

Livrable 2.5 V1

Une typologie de méthodes d'évaluation des interfaces appliquée aux tableaux de bord des EIAH

Auteur : Matthieu Cisel (Laboratoire STEF)

1. Introduction

L'objectif de ce document est de fournir au lecteur un guide pratique pour mener des évaluations des tableaux de bord conçus pour des EIAH. Les méthodes qui seront présentées sont pour la plupart issues de la psychologie ergonomique et des sciences du document, ces deux champs ayant contribué largement à l'avancement des travaux sur l'Interaction Homme-Machine. Les tableaux de bord constituent un type particulier d'interface, et à ce titre, les travaux de synthèse en évaluation ergonomique sont directement applicables à notre cas.

Nous avons fait le choix, pour rédiger ce document, de nous inspirer largement de la structure proposée dans l'ouvrage collectif dirigé par Chevalier et Tricot (2015), *l'ergonomie des documents électroniques*, ouvrage dont nous avons complété les propos par des articles consacrés à l'évaluation des EIAH (Nogry et al., 2004 ; Jamet et al., 2006). Nous avons, enfin, cherché à illustrer nos propos par quelques travaux consacrés à l'évaluation des tableaux de bord (Ali et al., 2012 ; Stephens-Martinez et al., 2014). Les méthodes que nous évoquerons peuvent aussi bien être qualitatives (entretiens individuels et de groupe, observations), que quantitatives (analyse de traces, questionnaires, etc.).

Le plan que nous suivrons est le suivant. Nous commencerons par définir les dimensions de l'évaluation que nous aborderons, pour reprendre le classique triptyque utilité, utilisabilité, acceptabilité. Après avoir défini ces termes, nous reviendrons brièvement sur la nécessité d'identifier les besoins des utilisateurs du tableau de bord, dans les premières phases du travail de conception. L'objet est ici de conceptualiser les besoins d'information de ces utilisateurs, autant pour la tâche actuelle, que pour la tâche future, c'est à dire la tâche qui sera instrumentée par l'EIAH. Cette première phase vise notamment à évaluer, en amont de tout développement, l'utilité des informations que le concepteur envisage de disposer sur le tableau de bord.

Viennent ensuite les modalités d'évaluation des interfaces, une fois celles-ci conçues. Nous avons choisi de reprendre la dichotomie entre évaluation analytique, qui implique notamment l'évaluation de l'interface sur la base d'un certain nombre de critères et qui ne sera qu'évoquée, et l'évaluation empirique, qui correspond aux tests utilisateurs. Ces tests utilisateurs peuvent se subdiviser entre méthodes d'évaluation on-line, qui sont mobilisées

pendant l'utilisation du tableau de bord, et méthodes d'évaluation offline, qui correspondent à une évaluation différée du tableau, notamment sur la base de questionnaires ou d'entretiens faisant suite au test.

2. Les différentes dimensions de l'évaluation

Avant de revenir sur le triptyque utilité/utilisabilité/acceptabilité, commençons par souligner que nous excluons de notre réflexion la question de l'accessibilité, qui porte sur la prise en compte des différents handicaps dans la conception de l'interface, et qui mériterait de plus amples développements. Nous nous pencherons avant tout sur les questions d'utilité et d'utilisabilité, que nous nous devons maintenant de définir.

Les critères relatifs à la catégorie "utilité" concernent la réponse aux besoins d'information de l'utilisateur du tableau de bord. L'utilité ne peut être définie dans l'absolu, mais par rapport à une situation et une tâche donnée, qu'il appartient à l'ergonome de caractériser dès les premières phases du travail de conception.

Vient ensuite la question de l'utilisabilité, qui relève des travaux sur l'interface homme-machine : de nombreuses typologies de critères ont été répertoriées. Pour reprendre les termes employés par Nogry et al. (2004), « D'après la définition ISO 9241-11, un logiciel est utilisable lorsque l'utilisateur peut réaliser sa tâche (efficacité), qu'il consomme un minimum de ressources pour le faire (efficience) et que le système est agréable à utiliser (satisfaction de l'utilisateur). ». À certains égards, l'utilisabilité concerne des qualités intrinsèques au tableau de bord, et peut être évaluée de manière décontextualisée, c'est-à-dire en dehors de la situation dans laquelle il a vocation à être insérée.

L'utilisabilité est évaluée selon un certain nombre de critères. Scapin et Bastien (1997) proposent, eux, huit critères ergonomiques pour l'évaluation des interfaces, dont celui de gestion des erreurs, dont il faut alors traduire la signification dans le contexte des tableaux de bord : « Le critère gestion des erreurs concerne tous les moyens permettant d'une part d'éviter ou de réduire les erreurs, et d'autre part de les corriger lorsqu'elles surviennent. [...] Trois sous-critères participent à la gestion des erreurs : protection contre les erreurs, qualité des messages d'erreurs et correction des erreurs. »

Nous concluons par les critères qui relèvent de l'adéquation « avec la situation de travail (en termes de temps, d'espace, d'organisation, etc.). Différents mots désignent cette catégorie, notamment acceptabilité et compatibilité, la notion d'acceptabilité intégrant les dimensions de motivation de l'utilisateur. » (Bastien et Tricot, 2015)

Les relations entre ces différentes dimensions ont fait l'objet d'un certain nombre de travaux, sur lesquels nous ne nous étendrons pas outre mesure, mais dont la diversité souligne la richesse des discussions relatives à leur articulation. Le modèle de Nielsen (1993) (Figure 1), qui établit une relation hiérarchique entre ces différentes dimensions, chacune conditionnant l'autre, est l'une des références les plus anciennes en la matière. Dans cette perspective, l'utilisabilité constitue un prérequis pour définir l'utilité. L'acceptabilité "pratique" est

conditionné non seulement à l'utilité pratique, mais aussi aux coûts de mis en place de la technologie, de fiabilité, etc.

