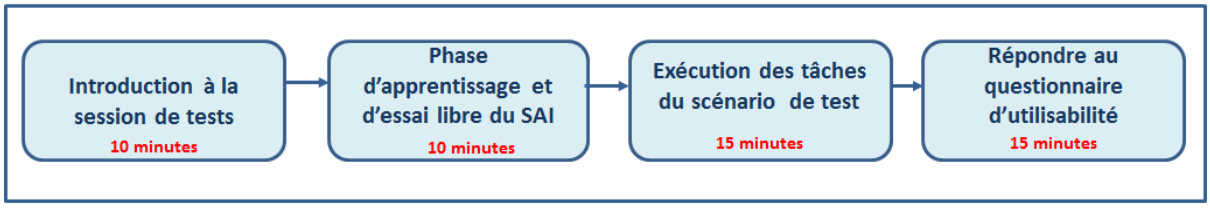


Figure 5. 3 : Plan d'une session de test d'utilisabilité

5.1.3 Exécution de l'expérimentation en utilisant l'outil SOUSEval

Après avoir défini la procédure expérimentale de notre étude de cas, nous sommes passés à la phase de conduite d'expérimentation. L'évaluation du SAI est effectuée via l'environnement intégré complet en s'appuyant principalement sur l'outil SOUSEval. Pour ce faire, nous avons suivi les trois étapes essentielles mises en œuvre dans l'outil SOUSEval qui portent sur la préparation, l'évaluation et l'analyse des résultats d'évaluation intégrée (*Chapitre 4, Section 4.5*).

Dans ce qui suit, nous présenterons chaque étape effectuée.

- **La préparation de l'évaluation du système SAI**

Nous avons commencé par la première étape qui consiste à la préparation de notre évaluation. Comme nous l'avons présenté dans le chapitre 4 (*Section 4.5*), cette phase comprend quatre sous-étapes. Nous avons spécifié d'abord les informations générales de l'évaluation en indiquant notre objectif visé (étude de faisabilité des indicateurs d'utilisabilité proposés (*Cf. chapitre 3*)). En se basant sur celui-ci, nous avons sélectionné les six critères d'utilisabilité pour lesquelles nous avons proposé des indicateurs d'utilisabilité (l'efficacité, l'efficience, la densité informationnelle, les actions minimales, le feedback immédiat et la lisibilité). Par la suite, nous avons sélectionné un ensemble de questions associées à ces critères. Enfin, nous avons identifié les dix utilisateurs à impliquer dans les sessions de tests d'utilisabilité et nous avons précisé leurs rendez-vous d'évaluation. La Figure 5.4 représente des aperçus de quelques interfaces du module de préparation qui portent sur les critères et les questions spécifiées lors de l'évaluation du système SAI.

- **L'exécution de l'évaluation du système SAI**

Dans la deuxième étape, nous avons abordé la conduite d'évaluation. En suivant les trois phases d'évaluation décrites dans le Chapitre 4 (*Section 4.5.2*), nous avons commencé par l'exécution de l'évaluation ergonomique des interfaces utilisateurs du SAI via le sous-module d'*évaluation ergonomique*. Suite à l'exécution de cette phase qui a permis l'analyse de code source du SAI, l'outil SOUSEval a capturé un ensemble des mesures d'utilisabilité relatives à chaque interface du SAI. Ces mesures portent sur les critères ergonomiques suivants : la lisibilité, la densité informationnelle et le feedback immédiat. La Figure 5.5 illustre un aperçu des résultats de mesures capturées de l'interface Agent_Etat_Ligne_Tram du SAI.

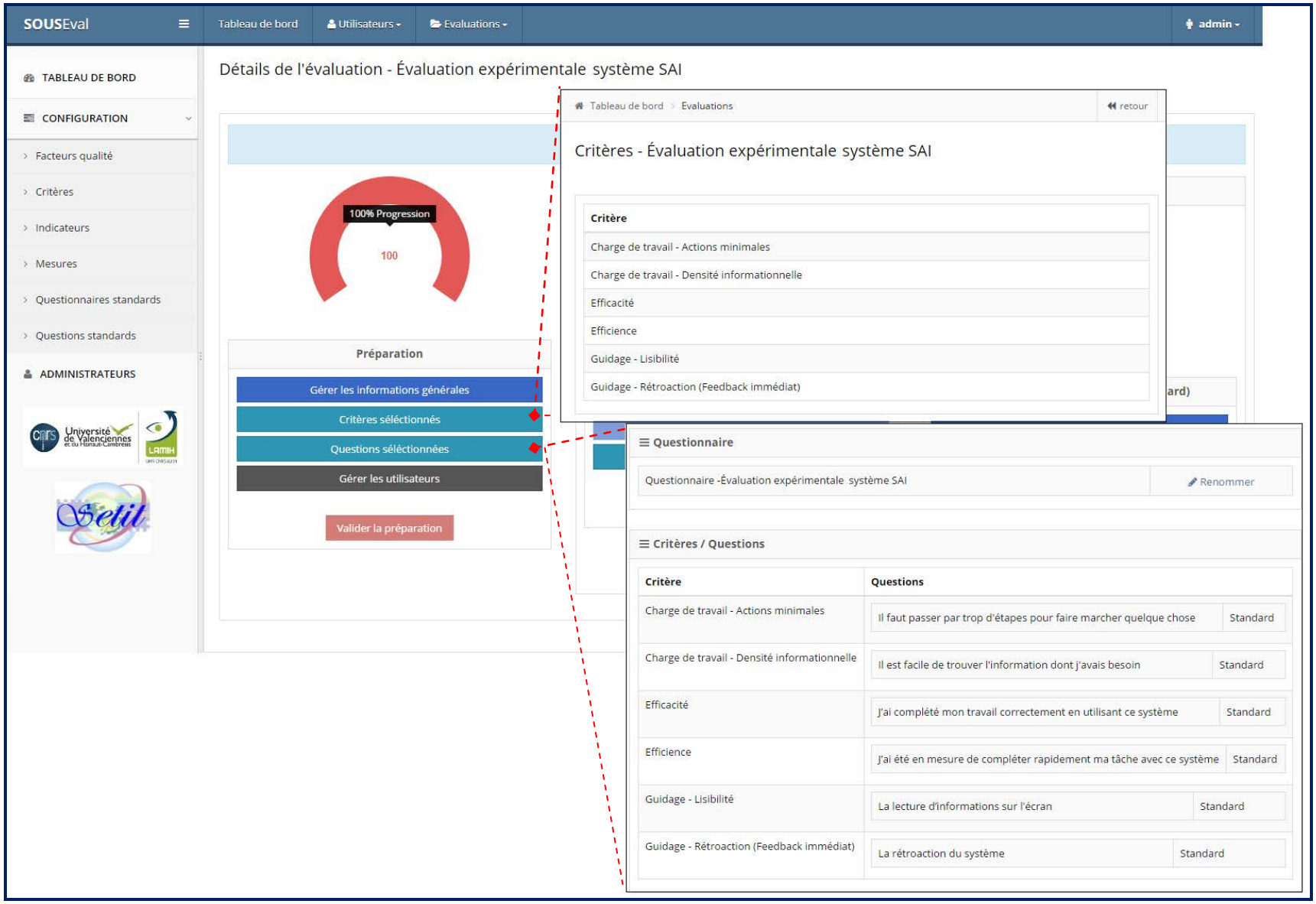


Figure 5. 4 : Aperçus de quelques interfaces du module de préparation relatives à notre étude de cas

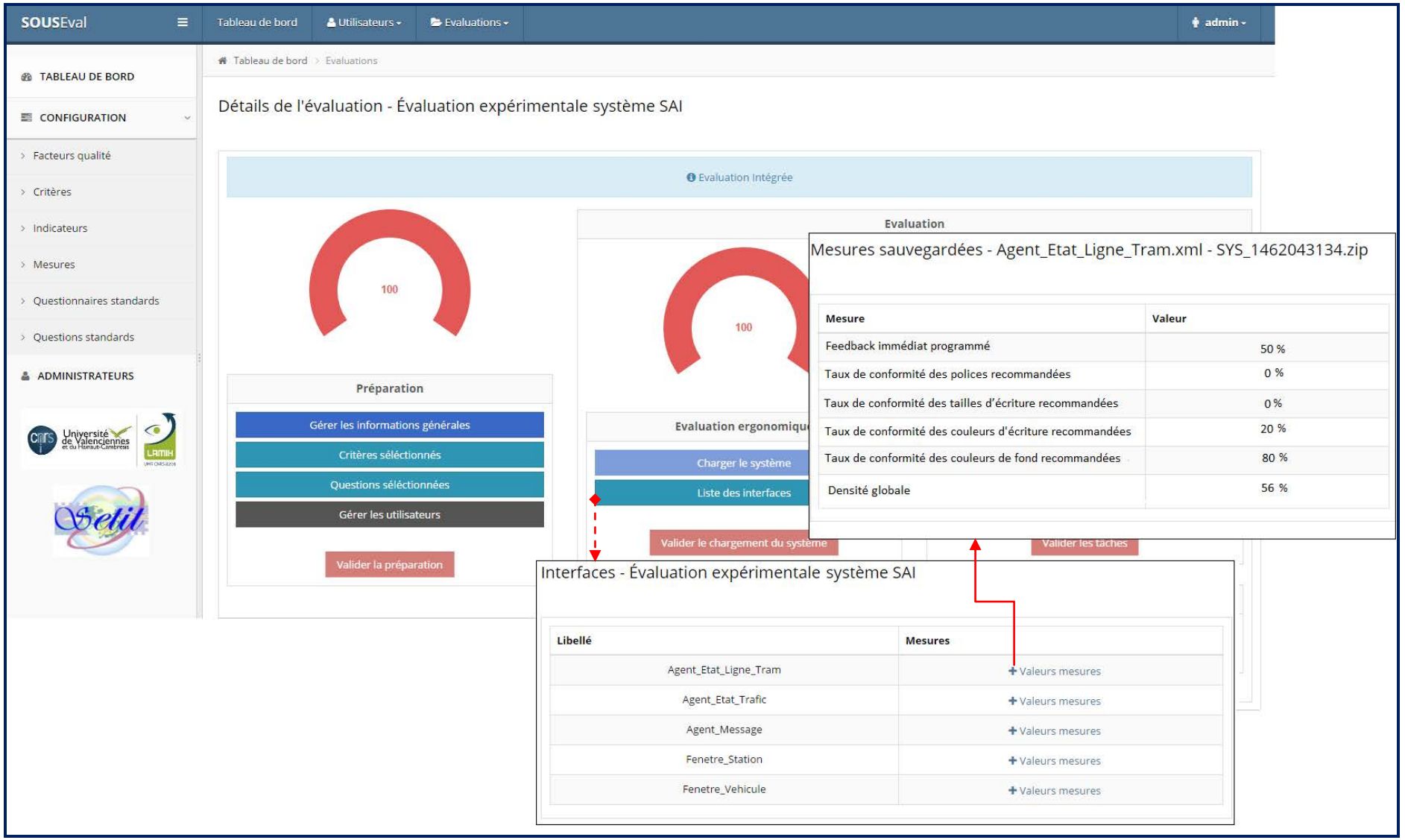


Figure 5. 5 : Aperçu des résultats de mesures capturées de l'interface Agent_Etat_Ligne_Tram

A l'issue de la première sous-étape de l'évaluation ergonomique des interfaces utilisateur du SAI, nous avons impliqué un utilisateur expert du SAI pour exécuter l'ensemble des tâches spécifiées du scénario défini pendant nos expérimentations. Par conséquent, cela nous a permis de capturer via la version étendue du mouchard électronique EISEval toutes les interactions de l'expert, ainsi que les mesures d'utilisabilité qui ont été enregistrées dans des fichiers xml. Ces mesures portent sur les critères d'efficacité, d'efficience, de feedback immédiat et d'actions minimales. En se basant sur ces fichiers capturés relatifs à chaque tâche (que nous les considérons comme étant de fichiers des références), nous avons passé par la suite à la spécification de chaque tâche dans l'outil SOUSEval via le sous-module *d'évaluation d'interactions*. Comme l'illustre la Figure 5.6, cela nous a permis de récupérer pour chaque tâche les résultats de deux mesures (nombre d'actions de l'expert et le temps de réalisation de tâche de l'expert). Après la validation de ces tâches, nous avons lancé le questionnaire d'utilisabilité déjà spécifié lors de la phase de préparation.

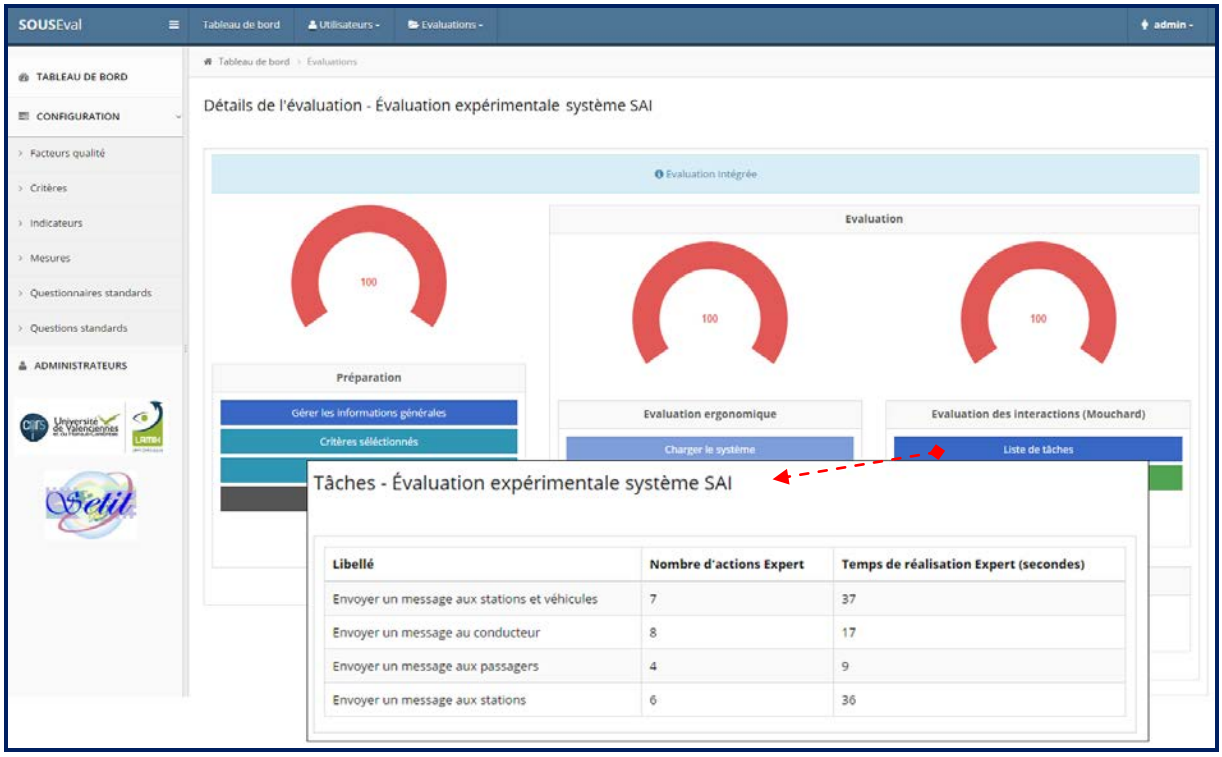


Figure 5. 6 : Aperçu de l'interface de la liste des tâches spécifiées via l'outil SOUSEval

En suivant le plan de test spécifié lors de la phase de planification d'expérimentation, nous avons démarré les sessions de tests d'utilisabilité en impliquant les dix utilisateurs. Durant chaque session de test, l'utilisateur exécute l'ensemble des tâches du scénario de l'expérimentation. Par la suite, il répond au questionnaire spécifié. A la fin de l'évaluation, et après avoir récupéré les fichiers de résultats des interactions de tous les utilisateurs via la version étendue du mouchard électronique EISEval, nous les avons transmis au sous-module *d'évaluation des interactions* dans notre outil SOUSEval.

- **La conclusion de l'évaluation du système SAI : analyses des résultats**

Après avoir achevé la phase de la conduite de l'évaluation qui a permis de capturer les mesures d'utilisabilité ainsi que les réponses des utilisateurs, nous sommes passés à la phase de conclusion de l'évaluation du SAI. Cette étape est effectuée via le sous-module appelé *indicateurs d'utilisabilité*. Elle consiste à analyser l'ensemble des résultats obtenus associés à chaque indicateur qui porte sur chaque critère d'utilisabilité sélectionné afin de détecter les problèmes et les niveaux de qualité du SAI.

Avant de procéder à l'analyse des résultats des indicateurs, nous avons impliqué l'expert en évaluation des systèmes de supervision du trafic urbain afin de sélectionner les seuils des mesures associées aux indicateurs d'utilisabilité.
Dans ce qui suit, nous présenterons les résultats et les analyses générés obtenus pour chaque indicateur d'utilisabilité sur la base des seuils fixés par l'expert lors de l'évaluation du SAI.

a) L'indicateur d'efficacité de réalisation des tâches du SAI

La Figure 5.7 présente les résultats de l'indicateur d'efficacité de réalisation des tâches (*Cf. Chapitre 3*) qui consiste à évaluer le critère d'efficacité des tâches réalisées par le système SAI. Cet indicateur affiche les résultats des utilisateurs qui intègrent la mesure de *Réponse à la question* (extraite de questionnaire d'utilisabilité) avec la mesure de *Taux moyen de réalisation des tâches d'un utilisateur* (extraite de la version étendue du mouchard électronique EISEval). Comme l'illustre cette figure, le résultat de corrélation entre les deux mesures indique une bonne corrélation (r = 0,82).

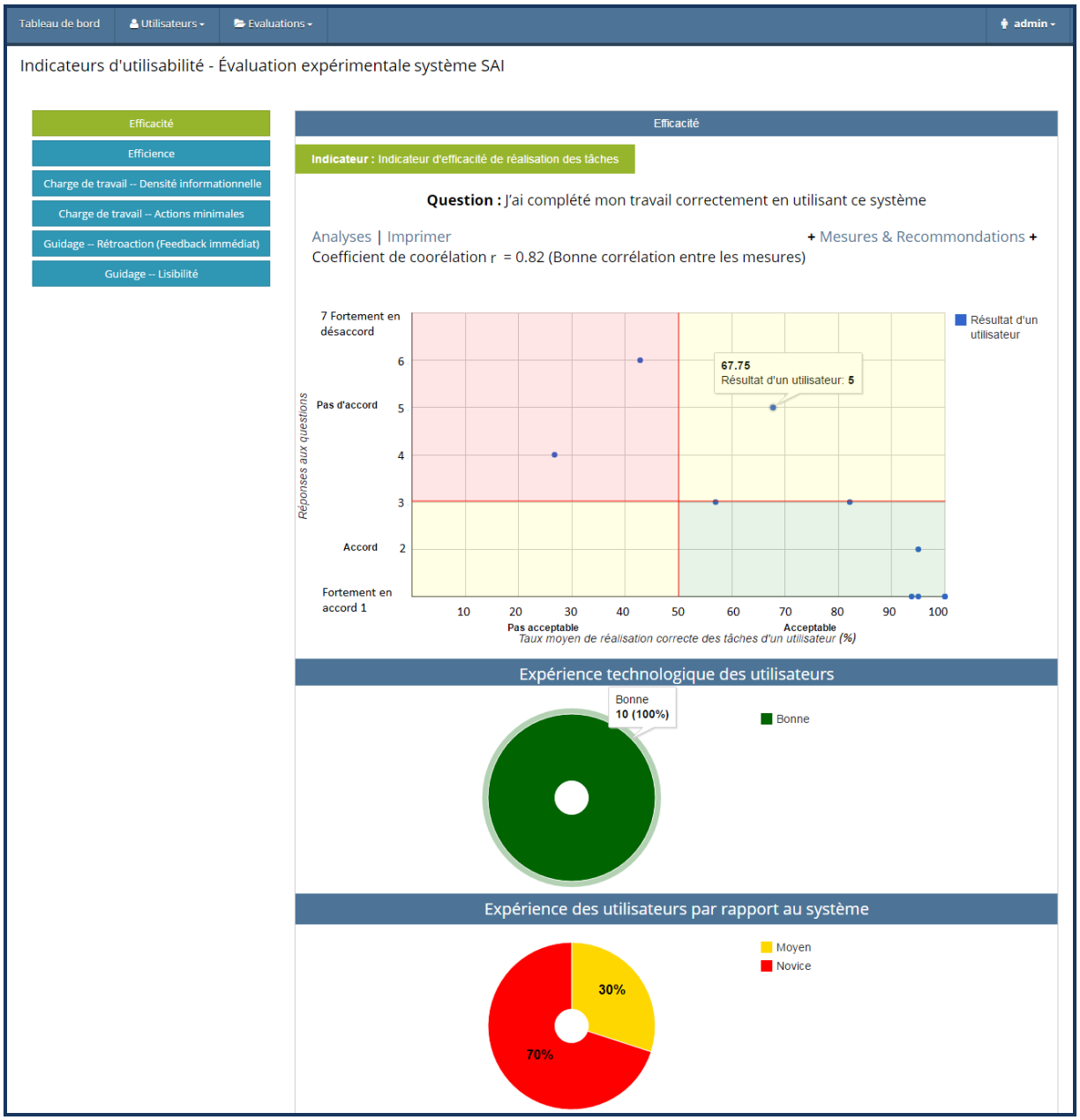


Figure 5. 7 : Aperçu de l'interface des résultats de l'indicateur d'efficacité de réalisation des tâches du SAI

Considérant la bonne corrélation entre les deux mesures, nous sommes passés à la phase d'analyse des résultats de l'indicateur.

Nous notons que l'expert a choisi la valeur médiane 50% comme un seuil pour la mesure du *Taux moyen de réalisation des tâches d'un utilisateur*. Il a considéré la valeur 3 comme un seuil pour la mesure de *réponse à une question*. Son choix a été justifié par la considération des études antérieures sur l'évaluation du même système (Trabelsi et Ezzedine, 2013) ainsi que le contexte d'utilisation pour l'évaluation : l'environnement (en laboratoire), l'équipement de test (PC et l'application WIMP), et les informations sur les profils des utilisateurs, leurs bonnes expériences de l'utilisation de la technologie, et leurs faibles expériences de l'utilisation du système SAI (70% des utilisateurs ne disposent pas d'expérience avec le système évalué et 30% d'entre eux l'ont utilisé au moins une fois), Figure 5.7.

Comme l'illustre la Figure 5.8, les analyses et les recommandations générées sur la base des seuils fixés par l'expert indiquent que la majorité des résultats des utilisateurs (avec une proportion de 70%) sont situés entre *Accord* et *Acceptable*, 20% des résultats sont situés entre *Pas d'accord* et *Pas acceptable* et 10% entre *Pas d'accord* et *Acceptable*. Cela indique un bon niveau de qualité de l'IHM en globale. Autrement dit, ces résultats prouvent aussi que 80% des utilisateurs ont exécuté leurs tâches d'une façon correcte. Néanmoins, l'insatisfaction de la minorité des utilisateurs (30%) peut être expliquée dans notre cas par le fait que l'évaluation subjective est probablement influencée par les participants impliqués qui ne sont pas des utilisateurs habituels du SAI. De plus, ce type de système de supervision du trafic urbain n'est pas facile pour l'utilisateur qui devrait habituellement avoir réçu une formation particulière.

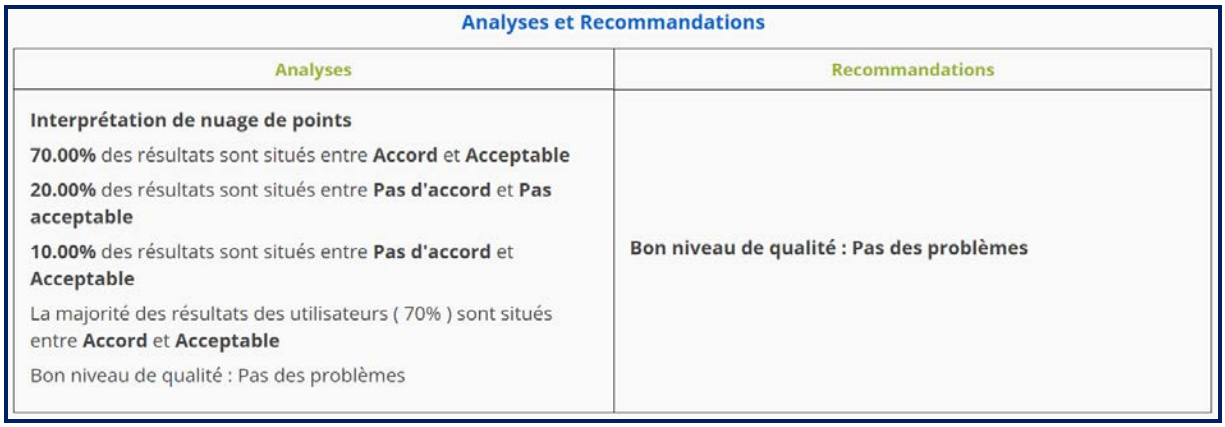


Figure 5. 8 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur d'efficacité de réalisation des tâches du système SAI

b) L'indicateur d'efficience de réalisation des tâches du SAI

La Figure 5.9 présente les résultats de l'indicateur d'efficience de réalisation des tâches (*Cf. Chapitre 3*) qui consiste à évaluer le critère d'efficience des tâches réalisées par le système SAI. Cet indicateur affiche les résultats des utilisateurs qui intègrent la mesure de *Réponse à la question* (extraite de questionnaire d'utilisabilité) avec la mesure de *Taux moyen d'efficience des tâches d'un utilisateur* (calculé par l'outil SOUSEval dont leurs mesures de base sont extraites à partir de la version étendue du mouchard électronique EISEval). Comme l'illustre cette figure, le résultat de corrélation entre les deux mesures indique une bonne corrélation (r = 0,64).

Comme l'illustre la Figure 5.9, concernant les seuils des mesures, l'expert a gardé la valeur 3 comme un seuil pour la mesure de *réponse à une question* compte tenu des informations sur les profils des utilisateurs, leurs faibles expériences en utilisant le système SAI (70% sont des novices) ainsi que leurs bonnes expériences par rapport à la technologie. Notons que notre outil SOUSEval permet de

déterminer le taux d'efficience d'utilisateur expert du SAI ainsi que le taux théorique qui a été déterminé à partir des recommandations d'Albert et Tullis (2013) (voir *Chapitre 3, Tableau 3.8*). De ce fait, pour définir le seuil de *Taux moyen d'efficience des tâches d'un utilisateur*, l'expert a considéré la recommandation de la valeur du taux théorique. Celui-ci est calculé en divisant le *Taux moyen de réalisation des tâches de l'expert* par le double du *Temps moyen d'exécution des tâches par l'expert.*

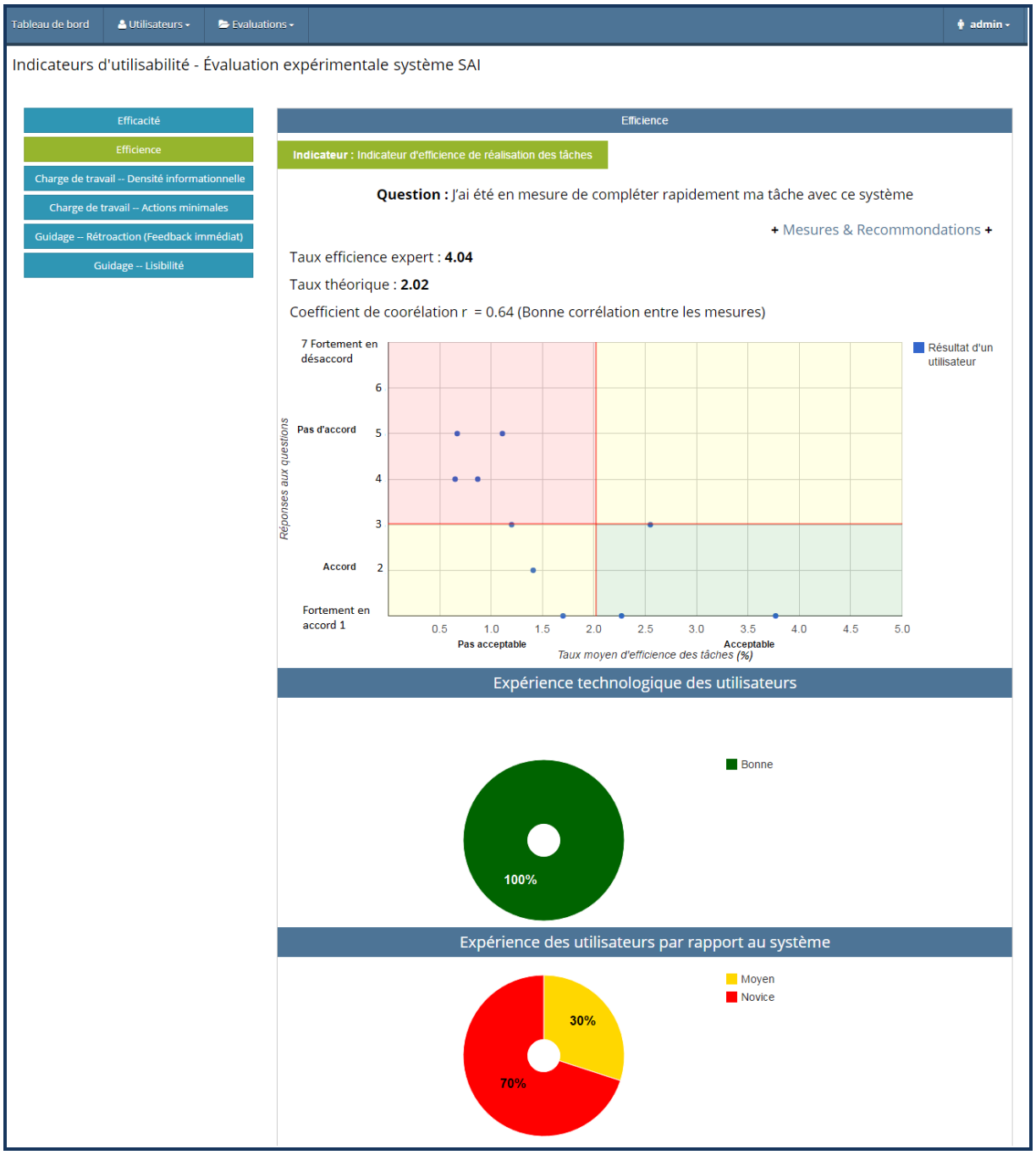


Figure 5. 9 : Aperçu de l'interface des résultats de l'indicateur d'efficience de réalisation des tâches du SAI

Comme l'illustre la Figure 5.10, les analyses et les recommandations générées sur la base des seuils fixés par l'expert indiquent que la majorité des résultats (70%) sont situés entre *Pas d'accord* et *Pas acceptable* (avec une proportion de 40%) et 30% entre *Accord* et *Pas acceptable*. En outre, ces analyses montrent que seulement 30% des résultats sont positionnés entre *Acceptable* et *Accord*. De ce

fait, l'indicateur a signalé un très mauvais niveau d'efficience de réalisation des tâches, cela signifie l'existence des problèmes relatifs aux deux mesures associées à l'indicateur. En outre, ces résultats montrent que les utilisateurs ont éprouvé des difficultés pour accomplir les tâches d'une manière efficace dans une courte période, ce qui est considéré comme acceptable puisque les participants impliqués ne sont pas les vrais utilisateurs du système SAI. De plus, l'indicateur génère quelques recommandations pour remédier aux problèmes détectés et il présente un graphique détaillé d'efficience par rapport à chaque tâche. Celui-ci (voir Figure 5.10) montre les résultats du *Taux moyen d'efficience des utilisateurs pour une tâche* par rapport aux deux mesures suivantes : le taux d'efficience d'une tâche réalisée par l'expert et le Taux théorique d'efficience d'une tâche[19].

Ce graphique montre que les valeurs de la mesure du *taux moyen d'efficience des utilisateurs* pour toutes les tâches sont inférieures à leurs valeurs cibles indiquées par l'expert. En outre, nous notons que pour les deux tâches (*Envoyer un message aux stations et aux véhicules* et *Envoyer un message aux stations*), les taux moyens d'efficience des utilisateurs sont plus proches de leurs valeurs cibles tandis que les deux autres tâches ont d'écarts plus élevés. Si notre évaluation a été effectuée par des utilisateurs typiques, cela pourrait indiquer l'existence des forts problèmes d'utilisabilité liés à l'efficience de réalisation de certaines tâches (*Envoyer un message au conducteur* et *Envoyer un message aux passagers*) et que probablement la ré-conception des interfaces utilisateur du SAI devrait être nécessaire.

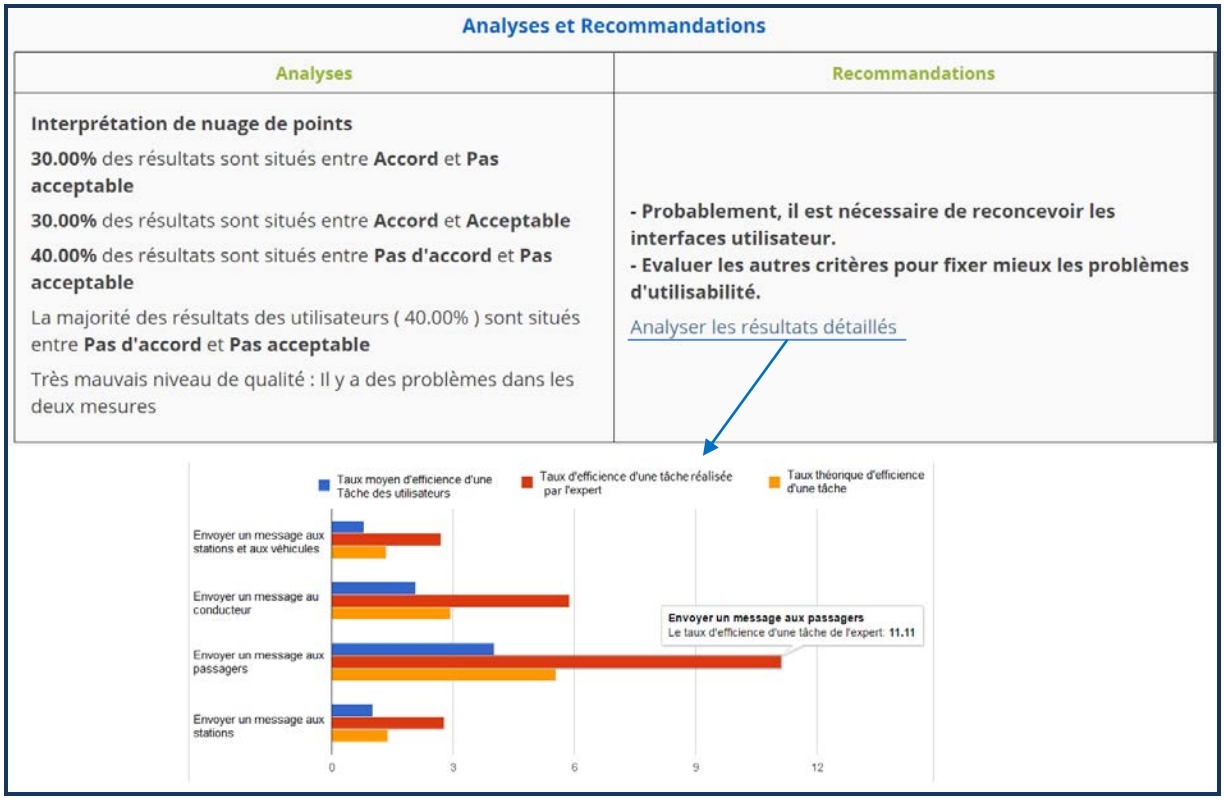


Figure 5. 10 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur d'efficience de réalisation des tâches du système SAI

c) L'indicateur de densité informationnelle des interfaces utilisateurs du SAI

La Figure 5.11 présente les résultats de l'indicateur de densité informationnelle (*Cf. Chapitre 3*) qui consiste à évaluer le critère de densité informationnelle des interfaces utilisateur du SAI. Cet indicateur génère les résultats de chaque interface utilisateur du SAI. Ceux-ci intègrent les résultats de

[19] Le taux théorique d'efficience d'une tâche = Taux de réalisation correcte d'une tâche de l'expert/ (Temps d'exécution de la tâche par l'expert * 2).

la mesure de la *Densité globale* d'une interface (extraite de la version étendue de l'inspecteur des règles ergonomiques via SOUSEval) avec la mesure de *Pourcentage des utilisateurs ayant des réponses positives de la question* (déterminée sur la base des réponses des utilisateurs capturées via le questionnaire d'utilisabilité). Afin de déterminer les résultats de cette mesure, l'expert en évaluation a considéré les 3-top réponses positives de la question relative à la densité informationnelle comme l'illustre la Figure 5.11. Compte tenu du contexte de l'évaluation, son choix a été justifié aussi par la faible expérience des utilisateurs en utilisant le système SAI.

Nous notons que l'expert a choisi 25% comme un seuil pour la *densité globale* sur la base des recommandations et des règles ergonomiques à propos de la densité globale pour les interfaces graphiques (Tullis, 1983 ; Vanderdonckt et Gillo, 1994 ; Galitz, 2007), ainsi que l'équipement de test (le PC et l'application WIMP). Pour la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question*, l'expert a choisi la valeur du milieu 50% comme un seuil à partir duquel les résultats sont considérés comme acceptables compte tenu des différents aspects du contexte d'utilisation adopté.

Sur la base de ces seuils, notre outil SOUSEval a généré l'ensemble des analyses qui portent sur chaque interface. Celles-ci déterminent les niveaux de qualité des interfaces utilisateur et fournissent des recommandations pour améliorer leurs qualités en cas de détection des problèmes. La Figure 5.12 présente l'ensemble des analyses et recommandations générées de l'indicateur de densité informationnelle des interfaces utilisateur évaluées du SAI.

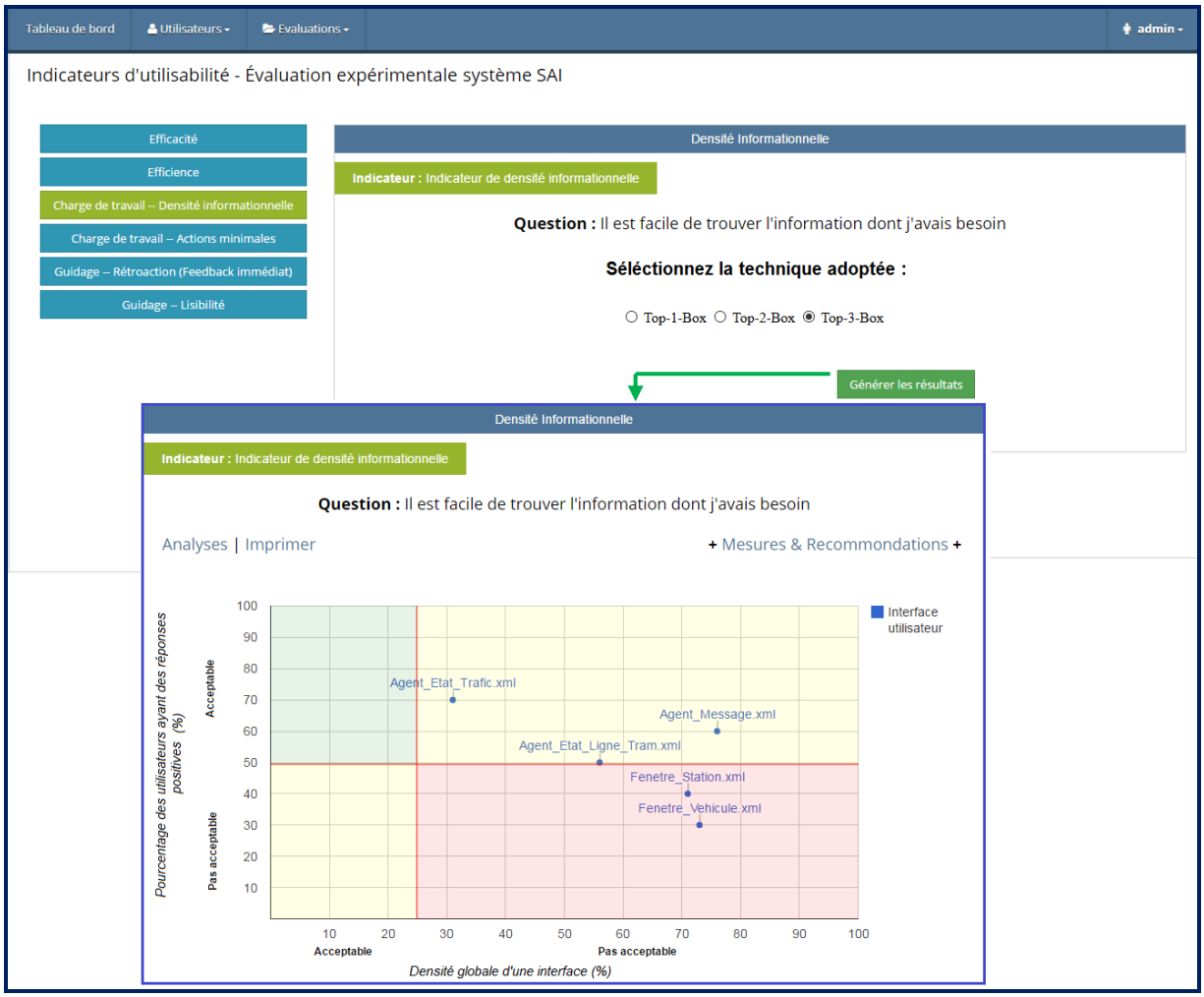


Figure 5. 11 : Aperçus des interfaces des résultats de l'indicateur de densité informationnelle des interfaces utilisateurs du SAI

Analyses des résultats des Interfaces utilisateur et Recommandations

Interface utilisateur	Niveau de qualité	Recommandations
Agent_Etat_Ligne_Tram.xml	Mauvais niveau de qualité : Il y a des problèmes de densité globale de l'interface	Des révisions sont nécessaires pour minimiser la densité élevée de l'interface.
Agent_Etat_Trafic.xml	Mauvais niveau de qualité : Il y a des problèmes de densité globale de l'interface	Des révisions sont nécessaires pour minimiser la densité élevée de l'interface.
Agent_Message.xml	Mauvais niveau de qualité : Il y a des problèmes de densité globale de l'interface	Des révisions sont nécessaires pour minimiser la densité élevée de l'interface.
Fenetre_Station.xml	Très mauvais niveau de qualité : Il y a des problèmes sur les deux mesures	Evaluer la possiblité de reconcevoir l'interface pour minimiser sa densité et faciliter la localisation de l'information par l'utilisateur
Fenetre_Vehicule.xml	Très mauvais niveau de qualité : Il y a des problèmes sur les deux mesures	Evaluer la possiblité de reconcevoir l'interface pour minimiser sa densité et faciliter la localisation de l'information par l'utilisateur

Figure 5. 12 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur de densité informationnelle sur les interfaces utilisateur du SAI

Comme l'illustre la Figure 5.12, nous notons que les trois interfaces utilisateur (*Agent_Etat_Ligne_Tram*, *Agent_Etat_Trafic* et *Agent_Message*) ont des mauvais niveaux de qualité qui sont dus à leurs densités globales élevées par rapport aux recommandations ergonomiques. Pour les deux interfaces *Fenetre_Station* et *Fenetre_Vehicule*, les analyses indiquent des très mauvais niveaux de qualité qui prouvent l'existence des problèmes relatifs aux deux mesures associées de l'indicateur. Dans ce cas, nous pourrions supposer que la faible expérience des participants en utilisant le système SAI justifie leurs difficultés à trouver les informations. Cependant, nous pensons que, dans ce cas, la faible expérience des utilisateurs confirme l'existence du problème, car il devrait être facile de trouver l'information à tout utilisateur du système (même les novices).

d) L'indicateur d'actions minimales des tâches réalisées avec le SAI

La Figure 5.13 présente les résultats de l'indicateur d'actions minimales des tâches réalisées avec le système SAI (*Cf. Chapitre 3*) qui consiste à évaluer le critère d'actions minimales. Cet indicateur génère les résultats de chaque tâche effectuée du scénario lors de l'utilisation du SAI. Il intègre les résultats de la mesure du *Pourcentage des participants réalisant les actions requises d'une tâche* (extraite à partir de la version étendue du mouchard électronique via SOUSEval) avec la mesure de *Pourcentage des utilisateurs ayant des réponses positives de la question* (déterminée sur la base des réponses des utilisateurs capturées via le questionnaire d'utilisabilité). Etant donné que la réponse *Pas d'accord* présente la réponse positive qui indique la satisfaction des utilisateurs, l'outil SOUSEval génère les valeurs de la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question* comme l'illustre la Figure 5.13.

Nous notons que l'expert a choisi les deux valeurs de milieu 50% comme seuils relatifs aux deux mesures de *Pourcentage des participants réalisant les actions requises d'une tâche* et le *Pourcentage des utilisateurs ayant des réponses positives à la question* à partir desquelles les résultats sont considérés comme acceptables sur la base des différents aspects du contexte adopté d'utilisation.

Par conséquent, l'ensemble des analyses et recommandations générées par l'indicateur, nous a permis de déterminer les différents niveaux de qualité des tâches réalisées. Comme l'illustre la Figure 5.14, nous notons des bons niveaux de qualité relatifs aux deux tâches suivantes : *Envoyer un message au conducteur* et *Envoyer un message aux stations*. Cela signifie que pour ces deux tâches, les utilisateurs ont pu réaliser les actions minimales requises effectuées par l'utilisateur expert du SAI.