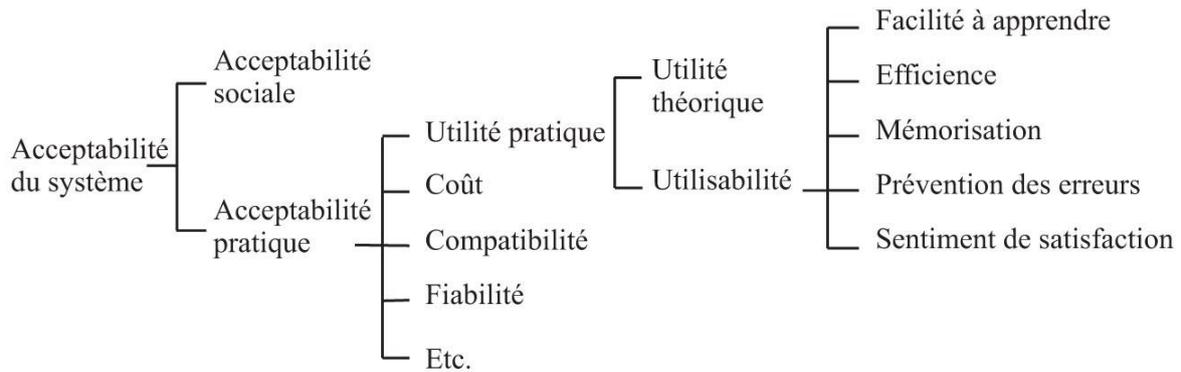


Figure 1. Le modèle de Nielsen relatif à l'articulation entre acceptabilité, utilité et utilisabilité (1993)

En présentant l'utilisabilité comme une condition nécessaire mais pas suffisante, une telle modélisation limite l'intérêt de cette dimension pour la compréhension des usages, ce qui a conduit des auteurs comme Dillon et Morris (1996) à présenter un modèle alternatif largement fondé sur le *Technology Acceptance Model* de Davis (1989), prenant en compte notamment la perception que l'utilisateur a de l'utilisabilité de l'artefact, qu'il faut distinguer de l'utilité "effective". Ce modèle hiérarchise moins les dimensions que sont l'utilité et l'utilisabilité, mais l'intention d'usage, rendue ici synonyme d'acceptabilité, dépend de ces deux paramètres.

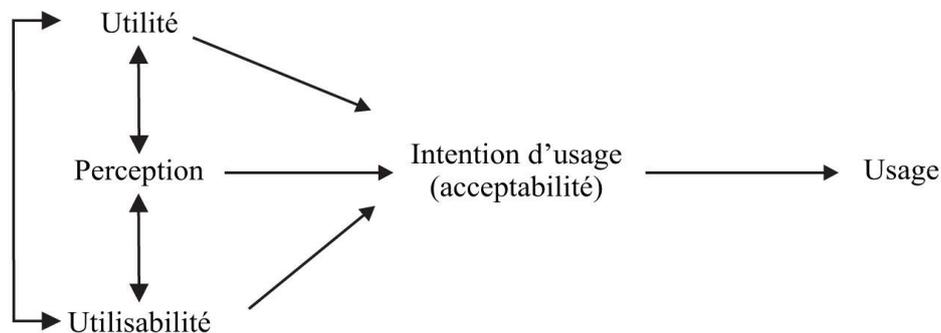


Figure 2. Le modèle simplifié de Dillon et Morris (1996)

Dans la littérature francophone, on pourra se référer à Tricot et al. (2003) pour des échos de ce débat. Les auteurs proposent un modèle offrant davantage de flexibilité dans les interrelations entre ces trois dimensions. Ils proposent enfin une typologie des méthodes d'évaluation des EIAH (Tableau 1), qui croise ces trois dimensions avec la dichotomie évaluation analytique (par inspection) et évaluation empirique, pour proposer six dimensions à l'évaluation.

Type d'évaluation	Empirique	Par inspection
Utilité	Adéquation entre objectif défini et apprentissage effectif Adéquation entre dispositif et format de la connaissance à acquérir Différence entre niveau de connaissances initial et terminal Mesures par des tâches de : - reconnaissance - rappel (contenu / structure) - résolution de problème - détection d'erreurs - production	Précision et présentation des objectifs Adéquation contenus /objectifs Précision du scénario didactique Adéquation scénario / objectifs / contenus Mise en œuvre des processus cognitifs et méta cognitifs Régulation Évaluation
Utilisabilité	Possibilité d'apprendre à utiliser le système Gestion et Prévention des erreurs Mémorisation du fonctionnement Efficience Sentiment de satisfaction Évaluation par : - observations - entretiens - analyse des parcours	Guidage et Incitation Groupement / Distinction des items par localisation ou format Feed-back immédiat et nature du feed-back Charge de travail Contrôle explicite Adaptabilité Gestion des erreurs Qualité des messages Homogénéité et cohérence Signifiante des codes et dénominations
Acceptabilité	Motivation Affects Culture Valeurs Évaluation par : - observations - entretiens - questionnaires	Acceptabilité en termes d'adéquation aux : - besoins ou objectifs de l'institution - attentes des apprenants - caractéristiques des apprenants Acceptabilité en termes de compatibilité avec : - l'organisation du temps - l'organisation des lieux Présence du matériel nécessaire Planification et suivis lisibles et cohérents Visibilité des résultats

Tableau 1. Une typologie des méthodes d'évaluation des EIAH par Tricot et al. (2013)

Cette classification fournit une synthèse intéressante des méthodes que nous présenterons dans la suite du document. L'un des principaux éléments qu'il convient de garder à l'esprit avant d'appliquer une telle grille sur l'évaluation des tableaux de bord réside notamment dans l'évaluation empirique de l'utilité. Dans le cas particulier des tableaux de bord, l'utilité n'est pas mesurée comme l'adéquation entre un objectif défini et un apprentissage effectif, sauf dans les cas particuliers où le tableau de bord est utilisé par l'apprenant dans une visée d'apprentissage, mais dans sa capacité à fournir des informations pertinentes à son utilisateur.

Son utilité n'est pas une qualité absolue, elle est relative aux besoins de l'utilisateur, qu'il convient d'identifier par l'analyse de son activité, question que nous allons aborder maintenant.

3. Analyse de l'activité à instrumenter dans les premières phases du travail de conception

Le travail d'identification des besoins lié au tableau de bord n'est qu'une dimension particulière de l'identification des besoins auxquels l'EIAH dans son ensemble est censé répondre. À ce titre, la littérature relevant de l'analyse de l'activité des enseignants pour l'identification de ces besoins peut fournir des pistes de travail intéressantes. Nous renverrons pour cela aux ouvrages sur l'ingénierie des EIAH (Tchounikine, 2009), ou à des références, plus anciennes, sur la conception des EIAO (Bruillard et Vivet, 1994). En substance, il est nécessaire de mieux comprendre la tâche dans laquelle le tableau de bord s'insère. Cela implique notamment de préciser les variables caractérisant les situations didactiques à instrumenter, et d'identifier les difficultés éventuelles des utilisateurs face à l'apprentissage à réaliser, difficultés dont le tableau de bord peut être censé rendre compte.