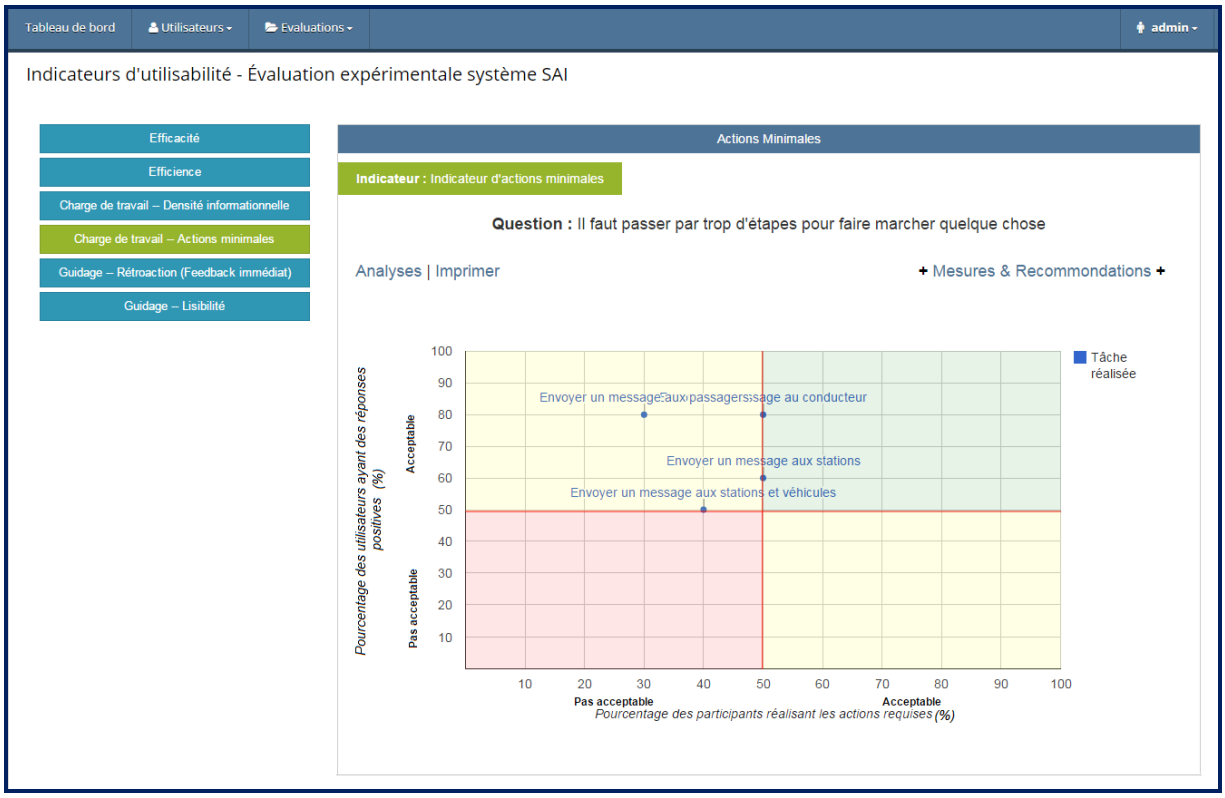


Figure 5. 13 : Aperçus des interfaces des résultats de l'indicateur des actions minimales du SAI

Analyses des résultats des tâches réalisées et Recommandations

Tâche réalisée	Niveau de qualité	Recommandations
Envoyer un message aux stations et véhicules	Mauvais niveau de qualité de tâche : Il y a des problèmes des actions minimales pour réaliser la tâche	Des révisions sont nécessaires pour améliorer les résultats
Envoyer un message au conducteur	Bon niveau de qualité de tâche : Pas des problèmes	--
Envoyer un message aux passagers	Mauvais niveau de qualité de tâche : Il y a des problèmes des actions minimales pour réaliser la tâche	Des révisions sont nécessaires pour améliorer les résultats
Envoyer un message aux stations	Bon niveau de qualité de tâche : Pas des problèmes	--

Figure 5. 14 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur d'actions minimales sur les tâches réalisées par SAI

Pour les deux autres tâches (*Envoyer un message aux stations et aux véhicules* et *Envoyer un message aux passagers*), les analyses indiquent des mauvais niveaux de qualité et l'existence des problèmes au niveau de la mesure *Pourcentage des participants réalisant les actions requises d'une tâche*. Ceux-ci sont dus aux faibles valeurs obtenues de cette mesure. Cela signifie que la majorité des utilisateurs n'ont pas pu effectuer les actions minimales requises lors de la réalisation de ces deux tâches. Avec des niveaux de satisfaction acceptables relatifs à ces deux tâches, ceci pourrait être lié à la faible expérience des utilisateurs avec le système SAI (70% sont novices). Celle-ci n'a pas pu leur permettre de suivre les chemins les plus courts pour réaliser ces tâches.

e) L'indicateur de feedback immédiat aux actions des utilisateurs avec les IHM

La Figure 5.15 présente les résultats de l'indicateur de feedback immédiat aux actions des utilisateurs avec les interfaces du système SAI (*Cf. Chapitre 3*) qui consiste à évaluer le critère de feedback immédiat. Cet indicateur génère les résultats de chaque utilisateur lors de l'utilisation du SAI. Ceux-ci

intègrent les résultats de la mesure du *Feedback aux actions d'un utilisateur* (extraite à partir de la version étendue du mouchard électronique) avec la mesure de *Réponse à la question* (extraite de questionnaire d'utilisabilité).

Etant donné que la mesure de *Feedback aux actions d'un utilisateur* consiste à vérifier la présence ou l'absence de Feedback par rapport à chaque action d'utilisateur, l'expert a suivi les recommandations qui indiquent qu'une valeur proportionnelle égale à 100% signifie que chaque action effectuée par l'utilisateur a au moins une rétroaction (voir Figure 5.15). De ce fait, il a choisi la valeur de 100% comme un seuil qui indique les résultats acceptables. Pour la mesure de *réponse à une question,* l'expert en évaluation a considéré la valeur 5 comme un seuil à partir duquel les résultats sont considérés acceptables compte tenu de la question et son échelle adoptée, ainsi que les différents aspects du contexte adopté d'utilisation (voir Figure 5.15).

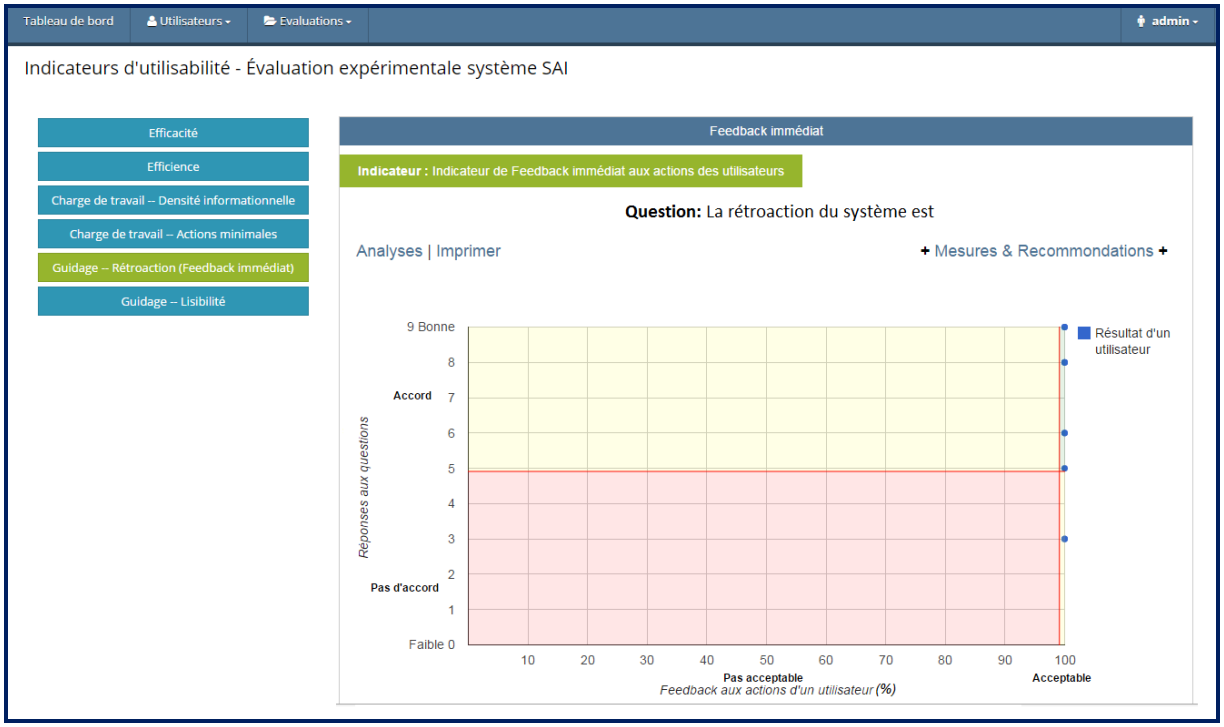


Figure 5. 15 : Aperçus des interfaces des résultats de l'indicateur de feedback immédiat aux actions des utilisateurs dans les interfaces du SAI

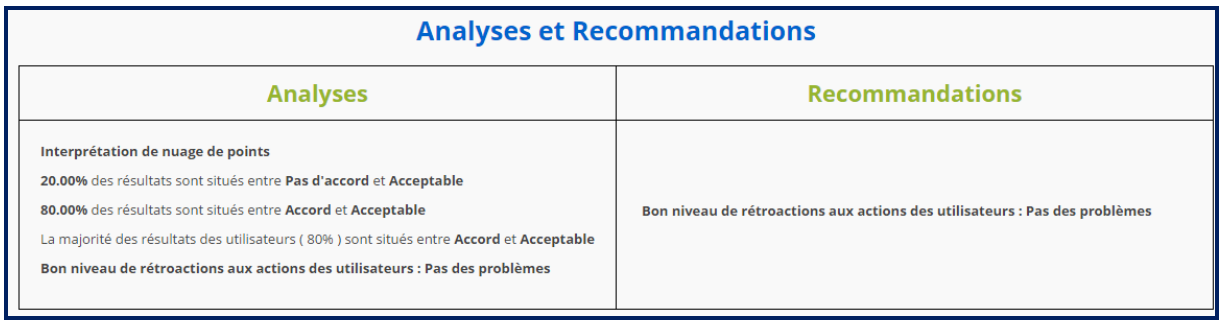


Figure 5. 16 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur de feedback immédiat aux actions des utilisateurs sur les interfaces du SAI

Comme l'illustre la Figure 5.16, les analyses et les recommandations générées sur la base des seuils fixés par l'expert indiquent que la majorité des résultats des utilisateurs (avec une proportion de 80%) sont situés entre *Accord* et *Acceptable* et 20% des résultats sont situés entre *Pas d'accord* et *Acceptable*. Cela indique un bon niveau de rétroaction immédiat par rapport aux actions des

utilisateurs. En outre, ces résultats prouvent aussi que chaque action des utilisateurs a au moins une rétroaction du système.

Néanmoins, l'insatisfaction de la minorité des utilisateurs (20%) peut être expliquée dans notre cas par le fait que l'évaluation subjective est probablement influencée par les participants impliqués qui ne sont pas des utilisateurs habituels du SAI.

f) L'indicateur de conformité des couleurs utilisées des interfaces utilisateur du SAI

La Figure 5.17 présente les résultats de l'indicateur de conformité des couleurs utilisées des interfaces du système SAI (*Cf. Chapitre 3*) qui consiste à évaluer le critère de lisibilité. Cet indicateur génère les résultats de chaque interface du SAI. Ceux-ci intègrent les résultats de deux mesures du *Taux de conformité des couleurs de fond recommandées* et le *Taux de conformité des couleurs d'écriture recommandées* (extraites à partir de la version étendue de l'inspecteur des règles ergonomiques via SOUSEval) avec la mesure du *Pourcentage des utilisateurs ayant des réponses positives à la question* (déterminée sur la base des réponses des utilisateurs capturées via le questionnaire d'utilisabilité).

En tenant compte de la question et son échelle adoptée, ainsi que la faible expérience des utilisateurs en utilisant SAI, l'expert en évaluation a considéré les 3 top réponses positives à la question pour déterminer le pourcentage de la satisfaction des utilisateurs, comme l'illustre la Figure 5.17. Cela a permis de générer les valeurs de la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question* (Figure 5.17).

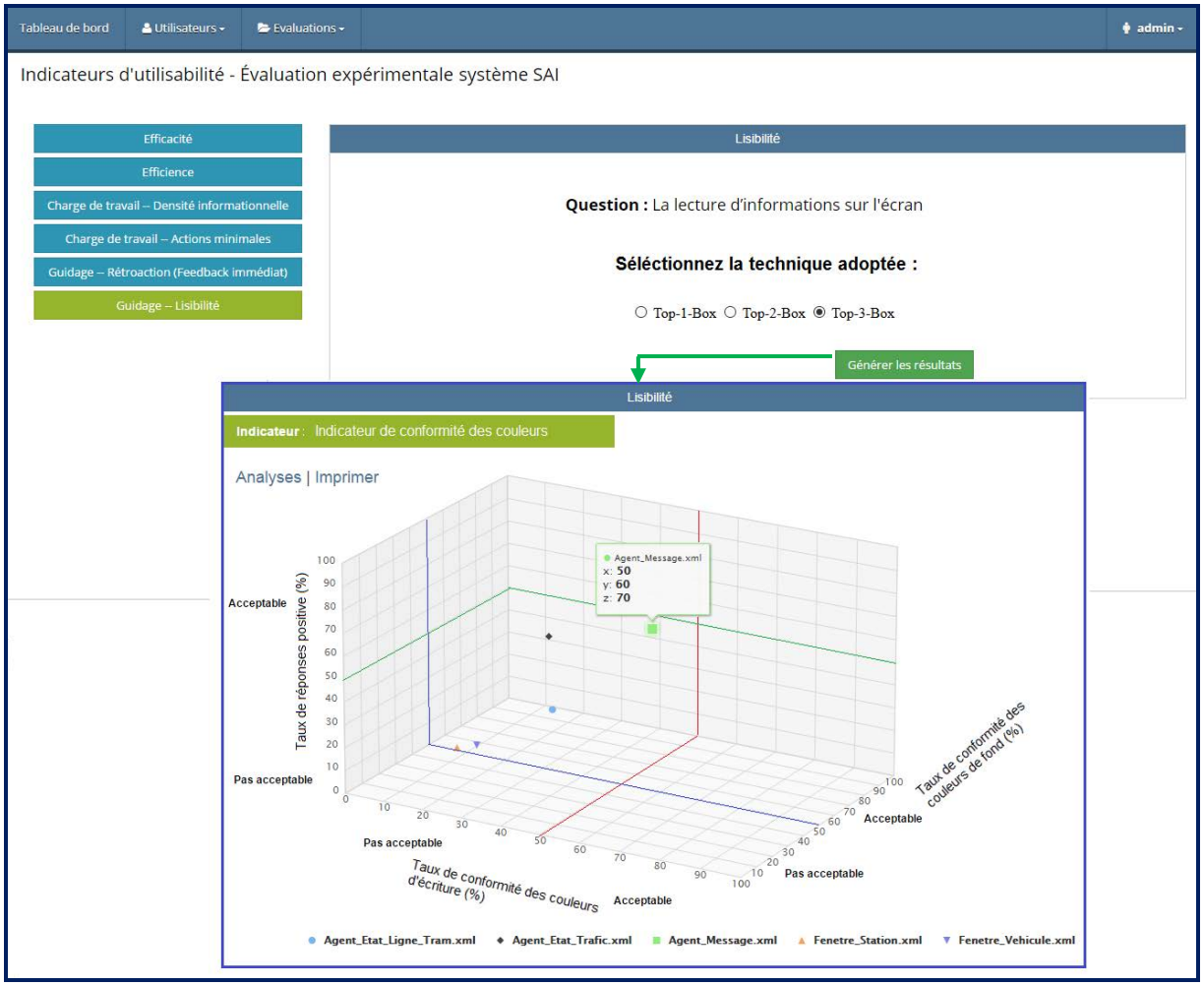


Figure 5. 17 : Aperçus des interfaces des résultats de l'indicateur de conformité des couleurs utilisées des interfaces utilisateur du SAI

Nous notons que l'expert a choisi les trois valeurs de milieu 50% comme seuils relatifs aux trois mesures à partir desquels les résultats sont considérés comme acceptables sur la base des différents aspects du contexte d'utilisation adopté.

Par conséquent, l'ensemble des analyses et recommandations générées par l'indicateur, nous a permis de déterminer les différents niveaux de qualité des interfaces utilisateur du SAI.

Comme l'illustre la Figure 5.18, nous notons un bon niveau de lisibilité par rapport aux couleurs utilisés uniquement au niveau de l'interface *Agent_Message.* Toutefois, ces analyses indiquent des mauvais niveaux de qualité pour les interfaces *Agent_Etat_Ligne_Tram* et *Agent_Etat_Trafic*. Pour l'interface *Agent_Etat_Ligne_Tram*, l'indicateur montre l'existence des problèmes qui portent sur la satisfaction des utilisateurs ainsi que la non-conformité des règles ergonomiques concernant les couleurs d'écriture utilisées. Cela nécessite des révisions pour améliorer la qualité de l'interface évaluée. Pour l'interface *Agent_Etat_Trafic,* les analyses confirment aussi l'existence des problèmes liés à la conformité des règles ergonomiques concernant les couleurs d'écriture utilisées.

Pour les deux autres interfaces, l'indicateur montre de très mauvais niveaux de lisibilité dus à l'existence des problèmes relatifs à toutes les mesures associées. Nous notons dans ce cas que l'insatisfaction des particiapnts qui ne sont pas les vrais utilisateurs du SAI confirme l'existence des problèmes. Ces interfaces requissent des révisions totales tout en respectant les règles et les recommandations ergonomiques.

Analyses des résultats des Interfaces utilisateur et Recommandations

Interface utilisateur	Niveau de qualité	Recommandations
Agent_Etat_Ligne_Tram.xml	Mauvais niveau de lisibilité: Il existe des problèmes sur la satisfaction des utilisateurs et les couleurs d'écriture utilisées de l'interface.	Des révisions sont nécessaires pour répondre aux règles ergonomiques de couleurs d'écriture. Réviser les besoins des utilisateurs. Évaluer d'autres critères
Agent_Etat_Trafic.xml	Mauvais niveau de lisibilité: Il existe des problèmes sur les couleurs d'écriture de l'interface	Certaines révisions sont nécessaires pour répondre aux règles ergonomiques de couleurs d'écriture.
Agent_Message.xml	Bon niveau de lisibilité de l'interface: Pas des problèmes	--
Fenetre_Station.xml	Très mauvais niveau de lisibilité: Il existe des problèmes dans les trois mesures	Des révisions totales sont nécessaires pour améliorer les caractéristiques textuelles en respectant les recommandations ergonomiques.
Fenetre_Vehicule.xml	Très mauvais niveau de lisibilité: Il existe des problèmes dans les trois mesures	Des révisions totales sont nécessaires pour améliorer les caractéristiques textuelles en respectant les recommandations ergonomiques.

Figure 5. 18 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur de conformité des couleurs utilisées dans les interfaces utilisateur du SAI

g) L'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur du SAI

La Figure 5.19 présente les résultats de l'indicateur de conformité des caractéristiques textuelles des interfaces du système SAI (*Cf. Chapitre 3*) qui consiste à évaluer le critère de lisibilité. Cet indicateur génère les résultats de chaque interface du SAI. Ceux-ci intègrent les résultats de deux mesures du *Taux de conformité des types de police recommandés* et *Taux de conformité des tailles de police recommandées* (extraites à partir de la version étendue de l'inspecteur des règles ergonomiques via SOUSEval) avec la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question* (déterminée sur la base des réponses des utilisateurs capturées via le questionnaire d'utilisabilité).

En tenant compte de la question et son échelle adoptée, ainsi la faible expérience des utilisateurs en utilisant SAI, l'expert en évaluation a considéré les 3-top réponses positives à la question pour déterminer le pourcentage de la satisfaction des utilisateurs, comme l'illustre la Figure 5.19. Cela a permis de générer les valeurs de la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question* (Figure 5.19).

Nous notons que l'expert a choisi les trois valeurs de milieu 50% comme seuils relatifs aux trois mesures à partir desquels les résultats sont considérés comme acceptables sur la base des différents aspects du contexte d'utilisation adopté.

Par conséquent, l'ensemble des analyses et recommandations générées par l'indicateur, nous a permis de déterminer les différents niveaux de qualité des interfaces utilisateur du SAI.

Comme l'illustre la Figure 5.20, nous notons un bon niveau de lisibilité par rapport aux caractéristiques textuelles utilisés uniquement au niveau de l'interface *Agent_Message.* Pour l'interface *Agent_Etat_Trafic,* l'indicateur indique l'existence des problèmes liés à la non-conformité des tailles de police utilisées par rapport aux recommandations ergonomiques. Cela nécessite des révisions pour améliorer la qualité de l'interface.

Figure 5. 19 : Aperçus des interfaces des résultats de l'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur du SAI

Par ailleurs, nous notons que les trois interfaces suivantes (*Agent_Etat_Ligne_Tram*, *Fenetre_Station* et *Fenetre_Vehicule*) ont des très mauvais niveaux de lisibilité. Cela indique l'existence des problèmes relatifs à toutes les mesures associées. En outre, les résultats de l'indicateur montrent des très mauvais résultats pour la mesure de Taux *de conformité des types de police recommandés* pour ces trois

interfaces (égale à 0%). Cela indique que les types de police de ces interfaces ne respectent pas les types de police recommandés. Cela peut justifier les résultats inacceptables au niveau de la mesure de *Pourcentage des utilisateurs ayant des réponses positives à la question* malgré leurs faibles expériences. Par conséquent, la majorité des utilisateurs ont éprouvé des difficultés de lecture de l'information présentée. De ce fait, ces interfaces nécessitent des révisions totales tout en respectant les règles et recommandations ergonomiques.

Analyses des résultats des Interfaces utilisateur et Recommandations

Interface utilisateur	Niveau de qualité	Recommandations
Agent_Etat_Ligne_Tram.xml	Très mauvais niveau de lisibilité: Il existe des problèmes dans les trois mesures	Des révisions totales sont nécessaires pour améliorer les caractéristiques textuelles en respectant les recommandations ergonomiques.
Agent_Etat_Trafic.xml	Mauvais niveau de lisibilité: Il existe des problèmes sur les tailles de police de l'interface	Certaines révisions sont nécessaires pour répondre aux règles ergonomiques de tailles de police.
Agent_Message.xml	Bon niveau de lisibilité de l'interface: Pas des problèmes	--
Fenetre_Station.xml	Très mauvais niveau de lisibilité: Il existe des problèmes dans les trois mesures	Des révisions totales sont nécessaires pour améliorer les caractéristiques textuelles en respectant les recommandations ergonomiques.
Fenetre_Vehicule.xml	Très mauvais niveau de lisibilité: Il existe des problèmes dans les trois mesures	Des révisions totales sont nécessaires pour améliorer les caractéristiques textuelles en respectant les recommandations ergonomiques.

Figure 5. 20 : Aperçu de l'interface des analyses et recommandations générées par l'indicateur de conformité des caractéristiques textuelles des interfaces utilisateur du SAI

- **Conclusion sur l'étude de cas**

Pour conclure sur cette étude de cas, nous synthétisons dans les Tableaux 5.2, 5.3 et 5.4, les résultats obtenus des indicateurs d'utilisabilité relatifs à l'ensemble des critères considérés lors de l'évaluation du SAI. Nous rappelons que notre objectif principal à travers cette étude de cas est la démonstration de la faisabilité des indicateurs proposés. L'évaluation du SAI nous a permis de détecter un ensemble de problèmes d'utilisabilité relatifs aux différents critères considérés (voir Tableaux 5.2, 2.3 et 5.3). Bien que nous ayons indiqué que le très mauvais niveau d'efficience pourrait être lié à la non implication des utilisateurs finaux du SAI, les analyses des indicateurs relatifs aux critères ergonomiques ont indiqué des mauvais et/ou très mauvais niveaux de qualité dans la majorité des IHM évaluées.

Cependant, ces résultats obtenus ont fortement dépendu du contexte d'utilisation adopté dans l'évaluation (utilisateurs non-finaux, technologie classique et évaluation dans un laboratoire). Cela signifie que les résultats peuvent être modifiés en utilisant d'autres contextes d'utilisation. En outre, le fait que les participants ne sont pas des utilisateurs finaux du système peut avoir un impact sur l'exactitude des résultats de l'évaluation. Cela signifie que cette session d'évaluation est insuffisante pour avoir une conclusion d'évaluation finale sur l'utilisabilité du SAI. Néanmoins, selon un expert d'évaluation pour ce type de système, les faibles expériences des utilisateurs pourraient confirmer les problèmes trouvés du fait que les utilisateurs devraient facilement lire et trouver les informations quel que soit leurs niveaux d'expérience.
En se basant sur cette application de l'évaluation du système de supervision du trafic (SAI) en utilisant l'environnement intégré, nous pouvons attester la faisabilité de l'utilisation des indicateurs d'utilisabilité pour l'intégration des résultats des évaluations subjectives et objectives d'une manière cohérente, effective et complémentaire. Ceux-ci permettent de détecter les niveaux de qualité du système évalué ainsi que de suggérer des recommandations pour remédier aux problèmes détectés.

Tableau 5. 2 : Tableau de synthèse des résultats des indicateurs d'utilisabilité pour l'évaluation globale du SAI

Critères d'utilisabilité	Efficacité	Efficience	Feedback immédiat
Niveau de qualité du SAI	Bon niveau	Très mauvais niveau	Bon niveau

Tableau 5. 3 : Tableau de synthèse des résultats des indicateurs d'utilisabilité pour l'évaluation ergonomique du SAI

Critères d'utilisabilité IU évaluées	Densité informationnelle	Lisibilité	
		Par rapport aux couleurs	*Par rapport aux caractéristiques textuelles*
Agent_Etat_Ligne_Tram	Mauvais niveau	Mauvais niveau	Très mauvais niveau
Agent_Etat_Trafic	Mauvais niveau	Mauvais niveau	Mauvais niveau
Agent_message	Mauvais niveau	Bon niveau	Bon niveau
Fenetre_Station	Très mauvais niveau	Très mauvais niveau	Très mauvais niveau
Fenetre_Vehicule	Très mauvais niveau	Très mauvais niveau	Très mauvais niveau

Tableau 5. 4 : Tableau de synthèse des résultats de l'indicateur d'actions minimales pour l'évaluation des tâches réalisées avec SAI

Critère d'utilisabilité Tâches réalisées	Actions minimales
Envoyer un message aux stations et véhicules	Mauvais niveau
Envoyer un message au conducteur	Bon niveau
Envoyer un message aux passagers	Mauvais niveau
Envoyer un message aux stations	Bon niveau

5.2 Discussion sur les risques de validité de notre étude réalisée

En s'inspirant de Wohlin et al. (2000), nous avons vérifié quatre types de risques de validité de notre étude menée, qui sont définis comme suit : la validité de construction, la validité interne, la validité externe et la validité de conclusion.

Les risques de validité de la construction découlent de la relation entre la théorie et l'observation et soulève les questions de savoir si le traitement reflète de manière adéquate les causes ; si le résultat reflète de manière adéquate les effets ; et dans notre cas, si les éléments à évaluer reflète de manière adéquate la qualité liée à l'utilisabilité. Pour minimiser ces risques, nous avons considéré un niveau de fiabilité supérieur ou égal à 0,92 pour sélectionner les meilleurs questionnaires d'utilisabilité standardisés à utiliser pour l'évaluation subjective (Lewis, 2014 ; Sauro, 2015). Plus particulièrement, nous avons utilisé les éléments des questionnaires d'une façon indépendante et individuelle, ce qui signifie que ceux-ci ne possèdent aucun degré de fiabilité globale des questionnaires. Cependant, nous avons considéré que chaque élément des questionnaires a été défini avec un but spécifique d'utilisabilité que nous avons exploré dans l'évaluation. En outre, nous avons utilisé des mesures

d'utilisabilité bien connues pour l'évaluation objective (Tullis, 1983 ; Vanderdonckt et Gillo, 1994 ; Seffah et al., 2006 ; Padda, 2003 ; Dix et al, 2004 ; ISO/IEC 25022, 2012). Nous avons supposé que ces mesures sont des bonnes mesures validées par les différents travaux qui les référencent.

Les risques de validité interne sont des influences qui peuvent affecter les valeurs des variables indépendantes en ce qui concerne la causalité à l'insu du chercheur. Dans notre étude de cas ce risque est associé aux participants impliqués dans les tests. Les participants ont été sélectionnés par convenance. Ils ont comme profils : étudiants de master, doctorant, ingénieurs et enseignants dans le même laboratoire où cette recherche est effectuée. Par conséquent, l'évaluation pourrait être biaisée par les utilisateurs lors de l'évaluation du système. Pour minimiser ce risque, nous avons invité tous les étudiants de doctorat, ingénieurs et enseignants au laboratoire pour effectuer l'évaluation et ils se sentaient libres d'accepter ou non d'une façon volontaire de participer à l'évaluation. Cependant nous ne pouvons pas contrôler ce biais et donc nous avons accepté ce risque.

Les risques de validité externe sont des conditions qui limitent la généralisation des résultats en dehors du champ de notre étude. Bien que toutes les mesures utilisées puissent être applicables dans tous les domaines, le domaine d'application actuel de notre approche est attaché aux systèmes de supervision du trafic urbain car les mesures de base sont générées à partir des outils spécifiques associés à ces systèmes. La généralisation de l'approche proposée dans d'autres domaines pourrait être réalisée. Toutefois, pour que notre outil SOUSEval soit générique, il faut généraliser les outils de l'environnement pour qu'ils puissent être appliqués à l'évaluation de n'importe quel type de systèmes. En outre, notre approche et outil réalisé sont spécifiques aux applications WIMP qui se caractérisent par les interfaces graphiques statiques. De ce fait, les modèles des mesures d'information des indicateurs proposés sont spécifiques pour cette technologie. Ils ne peuvent pas être appliqués pour évaluer tous les types des interfaces utilisateur (par exemple, vocal, audio, etc.) appliqués dans différents contextes d'utilisation. En plus, dans notre étude de cas, toutes les analyses effectuées sont établies sur la base du contexte d'utilisation et le type système interactif évalué (SAI) tel qu'il est présenté précédemment. Afin de généraliser notre travail, cela nécessite plus des études de recherche afin de générer des mesures pour d'autres types d'interfaces.

Enfin, les risques de validité de la conclusion sont celles qui affectent la capacité de tirer la bonne conclusion à propos de la relation entre le traitement et les résultats de notre étude (dans notre cas, l'intégration des résultats). Pour faire face à ce risque, d'une part pour les indicateurs d'efficacité et d'efficience, nous avons décidé de procéder à l'interprétation et l'analyse uniquement si les données fournissent un coefficient de corrélation $r \geq 0{,}50$ (c.-à-d r-squared, $r^2 \geq 0{,}25$) sur la base de (Cohen, 1988), même si certaines études ont déjà montré que, pour ce type de mesures, les mesures sont sensiblement corrélées (Sauro et Lewis, 2012, p.254-255 ; Sauro et Kindlund, 2005 ; Sauro et Lewis, 2009). Néanmoins, nous avons préféré confirmer la corrélation pour interpréter les résultats et fournir des recommandations.
D'une autre part, pour les indicateurs fondés sur les critères ergonomiques, nous avons proposé des recommandations bien connues adressées à l'évaluation de l'interface graphique utilisateur pour la définition de seuils étant donné que chaque mesure a été présentée d'une façon individuelle. Cependant, nous n'avons pas pu contrôler si l'expert en évaluation utilisera les recommandations ou non.

En plus de ces quatre types des risques, la *validité écologique* de l'étude effectuée peut être considérée comme étant un élément supplémentaire à l'évaluation. Selon (Brewer, 2000), dans la recherche, la validité écologique d'une étude signifie que les méthodes, les matériaux et le cadre de l'étude doivent se rapprocher du monde réel qui fait l'objet d'un examen. Dans notre étude, la validité écologique des tests réalisés porte sur le souci de reproduire en laboratoire le contexte réel d'utilisation du SAI. De ce

fait, pour minimiser ce risque nous avons effectué tous les tests dans la même salle de recherche en utilisant le même système SAI et en respectant les mêmes conditions d'expérimentation (e.g., une bonne luminosité, un bas niveau de bruit). Toutefois, il ne faut pas ignorer que l'environnement n'est pas réel et que nous n'avons pas pu le configurer de façon similaire à celui de la salle de contrôle réelle.
Par conséquent, nous avons accepté ce risque car notre principal objectif était plutôt de montrer la faisabilité de l'approche proposée.
Par ailleurs, Albert et Tullis (2013) ont cité sept sources de biais qui peuvent affecter une étude d'utilisabilité. Ceux-ci incluent les participants et l'environnement évoqués précédemment. Les cinq autres concernent les tâches, les méthodes d'évaluation, les artefacts, les modérateurs et les attentes.

En effet, *les tâches* choisies peuvent avoir un impact sur les problèmes identifiés. Nous nous sommes basés dans notre étude de cas sur un scénario bien défini de tâches. Il concerne une étude de cas réelle et critique qui consiste au déraillement du véhicule (tramway). Notre principal objectif était de détecter les principaux problèmes d'utilisabilité qui peuvent perturber le régulateur lors de l'exécution de ses tâches critiques. Néanmoins, nous savons que la criticité de la tâche peut influencer sur les résultats liés à l'efficacité et l'efficience, surtout dans ce cas où les participants impliqués dans l'évaluation ne sont pas les utilisateurs finaux du SAI. Nous avons assumé ce risque étant donné que nous avons considéré qu'il est important d'effectuer l'évaluation dans une situation dans laquelle la facilité d'utilisation du système est primordiale.

Les méthodes d'évaluation choisies peuvent également biaiser le type de problèmes identifiés. Ce biais a été minimisé par l'utilisation de différentes méthodes d'évaluation qui portent sur trois aspects différents d'utilisabilité liés à (i) la qualité ergonomique des interfaces utilisateur, (ii) les interactions entre les utilisateurs et le système à évaluer, et (iii) les perceptions des utilisateurs sur le système à évaluer.

Pour les *Artefacts*, Albert et Tullis (2013) argumentent que la nature du prototype ou du produit à évaluer a un impact énorme sur les résultats ; le type d'interaction peut varier, que ce soit un prototype sur papier, prototype fonctionnel ou semi-fonctionnel ou un système de production. Pour minimiser ce biais, nous avons appliqué notre approche dans l'évaluation d'un système industriel (un système de supervision du trafic urbain). Il a été développé dans un grand projet du Groupement Régional Nord-Pas de Calais pour la Recherche dans les Transports.

En outre, pour minimiser le biais de l'implication des différents modérateurs lors de la session d'évaluation, nous avons décidé d'impliquer un seul modérateur dans toutes les sessions d'évaluation. Bien que le modérateur a participé de façon impartiale, les participants ont pu se sentir influencé pendant l'évaluation. Considérant que l'évaluation de l'utilisabilité a porté sur un système développé par des tierces parties, et non pas par le modérateur, nous considérons cela comme un biais acceptable.

Enfin, concernant *les attentes* qui portent sur les zones les plus problématiques de l'interface utilisateur, nous n'en avons défini aucune. Par conséquent, les résultats ne sont pas influencés par cet aspect.

5.3 Conclusion

Dans ce chapitre, nous avons mis en pratique l'environnement intégré d'aide à l'évaluation des systèmes interactifs dans lequel nous avons implémenté notre approche proposée via l'outil réalisé SOUSEval. L'illustration du fonctionnement de l'environnement complet et l'application des indicateurs d'utilisabilité proposés ont été effectuées à travers une étude de cas. Celle–ci porte sur l'évaluation d'un système de supervision du Trafic (SAI). En se basant sur cette étude de cas, nous

avons pu montrer la faisabilité de l'utilisation des indicateurs d'utilisabilité proposés sur la base de la norme ISO/IEC 15939. Ceux-ci permettent d'assurer une bonne évaluation en intégrant les résultats d'évaluations subjective et objective d'une façon cohérente, unique et complémentaire. De plus, l'automatisation de notre approche à travers l'outil développé SOUSEval, nous a permis de la rendre plus pratique et plus aisée. Par ailleurs, un ensemble de ces indicateurs a été déjà validé à travers une autre étude de cas effectuée sur le même système évalué (SAI) (Assila et al., 2016b).

Conclusion Générale

Le travail réalisé dans cette thèse porte sur la thématique des interactions Homme-Machine et plus particulièrement sur l'évaluation de l'utilisabilité des interfaces utilisateur des systèmes interactifs.

L'évaluation est une tâche essentielle dans la production de systèmes interactifs dans n'importe quel domaine d'application. La supervision du trafic urbain est un des domaines qui exige un haut niveau de qualité du système interactif pour assurer la sécurité des transports. Le contrôle du trafic est un processus d'adéquation en temps réel des différents modes de transport à des conditions de fonctionnement réelles. Pour ce faire, l'opérateur humain interagit via une IHM avec le système pour prendre des décisions ayant un impact direct sur la régulation globale du réseau du transport public (Fayech, 2003). Généralement, la complexité des architectures de ces systèmes de supervision (par exemple, les systèmes interactifs à base d'agents) rend complexe leur évaluation (Tran et al, 2008 ; Trabelsi et Ezzedine, 2013).

Pour soutenir l'évaluation de l'utilisabilité de ces systèmes, différentes méthodes d'évaluation de l'utilisabilité peuvent être exploitées. Celles-ci se basent généralement sur des évaluations subjectives et objectives en tenant compte, respectivement, de la dépendance ou non de l'avis humain. Ces deux types d'évaluations sont complémentaires vu qu'elles fournissent des données du système évalué à partir de différentes perspectives. Grâce à cette complémentarité, plusieurs travaux ont recommandé largement leurs combinaison pour améliorer les résultats de l'évaluation (Tran et al., 2008 ; Al-Wabi et Al-Khalifa, 2009 ; Lewis, 2014 ; Charfi et al., 2015).

En suivant la même direction, Ezzedine (2002) a proposé un environnement intégré d'aide à l'évaluation de l'utilisabilité des systèmes interactifs de supervision du trafic. Cet environnement se focalise principalement sur l'intérêt d'assurer une intégration effective des résultats d'évaluations d'utilisabilité subjective et objective. Le développement de l'environnement a été débuté depuis une dizaine d'années. Les travaux de recherche effectués dans ce cadre ont permis de mettre en œuvre deux outils d'évaluation de l'environnement. Le premier est un mouchard électronique qui assure une évaluation des interactions des utilisateurs avec le système à évaluer (Tran, 2009). Quant au second, il concerne un inspecteur des règles ergonomiques qui consiste à effectuer une évaluation ergonomique des IHM du système à évaluer (Charfi, 2013).

L'objectif de cette thèse est de compléter le développement de l'environnement à travers la mise en œuvre des outils manquants et la mise en pratique de l'environnement d'une façon complète. Le premier présente un outil d'aide à la génération de questionnaire permettant d'assurer une évaluation subjective d'utilisabilité des interfaces utilisateur des systèmes interactifs. Le second présente un outil de synthèse qui vise à assurer l'intégration effective des résultats des évaluations subjective et objective. Ceux-ci sont capturés à partir de deux outils existants ainsi que l'outil d'aide à la génération de questionnaires. Pour atteindre cet objectif, l'intégration a été établie via la proposition des indicateurs d'utilisabilité définis sur la base de la norme ISO/IEC 15939.

Motivation du travail mené

Cette thèse nous a permis de compléter le développement de l'environnement intégré d'aide à l'évaluation des systèmes interactifs via la mise en œuvre de notre outil **SOUSEval** qui a contribué à :

- Assurer *le fonctionnement complet* de l'environnement en utilisant d'une façon complémentaire les trois outils suivants : le mouchard électronique EISEval, l'inspecteur des règles ergonomiques et l'outil d'aide à la génération de questionnaires.
- Inclure un outil d'aide à la génération de questionnaires favorisant *une évaluation subjective* de l'utilisabilité des interfaces utilisateur sur la base des avis des utilisateurs.

- Garantir *une évaluation d'utilisabilité intégrée subjective et objective* des interfaces utilisateur par l'implémentation des indicateurs d'utilisabilité proposés (*Cf. Chapitre 3*). Ceux-ci intègrent les résultats d'évaluations subjective et objective d'une manière cohérente et complémentaire. Cette évaluation porte sur trois aspects d'utilisabilité hétérogènes caractérisant les trois outils mentionnés auparavant : l'aspect ergonomique des interfaces utilisateur, l'aspect d'interaction entre l'utilisateur et le système à évaluer et l'aspect subjectif. Cette approche fournit une base pour l'aide à l'interprétation des résultats obtenus des indicateurs à travers des critères de décision. En outre, elle propose un processus générique à appliquer pour la construction de nouveaux indicateurs qui portent sur l'évaluation d'autres critères d'utilisabilité.
- Détecter des problèmes d'utilisabilité par rapport à des critères d'utilisabilité bien déterminés.
- Exploiter des mesures subjectives et objectives, relatives aux critères d'utilisabilité, très connues dans la littérature et mettre l'accent sur l'utilisation de questionnaires d'utilisabilité standardisés les plus connus et validés.
- Automatiser la collecte de données et mesures d'utilisabilité des trois outils ainsi que l'analyse et l'interprétation des résultats obtenus des indicateurs d'utilisabilité. Ces analyses considèrent le contexte d'utilisation de l'évaluation ainsi que les règles et les recommandations ergonomiques relatives aux interfaces utilisateur graphiques.
- Générer des recommandations pour l'aide à l'amélioration de qualité des interfaces utilisateurs des systèmes interactifs.

Comparaison de l'approche proposée par rapport aux autres existantes

Lors du Chapitre 2 (Section 2.2), nous avons présenté un ensemble des approches qui se basent sur la combinaison des évaluations subjectives et objectives. Une synthèse sur ces approches a été présentée en se référant à un ensemble des critères agissant sur l'objectif de la combinaison (voir Chapitre 2, Tableau 2.11). En outre, différents avantages et lacunes ont été énumérés, ce qui nous a permis de comparer notre approche par rapport à celles proposées.

Cette comparaison est illustrée dans le Tableau 1 portant sur les critères suivants :

- *La combinaison des évaluations subjective et objective* – l'ensemble des approches présentées dans le chapitre 2 ont assuré la combinaison des évaluations subjective et objective. D'une façon similaire, notre approche converge vers le même objectif.
- *L'intégration effective des évaluations subjective et objective* – Certaines approches n'ont pas assuré une intégration concrète des évaluations subjective et objective. Elles se limitent à présenter individuellement leurs résultats sans aucune proposition d'intégration des données. D'autres proposent l'intégration effective à travers l'agrégation de leurs résultats d'évaluations subjective et objective en une valeur globale d'utilisabilité. Généralement, ces résultats peuvent être très utiles pour comparer l'utilisabilité des différents systèmes ou pour montrer l'évolution de la qualité globale de différentes versions du système. D'autre part, cette valeur globale fournit une valeur à interpréter. Cependant, leur interprétation peut être difficile dans le cas où nous visons à déterminer le niveau de qualité lors d'une évaluation isolée d'une interface utilisateur et non pas dans le but de comparer plusieurs interfaces ou versions du système.
  Dans notre approche, l'agrégation des données en un résultat global est non bénéfique pour déterminer les niveaux de qualité et détecter les problèmes d'utilisabilité des interfaces utilisateur du système interactif évalué. Selon Albert et Tullis (2013, Chapitre 8), l'utilisation de « scorecard » présente une autre façon pour combiner les résultats d'utilisabilité. Celui-ci présente un type de graphique où différents résultats de mesures sont présentés graphiquement d'une façon

que les tendances globales et les aspects importants de données peuvent être détectées facilement. De ce fait, nous avons proposé de représenter les résultats des indicateurs d'utilisabilité en graphique du nuage des points. Cela permet de mieux illustrer l'intégration effective entre les mesures associées de l'indicateur ainsi que de rendre plus facile et compréhensible leur procédure d'interprétation. En effet, nous avons proposé un mode de représentation d'un indicateur d'utilisabilité sur la base de ses critères de décision. Bien que cette représentation soit utile, elle pourrait faire l'objet d'une analyse par rapport à d'autres références. Nous citons comme exemples les travaux de (Whitefield et al., 1991 ; Rohrer, 2014). Par rapport au travail de Whitefield et al. (1991), les auteurs ont défini un cadre pour classer différentes méthodes d'évaluation (les méthodes analytiques, les rapports des utilisateurs, les rapports de spécialistes et les méthodes d'observation) (voir Figure 6.1 (a)). Ce classement est effectué sur la base de deux dimensions : la première par rapport à l'utilisateur et la deuxième par rapport au composant de l'ordinateur. Concernant l'utilisateur, sa présence peut être réelle (i.e., présence physique) ou représentative (i.e., à travers des descriptions ou des modèles des utilisateurs). De la même façon, le composant de l'ordinateur (dans notre cas, l'interface utilisateur) comporte deux catégories : réel et représentatif. Dans notre approche, concernant les utilisateurs, nous avons impliqué des utilisateurs réels. Par rapport au composant de l'ordinateur, d'une part, nous pouvons nous positionner dans l'axe réel car les utilisateurs ont interagi d'une façon directe avec l'ordinateur afin de collecter des mesures objectives. D'autre part, nous considérons l'axe représentatif vu que les utilisateurs ont répondu au questionnaire d'utilisabilité après l'utilisation du système interactif à évaluer. Dans ce cas, ils ont eu besoin de se rappeler des caractéristiques des interfaces utilisateur ainsi que des tâches du scénario de l'expérimentation. De ce fait, la Figure 6.1 (b) illustre le positionnement de notre approche par rapport à ce travail.