Il faut néanmoins garder à l'esprit, dans le domaine des EIAH en particulier, l'effet structurant de l'artefact sur la recomposition de l'activité (Rabardel, 1995). En d'autres termes, on ne peut se cantonner uniquement, pour penser le tableau de bord, à la caractérisation de l'activité non instrumentée. Il faut penser le modèle de la tâche future, pour la situation d'apprentissage instrumentée, c'est à dire de la tâche telle qu'elle sera instrumentée par l'EIAH, prendre en considération la diversité des scénarios d'utilisation de l'artefact, et donc des besoins d'information correspondant. La situation d'apprentissage « s'entend comme une situation incluant divers acteurs contraints : les apprenants et les activités qu'ils sont censés effectuer, le maître et le rôle qu'il doit jouer ainsi que le système informatique et la place assignée à ce dernier. » (Bruillard et al., 2000). Nous portons ici une vision inclusive de ces situations, dans la mesure où nous prenons en considération les cas où le tableau de bord n'est pas utilisé de manière synchrone avec l'apprentissage, comme ce peut être le cas dans une logique de réingénierie d'un dispositif.

La modélisation de la tâche et l'identification des besoins peut être complexe lorsque l'on conçoit des tableaux de bord pour des EIAH ouverts. Des environnements « sont qualifiés d'ouverts dans le sens où ils ne restreignent ni les activités pédagogiques à un catalogue type, ni les formateurs à un style d'enseignement immuable. » (Bruillard et al., 2000). Le faible effet structurant d'un EIAH ouvert sur l'activité a pour corollaire la nécessité, pour penser le tableau de bord, de prendre en compte l'hétérogénéité des activités pédagogiques potentielles correspondantes. La diversité des modalités de réalisation des tâches à instrumenter est fonction des caractéristiques de l'enseignant, ou plus généralement de l'utilisateur médiateur ou primaire de l'EIAH. Il peut être utile à cet égard de chercher à appréhender les déterminants de l'action de l'utilisateur médiateur de l'EIAH, généralement l'enseignant, notamment par des entretiens.

L'une des principales difficultés de la tâche est de fonder son analyse sur un échantillon suffisamment représentatif des utilisateurs de l'EIAH, tout en rationalisant le travail de recommandation qui fera suite à ces analyses, et qui conduira à une réduction de la complexité des situations. Ce travail de recommandation, qui amènera à proposer des indicateurs ou des visualisations qui seront proposés dans le tableau de bord, contient nécessairement une part d'arbitraire, qu'il convient de justifier tant sur le plan théorique qu'empirique par l'analyse de l'activité.

Nous ne pourrions entrer ici dans le détail des méthodologies afférentes, caractéristiques des phases préliminaires de la conception, observation de terrain, entretiens, questionnaires, ni dans celui des cadres théoriques, comme la théorie de l'activité, mobilisés pour cet aspect de la conception. Laissons donc de côté la phase de conception où, à partir du travail d'analyse de l'activité de l'utilisateur, sommaire ou détaillé, une première proposition de tableau de bord est réalisée, pour nous attarder davantage sur la phase où ce prototype va être évalué. Nous commencerons par l'évaluation analytique, qui peut porter aussi bien sur l'utilité que sur l'utilisabilité.

4. Évaluation analytique du tableau de bord

L'évaluation analytique ne sera évoquée que brièvement dans le cadre de ce document. Elle correspond à un ensemble de méthodes, dont les méthodes d'inspection de l'utilisabilité, qui nécessitent ou non le recours à un expert. Bastien et Tricot (2015) rapportent quatre types d'inspections ergonomiques : « l'analyse de la conformité à un ensemble de recommandations (*guideline reviews*), l'analyse de la conformité à des normes (*standards inspection*), l'utilisation d'heuristiques et l'utilisation de critères. »

Les critères ergonomiques de Scapin et Bastien (1997) constituent eux aussi une approche intégrée de l'évaluation ergonomique, mais en prenant en compte l'évaluation de certains aspects de l'utilité, plus particulièrement la « compatibilité » avec les tâches. Rappelons avec Nogry et al. (2004) que dans le domaine de l'évaluation analytique, « il est possible d'utiliser les critères ergonomiques pour évaluer l'utilisabilité des EIAH, mais ces critères doivent parfois être adaptés en fonction de l'objectif pédagogique du logiciel, des utilisateurs des EIAH et de la tâche. ». L'un des points d'attention, selon cet auteur, réside dans la nécessaire distinction entre « erreurs dans l'utilisation du logiciel et les erreurs au sens de réponses erronées (erreurs conceptuelles) ». Néanmoins, cette remarque ne vaut pour les tableaux de bord qu'à partir du moment où l'apprenant, élève ou étudiant, en représente l'un des utilisateurs.

Elle indique également sept principes pour la présentation des informations : « 1 / clarté : le contenu s'affiche rapidement et avec précision ; 2 / discriminabilité : les informations peuvent être distinguées avec précision ; 3 / concision : seules les informations nécessaires à la tâche sont affichées ; 4 / cohérence : les mêmes informations sont présentées de manière identique sur toute l'application ; 5 / détectabilité : les informations sont codées de façon adéquate au

bon endroit ; 6 / lisibilité : l'information est facile à lire ; 7 / compréhensibilité : la signification des termes est clairement compréhensible. »

Le lecteur qui voudra approfondir la question pourra trouver dans la littérature sur les critères ergonomiques un ensemble de normes sur lesquels adosser sa réflexion. Nous ne nous attarderons pas sur les méthodes d'évaluation analytique, pour nous concentrer sur l'évaluation empirique.

5. Évaluation empirique : les tests utilisateurs

Avant d'entrer dans le détail des méthodologies mises en œuvre pour réaliser un test utilisateur, introduisons plusieurs dimensions visant à les caractériser : la question de la finalité de ce test (Bastien et Tricot, 2015), le nombre d'utilisateurs, l'identification des composantes du tableau de bord à analyser, ou le lieu de déroulement du test.

La fonction d'un test dépend avant tout de sa place dans le cycle de conception. Un test **exploratoire** intervient tôt dans le processus de conception; il peut se fonder sur des maquettes ou des prototypes et vise avant tout à tester des concepts, la notion d'utilité. Un test d'**évaluation** intervient plus en aval, mais toujours tôt dans les phases de conception, il implique des versions plus détaillées de l'interface aient été produites. Le test de **validation** intervient peu avant la finalisation de l'EIAH, et donc du tableau de bord, tandis que le test de **comparaison** vise à produire des alternatives d'interface, dans notre cas, de tableau de bord, parmi lesquelles il s'agira de trancher.