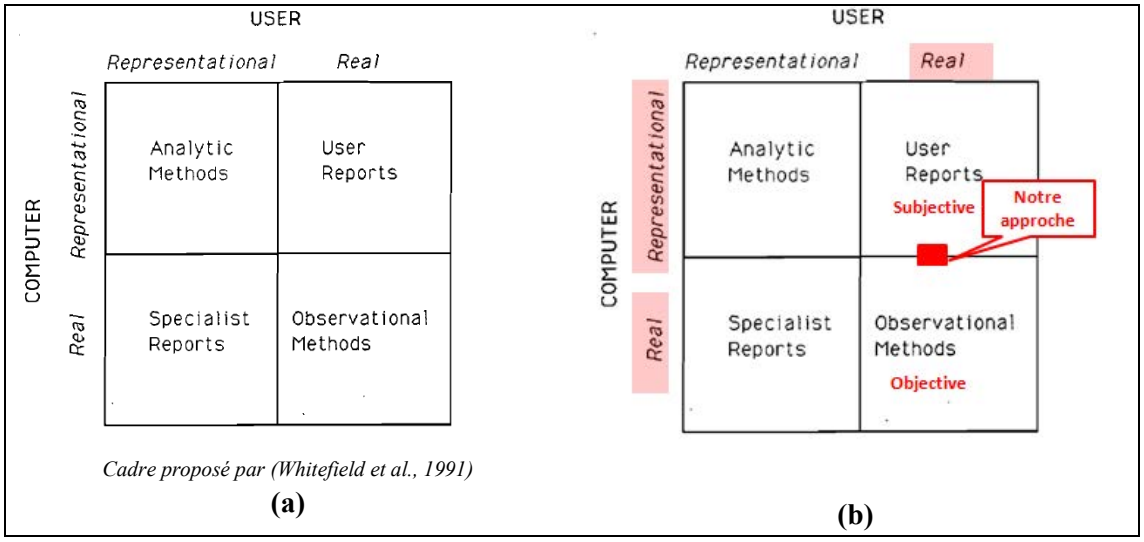
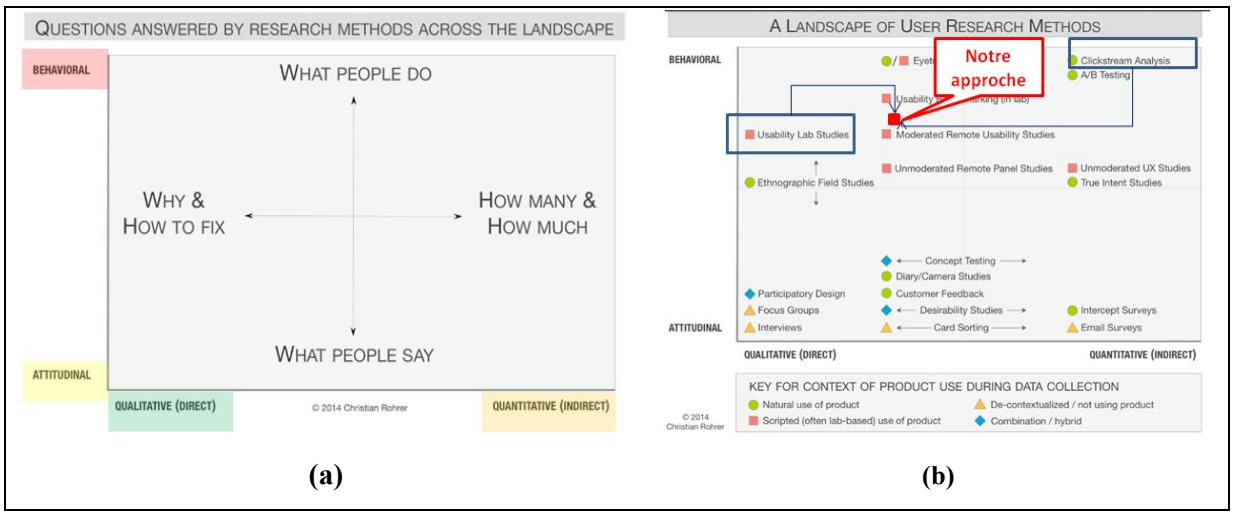


Figure 6. 1 : Positionnement de notre approche par rapport au cadre de référence de (Whitefield et al.,1991)

En outre, Rohrer (2014) a classifié les méthodes d'évaluation sur la base de trois dimensions: attitude versus comportementale, qualitative versus quantitative et le contexte d'utilisation (voir Figure 6.2 (a)). En termes de la dimension d'attitude versus comportementale, il les distingue comme *« ce que les gens disent »* contre « *ce que les gens font »*. En outre, il affirme que la dimension qualitative porte sur la question *« Pourquoi et comment résoudre le problème ? »* alors que celle quantitative répond à la question *« Combien de problèmes ? »*. Par rapport à ce travail, notre approche est positionnée comme l'illustre la Figure 6.2 (b) sur la base de *Usability Lab Studies* vu l'utilisation du questionnaire, et aussi sur la base de *Clickstream Analysis* vu

l'utilisation du mouchard électronique EISEval. Par ailleurs, elle converge plus vers l'aspect comportemental en utilisant non seulement des mesures mais aussi des données qualitatives en considérant les avis personnels d'experts lors de l'analyse des résultats.



(a) **(b)**

Figure 6. 2 : Positionnement de notre approche par rapport au travail de (Rohrer, 2014)

- *La combinaison des évaluations subjective et objective d'une façon complémentaire* – Bien que toutes les approches présentées aient été fondées sur la complémentarité entre les évaluations subjective et objective, cette complémentarité diffère d'une approche à une autre. La majorité de ces évaluations est basée sur l'utilisation d'un ensemble de critères d'utilisabilité hétérogènes dont certains sont relatifs à l'évaluation objective et d'autres à l'évaluation subjective. Certaines autres approches (Kerzazi et Lavalée, 2011 ; Hentati et al., 2013) se contentent d'évaluer un ensemble de critères d'utilisabilité en se basant à la fois sur l'évaluation subjective et objective. De sorte qu'un critère peut être évalué d'une façon complémentaire subjectivement et objectivement. Dans notre approche, nous avons adopté cette façon qui était sans doute plus adaptée pour contribuer à assurer l'intégration d'une façon complémentaire et atteindre notre objectif.
- *La détection avec précision des problèmes d'utilisabilité des interfaces utilisateur* – Pour certaines approches (Al-Wabil et Al Khalifa, 2009 ; Kerzazi et Lavalée, 2011 ; Charfi et al., 2015), bien qu'elles n'aient pas assuré une intégration effective des évaluations subjective et objective, elles se contentent de détecter les problèmes d'utilisabilité des interfaces utilisateur. Toutefois, pour les approches qui se focalisent sur l'intégration des résultats via des valeurs globales d'utilisabilité, celles-ci sont plutôt plus bénéfique dans le cas d'un test d'analyse comparative (benchmarking) ou d'une évaluation sommative (Sauro, 2005). De ce fait, ces résultats ne sont pas utiles pour découvrir et résoudre les problèmes d'utilisabilité d'une interface utilisateur (Sauro, 2005). A cette fin, pour effectuer la détection, l'amélioration et les corrections de problèmes d'utilisabilité, nous avons proposé à travers notre approche de revenir aux données brutes (subjective et objective) et considérer le but spécifique pour lequel elles ont été créées afin de savoir ce qu'il faut corriger.
- *La proposition des interprétations des résultats d'évaluation* – Contrairement à la majorité des approches présentées, nous avons proposé dans notre approche de fournir une base pour l'aide à l'interprétation des résultats des indicateurs d'utilisabilité proposés à travers les critères de décision. Ceux-ci présentent un des concepts de base définis par la norme ISO/IEC 15939 pour assurer la procédure d'interprétation.
- *La suggestion de recommandations* – à l'exception de l'approche de (Charfi et al., 2015) qui a généré des recommandations lors de l'évaluation ergonomique des interfaces utilisateur, les autres

approches n'ont rien fourni. Dans notre approche, nous avons proposé un ensemble de recommandations afin d'aider l'évaluateur à corriger les problèmes détectés et améliorer la qualité des interfaces utilisateur du système évalué. Ces recommandations sont appropriées à chaque indicateur d'utilisabilité.

Tableau 6. 1 : Tableau comparatif de notre approche par rapport aux approches présentées dans le chapitre 2

	WebUSE (Nikov et al., 2003)	SUM (Sauro et Kindlund, 2005)	(Al-Wabil et Al Khalifa, 2009)	(Kerzazi et Lavalée, 2011)	(Hentati et al., 2013)	(Charfi et al., 2015)	**Notre approche (Assila et al., 2016)**
Combinaison des évaluations subjective et objective	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Intégration effective des évaluations subjective et objective	**X**	**X**			**X**		**X**
Combinaison des évaluations subjective et objective d'une façon complémentaire et cohérente				**X**	**X**		**X**
Détection des problèmes d'utilisabilité des interfaces utilisateur			**X**	**X**		**X**	**X**
Proposition des interprétations à leurs résultats	**X**	**X**					**X**
Suggestion des recommandations						**X**	**X**

Limites et Perspectives

Comme présenté auparavant, ce travail nous a offert un certain nombre d'avantages en termes d'évaluation de l'utilisabilité des systèmes interactifs. Néanmoins, nous avons distingué quelques faiblesses à considérer lors de l'utilisation de notre approche. Ceux-ci se résument comme suit :

- L'environnement est limité aux systèmes de supervision du trafic complexes à base d'agents. Il n'évalue actuellement que les systèmes WIMP. Cela signifie que les mesures définies et les procédures d'interprétation sont très spécifiques pour cette technologie. De ce fait, la personnalisation de certains indicateurs proposés est nécessaire pour les appliquer avec d'autres technologies (web, mobile, tangible, tactile, etc.).
- L'utilisation de cet environnement intégré nécessite l'introduction du code source du système à évaluer pour la capture des mesures relatives aux attributs d'éléments graphiques. Néanmoins, le code source n'est pas toujours disponible. En outre, lors de l'utilisation de l'outil SOUSEval, l'implication d'un expert d'évaluation, dont son rôle est de choisir les seuils des mesures d'utilisabilité, constitue une difficulté qui peut engendrer des coûts élevés d'évaluation.
- Les mesures subjectives adoptées sont basées uniquement sur l'utilisation des questionnaires d'utilisabilité standardisés, qui se caractérisent par des questions fermées avec des types d'échelles

spécifiques (tels que l'échelle de Likert et l'échelle dichotomique). De ce fait, les questionnaires non normalisés (ou prédéfinis) basés sur des questions ouvertes ne peuvent pas être appliqués dans notre approche.

- Bien que notre approche permette d'identifier les problèmes d'utilisabilité spécifiques, elle ne propose pas une façon de priorisation des efforts de correction dans le cas où il existe beaucoup de problèmes à aborder (c-à-dire, lorsque plusieurs indicateurs indiquent des mauvais ou très mauvais niveaux de qualité). Cela est dû au fait que notre approche catégorise les résultats selon trois niveaux de qualité (très mauvais, mauvais ou bon). De cette façon, nous pouvons obtenir toute une gamme de valeurs dans la même catégorie.
  Lorsque nous considérons chaque indicateur séparément, notre approche suggère une façon permettant de prioriser les corrections à effectuer pour chaque critère d'utilisabilité évalué. Celle-ci est indiquée dans la partie de recommandation de l'indicateur en générant des analyses détaillées des résultats (voir par exemple, *Chapitre 3, Tableau 3.7*). Pour ce faire, nous regardons à l'intérieur de la valeur classée de l'indicateur (qui a comme niveau de qualité mauvais ou très mauvais) et nous distinguons le pire des meilleurs résultats.
  Cependant, lorsque nous analysons tous les indicateurs ensemble, nous avons seulement les catégories finales. S'ils sont dans le même niveau de qualité (par exemple, mauvais), il est difficile de donner la priorité pour décider l'ordre de correction. Dans ce cas, l'évaluateur doit donner la priorité aux critères évalués en tenant compte de la priorisation des exigences de qualité définies dans le cahier de charge dans lequel les besoins des clients sont généralement décrits.
- Dans notre travail actuel, nous n'avons pas défini des indicateurs d'utilisabilité qui permettent de répondre à tous les besoins d'information par rapport à tous les critères ergonomiques de (Scapin et Bastien, 1997). Nous nous sommes limités à un premier groupe de critères pour mettre en pratique l'environnement intégré en appliquant notre approche de la définition des indicateurs d'utilisabilité.

Sur la base de ces limites rencontrées, nous introduisons par la suite, nos principales perspectives de recherche :

- *L'évaluation du système SAI* – Après avoir montré la faisabilité des indicateurs d'utilisabilité proposés à travers l'évaluation du système SAI, il est nécessaire d'effectuer une nouvelle évaluation sur le terrain avec des utilisateurs finaux de ce système. Cette évaluation sera réalisée une fois que le concepteur du SAI aura effectué les améliorations nécessaires dans les interfaces utilisateur en tenant compte des recommandations suggérées lors de l'évaluation effectuée.
- *Definition des nouveaux indicateurs* – Nous envisagerons dans nos futurs travaux de proposer des indicateurs qui répondent aux autres besoins d'information associés à d'autres critères ergonomiques. Pour cela, nous pensons utiliser l'approche proposée comme base en ajoutant des nouvelles étapes si necessaires.
- *La généralisation de l'approche et de l'outil proposés* – Pour rendre possible une large utilisation de notre approche proposée, il est essentiel d'évoluer et de généraliser l'environnement intégré. Etant donné qu'à son état actuel, l'environnement est dédié aux systèmes interactifs de supervision du Trafic du type WIMP, sa généricité peut être établie d'une part par leur exploitation dans d'autres domaines d'application, et d'autre part, pour l'évaluation de différents types d'interfaces utilisateur (tels que l'UI web, mobile, tangible, tactile, etc.) compte tenu de différents contextes d'utilisation. Toutefois, cela requiert un certain effort pour la mise en œuvre des outils d'évaluation appropriés permettant de capturer automatiquement les mesures d'utilisabilité des

applications spécifiques à évaluer. Ainsi, il s'avère très important d'adapter ou définir des modèles de spécifications d'indicateurs selon les types d'interfaces utilisateur à évaluer.

- *La définition des seuils des mesures associées aux indicateurs* – Généralement la définition des seuils des mesures associées aux indicateurs est une tâche difficile qui requiert l'implication d'un expert en évaluation pour décider la valeur à partir de laquelle (jusqu'à lequel) les résultats sont considérés acceptables. Dans notre travail, nous avons considéré un seul seuil par mesure sur la base des travaux de (Albert et Tullis, 2013). En effet, l'expert doit considérer le contexte d'utilisation de l'évaluation, le type d'interface à évaluer ainsi que l'ensemble des règles et recommandations ergonomiques existantes. Pour l'aide à la définition de ce seuil, nous souhaiterions définir et construire une base de données historique des évaluations permettant d'enregistrer les mesures collectées ainsi que leurs seuils pour différents contextes d'évaluation. Cependant, pour être réellement utile, ce travail nécessite une équipe de recherche plus diversifiée pour effectuer l'évaluation dans différents domaines, dans le cadre d'un projet à long terme. En outre, nous envisageons de rendre flexible la définition des seuils en considérant un intervalle des valeurs qui indique les valeurs acceptables.
- Dans notre travail actuel, nous avons proposé des indicateurs d'utilisabilité qui se basent sur l'intégration des résultats des questionnaires avec ceux extraits de l'un des outils d'évaluation objective (mouchard électronique ou inspecteur des règles ergonomiques). Dans nos futurs travaux, nous envisageons d'améliorer ces indicateurs afin d'établir une évaluation subjective et objective plus complète de sorte que l'indicateur pourrait inclure des données provenant des questionnaires, de l'inspecteur des règles ergonomiques et du mouchard électronique d'une façon complémentaire. Généralement, ce travail requiert beaucoup d'efforts en matière de développement et de la recherche pour définir et implémenter des nouveaux indicateurs qui intègrent au minimum trois mesures d'utilisabilité ainsi que la spécification de leurs interprétations.
- *La représentation graphique des indicateurs en exploitant les concepts de visualisation de l'information* – la visualisation de l'information est un domaine vaste et intéressant permettant la représentation visuelle de données en exploitant différents types de graphiques (tels que le treemap, les nuages de mots, etc.) (Kerren et al., 2008). Dans notre travail, nous nous sommes focalisés sur le domaine de la mesure en exploitant la littérature provenant du génie logiciel. De ce fait, nous avons adopté les graphiques existants, de type nuage des points, qui permettent d'illustrer effectivement et concrètement l'intégration des évaluations subjective et objective. En outre, ces graphiques étaient très utiles pour l'intégration de deux ou trois mesures. Néanmoins, pour le cas des indicateurs qui traitent plus que trois mesures, nous envisageons d'exploiter, dans nos futurs travaux, les concepts de visualisation de l'information, notamment dans le but de représenter ces indicateurs et analyser leurs résultats.

Enfin, nous affirmons que toutes ces perspectives ouvrent des voies de recherche pour aboutir à une meilleure évaluation de l'utilisabilité des systèmes interactifs.

Bibliographie

Abran, A., Al-Qutaish, R. et Cuadrado-Gallego, J. (2006). Analysis of the ISO 9126 on Software Product Quality Evaluation from the Metrology and ISO 15939 Perspectives. *WSEAS Transactions on Computers,* 5(11): 2778-2786, World Scientific & Engineering Academy and Society.

Abran, A., Khelifi, A. et Suryn, W. (2003).Usability meanings and interpretations in ISO standards.*Software Quality Journal*, 11: 325–338.

Albert, B. et Tullis, T. (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Second Edition. Morgan Kaufmann.

Albert, W. et Dixon, E. (2003). Is this what you expected? The use of expectation measures in usability testing. *In Proceedings of the Usability Professionals Association Annual Conference*. UPA, Scottsdale, AZ.

Allal, L. (1991). Vers une pratique de l'évaluation formative : matériel de formation continue des enseignants. Bruxelles : De Boeck-Wesmael.

Al-Wabil, A. et Al-Khalifa, H. (2009). A Framework for Integrating Usability Evaluations Methods: The Mawhiba Web Portal Case Study. *In Proceedings of the International conference on the Current Trends in Information Technology*, CTIT'09, pages 1-6, Dubai.

ANSI/NCITS 354-2001. (2001). American National Standard Institute. *Common Industry Format for Usability Test Reports*. Washington, DC: Author.

Assila, A., Ezzedine, H. et Bouhlel, M.S. (2013a). A Web questionnaire generating tool to aid for interactive systems quality subjective assessment. *In Proceedings of the IEEE International Conference on Control, Decision and Information Technologies*, CODIT'13, pages 1-7, Tunisie.

Assila, A., Ezzedine, H. et Bouhlel, M.S. (2013b). Towards improving the subjective quality evaluation of Human Computer Interfaces using a questionnaire tool. *ICALT, International Conference on Advanced Logistics and Transport* (29-31 May), IEEE, pages 275-283, Sousse, Tunisia,

Assila, A., Oliveira, K. et Ezzedine, H. (2014a). Towards Qualitative and Quantitative Data Integration Approach for Enhancing HCI Quality Evaluation. *In* Kurosu, M (eds.): Human-Computer Interaction. Theories, Methods, and Tools - *16th HCI International Conference*, Proceedings, Part I, Springer, Lecture Notes in Computer Science, pages 469–480. Heraklion, Crete, Greece.

Assila, A., Oliveira, K. et Ezzedine, H. (2014b). Towards indicators for HCI quality evaluation support. *In* Indulska, M., Purao, S (eds.): Advances in Conceptual Modeling - *ER 2014 Workshops, ENMO, MoBiD, MReBA, QMMQ, SeCoGIS, WISM, and ER Demos, Proceedings*, Springer, Lecture Notes in Computer Science, pages 178–187. Atlanta, GA, USA.

Assila, A., Oliveira, K. et Ezzedine, H. (2014c). Towards quality indicators for supporting the evaluation of interactive systems HCI. *Proceedings of the 2014 Ergonomie et Informatique Avancée Conférence - Design, Ergonomie et IHM : quelle articulation pour la co-conception de l'interaction*, ACM, pages 66-69, Biarritz, France.

Assila, A., Oliveira, K. et Ezzedine, H. (2016a). Standardized Usability Questionnaires: Features and Quality focus. *electronic Journal of Computer Science and Information Technology*. (Acceptée le 7 Mars 2016).

Assila, A., Oliveira, K. et Ezzedine, H. (2016b). An Environment for integrating subjective and objective Usability findings based on Measures. *IEEE 10th International Conference on Research Challenges in Information Science*. IEEE, pages 645-656, Grenoble, France.

Assila, A., Oliveira, K. et Ezzedine, H. (2016c). Integration of Subjective and Objective Usability Evaluation based on ISO/IEC 15939: a Case Study for Traffic Supervision Systems. *International Journal of Human-Computer Interaction*, DOI:10.1080/10447318.2016.1220068.

Bach, C. et Scapin, D. L. (2005). Critères ergonomiques pour les interactions homme-environnements virtuels définitions, justifications et exemples. Rapport technique, INRIA, N 5531.

Basili, V., Caldiera, G., et Rombach, H. (1994). The goal question metric approach. In Encyclopedia of software engineering, pages 528-532, Wiley.

Bastien, J. M. C et Scapin, L. (2002). Les méthodes ergonomiques : de l'analyse à la conception et à l'évaluation. *In Proceedings of Ergonomie et Informatique Avancée*, Ergo-IA'2002, Biarritz.

Bastien, J. M. C. et Scapin, L. (2001). Evaluation des systèmes d'information et critères ergonomiques. In Kolski, C. (eds.) : *Environnement évolués et évaluation de l'IHM. Interaction Homme Machine pour les SI*, vol(2), pages 53-80, Paris : Hermès.

Bastien, J. M. C. et Scapin, L. (1993). Critères ergonomiques pour l'évaluation d'interfaces utilisateurs. Rapport technique, INRIA : Le Chesnay, N 156.

Benedek, J. et Miner, T. (2002). Measuring desirability: New methods for evaluating desirability in a usability lab setting. *In Proceedings of UPA Conference*, Orlando.

Bentley, R., Hughes, J. A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D. et Sommerville, I. (1992).Ethnographically informed systems design for air traffic control. *In the Conference on Computer Supported Cooperative Work*, pages 123-129, Toronto.

Bernsen, N. O. et Dybkjær, L. (2009). *Multimodal usability*. London: Springer.

Bevan, N. (1997). Quality in Use: Incorporating Human Factors into the software engineering lifecycle. *In Proceedings of the Third International Symposium and Forum on Software Engineering Standards*, ISESS'97.

Bevan, N. (1995). Measuring usability as quality of use. *Journal of Software Quality*, 4: 115-140.

Bevan, N. et Macleod, M. (1994). Usability measurement in context. *Behavior and Information Technology*, 13: 132-145.

Beyer, H. et Holtzblatt, K. (1998). *Contextual design: Defining Customer-Centered Systems*, volume 6. Sans Francisco: Morgan Kaufmann Publishers, New York, USA.

Bonte, T., Ezzedine, H., Hammadi, S., Hayat, S., Kolski, C., Ould Sidi, M.,Tahon, C., Trabelsi, A. et Malca, F. (2007). Rapport final du projet SART : Système d'aide à la régulation de trafic du réseau de transport valenciennois et de ses pôles d'échanges, décembre 2007. Rapport Technique, SART.

Borsci, S., Federici, S., Gnaldi, M., Bacci , S. et Bartolucci, F. (2015). Assessing user Satisfaction in the era of User Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a function of Product Experience. *International Journal of Human-Computer Interaction*, 31(8): 484-495.

Bowman, D. A., Gabbard, J. L. et Hix, D. (2002). A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence*, 11(4): 404–424.

Bradner, E., Dawe, M. (2012). Parts of the SUM: a case study of usability benchmarking using the SUM Metric. *In Proceedings of UPA International Conference*, Henderson, Nevada, USA.

Brangier, E. (2007). Besoin et interface. *In* Akoka, J. et Comyn-Wattiau, I. (eds) : *Encyclopédie de l'informatique et des systèmes d'information*, pages 1070-1084, Paris : Vuibert.