Le nombre d'utilisateurs, ainsi que la présence du testeur auprès des utilisateurs embarqués pour le test (Dumas & Redish, 1993) constituent deux des variables clé qu'il faut déterminer. Rassembler un groupe de testeurs représentatif de la population cible peut se révéler particulièrement coûteux ou difficile, y compris pour des tests qui n'impliquent pas la présence de l'observateur. Des auteurs comme Nielsen et Laudauer (1993) suggèrent de réaliser les tests avec des petits groupes de personnes, une demi-douzaine, et ce de manière itérative au fur et à mesure du déroulement des différentes phases de conception. Ainsi, une quinzaine d'utilisateurs peuvent être utilisés par groupe de cinq. La solution intermédiaire peut consister à ajouter des utilisateurs au fil de l'eau, jusqu'à ce que l'ensemble des problèmes aient été résolus, ou, en d'autres termes, jusqu'à ce que la saturation ait été atteinte. On renverra à Lewis (2006) et Lewis & Sauro (2006) pour des écrits sur le nombre d'utilisateurs à rassembler.

Il faudra ensuite se demander si l'on souhaite évaluer le tableau de bord dans son ensemble ou si, au contraire, on préférera évaluer séparément les différentes composantes du tableau de bord (Brinkman, Haakma, & Bouwhuis, 2007). Dernière variable à prendre en compte : le lieu de déroulement du test. Faut-il sortir l'utilisateur final de son contexte, ou tester en situation écologique, en situation de travail, ce qui, dans le cas des EIAH, signifie bien souvent, en classe ? Envisager la seconde situation peut conduire à la mise en œuvre de méthodes ethnographiques, sur lesquelles nous reviendrons brièvement dans la partie qui suit, consacrées aux techniques « on-line », qui visent à étudier en temps réel le

comportement des utilisateurs qui mobiliseraient le tableau de bord, pour aborder dans un second temps le cas des méthodes off-line.

A. Techniques d'évaluation on-line

Dans cette partie consacrée aux techniques d'évaluation on-line, nous nous consacrerons avant tout sur l'analyse de parcours (Amadiou, Bastien et Tricot, 2015), qui vise à appréhender l'utilisabilité de l'interface, et passerons brièvement sur les méthodes plus qualitatives, qui peuvent être utilisées pour mesurer les trois dimensions qui nous intéressent, à savoir l'utilité, l'utilisabilité, et l'acceptabilité. Ce faisant, nous ne nous attarderons pas sur les méthodes oculométriques, qui permettent de suivre, grâce à une instrumentation spécifique, le trajet du regard sur l'interface. Nous renverrons à Strahm et al. (2015) pour une réflexion sur la question.

a) Analyse des parcours

Les techniques d'analyse des parcours visent à documenter et analyser la trajectoire de l'utilisateur au sein de l'interface qu'est le tableau de bord, ou, plus largement, au sein de l'application qui comprend le tableau de bord. Dans ce dernier cas, l'objectif peut consister à mieux comprendre comment le tableau de bord s'insère dans un usage plus général de l'application. L'analyse des parcours peut se fonder sur des enregistrements vidéos de l'usage de l'interface, réalisés ou non par des **logiciels de capture d'écran** (on parle alors parfois de *screenflow*), ou sur des **analyses de traces d'interaction**, si celles-ci sont disponibles. Cette dernière approche peut être intéressante si les traces reflètent de manière précise le comportement de l'utilisateur, car elle économise un temps conséquent si l'alternative consiste à coder manuellement des événements capturés sur une vidéo.

On pourra citer par exemple le cas d'une expérience menée par Faber et al. (2017), où l'analyse de traces d'interaction a permis d'évaluer l'utilisation d'un tableau de bord dans un contexte où plusieurs centaines d'utilisateurs testaient l'EIAH, dans près de 80 établissements. Un tel travail aurait été particulièrement laborieux s'il avait fallu analyser le flux vidéo de dizaines d'utilisateurs. Le succès de l'expérience tient en grande partie en fait qu'il ne s'agissait pas d'appréhender uniquement la navigation au sein du tableau de bord, mais bien l'utilisation du tableau de bord au sein de l'EIAH testé en conditions écologiques, i.e. sans observateur. Il faut néanmoins prendre garde au fait que les événements les plus signifiants d'une navigation au sein d'un tableau de bord peuvent ne pas générer de traces.

Au-delà de la méthode de collecte de données, une première fonction de l'analyse de parcours peut consister à mettre en place des **mesures d'efficacité** des parcours, comme les taux de rappel et de précision (Buckland & Gey, 1994), sans se soucier nécessairement de la séquence des actions réalisées par les utilisateurs. La seconde prend en considération cette séquence, et peut impliquer la mise en place d'une **grammaire de parcours**. Cette dernière consiste à mettre en place un codage des interactions avec l'interface, pour suivre des séquences d'actions.

Les mesures d'efficacité du parcours peuvent se fonder sur des ratios entre différents indices. Illustrons-en quelques-uns sur la base de la revue d'Amadiou et al. (2015), qui nous impose de définir le terme « Nœud ». Celui-ci est utilisé comme terme générique pour désigner une unité documentaire. Pour un tableau de bord, ce peut être une sous-composante du tableau, en particulier lorsque l'utilisateur peut, à partir d'une même interface, naviguer entre ses différentes composantes.

Sans souci d'exhaustivité quant à ces mesures d'efficacité, illustrons comment un indice peut être construit. Soit O , le nombre de cibles effectivement ouvertes par l'utilisateur. On pose R , comme le nombre de nœuds pertinents ou « cibles » (une cible est un nœud qui correspond au besoin de l'utilisateur, ou, dans le cadre d'une expérimentation, au but de la tâche, le « besoin de l'utilisateur » étant alors défini par l'expérimentateur) ; ou R' , le nombre de nœuds du parcours optimal (c'est-à-dire le nombre de nœuds nécessaires et suffisants à la réalisation de la tâche). » (Amadiou et al., 2015). Le **taux de rappel** est défini comme O/R par Buckland & Gey (1994). D'autres indices peuvent être employés pour définir le taux de précision et du taux de rappel. De même, il est fréquent que d'autres mesures soient utilisées pour compléter ces approches, comme les mesures de temps, de redondance. Leur détail n'entre pas dans le périmètre de ce document, et nous renvoyons à l'ouvrage collectif pour un recensement des méthodes de calcul de ces indicateurs.