Brooke, J. (1996). SUS: A –quick and dirty" usability scale. *In* Jordan, P., Thomas, B. et Weerdmeester, B. (eds.): *Usability Evaluation in Industry,* pages 189-194, London: Taylor & Francis.

Buanga, P. M. (2011). Automated evaluation of graphical user interface metrics. Mémoire de l'obtention de maître en Informatique. Université de Louvain-la-Neuve. http://www.usixml.org/servlet/Repository/mbenza-msc2011-323.pdf?ID=307.

Burbeck, S. (1992). Application Programming in Smalltalk-80: How to use Model-View-Controller (MVC). *University of Illinois in Urbana-Champaign (UIUC) Smalltalk Archive.* Available at: http://st-www.cs.illinois.edu/users/smarch/st-docs/mvc.html.

Carifio, J. et Rocco, J. P. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3(3): 106-116.

Castro, F. S., Dias, D. M., Higarashi, I. H., Scochi, C. G. et Fonseca, L. M. (2015). Evaluation of digital educational student technology interaction in neonatal nursing. *Revista Da Escola De Enfermagem Da USP*, 49(1): 114-121.

Centeno, V. L., Kloos, C. D., del Toro, B. C. et Gaedke, M. (2007). Web accessibility evaluation via xslt. *In* Weske, M., Hacid, Mohand-Saad et Godart, C. (eds.): *Web Information Systems Engineering WISE*, volume 4832, Lecture Notes in Computer Science, pages 459–469. Springer Berlin / Heidelberg.

Centeno, V. L., Kloos, C. D., Gaedke, M. et Nussbaumer, M. (2006). Waex: Web accessibility evaluator in a single xslt file. *Automated Specification and Verification of Web Systems*, pages 29-36.

Charfi, S., Ezzedine, H. et Kolski, C. (2015). RITA: a useR Interface evaluation framework. *Journal of Universal Computer Science*, 21 (4): 526-560.

Charfi S., Trabelsi, A., Ezzedine, H. et Kolski, C. (2014). Widgets dedicated to user interface evaluation. *International Journal of Human-Computer Interaction*, 30(5): 408-421.

Charfi, S. (2013). Définition et test d'un outil d'aide à l'évaluation de systèmes à architecture à base d'agent, par approche à base de connaissances. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis.

Charfi, S. et Ezzedine, H. (2011). Proposal of an automatic tool for Guidelines validation to aid for interactive systems ergonomic quality assessment. *In Proceedings of the International Conference on Communications, Computing and Control Applications*, CCCA'2011, Tunisia.

Chattratichart, J. et Lindgaard, G. (2008). A Comparative Evaluation of Heuristic-Based Usability Inspection Methods. *In Proceedings of the Conference on Human Factors in Computing Systems*, CHI'08, pages 2213-2220.

Chiew, K. T. et Salim, S. S. (2003). WEBUSE: Website usability evaluation tools. *Malaysian Journal of Computer Science*, 16(1): 47-57.

Chin, J. P., Diehl, V. A., et Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human–computer interface. *In Proceedings of CHI*, pages 213-218.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Dalton, P., Maute, C., Oshida, A., Hikichi, S. et Izumi, Y. (2008). The Use of Semantic Differential Scaling to Define the Multi-Dimensional Representation of Odors. *Journal of Sensory Studies*, 23(4): 485-497.

Davis, F. D. (1989). Perceived Usefulness, Percevied Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13: 319-340.

De Haan, G., Van Der Veer, G. C. et Van Vliet, J. C. (1992). Formal modeling techniques in human computer interaction. In van der Veer, G. C., Bagnara, S. et Kempen, G. A. M. (eds): *Cognitive Ergonomics: Contributions from Experimental Psychology, Theoretical Issues*, pages 27–67. Amsterdam, The Netherlands: Elsevier Science Publishers.

Dix, A., Finlay, J., Abowd, G. et Beale, R. (1998). *Human-Computer Interaction,* Second Edition, London: Prentice Hall International.

Dix, A., Finlay, J., Abowd, G. et Beale, R. (2004). *Human-Computer Interaction,* Third Edition, London: Prentice Hall International.

Donyaee, M. (2001). Towards an Integrated Model for Specifying and Measuring Quality in Use. *In the Webster library*, Concordia University. Montreal, Canada.

Donyaee, M., Seffah, A. et Khelifi, A. (2001). QUIM: An Integrated Framework for Modeling and Measuring Quality in Use. *In Proceedings of INTERACT 2001,* pages 715-716.Tokyo, Japan.

Dumas, J. S. et Fox, J. E. (2009). Usability Testing: Current Practice and Future Directions. *In* Sears, A. L.et Jacko, J. A. (eds.): *The Human-Computer Interaction Handbook*, pages 1129-1149, USA: Taylor and Francis.

Duquia, R. P., Bastos, J. L., Bonamigo, R. R., González-Chica, D. A. et Martínez-Mesa, J. (2014). Presenting data in tables and charts. *PMC Journal*, 89(2): 280-285.

Elling, S., Lentz L. et Jong M. (2007). Website Evaluation Questionnaire: Development of a Research-Based Tool for Evaluating Informational Websites. *In* Wimmer, M. A., Scholl, H. J. et Grönlund, A. (eds.): *Lecture Notes in Computer Science,* vol 4656, pages 293-304), Berlin Heidelberg: Springer-Verlag.

Erdinç, O. et Lewis, J. R. (2013). Psychometric Evaluation of the T-CSUQ: The Turkish Version of the Computer System Usability Questionnaire. *International Journal of Human-Computer Interaction*, 29(5): 319-326.

Ezzedine, H., Bonte, T., Kolski , C. et Tahon, C. (2008). Integration of traffic management and traveller information systems: basic principles and case study in intermodal transport system management. *International Journal of Computers, Communications and Control* (IJCCC), 3: 281–294.

Ezzedine H., Trabelsi, A. (2005). From the design to the evaluation of an agent-based human-machine interface. Application to supervision for urban transport system. *In Proceedings of Scientific Computation, Applied Mathematics and Simulation*, pages 717-725.

Ezzedine, H. (2002). Méthodes et Modèles de Spécification et Evaluation des Interfaces Homme-Machine dans les systèmes industriels complexes. Mémoire d'Habilitation à diriger des recherches. Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes, décembre.

Ezzedine, H., Abed, M. (1997). Une méthode d'évaluation d'Interface Homme Machine de supervision d'un procédé industriel. *Journal Européen des Systèmes Automatisés* (RAIRO-APII-JESA), 7: 1078-1110.

Faraday, P. (2000). Visually critiquing web pages. *In Proceedings of 6th Conf. on Human Factors and the Web*, Austin, Texas.

Farenc, C., Palanque, P., and Vanderdonckt, J. (1995). User Interface Evaluation: is it Ever Usable?. Proc of 6th International Conference on Human-Computer Interaction HCI International'95 (Yokohama, 9-14 July 1995), pages 329-334, Elsevier Science Publishers, Amsterdam.

Farenc, C., Palanque, P., Bastien, J. M. C., Scapin, D. L. et Winckler, M. (2001).Towards a General Guidance and support Tool for Usability Optimisation. *In Proceedings of the HCI International*, New Orleans, USA.

Farenc, C. et Palanque, P. (1999). A generic framework based on ergonomics rules for computer aided design of user interface. *In Proceedings of the Third International Conference on Computer-Aided Design of User Interfaces*.

Farenc, C. (1997). ERGOVAL : une méthode de structuration des règles ergonomiques permettant l'évaluation automatique d'interfaces graphiques. Thèse de doctorat, Université Toulouse 1.

Fayech, B. (2003). Régulation des réseau de transport multimodal : Système muliti-agents et algorithmes évolutionnistes. Thèse de doctorat, Université de Lille.

Filho, A. S., Novaes, M. A. et Gomes, A. S. (2015). A 3D visualization framework to social network monitoring and analysis. *Computers in Human Behavior*, 49: 623-634.

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22(5): 323-327.

Gabillon, Y., Lepreux, S. et Oliveira, K. (2013). Towards ergonomic User Interface composition: A study about information density criterion. *In Proceedings of 15th International Conference HCI*, pages 211-220).

Galitz, W. O. (2007). *Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*, John Wiley & Sons.

Garcia, F., Bertoa, M. F., Calero, C., Vallecillo, A., Ruiz, F., Piattini, M. et al. (2006). Towards a consistent terminology for software measurement. *Information and Software Technology*, 48(8): 631-644.

Gerard, F. M. (2013). L'évaluation au service de la régulation des apprentissages : enjeux, nécessités et difficultés. *Revue française de linguistique appliquée*, XVIII-1 : 75-92.

Grammenos, D., Akoumianakis, D. et Stephanidis, C. (2000). Integrated support for working with guidelines: the sherlock guideline management system. *Interacting with Computers*, 12(3): 281-311.

Grislin, M. et Kolski, C. (1996). Evaluation des interfaces homme-machine lors du développement de système interactif. *Technique et Science Informatiques TSI*, 3: 265-296.

Hamborg, K., Vehse, B. et Bludau, H. (2004). Questionnaire based usability evaluation of hospital information systems. *Electronic Journal of Information Systems Evaluation*, 7(1): 21-30.

Hartson, H. R., Andre, T. S. et Will, R. C. (2003).Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, 15(1): 145-181.

Hartson, R. et Pyla, P. S. (2012).*The UX Book: Process and Guidelines for ensuring a quality user experience*. Amsterdam, The Netherlands: Morgan Kaufmann.

Hassenzahl, M., Burmester, M. et Koller, F. (2003). AttrakDiff: questionnaire to measure perceived hedonic and pragmatic quality. *In* Ziegler, J. et Szwillus, G. (eds.): *Mensch and Computer*, pages 187-196.

Hentati, M., Trabelsi, A. et Charfi, S. (2013). A decision-making approach for the synthesis of results issued from different user interface evaluation methods. *In Proceedings of the Fourth International Conference on Information and Communication Technology and Accessibility*, ICTA, pages 1-6, Tunisia.

Hilbert, D. M. et Redmiles, D. F. (2000). Extracting usability information from user interface events. *ACM Computing Surveys (CSUR)*, 32(4): 384-421.

Holzinger, A. (2005). Usability engineering methods for software developers. *Communications of the ACM Commun*, 48(1): 71-74.

Holzinger, A. (2004). Rapid prototyping for a virtual medical campus interface. *IEEE Software*, 21(1): 92–99.

Hong, J. I., Heer, J., Waterson, S. et Landay, J. A. (2001). WebQuilt: A Proxy-based Approach to Remote Web Usability Testing. *ACM Transactions on Information Systems*, 19(3): 263-385.

Hongquan, X. (2004). Chapter 12 Relationships Between Quantitative Variables: Regression and Correlation. *In Lecture Note Introduction to Statistical Methods for Business and Economics*.

Hornbæk, K. et Law, E. (2007). Meta-analysis of Correlations among Usability Measures. *In Proceedings of Conference on Human Factors in Computing Systems*, CHI 2007, pages 617-626.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2): 79–102.

IEEE Std 1061. (1998). IEEE Standard for a Software Quality Metrics Methodology.

IEEE Std 610.12. (1990). Standard Glossary of Software Engineering Terminology, The Institute of Electrical and Electronics Engineers, Inc., NY.

ISO 9241-11. (1998). ISO 9241-11, Ergonomic requirements for office work with visual display terminals (VDT) s-Part 11 Guidance on usability, 1998.

ISO 9241-171. (2008). Ergonomics of human-system interaction -- Part 171: Guidance on software accessibility. Geneva, Switzerland.

ISO/IEC 14598-1. (1999). Information technology -- Software product evaluation -- Part 1: General overview.

ISO/IEC DIS 14598-5. (1998). Information technology - Evaluation of software product - Part 5: Process for evaluators.

ISO/IEC 15939. (2007). International Organization for Standardization/ International Electrotechnical Commission. (2007).Systems and software engineering — Measurement process (ISO / IEC 15939:2007 (E)).Geneva, Switzerland.

ISO/IEC 25000. (2014).Ingénierie des systèmes et du logiciel -- Exigences de qualité des systèmes et du logiciel et évaluation (SQuaRE) -- Guide de SQuaRE.

ISO/IEC 25010. (2011). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. International Organization for Standardization, Geneva, Switzerland.

ISO/IEC 25021. (2012). Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Quality measure elements.

ISO/IEC 25022. (2012). Systems and software engineering - Systems and software quality requirements and evaluation (SQuaRE) - Measurement of quality in use.

ISO/IEC 25040. (2011). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process.

ISO/IEC 9126-1. (2001). Software engineering - Product quality - Part 1: Quality model.

ISO/IEC 9241-151. (2008). Ergonomics of human-system interaction -- Part 151: Guidance on World Wide Web user interfaces. Geneva, Switzerland.

ISO/IEC 9241-920. (2009). Ergonomics of human-system interaction -- Part 920: Guidance on tactile and haptic interactions. Geneva, Switzerland.

ISO/IEC TR 9126-2. (2003). Software engineering -- Product quality -- Part 2: External metrics.

ISO/IEC TR 9126-4. (2004). Software engineering -- Product quality -- Part 4: Quality in use metrics.

ISO/WD 9241-112. (2013). Ergonomics of human-system interaction — Part 112: Principles for the presentation of information.

Ivory, M. Y. et Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, 33(4): 470-516.

Jaaskelainen, R. (2000). Focus on Methodolody in think-aloud studies on translating. *In* Tirkkonen-Condit, S. et Jaaskelainen, R. (eds.): *Tapping and mapping the processes of translation and interpreting*: *Outlooks on empirical research*, pages 71-82, Benjamins Translation Library.

Jolliffe, I. T. (2002). *Principal Component Analysis*. 2nd Edition, Springer, New York

John, B. E. et Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction* 3(4): 320–351.

Kantner, L. et Rosenbaum, S. (1997). *Usability Studies of WWW Sites: Heuristic Evaluation vs. Laboratory Testing SIGDOC*, pages 153-160.

Kasik, D. J. et George, H. G. (1996). Toward automatic generation of novice user test scripts. *In Proceedings of the Conference on Human Factors in Computing Systems*, pages 244-251.

Keith, N. V. (2005). Using guidelines to assist in the visualization design process. *In proceedings of the Asia-Pacific symposium on Information visualization*, APVis'05, pages 115-123.

Kerren, A., Stasko, J.T., Fekete, J., North, C. (2008). *Information Visualization : Human-Centered Issues and Perspectives*. Lecture Note in Computer Science, Springer, pages 183.

Kerzazi, N. et Lavallee, M. (2011). Inquiry on usability of two software process modeling systems using ISO/IEC 9241. *In Proceedings of 24th Canadian Conference on Electrical and Computer Engineering*, pages 773-776.

Kirakowski, J. et Cierlik, B. (1998). Measuring the usability of websites. *In Proceedings of the Human Factors and Ergonomics Society*, pages 424-428).

Kirakowski, J. et Corbett, M. (1993). SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3): 210-212.

Kolski, C. (1989). Contribution à l'ergonomie de conception des interfaces graphiques homme-machine dans les procédés industriels, application au système expert SYNOP. *Mémoire de Doctorat*, Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes, janvier.

Kolski, C. et Ezzedine, H. (2003). Conception et évaluation des ihm de supervision : éléments méthodologiques. *Génie Logiciel*, 65 : 2-11.

Kuusela, H. et Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology*, 113(3): 387-404.

Lewis, C., Polson, P., Wharton, C. et Rieman, J. (1990c). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. *In Proceedings of CHI '90*. ACM

Lewis, J. R. (2014). Usability: Lessons Learned . . . and Yet to Be Learned. *International Journal of Human-Computer Interaction*, 30: 663–684.

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human Computer Interaction*, 14: 463-488.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human Computer Interaction*, 7: 57-78.

Lewis, J. R. (1990b). Psychometric Evaluation of an After-Scenario Questionnaire for Computer Usability Studies: The ASQ (Tech. Report 54.541). International Business Machines Corp, Boca Raton, FL.

Lewis, J. R. (1990a). Psychometric Evaluation of a Post-study System Usability Questionnaire: The PSSUQ (Tech. Report 54.535). International Business Machines Corp, Boca Raton, FL.

Lewis, J. R., Brown, J. et Mayes, D. K. (2015b). Psychometric Evaluation of the EMO and the SUS in the context of a Large-Sample Unmoderated Usability study. *International Journal of Human-Computer Interaction*, 31(8): 445-553.

Lewis, J. R. et Hardzinski, M. L. (2015). Investigating the psychometric properties of the Speech User Interface Service Quality questionnaire. *International Journal of Speech Technology*, 18(3): 479-487.

Lewis, J. R., Utesch, B. S. et Maher, D. E. (2015a). Measuring Perceived Usability: The SUS, UMUX-LITE, and Alt Usability. *International Journal of Human-Computer Interaction*, 31(8): 496-505.

Lewis, J. R. et Mayes, D. K. (2014). Development and psychometric evaluation of the Emotional Metric Outcomes (EMO) questionnaire. *International Journal of Human-Computer Interaction*, 30: 685-702.

Lewis, J. R., Utesch, B. S. et Maher, D. E. (2013). UMUX-LITE – When There's No Time for the SUS. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2099-2102.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140: 1-55.

Lin, H. X., Choong, Y. Y. et Salvendy, G. (1997). A Proposed Index of Usability: A Method for Comparing the Relative Usability of Different Software Systems Usability Evaluation Methods. *Behavior and Information Technology*, 16(4/5): 267-278.

Lund, A. (2001). Measuring usability with the USE questionnaire. *Usability and User Experience Newsletter, STC Usability SIG*, 8 (2): 1-4.

Lund, A. (1998). USE Questionnaire Resource Page. Available at http://usesurvey.com .

Macleod, M., Bowden, R., Bevan, N. et Curson, I. (1997).The MUSIC performance measurement method. *Behaviour and Information Technology*, 16 (4/5): 279-293.

Mahatody, T., Sagar, M. et Kolski, C. (2010). State of the art on the cognitive walkthrough method, its variants and evolutions. *International Journal of Human-Computer Interaction*, 26(8):741-785.

Mariage, C. (2005). MetroWeb - logiciel de support à l'évaluation de la qualité ergonomique des sites web. Thèse de doctorat, Université Catholique de Louvain.

Mariage, C., Vanderdonckt, J., Beirekdar, A. et Noirhomme, M. (2004). Destine : outil d'aide à l'évaluation de l'ergonomie des sites web. *In Proceedings of the 16th conference on Association Francophone d'Interaction Homme-Machine*, IHM 2004, pages 117–124.

May, J. et Barnard, P. J. (1994). Supportive evaluation of interface design. *In Proceedings of the First Interdisciplinary Workshop on Cognitive Modeling and User Interface Design*, Vienna, Austria.

McGarry, J., Card, D., Jones, C., Layman, B., Clark, E., Dean, J., et Hall, F. (2002). Practical software measurement: Objective information for decision makers. Boston, MA: Addison-Wesley.

McGee, M. (2004). Master usability scaling: Magnitude estimation and master scaling applied to usability measurement. *In Proceedings of CHI*, pages 335-342, Vienna, Austria.

McGee, M. (2003). Usability magnitude estimation. *In Proceedings of the Human Factors and Ergonomics Society*, HFES, pages 691-695, Santa Monica, CA.

McNamara, N. et Kirakowski, J. (2011). Measuring user-satisfaction with electronic consumer products: The consumer products questionnaire. *International Journal of Human-Computer Studies*, 69(6): 375-386.

Microsoft : Windows user experience interaction guidelines. http://download.microsoft.com/download/e/1/9/e191fd8c-bce8-4dba-a9d5-2d4e3f3ec1d3/UXGuide.pdf, December 2011.

Molich, R., Thomsen, A. D., Karyukina,B., Schmidt, L., Ede, M., van Oel, W. et Arcuri, M. (1999). Comparative evaluation of usability tests. In Proceedings of CHI EA'99, pages 83–84, NewYork, USA.

Morandini, M., de Moraes, R., Leite, R., Cerrato, M. V. et Lordello, C. M. (2011). Project and development of ergocoin version 2.0. *In Proceedings of the 14th international conference on Human-computer interaction: design and development approaches*, Vol. Part I, HCII'11, pages 471-479, Berlin, Heidelberg.

Morasca, S. (2001). *Software Measurement*, Handbook of Software Engineering and Knowledge Engineering, (S.K. Chang, ed.), Chapter 2: Software Measurement, World Scientific, pages 239-276.

Moussa, F., Kolski, C. et Riahi, M. A. (2006). Analyse des dysfonctionnements des systèmes complexes en amont de la conception des ihm : apports, difficultés, et étude de cas. *Revue d'Interaction Homme Machine* (RIHM), 7 : 79-11.

Nielsen, J. (1997). The use and misuse of focus groups. *Software IEEE*, 14(1): 94-95.

Nielsen, J. (1993). *Engineering Usability*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *In Proceedings of CHI*, pages 373-380.

Nielsen, J. et Mack, R. L. (1994). *Usability Inspection Methods*. New York : John Wiley & Sons.

Nielsen, J. et Molich, R. (1990). Heuristic evaluation of user interfaces. *In Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, CHI '90, pages 249-256.

Nielsen, J. et Tahir, M. (2001). *Homepage Usability: 50 Websites Deconstructed*. New Riders Publishing, Thousand Oaks, CA, USA.

Nikov, A., Vassileva, S., Angelova, S., Tzvetanova, S. et Stoeva, S. (2003). WebUse: An approach for web usability evaluation. *In Proceedings of 3rd Symposium on Production Research*, pages 511-518, Istanbul.

Nikov, A. et Stoeva, S. (2001). Quick fuzzy backpropagation algorithm. *Neural Networks,* 14(2): 231-244.

Nishiuchi,N. et Takahashi, Y. (2015). Objective Evaluation Method of Usability Using Parameters of User's Fingertip Movement. *Transactions on Computational Science XXV*. Lecture Notes in Computer Science, vol 9030, pages 77-89.

Nunnally, J. C. (1978).*Psychometric theory*. New York: McGraw-Hill.

Oliveira, K., Thion, V., Dupuy-Chessa, S., Gervais, M., Cherfi Si-Said, S. et Kolski, C. (2012). Limites de l'évaluation d'un système d'information : une analyse fondée sur l'expérience pratique. In Actes XXXème Congrès INFORSID, INFormatique des ORganisations et Systèmes d'Information et de Décision, pages 411-427.

Olmsted-Hawala, E. L. Elizabeth, D., Hawala, S. et Ashenfelter, K. T. (2010).Think aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. *In Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 2381–2390, NewYork, USA. ACM Press.

Olsen, H. (2002). An evaluation of danish qualitative interview investigations. *Nordisk Psykologi*, 54(2): 145-172.

Olsen, Jr. et Dan, R. (2007). Evaluating user interface systems research. *In Proceedings of the 20th annual ACM symposium on User interface software and technology*, UIST '07, pages 251–258, New York, USA. ACM. ISBN 978-1-59593-679-0.

Olsina, L., Papa, F., et Molina, H. (2008). How to Measure and Evaluate Web Applications in a Consistent Way. In HCIS book Web Engineering: Modeling and Implementing Web Applications, Rossi, G., Pastor, Ó., Schwabe, D., and Olsina, L. (Eds.), pages 385-420, Springer.

Padda, H.K. (2003). QUIM map: a repository for usability/quality in use measurement. Masters thesis, Concordia University.