Dans le cas où l'on s'intéresse à la séquence des actions réalisées sur le tableau de bord, deux approches peuvent être suivies. Dans le premier cas, on identifie ce que l'on considère comme étant le parcours idéal d'utilisation de l'interface, et on mesure une distance à ce parcours. Cette approche impose d'identifier des patterns de navigation chez l'utilisateur, par exemple entre les différentes sous-composantes d'un tableau de bord, puis de les comparer à des modèles de navigation. Cette approche nécessite néanmoins d'avoir défini précisément ce que représente le parcours le plus efficace, ce qui n'est pas toujours faisable. L'approche par **grammaire de parcours** permet de contourner l'obstacle, en « décrivant les parcours en fonction des questions que l'on se pose. ». Nous renvoyons à Tricot et Coste (1995) pour une plus ample réflexion méthodologique sur la question.

b) Recueil des verbalisations et méthodes ethnographiques d'observation

Abordons maintenant les méthodes qualitatives, qui, bien que faisant partie des méthodes on-line, ne sont pas développées outre mesure dans l'ouvrage qui structure notre synthèse, plus focalisé sur les questions d'utilisabilité. Ces méthodes peuvent servir aussi bien à étudier l'utilité, que l'utilisabilité ou l'acceptabilité de l'application.

Les remarques qui suivent sont issues avant tout des articles de de Nogry et al. (2004), et de Jamet et al. (2006). Les méthodes qualitatives viennent combler certains des défauts des méthodes quantitatives, à commencer par les difficultés à généraliser les résultats qu'elles produisent.

Les méthodes qualitatives développées en sciences humaines offrent des outils pour appréhender la globalité du phénomène étudié. Le **recueil des verbalisations** consiste à

“demander à l'utilisateur de penser à haute voix durant la tâche proposée afin d'identifier les raisonnements qu'il met en œuvre pour réaliser la tâche demandée.”. En d'autres termes, il s'agit d'identifier les raisons qui poussent un utilisateur utilise telle ou telle partie du tableau de bord, notamment pour chercher à identifier ce qu'il pense que chacune des composantes lui apporte comme information.

Comme le souligne Jamet (2006) dans sa réflexion sur les **protocoles verbaux**, « Ces verbalisations peuvent être concurrentes à la tâche mais elles ralentissent le traitement. Elles peuvent être différées après la tâche, par exemple en diffusant la vidéo du protocole au sujet et en lui demandant de la commenter. ». La question se pose en particulier si le tableau de bord a vocation à être utilisé en classe de manière synchrone avec les activités d'apprentissage, une verbalisation pouvant interférer de manière conséquente avec le déroulement de la séance. Cela classe alors ces protocoles verbaux dans la catégorie des méthodes offline.

Si le tableau de bord a vocation à être utilisé de manière synchrone avec l'activité des utilisateurs de l'EIAH, qu'il soit mobilisé par un enseignant et/ou des étudiants, il peut être pertinent de mettre en œuvre des **méthodes ethnographiques**, qui consistent à observer une situation « de l'intérieur ». Cette approche vise par exemple à appréhender une classe dans son ensemble, pour mieux identifier les problèmes que pose l'utilisation du tableau de bord en temps réel, et qui auraient pu à échapper à une session d'observation *ex situ*.

L'observation peut être **expérientielle** (Postic et de Ketele, 1988), ou **systématique** si une grille d'observation est mise en place pour coder, le plus souvent en temps réel, les événements qui sont recensés. On distingue les grilles selon qu'elles sont basées sur un **système de catégories** ou un **système de signes**. Dans le premier cas, l'ensemble des événements observés doivent être rangés dans une catégorie. Dans le second cas, on se focalise uniquement sur des événements signifiants, tels que des incidents, qui sont eux-mêmes classifiés.

En cas d'impossibilité d'enregistrement vidéo ou d'analyse de traces d'interaction¹, une grille d'observation systématique peut se révéler utile pour suivre l'activité de l'utilisateur, même si elle impose de se cantonner à un nombre restreint d'utilisateurs. Au-delà du fait que l'on économise le temps de codage a posteriori de l'enregistrement vidéo, elle présente également quelques avantages au regard d'une capture du flux d'image, dont notamment celui de pouvoir coder les éventuelles interactions entre les utilisateurs en plus des actions réalisées sur l'artefact. Cet avantage peut être appréciable lorsque l'EIAH est testé en conditions écologiques, et que le tableau de bord fait partie des instruments mobilisés pendant l'activité pédagogique. Un enregistrement vidéo peut échouer à capturer cette diversité des interactions. Ces remarques concluent notre présentation des méthodes online. Nous allons

¹ Il est des situations, en particulier en contexte scolaire, dans lesquelles l'enregistrement vidéo peut être complexe pour des raisons légales.

maintenant nous concentrer sur les méthodes offline, que nous avons déjà évoquées plus tôt dans ce document.

B. Techniques d'évaluation Offline

Nous commencerons par traiter des modalités de collecte de données : entretiens ou questionnaires, pour conclure sur leurs finalités. Nous évoquerons ce faisant les méthodes de transfert et de rappel d'une part, et les méthodes d'évaluation de la charge mentale et des états émotionnels d'autre part (Jamet, Lemerrier et Février, 2015).

a) Modalités de collecte de données

Les entretiens peuvent faire suite à des observations, ou tout du moins à des utilisations en situation écologique du tableau de bord. On distingue les entretiens selon qu'ils sont ouverts, structurés ou semi-structurés, selon le degré de directivité de l'enquêteur dans la conduite de l'entretien. Dans le cas de l'entretien structuré, les questions sont posées dans le même ordre, et formulées de la même manière, pour l'ensemble des participants, contrairement aux autres cas. Comme le soulignent Nogry et al. (2004) « Dans le cadre d'une évaluation en classe, il peut être utile de faire en complément des entretiens avec les enseignants pour confronter les différents points de vue sur le déroulement de l'évaluation. ». Si l'instrument servant de base à l'entretien structuré peut offrir de nombreuses convergences avec un questionnaire, il présente l'intérêt de collecter des réponses ouvertes bien plus riches et développées que ne le permettrait de le faire le même questionnaire.