Paganelli, L. et Paternò, F. (2002). Intelligent Analysis of User Interactions with Web Applications. *In Proceedings of ACM IUI*, San Francisco, CA

Paganelli, L., Paternò, F. (2003). Tools for remote usability evaluation of Web applications through browser logs and task models. *Behavior Research Methods, Instruments and Computers*, 35(3): 369-378.

Park, R. E., Goethert, W. B., et Florac, W. A. (1996). *Goal-driven software measurement: A guidebook*. Pittsburgh, PA: Carnegie Mellon University, Software Engineering Institute.

Parush, A. (2000). A Database Approach to Building and Using Online Human Computer Interaction Guidelines. In Vanderdonckt, J. et Farenc, C. (eds): *Tools for Working with Guidelines*, pages 77-84, London: Springer-verlag.

Paternò, F., Russino, A. et Santoro, C. (2007). Remote evaluation of Mobile Applications. *In Proceedings of 6th International Workshop of Task Models and Diagrams for User Interface Design*, TAMODIA 2007.

Polkosky, M. D. (2008). Machines as mediators: The challenge of technology for interpersonal communication theory and research. *In* Kojin, E. (ed.): *Mediated interpersonal communication*, pages 34-57, New York: Routledge.

Privitera, G. J. (2015). *Statistics for the behavioral sciences*. 2nd Revised edition. SAGE Publications Inc.

Purchase, H. C. (2012). *Experimental Human-Computer Interaction: A Practical Guide with Visual Examples*. Publisher Cambridge University Press.

Rauterberg, M. et Aeppili, R. (1995). Learning in man-machine systems: the measurement of behavioural and cognitive complexity. *In Proceedings of the IEEE Conference on Systems, Man and Cybernetics couver*, pages 4685-4690. Institute of Electrical and Electronics Engineers.

Rieman, J., Davies, S., Hair, D. C., Esemplare, M., Polson, P. et Lewis, C. (1991). An automated cognitive walkthrough. *In Proceedings of the Conference on Human Factors in Computing Systems*, pages 427-428.

Rohrer, C. (2014). When to use which user-experience research methods. Retrieved from https://www.nngroup.com/articles/which-ux-research-methods/.

Rosenbaum, S., Cockton, G., Coyne, K., Muller, M., et Rauch, T. (2002). Focus groups in HCI: wealth of information or waste of resources. In Proceedings of Human Factors in Computing Systems, CHI '02 Extended Abstract, pages 702-703.

Rubin, J. et Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Second Edition, Wiley.

Rukshan, A. et Baravalle, A. (2011). A quantitative approach to usability evaluation of web sites. *In Proceedings of Advances in Computing Technology*.

Russ-Eft, D. et Preskill, H. (2001). Evaluation in organizations: A systematic approach to enhancing learning, performance, and change. New York, NY: Basic Books.

Ryu, Y. S. (2009). Mobile phone usability questionnaire (MPUQ) and automated usability evaluation. *In Proceedings of 13th International Conference Human-Computer Interaction new trends*, pages 349-351.

Ryu, Y. S. et Smith-Jackson, T. L. (2005). Usability Questionnaire Items for Mobile Products and Content Validity. *In Proceedings of HCI International*, pages 22-27, Las Vegas.

Sauro, J. (2015). SUPR-Q: A Comprehensive Measure of the Quality of the Website User Experience. *Journal of usability studies*, 10(2): 68-86.

Sauro, J. (2011a). *A Practical Guide to the System Usability Scale (SUS): Background, Benchmarks & Best Practices*. Measuring Usability LLC, Denver.

Sauro, J. (2011b). *The Standardized Universal Percentile Rank Questionnaire (SUPR-Q)*. Available at www.suprq.com/.

Sauro, J. (2005). *SUM: Single Usability Metric*. Retrieved from http://www.measuringu.com/SUM/index.htm.

Sauro, J. et Kindlund, E. (2005). A Method to Standardize Usability Metrics Into a Single Score. *In Proceedings of CHI*, pages 401-409.

Sauro, J. et Lewis, J. R. (2012). *Quantifying the user experience: Practical statistics for user research*. Amsterdam: Elsevier/Morgan Kaufmann

Sauro, J. et Dumas, J. S. (2009). Comparison of three one-question, post-task usability questionnaires. *In Proceedings of CHI*, pages 1599-1608. ACM, Boston.

Sauro, J. et Lewis, J. R. (2009). Correlations among prototypical usability metrics: Evidence for the construct of usability. *In Proceedings of CHI*, pages 1609-1618, Boston.

Scapin, D. L. et Bastien, J. M. C. (1997). Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour and Information Technology*, 16: 220-231.

Schofield, J. (2005).The Statistically Unreliable Nature of Lines of Code, *CrossTalk*, 18(4): 29-33, April 2005. https://samate.nist.gov/index.php/Metrics_and_Measures.html

Sears, A. et Jacko, J. (2009). *Human-Computer Interaction: Development Process*.CRC Press.

Seffah, A. et Habieb-Mammar, H. (2009). Usability engineering laboratories: limitations and challenges toward a unifying tools/practices environment. *Behaviour and Information Technology*, 28(3): 281-291.

Seffah, A. et Metzker, E. (2008). *Adoption-centric Usability Engineering: Systematic Deployment, Assessment and Improvement of Usability Methods in Software Engineering*. Springer Publishing Company, Incorporated.

Seffah, A., Donyaee, M., Kline, R. B. et Padda, H. K. (2006). Usability measurement and metrics: A consolidated model. *Software Quality Journal*, 14(2): 159-178.

Senach, B. (1990). Evaluation ergonomique des IHM : Une revue de la littérature. Rapport technique, INRIA N 1180.

Shackel, B. (1991). Usability – context, framework, design and evaluation. *In* Shackel, B. et Richardson, S. (eds.): *Human Factors for Informatics Usability*, pages 21-38, Cambridge University Press, Cambridge.

Shneiderman, B. (1998). Tree maps for space-constrained visualization of hierarchies. *Human Computer Interaction Lab,* University of Maryland.

Shneiderman, B. (1992). *Designing the user interface: strategies for effective human computer interaction*. Boston: Addison-Wesley Longman Publishing.

Shneiderman, B. et Plaisant, C. (2005). Chapter 14.5: Information Visualization. In *Designing the User Interface*, pages 580-603, Boston: Pearson.

Sodiya, A. S. (2009). User interface design and ergonomics. Rapport CIT 711. National Open University of Nigeria. http://documents.mx/documents/cit-711.html.

Staron, M., Meding, W., et Nilsson. C. (2009). A framework for developing measurement systems and its industrial evaluation. *Information and Software Technology*, 51: 721–737.

Suetendael, N. V., et Elwell, D. (1991). *Software quality metrics*. Atlantic City International Airport, NJ: Federal Aviation Administration Technical Center.

Tardieu, J., Misdariis, N., Langlois, S., Gaillard, P. et Lemercier, C. (2015). Sonification of in-vehicle interface reduces gaze movements under dual-task condition. *Applied Ergonomics*, 50: 41-49.

TechSmith. (2007). Morae: Usability testing solution for web sites and software. http ://www.techsmith.com/Morae.asp., December 2012.

Tedesco, D. P. et Tullis, T. S. (2006). A comparison of methods for eliciting post-task subjective ratings in usability testing. Proceedings of the Usability Professionals Association Annual Conference.

Tennant, G. (2001). *SPC and TQM in manufacturing and services*, Gower, Hampshire.

Thimbleby, H. (2007). Interaction walkthrough: Evaluation of safety critical interactive systems. *In* Doherty, G.et Blandford, A. (eds): DSVIS 2006, volume 4323 of LNCS, pages 52–66, Springer Verlag.

Trabelsi, A. (2006). Contribution à l'évaluation des systèmes interactifs orientés agents, application à un poste de supervision de transport urbain. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis.

Trabelsi, A. et Ezzedine, H. (2013). Evaluation of an Information Assistance System based on an agent-based architecture in transportation domain: first results. *International Journal of Computers*, *Communications and Control*, 8(2): 320-333.

Trabelsi, A., Ezzedine, H. et Kolski, C. (2006). Un mouchard électronique orienté agent pour l'évaluation de systèmes interactifs de supervision. *In Proceedings of the Conférence Internationale Francophone d'Automatique*, CIFA'2006, Bordeaux, France.

Tran, C. (2009). Vers un environnement générique et configurable pour l'aide à l'évaluation des systèmes interactifs à base d'agents. Application à un Système d'Aide à l'Information Voyageur. Thèse de doctorat, Université de Valenciennes et du Hainaut Cambrésis.

Tran, C. D., Ezzedine, H. et Kolski, C. (2013). EISEval, a generic reconfigurable environment for evaluating agent-based interactive systems. *International Journal of Human-Computer Studies*, 71(6): 725-761.

Tran, C., Ezzedine, H. et Kolski, C. (2008). Evaluation of agent-based interactive systems: proposal of an electronic informer using Petri Nets. *Journal of Universal Computer Science*, 14(9): 3202-3216.

Travis , D. (2014). 247 web usability guidelines. www.userfocus.co.uk/resources/guidelines.html, Octobre 2014.

Tufte, E. R. (2002).*The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press)

Tullis, T. S. (1997). Screen design. *In* Helander, M. Landauer, T. K. et Prabhu, P. (eds.): *Handbook of human computer interaction,* 2nd ed., pages 503-531, NY: Elsevier.

Tullis, T. S. (1984). Predicting the Usability of Alphanumeric Displays. PhD.Dissertation, Rice University. Lawrence, Kansas: The Report Store.

Tullis, T. S. (1983). The Formatting of Alphanumeric Displays: A Review and Analysis. *Human Factors*, 25(6): 657-682.

Tullis, T. et Albert, B. (2008). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Amsterdam: Elsevier/Morgan Kaufmann.

Tullis, T. S. et Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. *In Proceedings of the Usability Professionals Association Annual Conference*, pages 1-12, Minneapolis.

Usabilla. (2012). Usabilla. www.usabilla.com, September 2012.

Väänänen-Vainio-Mattila, K. et Segerståhl, K. (2009). A Tool for Evaluating Service User eXperience (ServUX): Development of a Modular Questionnaire. *In User Experience Evaluation Methods in Product Development,* UXEM'09, Workshop in Interact'09 conference, Uppsala, Sweden.

Vanderdonckt, J. (2003). Visual Design Methods in Interactive Applications. *In* Albers, M. et Mazur, B. (eds.): *Content and Complexity: Information Design in Technical Communication,* pages 187-203, Lawrence Erlbaum Associates, Mahwah.

Vanderdonckt, J. (1999). Development milestones towards a tool for working with guidelines. *Interacting with Computers*, 12(2): 81-118.

Vanderdonckt, J. (1994). *Guide ergonomique des interfaces homme-machine*. Presses Universitaires de Namur.

Vanderdonckt, J. et Gillo, X. (1994). Visual techniques for traditional and multimedia layouts. *In Proceedings of the Workshop on Advanced Visual Interfaces*, AVI '94, pages 95-104.Bari, Italy.

Waterson, S. (2002). Web Quilt: A Visual Analysis Tool for Understanding Web Usability Click stream Data. Masters Report, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley.

Whitefield, A., Wilson, F., et Dowell, J. (1991). A framework for human factors evaluation. *Behaviour & Information Technology*, 10(1): 65-79.

Wickens, C. D., John, D. L., Liu, Y. et Becker, S. E. G. (2004). *An Introduction to Human Factors Engineering*. Upper Saddle River, NJ: Pearson Prentice Hall.

Witte, R. S. et Witte, J. S. (2009) *Statistics*. John Wiley & Sons.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B. et Wesslén, A. (2000). *Experimentation in software engineering: An introduction*. Boston: Kluwer Academic.

Yang, T., Linder, J. et Bolchini, D. (2012). DEEP: Design-Oriented Evaluation of Perceived Usability. *International Journal of Human-Computer Interaction*, 28(5): 308-346.

Zen, M. et Vanderdonckt, J. (2014). Towards an evaluation of graphical user interfaces aesthetics based on metrics. *In Proceedings of IEEE Eighth International Conference on Research Challenges in Information Science*, RCIS, pages 1-12.

Zettlemoyer, L. S., Amant, R. et Dulberg, M. S. (1999). Ibots: agent control through the user interface. *In Proceedings of the 4th international conference on intelligent user interfaces*, IUI '99, pages 31-37.

Zhang, Z., Basili, V. et Shneiderman, S. (1999). Perspective-based usability inspection : An empirical validation of efficacy. *Empirical Software Engineering: An International Journal*, 4: 43–70.

Zijlstra, R., et van Doorn, L. (1985). The Construction of a Scale to Measure Subjective Effort (Tech. Rep.). Delft University of Technology, Department of Philosophy and Social Sciences, Delft, Netherlands.

Zülch, G et Stowasser, S. (2000). Usability Evaluation of User Interfaces with the Computer-aided Evaluation Tool PRO-KUS. *MMI Interaktiv*, 1(3) : 1-17.

Annexe A

Tableau A. 1 : Synthèse globale sur les échelles des questionnaires standardisés existants

Questionnaire et sa date de création	Échelle adoptée
QUIS	Echelle sémantique de 10 points numériques allant de 0 à 9.
TAM	Echelle sémantique de 7 points nominales allant de « probable » à « improbable »
ASQ	Echelle de Likert de 7 points allant de « 1 : fortement en accord » jusqu'à « 7 : fortement en désaccord »
PSSUQ	Echelle de Likert de 7 points allant de «1 : fortement en accord» jusqu' à «7 : fortement en désaccord»
SUMI	Echelle dichotomique de 3 points (d'accord, ne sait pas et pas d'accord)
CSUQ	Echelle de Likert de 7 points[20] allant de «1 : fortement en accord» jusqu' à «7 : fortement en désaccord»
SUS	Echelle de Likert de 5 points allant de «1 : Fortement en désaccord» jusqu'à « 5 : Fortement en accord »
PUTQ	Echelle de Likert de 7 points allant de «1 : Mauvais» jusqu'à «7 : Bien»
WAMMI	Echelle de Likert de 5 points allant de «fortement en accord» jusqu'à «fortement en désaccord».
USE	Echelle de Likert de 7 points allant de «1 : Désaccord» jusqu'à «7 : Accord»
ER	Echelle de Likert de 7 points (version 1) et 5 points (Version 2) allant de « très facile » à « très difficile »
WEBUSE	Echelle de Likert de 5 points nominaux : « fortement en accord», «accord », «Passable », « Désaccord », «fortement en désaccord»
UME	Echelle spécifique UME en attribuant une valeur supérieure à zéro
MPUQ	Echelle de Likert de 7 points allant de «1 : fortement en désaccord» à «7 : fortement en accord»
SEQ	Echelle de Likert de 5 points (Version 1) et 7 points (Version 2) allant de « très difficile » jusqu'à « très facile »
WEQ	Echelle de Likert de 5 points allant de «fortement en désaccord» à «fortement en accord»
SMEQ	Échelle spécifique graduée de 0 à 150 avec neuf étiquettes de «Pas du tout difficile à faire» à «Extrêmement difficile à faire »
UMUX	Echelle de Likert de 7 points allant de «1 : fortement en désaccord» jusqu'à «7 : fortement en accord»
SUPR-Q	Echelle de Likert de : - 5 points allant de «1 : fortement en désaccord» à «5 : fortement en accord» (pour les 12 premiers éléments) - 11 points allant de « 0 : Pas de tout probable » à « 10 : Extrêmement probable » (pour le dernier élément)
DEEP	Echelle de Likert de 5 points nominaux allant de «fortement en désaccord» à «fortement en accord»
T-CSUQ	Echelle de Likert de 7 points allant de «1 : fortement en accord» jusqu' à «7 : fortement en désaccord»
UMUX-LITE	Echelle de Likert de 7 points allant de «1 : fortement en désaccord» à «7 : fortement en accord»
SUISQ **SUISQ-R** **SUISQ-MR**	Echelle de Likert de 5 points nominaux : « Extrêmement insatisfait, insatisfait, ni satisfait ni insatisfait, satisfait, et extrêmement satisfait »
AltUsability	Echelle de Likert de 7 points « 1 : fortement en désaccord » jusqu'à « 7 : fortement en accord »

[20] Notant que dans les versions originales de CSUQ et PSSUQ, les auteurs ont défini la valeur 1 pour indiquer un plus grand degré de satisfaction. Tandis que certains ont préféré échanger les étiquettes de «fortement en accord» et «fortement en désaccord » (voir Tullis et Albert, 2008, p. 140). Selon Sauro et Lewis : *"From a strict interpretation of standardization, it's best to avoid this type of manipulation unless there is evidence that it does not affect the factor structure of the items"(Sauro et Lewis, 2012, p. 193).* Dans notre travail, nous avons adopté les échelles des versions originales.

Tableau A. 2 : Aspects de qualité de questionnaires universels d'utilisabilité

Aspects de qualité	QUIS	TAM	PSSUQ / CSUQ		Alt_ Usability	T-CSUQ	SUMI	SUS	PUTQ	USE	UMUX	UMUX-LITE	ASQ	ER	UME	SMEQ	SEQ
Satisfaction										X	X						
Réaction globale au logiciel/ Système global	X		X	X		X											
Facteurs d'écran	X																
Terminologie et les informations du système	X																
Facteurs d'apprentissage/ Apprentissage/ Facilité d'apprentissage	X						X	X	X	X							
Capacités du système	X																
Utilisabilité / Facilité d'utilisation		X						X		X		X					
Utilité du système/ Adéquation fonctionnelle du système		X	X	X		X				X		X					
Qualité de l'information			X	X		X											
Qualité de l'interface			X	X		X											
Efficience					X		X				X						
Affect							X										
Serviabilité							X										
Contrôle					X		X										
Compatibilité									X								
Cohérence									X								
Flexibilité									X								
Action minimale									X								
Charge minimale de la mémoire									X								
Limitation de la perception									X								
Guidage de l'utilisateur									X								
Efficacité											X						
Facilité globale d'achèvement de la tâche													X	X	X	X	X
Satisfaction avec le temps d'achèvement													X				
Satisfaction à l'égard des informations d'aide													X				
Facilité de navigation					X												
Capacité de recherche					X												
Familiarité					X												
Le besoin					X												
Attractivité					X												

Tableau A. 3: Aspects de qualité de questionnaires spécifiques à l'utilisabilité

Aspects de qualité	WAMMI	WEBUSE	WEQ	SUPR-Q	DEEP	MPUQ	SUISQ
Attractivité	X						
Contrôlabilité	X						
Efficience	X						
Apprentissage	X						
Serviabilité	X						
Contenu, organisation et lisibilité		X					
Navigation et liens		X					
Conception de l'interface utilisateur		X					
Performance et efficacité		X					
Contenu –pertinence			X				
Contenu- compréhensibilité			X				
Contenu- exhaustivité			X				
Navigation - convivialité			X				
Navigation - structure			X				
Navigation- hyperliens			X				
Navigation - vitesse			X				
Navigation – option de recherche			X				
Mise en page (Layout)			X				
Apparence				X			
Fidélité				X			
Facilité d'utilisation				X			
Crédibilité / Confiance				X			
Contenu					X		
Structure et architecture de l'information					X		
Navigation					X		
Effort cognitif perçu					X		
Mise en cohérence					X		
Guidage visuel					X		
Facilité d'apprentissage et d'utilisation						X	
Assistance à l'exploitation et à la résolution de problèmes						X	
Aspect émotionnel et capacités multimédia						X	
Commandes et charge minimale de la mémoire						X	
Efficience et contrôle						X	
Tâches typiques pour les téléphones mobiles						X	
Orientation vers un but de l'utilisateur							X
Comportement de service de la clientèle							X
Le facteur des caractéristiques de la parole							X
Verbosité							X

Tableau A. 4: Exemples des éléments de questionnaires universels d'utilisabilité en fonction des critères standards

Critères d'utilisabilité		Items	Questionnaire
Critères de l'ISO 9421-11	Efficacité	Tasks can be performed in a straight-forward manner	QUIS
		I can effectively complete my work using this system	CSUQ/PSSUQ/T-CSUQ
		Tasks can be performed in a straight forward manner using this software	SUMI
		I can use it successfully every time	USE
		[This system's] capabilities meet my requirements (Finstad, 2010 ; Kerzazi et Lavallée, 2011)	UMUX/UMUX-Lite
		Overall, I am satisfied with the ease of completing the tasks in this scenario (Lewis, 1995)	ASQ
	Efficience	I am able to complete my work quickly using this system	CSUQ/PSSUQ/T-CSUQ
		I found the system very cumbersome to use	SUS
		This software responds too slowly to inputs	SUMI
		I have to spend too much time correcting things with [this system] (Kerzazi & Lavallée, 2011)	UMUX
		Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario (Lewis, 1995)	ASQ
		Using [this product] in my job would enable me to accomplish tasks more quickly	TAM
		This system helps me to do my job more efficiently	AltUsability
	Satisfaction	Overall reactions to the software frustrating satisfying	QUIS
		Overall, I am satisfied with this system	CSUQ/PSSUQ/T-CSUQ
		I think that I would like to use this system	SUS
		Working with this software is satisfying	SUMI
		I am satisfied with it	USE
		Using [this system] is a frustrating experience (Finstad, 2010 ; Kerzazi et Lavallée, 2011)	UMUX
		Overall, I am satisfied with the ease of completing the tasks in this scenario	ASQ
Critères de l'ISO/WD 9241-112	Detectabilité	Characters on the computer screen (hard to read,…, easy to read)	QUIS
		Either the amount or quality of the help information varies across the system	SUMI
		Are selected data highlighted?	PUTQ
		My interaction with [this product] would be clear and understandable	TAM
	Discriminabilité	Organization of information on screen (confusing,…, very clear)	QUIS
		The organization of information on the system screens is clear	CSUQ/PSSUQ/T-CSUQ
		The way that system information is presented is clear and understandable	SUMI
		Are menus distinct from other displayed information?	PUTQ
	Adéquation (Appropriateness)	I found the various functions in the system were well integrated	SUS
		The software documentation is very informative	SUMI
		Are data items kept short?	PUTQ
	Compréhensibilité	My interaction with [this product] would be clear and understandable	TAM
		I can understand and act on the information provided by this software	SUMI
		Do the commands have distinctive meanings?	PUTQ
		Messages on screen which prompt user for input (confusing,...,clear)	QUIS
	Cohérence (des informations présentées)	Use of terms throughout system (inconsistent,…, consistent)	QUIS
		I thought there was too much inconsistency in this system	SUS
		I think this software is inconsistent	SUMI
		Is the display format consistent?	PUTQ
		I don't notice any inconsistencies as I use it	USE
Critères de Scapin et Bastien	Guidage	Highlighting on the screen simplifies task	QUIS
		Computer keeps you informed about what it is doing (never always)	QUIS
		Help messages on the screen (unhelpful,…, helpful)	QUIS
		The information (such as online help, on-screen messages and other documentation) provided with this system is clear	CSUQ/PSSUQ/T-CSUQ
		The organization of information on the system screens is clear	CSUQ/PSSUQ/T-CSUQ
		The organization of the menus seems quite logical	SUMI
		It makes the things I want to accomplish easier to get done	USE
		Are groups of information demarcated?	PUTQ
		Is the guidance information always available?	PUTQ
		Is HELP provided?	PUTQ
		Overall, I am satisfied with the support information (online help, messages, documentation) when completing the tasks	ASQ
	Charge de travail	It is easy to find the information I needed	CSUQ/PSSUQ/T-CSUQ
		There is never enough information on the screen when it's needed	SUMI
		There are too many steps required to get something to work	SUMI
		It requires the fewest steps possible to accomplish what I want to do with it	USE
		Is the screen density reasonable?	PUTQ
	Contrôle explicite	I feel in command of this software when I am using it	SUMI
		Does it provide CANCEL option?	PUTQ
		I feel in control when I work within this system	AltUsability

Tableau A.4 (suite): Exemples des éléments de questionnaires universels d'utilisabilité en fonction des critères standards

Critères d'utilisabilité		Items	Questionnaire
	Cohérence (des choix de conception d'interface)	Use of terms throughout system (inconsistent,…, consistent) I thought there was too much inconsistency in this system I think this software is inconsistent Is the display format consistent? I don't notice any inconsistencies as I use it	QUIS SUS SUMI PUTQ USE
	Adaptabilité	Experienced and inexperienced users' needs are taken into consideration It is obvious that user needs have been fully taken into consideration It is easy to make the software do exactly what you want I would find it easy to get [this product] to do what I want it to do. I would find [this product] to be flexible to interact with It is flexible Both occasional and regular users would like it Does system provide good training for different users? Can user name displays and elements according to their needs? This system offers capabilities familiar to me	QUIS SUMI SUMI TAM TAM USE USE PUTQ PUTQ AltUsability
	Gestion des erreurs	Error messages (unhelpful,…, helpful) /Correcting your mistakes difficult The system gives error messages that clearly tell me how to fix problems Whenever I make a mistake using the system, I recover easily and quickly Error messages are not adequate I can recover from mistakes quickly and easily Are erroneous entries displayed? Are error messages non-disruptive/informative?	QUIS CSUQ/PSSUQ/T-CSUQ CSUQ/PSSUQ/T-CSUQ SUMI USE PUTQ PUTQ
	Signifiance des codes et dénominations	Computer terminology is related to the task you are doing I sometimes wonder if I am using the right function Are the command names meaningful? Do the commands have distinctive meanings?	QUIS SUMI PUTQ PUTQ
	Compatibilité	The software hasn't always done what I was expecting Is the control of cursor compatible with movement? Are the results of control entry compatible with user expectations?	SUMI PUTQ PUTQ

Tableau A. 5 : Exemples des éléments de questionnaires spécifiques d'utilisabilité en fonction des critères standards

Critères d'utilisabilité		Items	Questionnaire
Critères de l'ISO 9421-11	Efficacité	It is efficient to use this website. Does the product support the operation of all the tasks in a way that you find useful?	WEBUSE MPUQ
	Efficience	I need not wait too long to download a file or open a page I think it takes a long time to download a new web page from this site. I am able to find what I need quickly on this website I could quickly get to know the structure of the website by skimming its home page. Does this product enable the quick, effective, and economical performance of tasks?	WEBUSE WEQ SUPR-Q DEEP MPUQ
	Satisfaction	I don't like using this website I enjoy using the website Do you feel excited when using this product?	WAMMI SUPR-Q MPUQ
Critères de l'ISO/WD 9241-112	Detectabilité	This website helps me find what I am looking for Reading content at this website is easy Placement of links or menu is standard throughout the website and I can easily recognize them It is clear which hyperlink will lead to the information I am looking for. The wording of the text was clear. The highlighted areas of a page helped me locate the information I needed Are the characters on the screen easy to read?	WAMMI WEBUSE WEBUSE WEQ DEEP DEEP MPUQ
	Discriminabilité	This website seems logical to me. The content of this website is well organized I find the structure of this website clear The website has a clean and simple presentation Under each section of the website, the web pages were well organized. Is the organization of information on the product screen clear?	WAMMI WEBUSE WEQ SUPR-Q DEEP MPUQ

Tableau A. 5 (suite): Exemples des éléments de questionnaires spécifiques d'utilisabilité en fonction des critères standards

Critères d'utilisabilité		Items	Questionnaire
Critères de l'ISO/WD 9241-112	Adéquation (Appropriateness)	I can quickly find what I want on this website. I can easily find what I want at this website I find the information in this website precise The information on this website is valuable It was easy to find the information I needed on the website Is the amount of information displayed on the screen adequate?	WAMMI WEBUSE WEQ SUPR-Q DEEP MPUQ
	Compréhensibilité	Everything on this website is easy to understand I am comfortable and familiar with the language used I find the information in this website easy to understand The content (including text, pictures, audios, and videos etc.) was easy to understand Is the interface with this product clear and understandable?	WAMMI WEBUSE WEQ DEEP MPUQ
	Cohérence	This website has a consistent feel and look The layout under each section of the website was consistent Is the data display sufficiently consistent?	WEBUSE DEEP MPUQ
Critères de Scapin et Bastien	Guidage	This website helps me find what I am looking for. This website always provides clear and useful messages when I don't know how to proceed I can easily know where I am at this website I always know where I am on this website The website has a clean and simple presentation This website helped me find what I was looking for Is the backlighting feature for the keyboard and screen helpful?	WAMMI WEBUSE WEBUSE WEQ SUPR-Q DEEP MPUQ
	Charge de travail	I can quickly find what I want on this website I can easily find what I want at this website I find the information in this website precise The information on this website is valuable It was easy to find the information I needed on the website Are data items kept short?	WAMMI WEBUSE WEQ SUPR-Q DEEP MPUQ
	Contrôle explicite	I feel in control when I'm using this website. It is easy to move around at this website by using the links or back button of the browser Can you regulate, control, and operate the product easily?	WAMMI WEBUSE MPUQ
	Homogénéité/ Cohérence	This website has a consistent feel and look The layout under each section of the website was consistent Is the data display sufficiently consistent?	WEBUSE DEEP MPUQ
	Adaptabilité	Have the user needs regarding this product been sufficiently taken into consideration?	MPUQ
	Gestion des erreurs	This website does not contain too many web advertisements Are the messages aimed at prevent you from making mistakes adequate? Are the error messages effective in assisting you to fix problems?	WEBUSE MPUQ MPUQ
	Signifiance des codes et dénominations	I get what I expect when I click on things on this website I am comfortable and familiar with the language used I find many words in this website difficult to understand I got what I expected when I clicked on things on this website Are the command names meaningful?	WAMMI WEBUSE WEQ DEEP MPUQ
	Compatibilité	Are the color coding and data display compatible with familiar conventions?	MPUQ

Les outils utilisant les règles ergonomiques

Dans la littérature, nombreux des outils d'évaluation portant sur l'inspection de la conformité des IHM en exploitant les connaissances et les règles ergonomiques ont été proposés. Une synthèse sur quelques outils est présentée dans le Tableau A.6 en se basant sur les critères suivants :

- **Nature de l'outil :** logiciel, service web
- **Interface utilisateur visée :** spécifie le type d'interface que l'outil peut évaluer, citons par exemples : l'interface WIMP, WUI ou l'interface utilisateur des applications mobiles.

- **Intervenants dans le processus d'évaluation :** expert, évaluateur, développeur, concepteur, utilisateur.
- **Aspects de qualité à inspecter :** concerne les aspects de qualité à inspecter tels que l'utilisabilité, l'accessibilité, etc.
- **Les règles qui caractérisent l'outil et leurs structures et mises en place :** concerne les règles utilisées et la manière adoptée pour les structurer et les mettre en place.
- **La gestion des règles :** concerne l'ajout, l'accès, la modification et la suppression des règles.
- **Résultats de l'outil :** concerne les services fournis par l'outil, nous citons par exemples : rapport d'évaluation, recommandations sur les interfaces utilisateur établies par l'outil, suggestions d'amélioration de l'interface utilisateur, règles ergonomiques non respectées.

Les outils de recueil des données d'interaction entre l'utilisateur et le système interactif à évaluer

Nous résumons dans le Tableau A.7 quelques exemples de ces outils en tenant compte des aspects suivants :

- **Nature de l'outil :** logiciel, service web.
- **Nature de système à évaluer :** précise la nature du système interactif à évaluer : (tels que, les applications WIMP, Web, Mobile, etc.).
- **Technique de capture des données et leurs natures :** Spécifie la technique de capture adoptée par l'outil et la nature des évènements capturés de l'interaction. Selon Hilbert et Redmiles (2000), il existe six niveaux d'évènements :
  - Les événements physiques (nommés aussi les événements de plus bas niveau, par exemple, le clic souris).
  - Les événements des périphériques d'entrée (par exemple, le matériel généré par les interruptions clés ou de la souris).
  - Les événements interface utilisateur (tel que, les changements dans la focalisation d'entrée)
  - Le niveau abstrait d'interaction (par exemple, fournir des valeurs dans les champs d'entrée).
  - Le domaine (par exemple, fournir des informations d'adresse).
  - Le but/ la tâche (par exemple, passer une commande)
- **Intervenants dans le processus d'évaluation :** indique les intervenants lors de l'évaluation et l'exécution des tâches, tels que : l'expert, l'évaluateur, le développeur, les utilisateurs, etc.
- **Niveau d'automatisation de l'outil :** concerne les trois phases de capture, analyse, et critique. En s'inspirant d'Ivory et Hearst (Ivory et Hearst, 2001), nous distinguons trois niveaux, respectivement : Manuel (M), Semi-Automatique (SM), et Automatique (A).
- **Résultats d'évaluation fournis :** indique les problèmes détectés et la capacité de l'outil à suggérer des améliorations de l'interface.
- **Présentation des résultats d'analyse :** décrit la façon avec laquelle les résultats d'analyse sont présentés.

Les outils présentés seront ordonnés dans l'ordre croissant de leurs apparitions. Pour plus de détails le lecteur peut se référer aux thèses de (Charfi, 2013), (Tran, 2009), (Trabelsi, 2006) et (Ivory et Hearst, 2002).

Tableau A. 6 : Synthèse globale sur quelques outils d'évaluation de l'utilisabilité utilisant les connaissances et les règles ergonomiques

Outil	Nature de l'outil	Interface utilisateur visée	Intervenants	Aspects de qualité à inspecter	Les règles et leurs structures	Gestion des règles	Résultats fournis par l'outil
Sherlock (Grammenos et al, 2000)	Logiciel	Interfaces WIMP	Evaluateur Experts d'utilisabilité Concepteur Analyste d'utilisabilité	Utilisabilité	Il concerne n'importe quel type des règles relatives aux interfaces WIMP. Ces règles sont sauvegardées dans des bases de règles de système sous un format spécifique structuré selon différentes propriétés telles que la clef, le titre, la classe de règles, type d'inspection, l'auteur, etc.	Accéder, parcourir et choisir des règles et ajouter des nouvelles de règles.	Rapport d'évaluation contenant les violations d'utilisabilité détectées sur l'interface évaluée
Système de gestion de règles (Parush, 2000)	Logiciel	Différentes plateformes spécifiques ou générales	Evaluateur	Utilisabilité	Une règle est caractérisée par trois niveaux (section, sous-section, et titre). Quatre modules complémentaires ont été mis en place dans une règle (la recommandation, les notes additionnelles, le fichier d'image, et la liste des références aux autres règles).	Créer des nouvelles règles, et modifier les règles existantes	Les informations liées à la règle
ERGOVAL (Farenc et al., 2001)	Logiciel	Interfaces WIMP	Evaluateur	Utilisabilité Accessibilité	Les règles ergonomiques sont implémentées dans la base de connaissances d'ERGOVAL et ne nécessitent pas l'utilisation des autres connaissances au cours de l'évaluation. Ces règles sont liées aux interfaces graphiques selon les formes des objets et leurs placements.	Aucune	Rapport d'évaluation qui comprend les règles violées
Web Tango (Ivory et Hearst, 2001)	Logiciel	Interfaces web	Concepteurs non professionnels	Utilisabilité	Les règles (nommées profils) sont construites en utilisant un ensemble des mesures quantitatives. Elles sont établies à partir de sites de haut rang.	Aucune	Qualité de l'interface (bonne, moyenne, ou faible)

Tableau A.6 (suite) : Synthèse globale sur quelques outils d'évaluation de l'utilisabilité utilisant les connaissances et les règles ergonomiques

Outil	Nature de l'outil	Interface utilisateur visée	Intervenants	Aspects de qualité à inspecter	Les règles et leurs structures	Gestion des règles	Résultats fournis par l'outil
DESTINE (Mariage et al, 2004)	Logiciel	Interfaces web	Evaluateur avec des bonnes connaissances en ergonomie	Utilisabilité Accessibilité	Les règles sont structurées d'une façon dynamique sous forme de fichiers XML en adoptant le langage GDL (~~–G~~uideline Definition Langage")	Les règles sont flexibles (ajouter, modifier, supprimer des règles)	Rapport d'évaluation détaillé qui inclut les corrections à effectuer
Outil d'évaluation ergonomique (Charfi et Ezzedine, 2011)	Logiciel	Interfaces WIMP	Evaluateur	Utilisabilité	Les règles sont établies sous forme textuelle et elles sont sauvegardées dans un modèle de référence.	Créer des nouvelles règles	Rapport général et rapport détaillé sous forme PDF contiennent l'ensemble de défaillances et de recommandations ergonomiques relatives aux composants graphiques de l'interface.
AWebHUT (Rukshan et Baravalle, 2011)	Service web	Interfaces web	Evaluateur	Utilisabilité	Les règles sont bien définies et sont associées à des différentes catégories telles que la cohérence de la présentation et le contrôle, le feedback adéquat, la navigation contextuelle.	Aucune	Rapport de synthèse indique le niveau de l'utilisabilité de l'interface (Neutre, Violer, Respecter, et Erreur) Rapport détaillé présente chaque règle d'utilisabilité et précise où elle est violée
Inspecteur des règles ergonomiques (Charfi et al., 2014, 2015)	Logiciel	Interfaces WIMP, Web et Mobiles	Expert en ergonomie Evaluateur	Utilisabilité	Les règles sont présentées sous forme des fichiers XML. Une règle est structurée selon différents attributs, à titre d'exemples nous citons : (1) un identifiant numérique, qui est attribué automatiquement aux nouvelles règles, (2) une description générale de la règle, (3) les erreurs associées et les recommandations dans le cas de non-conformité avec la règle, etc.	Ajouter une nouvelle règle, consulter une règle, modifier une règle, évaluer la configuration, et choisir une règle	Rapport d'évaluation qui comprend les problèmes de conception détectés, les recommandations pour améliorer l'interface utilisateur et un aperçu du processus d'évaluation. Ce rapport peut être généré sous différents formats (PDF, HTML et TXT).

Tableau A. 7 : Synthèse globale sur quelques outils d'évaluation de recueil des données d'interaction entre l'utilisateur et le système interactif à évaluer

Outil	Nature de l'outil	Nature de système à évaluer	Technique de capture et Nature des données capturées	Intervenants	Automatisation*			Résultats d'évaluation fournis	Présentation des résultats d'analyse
					Capture	Analyse	Critique		
WebQuilt (Hong et al., 2001 ; Waterson, 2002)	Service web	Web	Il utilise un Proxy pour capturer des données portant sur les actions des utilisateurs (événements physiques, tel que le clic sur les pages web) et les sauvegarde dans des fichiers log.	Utilisateurs Evaluateur	A	A	M	Les résultats indiquent un ensemble des caractéristiques des pages web (tels que les pages visitées, la fréquence des chemins suivis, les patterns de navigation)	Les résultats de l'analyse sont générés à travers un graphe interactif qui représente les pages visitées et les transitions entre les pages.
Web RemUSINE (Paganelli et Paternò, 2002 ; 2003)	Logiciel	Web	La capture des données est établie via l'outil QC/ Replay (log file, Javascript et les Applets). Ces données portent sur trois types d'événements : les actions d'interactions de l'utilisateur sur le navigateur web, les événements internes du navigateur, et les événements portant sur le changement de la tâche cible de l'utilisateur. Il capture aussi quelques données additionnelles telles que l'adresse IP, la date, etc.	Utilisateurs Evaluateur	A	A	M	Les résultats indiquent des informations concernant le contenu, la taille des pages web, les erreurs de pré-conditions ou les actions inutiles commises, des mesures quantitatives telles que le temps d'achèvement d'une tâche, temps pris pour visiter ou télécharger chaque page, etc.	Les résultats d'analyse sont représentés sous forme textuelle et/ou graphique.

***** A = Automatique ; SM = Semi-Automatique ; M= Manuel

Tableau A. 7 (suite) : Synthèse globale sur quelques outils d'évaluation de recueil des données d'interaction entre l'utilisateur et le système interactif à évaluer

Outil	Nature de l'outil	Nature de système à évaluer	Technique de capture et Nature des données capturées	Intervenants	Automatisation*			Résultats d'évaluation fournis	Présentation des résultats d'analyse
					Capture	Analyse	Critique		
MESIA (Trabelsi et al, 2006 ; Trabelsi et Ezzedine, 2012)	Logiciel (Mouchard électronique)	WIMP	Il capture un ensemble d'événements via l'injection de code dans la source. Ces évènements portent sur les actions des utilisateurs sur les dispositifs d'interaction (tels que, clavier, souris) ou les éléments interactifs graphiques à titre d'exemple citons, le nombre de clics de souris sur le bouton « OK ».	Utilisateurs Evaluateur Concepteur	A	A	M	Les résultats indiquent un ensemble des informations concernant des mesures telles que, le temps d'affichage de chaque interface du système évalué et le nombre des événements de bas niveau du clavier (citons par exemples, le nombre d'annulations, le nombre d'utilisations d'aide, etc.)	Les résultats d'analyse sont enregistrés dans des fichiers sous format Excel et seront utilisés pour la génération de RdP (Réseau de Pétri) concernant les modèles des tâches réellement effectuées par l'opérateur et les modèles des tâches à effectuer et finalement fournit des résultats sous forme textuelle.
Multi Device RemUSINE (Paterno et al., 2007)	Logiciel	Mobile	La capture des données est établie via l'outil « Mobile Logger » qui sont sauvegardées dans des fichiers XML. Il capture les trois types d'événements similaires à ceux de Web RemUSINE mais portant sur les applications mobiles. Ainsi, il permet de capturer quelques données contextuelles relatives aux conditions environnementales et du dispositif mobile, citons par exemple la puissance du signal réseau, la lumière de l'environnement, etc.	Utilisateurs Evaluateur	A	A	M	Les résultats d'analyse permettent d'indiquer des informations sur des problèmes plus évolués que celles cités par Web RemUSINE portant sur les conditions qui ont affecté les activités de l'utilisateur lors de leurs interactions avec les interfaces Mobiles.	Les résultats d'analyse sont représentés sous forme textuelle et/ou graphique.

***** A = Automatique ; SM = Semi-Automatique ; M= Manuel

Tableau A.7 (suite) : Synthèse globale sur quelques outils d'évaluation de recueil des données d'interaction entre l'utilisateur et le système interactif à évaluer

Outil	Nature de l'outil	Nature de système à évaluer	Technique de capture et Nature des données capturées	Intervenants	Automatisation*			Résultats d'évaluation fournis	Présentation des résultats d'analyse
					Capture	Analyse	Critique		
EISEval (Tran et al., 2013)	Logiciel (Mouchard électronique)	WIMP	La capture des données est établie via l'injection du code dans le code source. Ces données portent sur les événements de niveau bas et intermédiaire portant sur les interactions entre les utilisateurs et le système interactif à évaluer (sous forme des événements UI et des événements des dispositifs d'interaction) et entre les agents eux-mêmes (sous forme des services). Les données capturées sont par la suite sauvegardées dans la base de données.	Utilisateurs Evaluateur	A	A	M	Les résultats d'analyse présentent un ensemble des mesures calculables relatives aux services et aux tâches telles que le nombre total de tâches réalisées, le nombre des services exécutés, etc.	Les résultats d'analyse sont affichés dans l'outil. Ce dernier permet de générer les réseaux de Pétri associés aux tâches et les séquences d'interaction qui peuvent être enregistrés dans des fichiers PNML.

***** A = Automatique ; SM = Semi-Automatique ; M= Manuel

Une approche et un outil pour l'évaluation subjective et objective de l'utilisabilité des systèmes interactifs : Application à un Système d'Aide à l'Information Voyageurs

Résumé

L'évaluation de l'utilisabilité des systèmes interactifs est l'un des plus grands intérêts de la communauté des Interactions Homme-Machine. Pour réussir cette évaluation, plusieurs méthodes et outils d'évaluation subjective et objective ont été proposés et appliqués dans l'académie et l'industrie. Dans ce contexte, les évaluateurs d'utilisabilité travaillent habituellement avec beaucoup et différents types de données. Ceux-ci sont générés à partir de diverses méthodes et outils et liés à différents critères d'utilisabilité. En général, la compréhension et l'interprétation de ces données hétérogènes ne sont pas des tâches faciles. Dans cette thèse, nous proposons d'intégrer des données d'évaluation subjective et objective afin de mieux aider les évaluateurs à la prise de décision sur la qualité des systèmes interactifs dans le domaine des systèmes de supervision du trafic. En effet, nous avons appliqué le modèle de mesure d'information défini par la norme ISO/IEC 15939 avec l'utilisation des indicateurs dans un environnement d'évaluation des systèmes interactifs. Le développement de cet environnement a débuté depuis une décennie. Ainsi, pour compléter sa mise en œuvre et le mettre en pratique, nous avons appliqué plusieurs indicateurs d'utilisabilité proposés dans l'évaluation d'un système de supervision du trafic. Ces indicateurs intègrent les résultats d'évaluation subjective et objective des questionnaires d'utilisabilité, un inspecteur des règles ergonomiques et un mouchard électronique. Ils fournissent une base pour aider à l'interprétation des résultats obtenus. De plus, ils permettent la détection des problèmes d'utilisabilité et donnent des recommandations pour améliorer les interfaces utilisateur. Enfin, cette thèse propose des perspectives de recherche sur la base des limites rencontrées.

Mots-clés. Interaction Homme-Machine, Interface Utilisateur, évaluation, utilisabilité, mesures, critères, intégration, subjective, objective.

An approach and a tool for subjective and objective evaluation of the usability of interactive systems: Application to passengers Information Assistance System

Abstract

Usability evaluation for interactive systems is one of the greatest interests for Human-Computer Interaction community. To the success of this evaluation, several subjective and objective evaluation methods and tools have been proposed and applied in academy and industry. In this context, usability evaluators usually work with a lot of and different kind of data. These data are generated from various methods and tools and related to different usability criteria. In general, understanding and interpreting these heterogeneous data are not easy tasks. In this thesis, we propose to integrate subjective and objective evaluation data to better support the quality evaluators' decision-making for traffic supervision systems domain. In fact, we applied the measurement information model defined by ISO/IEC 15939 with the use of indicators in an interactive system evaluation environment. The development of this environment has started in a decade. Thus, for completing implementation and putting it into practice, we have applied several proposed usability indicators in the evaluation of a traffic supervision system. These indicators integrate subjective and objective evaluation results from usability questionnaires, an ergonomic guidelines inspector and an electronic informer. They provide a basis for aiding the interpretation of their results obtained. Furthermore, they allow detecting usability problems and give recommendations for enhancing the user interfaces. Finally, this thesis propose research perspectives based on the encountered limits.

Keywords. Human-Computer Interaction, User interface, evaluation, usability, measures, criteria, integration, subjective, objective.