L'entretien peut être complété par des questionnaires, dont l'objet est de mesure est de mesurer, via des échelles normées, l'utilité et l'utilisabilité perçue de différentes composantes du tableau de bord. Le questionnaire peut être administré en présence de l'enquêteur comme à distance. Dans ce dernier cas, il nous faut évoquer le cas où des maquettes d'interface sont présentées à des utilisateurs potentiels, en particulier dans le cas des tests exploratoires. Ces utilisateurs doivent alors se prononcer sur ces interfaces via des mesures subjectives. Nous allons dans les paragraphes qui suivent illustrer les finalités de ces entretiens et questionnaires, en commençant par les méthodes de transfert, de rappel et de reconnaissance.

b) Méthodes de transfert, de rappel et de reconnaissance

Les tests de transfert, de rappel et de reconnaissance peuvent être réalisés immédiatement après, ou plusieurs jours après l'utilisation du tableau de bord. Les méthodes de rappel visent à chercher à évaluer si le participant est capable de se souvenir de la structure du tableau de bord une fois qu'il ne l'a plus sous les yeux; l'objectif n'est pas ici de tester les capacités de mémorisation de l'observateur, mais d'évaluer l'utilité et l'utilisabilité du tableau de bord, dont on postule qu'elles impactent le rappel. Ils peuvent être accompagnés de tests de reconnaissance, où l'enquêteur présente des composantes du tableau de bord, qui comportent notamment des pièges, pour mesurer la capacité de l'utilisateur à discriminer entre les deux. Les techniques de transfert visent quant à elles à mesurer la compréhension que l'utilisateur a du contenu du tableau de bord, en lui demandant de mobiliser l'information qu'il contient

pour résoudre des problèmes particuliers, liés à l'utilité de l'EIAH. « Cette évaluation peut passer par l'utilisation de questions d'inférence, qui sont caractérisées par l'absence de réponse explicite dans le document ». Jamet et al. (2015) estiment que dans le cadre d'un EIAH, les mesures de mémorisation et de compréhension constituent des mesures de performance relatives avant tout à l'utilité de l'interface. Dans la mesure où le tableau de bord n'a pas nécessairement vocation à servir de base à un apprentissage, il faut nuancer ce propos dans le cas qui nous intéresse ici. Les dimensions qui suivent sont quant à elles mesurées par des questionnaires subjectifs et sont relatives avant tout à l'utilisabilité du tableau de bord.

c) Méthodes d'évaluation de la charge mentale et des états émotionnels

Les méthodes d'évaluation de la charge mentale visent notamment à évaluer d'une part la difficulté de la tâche impliquant l'utilisation du tableau de bord, et d'autre part la complexité de celle-ci. « La difficulté de la tâche renvoie aux caractéristiques internes du sujet, c'est-à-dire à son âge, son expertise dans le domaine d'étude, sa motivation et les stratégies qu'il met en œuvre pour mener à bien son apprentissage. La complexité de la tâche renvoie cette fois aux caractéristiques de la tâche, intégrant la demande, le contenu de l'apprentissage, l'organisation de l'information sur l'interface. ». Il existe des échelles de mesures subjectives de la charge mentale où le sujet doit indiquer sur un axe vertical « à quel point des efforts ont été requis pour réaliser la tâche qu'on lui a demandée. » (Jamet et al., 2015).

Méthodes d'évaluation des états émotionnels

Les méthodes d'auto-estimation de l'état émotionnel se fondent généralement sur des « questionnaires qui proposent une liste d'adjectifs connotés émotionnellement, soit positivement soit négativement. » (Jamet et al., 2015). On notera que de nombreux questionnaires subjectifs d'utilisabilité ont été utilisés au cours des dernières décennies, et notamment de la dimension liée à la satisfaction qu'elle inclut. On reprendra ici l'exemple fourni par Jamet et al. (2015), le SUMI (*Software Usability Measurement Inventory*). Il est constitué de 50 items mesurant différentes dimensions de l'utilisabilité : le sentiment d'être aidé par le logiciel, les affects, le sentiment de contrôle, l'apprenabilité du système et l'utilité des aides fournies. À nouveau, nous renverrons à Jamet et al. (2015) pour une discussion plus approfondie de la question. Avant de conclure cette synthèse, nous proposerons quelques illustrations issues de la littérature sur l'évaluation des tableaux de bord.

6. Quelques illustrations issues de la littérature

Le premier exemple que nous avons choisi est fondé sur un test exploratoire (Stephens-Martinez et al., 2014), où il s'agit pour l'enquêteur d'estimer, via des maquettes, l'utilité de différentes composantes d'un tableau de bord. Le questionnaire inclut des informations sur l'utilisateur (expérience dans l'enseignement, âge, etc.), afin d'interpréter d'éventuelles variations dans les réponses fournies. Les enquêteurs présentent ensuite à l'utilisateur une maquette de l'interface (Figure 3)

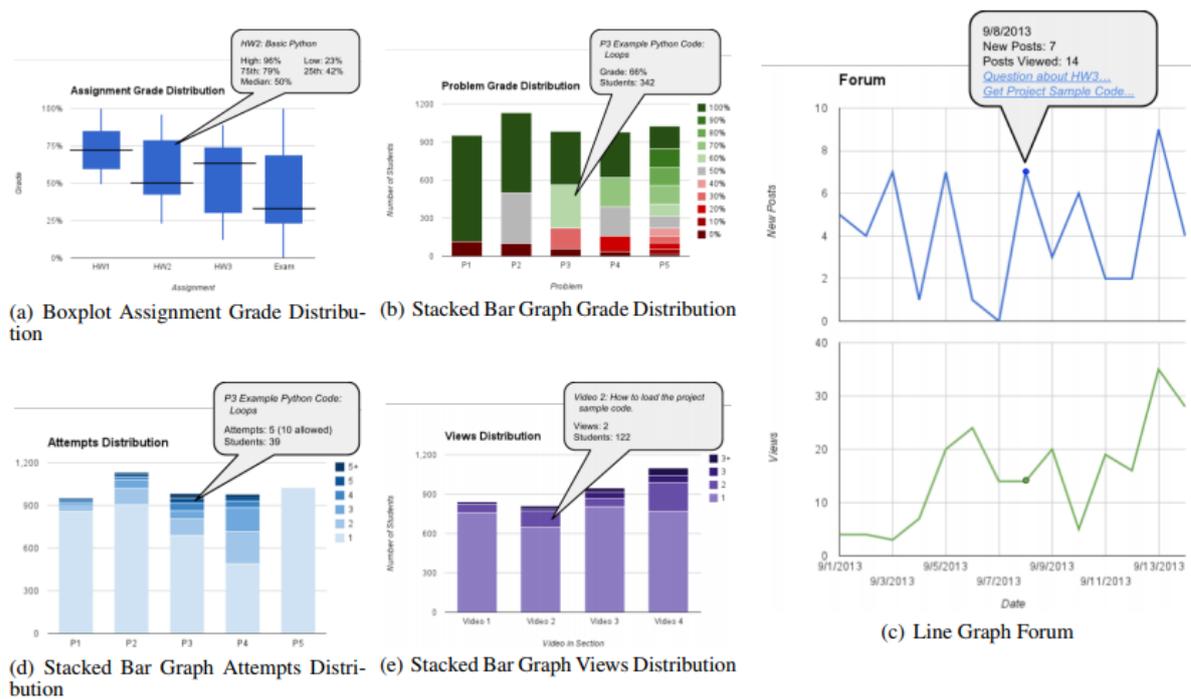


Figure 2. Mockups shown to survey participants. The call out bubbles represents what will be seen if the user hovers the mouse over a part of the graph.

Figure 3 Une maquette de tableau de bord présentée aux enquêtés

Les auteurs ont utilisé pour évaluer l'intention d'usage des différentes composantes de ce tableau de bord une échelle (Figure 4), où les enquêtés doivent estimer s'ils l'utilisent déjà la sous-composante du tableau de bord, s'ils utiliseraient si d'aventure celle-ci était disponible, ou s'ils ne l'utiliseraient pas.

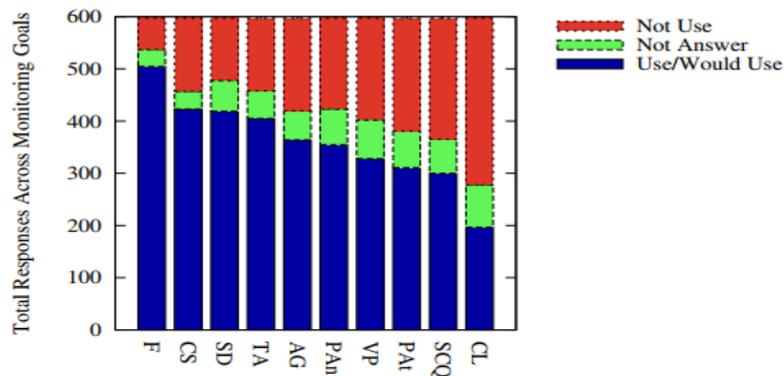


Figure 4. For each information source, the number of participants who use it or would use if available (combined into a single category), do not use, or did not answer. Did not answer means that the participant chose an answer for a subset of the information sources for that Monitoring Goal. It is shown here to see the relative rate of responses.

Figure 4. Analyse des réponses sur l'intention d'usage de différentes composantes d'un tableau de bord.

Une autre manière d’aborder la question de l’utilité a consisté pour ces auteurs à utiliser une échelle de likert classique, où les participants doivent déterminer dans quelle mesure ils sont d’accord avec le fait que la visualisation proposée est utile (Figure 5 Gauche), puis de déterminer à quelle étape de la préparation de leur cours ils pensaient qu’une telle visualisation leur serait la plus utile (Figure 5 Droite).

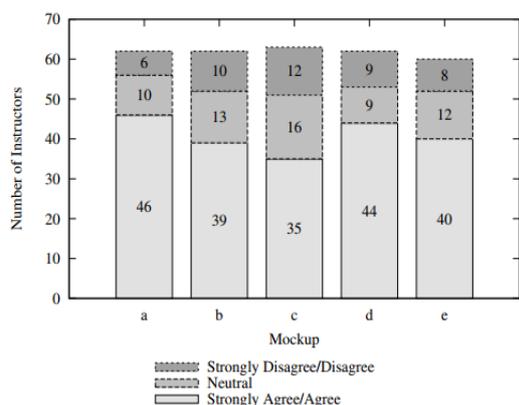


Figure 6. Likert scale responses to the statement “This visualization is useful.” Letters correspond to Figure 2’s subfigures.

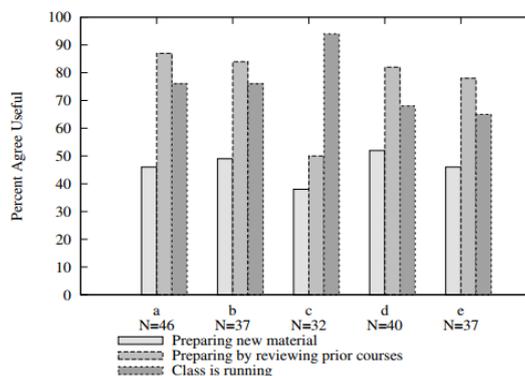


Figure 8. Responses to the question “When would you use this mockup,” with choices in terms of three phases in a “MOOC cycle”: preparing new material, preparing by reviewing previous course runs, and while the class is running. Letters correspond to Figure 2’s subfigures.

Figure 5. Deux mesures de l’utilité de différentes composantes du tableau de bord

Dans cette autre étude menée sur le tableau de bord LOCO-Analyst (Ali et al., 2012), les auteurs présentent un ensemble de questions offrant aux participants des réponses ouvertes (Tableau 2), qui sont ensuite codées à la main par les enquêteurs.

Category	Question Description in the questionnaire	Asked in 2006	Asked in 2009
Perceived usefulness of the tool for improving the course content/instruction	Q1†: The tool enables me to get an insight into the students’ interactions with the learning content	+	+
	Q2†: The information the tool provides helps me identify what needs to be improved in the learning content	+	+
	Q3†: The tool provides relevant information regarding the students’ interactions within the online learning environment.	+	+
	Q4†: The information provided by the tool helps me determine how to improve the students’ online interactions.	+	+
	Q5: The tool helps me identify the domain topics the students were having difficulties with.	-	+
Perceived value of the tool’s GUI	Q6: LOCO-Analyst’s GUI (Graphical User Interface) is intuitive enough	+	+
	Q7†: LOCO-Analyst’s GUI is overburdened with information	+	+
	Q8: My general opinion about the GUI	+	+
General perception of the tool	Q9: I would like to be able to use LOCO-Analyst in my teaching practice	-	+
	Q10: LOCO-Analyst provides me with more useful feedback than other similar tool(s) I have used/tried	+	+
	Q11: LOCO-Analyst is more intuitive than the other tools capable of for feedback provisioning I have used/tried	-	+

Tableau 2. Un questionnaire visant à évaluer deux versions successives du tableau de bord LOCO-Analyst

De nombreuses autres publications de cet acabit sont disponibles dans la littérature scientifique, mais leur recensement dépasse les objectifs de ce document.

Conclusion

Dans ce document, nous avons cherché à recenser une certaine diversité de techniques d'évaluation des interfaces, issues majoritairement de la psychologie ergonomique et des sciences du document. Nous les avons complétées avec quelques éléments issus des sciences humaines. Il va de soi que si certaines de ces techniques sont complémentaires, une sélection est nécessaire. Ce choix est à la discrétion du chercheur ou du praticien, qui sélectionnera les méthodes qui lui paraissent le plus appropriées au regard de ses objectifs, son expertise, et des moyens dont il dispose.

Références

- Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, 58(1), 470–489. <https://doi.org/10.1016/j.compedu.2011.08.030>
- Amadiou, F., Bastien, C., & Tricot, A. (2015). *Les méthodes on-line 1 : analyse des parcours*. Presses Universitaires de France. Retrieved from <http://www.cairn.info/revue-le-travail-humain-puf-2008--p-251.htm>
- Bastien, J. M. C., & Tricot, A. (2015). *L'évaluation ergonomique des documents électroniques*. Presses Universitaires de France. Retrieved from <http://www.cairn.info/revue-le-travail-humain-puf-2008--p-205.htm>
- Brinkman, W. P., Haakma, R., & Bouwhuis, D. G. (2007). Towards an empirical method of efficiency testing of system parts : A methodological study. *Interacting with Computers*, 19, 342-356.
- Bruillard, É., Delozanne, E., Leroux, P., Delannoy, P., Dubourg, X., Jacoboni, P., ... Teutsch, P. (2000). Quinze ans de recherche informatique sur les sciences et techniques éducatives au LIUM. *Sciences et Techniques éducatives*, 7(1), 87–145. Retrieved from <http://cat.inist.fr/?aModele=afficheN&cpsid=1155601>
- Bruillard, E., & Vivet, M. (1994). Concevoir des EIAO pour des situations scolaires. Approche méthodologique. *Recherches En Didactique Des Mathématiques*, 14(1-2), 275–304. Retrieved from <http://rdm.penseesauvage.com/Concevoir-des-EIAO-pour-des.html>
- Buckland, M., & Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science*, 45, 12-19.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319- 340.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 982–1003. <https://doi.org/10.1287/mnsc.35.8.982>

Dillon, A., & Morris, M. (1996). User acceptance of new information technology – theories and models. In M. Williams (Ed.), *Review of Information Science and Technology* (pp. 3-32). Medford : Information Today.

Dumas, J. S., & Redish, J. C. (1993). *A Practical Guide to Usability Testing*. Norwood, NJ : Ablex.

Faber, Janke M., Hans Luyten, and Adrie J. Visscher (2017). “The Effects of a Digital Formative Assessment Tool on Mathematics Achievement and Student Motivation: Results of a Randomized Experiment.” *Computers & Education* 106: 83–96. <http://www.sciencedirect.com/science/article/pii/S0360131516302433>.

HOC, J.M., AND LEPLAT, J. (1983) Evaluation of different modalities of verbalization in a sorting task. *International Journal of Man-Machine Studies* 18, p. 283-306.

Jamet, E. (2006). Une présentation des principales méthodes d'évaluation des EIAH en psychologie cognitive. *Sciences et Technologies de l'Information et de La Communication Pour l'Éducation et La Formation (STICEF)*, 13, 13 pages. Retrieved from <https://hal.archives-ouvertes.fr/hal-00696350>

Jamet, É., Lemercier, C., & Février, F. (2015). *Évaluer l'utilisation d'un document a posteriori: les méthodes off-line*. Presses Universitaires de France. Retrieved from <http://www.cairn.info/revue-le-travail-humain-puf-2008--p-231.htm>

Lewis, J. R. (2006). Sample sizes for usability tests : Mostly math, not magic. *Interactions*, XIII (6), 29-33.

Lewis, J. R., & Sauro, J. (2006). When 100 % really isn't 100 % : Improving the accuracy of small-sample estimates of completion rates. *Journal of Usability Studies*, 1 (3), 136-150.

Morris, M. G., & Dillon, A. (1997). How user perceptions influence software use. *IEEE Software*, 14(4), 58–65. <https://doi.org/10.1109/52.595956>

Nielsen, J. (1993). *Usability Engineering*. Boston : Academic Press.

Nielsen, J., & Landauer, T. K. (1993, 24-29 April). A Mathematical Model of the Finding of Usability Problems. Paper presented at the INTERCHI'93 Conference on Human Factors in Computing Systems, New York, NY

Nogry, S., Jean-Daubias, S., & Ollagnier-Beldame, M. (2004). Évaluation des EIAH : une nécessaire diversité des méthodes (pp. 265–271). Presented at the Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et l'Industrie, Université de Technologie de Compiègne. Retrieved from <https://edutice.archives-ouvertes.fr/edutice-00000729/document>

Postic, De Ketele (1988). *Observer les situations éducatives*. Paris, PUF

Rabardel, P. (1995). *Les hommes et les technologies; approche cognitive des instruments contemporains*. Armand Colin. Retrieved from <https://hal.archives-ouvertes.fr/hal-0101746>

Scapin, D. L., & Bastien, J.-M.-C. (1997). Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology*, 16, 220-231.

Stephens-Martinez, K., Hearst, M. A., & Fox, A. (2014). Monitoring MOOCs: Which Information Sources Do Instructors Value? In *Proceedings of the First ACM Conference on Learning @ Scale Conference* (pp. 79–88). New York, NY, USA: ACM. <https://doi.org/10.1145/2556325.2566246>

Strahm, M., Kicka, M., & Baccino, T. (2015). *Les méthodes on line 2: mouvements oculaires*. Presses Universitaires de France. Retrieved from <http://www.cairn.info/revue-le-travail-humain-puf-2008--p-271.htm>

Tchounikine, P. (2009, June). *Précis de recherche en Ingénierie des EIAH*. Retrieved from <https://hal.archives-ouvertes.fr/hal-00413694>

Tricot, A., Plégat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G., & Morcillo, A. (2003). Utilité, utilisabilité, acceptabilité: interpréter les relations entre trois dimensions de l'évaluation des EIAH (pp. 391–402). Presented at the Environnements Informatiques pour l'Apprentissage Humain 2003, ATIEF; INRP. Retrieved from <https://edutice.archives-ouvertes.fr/edutice-00000154/document>

Tricot, A., & Coste, J.-P. (1995). Evaluating complex learner-computer interaction : What criteria for what task ? EARLI'95 Conference. Nijmegen, August 26-3