

Évaluation des EIAH : une nécessaire diversité des méthodes

Sandra Nogry, Stéphanie Jean-Daubias

LIRIS
Université Claude Bernard Lyon 1 - CNRS
Nautibus - 8 bd Niels Bohr - Campus de la Doua
69622 Villeurbanne Cedex
France
Tel : 04 72 44 82 75 - Fax : 04 72 43 15 36
{Sandra.Nogry, Stephanie.Jean-Daubias}@liris.cnrs.fr

Communication "Recherche"

Résumé

Il est maintenant bien établi que l'évaluation des Environnements Informatiques pour l'Apprentissage Humain est essentielle. Mais quelle méthode d'évaluation choisir ? Que doit-on évaluer ? L'objectif de cet article est de présenter différentes méthodes d'évaluation adaptées aux EIAH correspondant à deux types d'évaluation : évaluation de l'utilisabilité et évaluation de l'utilité. L'évaluation de l'utilisabilité d'un EIAH peut être faite en adaptant les techniques classiques développées en IHM. À l'inverse, l'évaluation de l'utilité d'un EIAH, qui consiste à évaluer son impact sur l'apprentissage, nécessite l'utilisation de méthodes spécifiques. Dans cet article, nous présentons ces différentes méthodes puis nous abordons la question du choix de la méthode en fonction de l'EIAH à évaluer. Nous mettons ensuite en évidence à travers deux exemples l'intérêt de combiner plusieurs méthodes afin d'évaluer à la fois l'impact de l'EIAH sur apprentissage et la complexité de la situation d'apprentissage.

Mots clés : EIAH, évaluation, méthodologies, utilisabilité, apprentissage.

Abstract

It is now well established that the evaluation of Interactive Learning Environments is crucial. But how to evaluate an ILE? This paper presents various methods for ILE evaluation distinguishing between two kinds of evaluation: evaluation of usability versus evaluation of utility. Evaluation of the usability of an ILE can be realised by adapting classical methods developed by HCI researchers. Conversely, evaluation of the utility of an ILE that consists in evaluating the impact of the ILE on learning requires the use of specific methods. We present here these methods and then, we discuss about the choice of the method depending on the ILE we want to evaluate. Finally, we present two examples to illustrate the relevance of combining several methods in order to evaluate both the impact of the ILE on learning and the complexity of the learning situation.

Keywords: ILE, assessment, methodology, usability, learning.

Introduction

En EIAH (Environnements Informatiques pour l'Apprentissage Humain), l'évaluation des systèmes tient une place de plus en plus importante et fait désormais souvent partie intégrante de leur cycle de conception. Cette évaluation, nécessairement pluridisciplinaire (sciences de l'éducation, didactique, psychologie cognitive, informatique, ergonomie), comporte différentes facettes : nous nous intéressons ici uniquement à l'évaluation de l'EIAH au cours de sa conception, c'est-à-dire après sa validation interne [15] et avant l'évaluation de ses usages.

Pour ce, comme Senach, nous distinguons l'utilisabilité d'un système informatique (sa capacité à permettre à l'utilisateur d'atteindre facilement ses objectifs) de son utilité (son adéquation aux objectifs de haut niveau du client) [22]. En EIAH, l'utilisateur est généralement l'apprenant, le client est l'enseignant ou le système éducatif et les objectifs de haut niveau concernent l'apprentissage. Les EIAH dont nous traitons dans cet article s'adressent aux apprenants et visent à favoriser l'apprentissage (notre propos ne concerne directement ni les EIAH destinés à l'évaluation des compétences des apprenants, ni les systèmes s'adressant aux enseignants). Si l'évaluation de l'utilisabilité des EIAH fait appel, en les adaptant, à des techniques classiques d'évaluation de systèmes, l'évaluation de l'utilité concerne l'appréciation de l'impact de l'enseignement dispensé sur les connaissances ou compétences des apprenants, elle est plus spécifique et doit faire appel à des méthodes issues de différentes disciplines.

Dans cet article nous présentons tout d'abord en quoi consiste l'évaluation de l'utilisabilité d'un système et nous montrons comment les techniques existantes doivent être adaptées pour évaluer un EIAH. Nous présentons ensuite différentes méthodes existantes pouvant permettre d'évaluer l'apprentissage et nous mettons ces méthodes en regard des objectifs du système. Enfin, nous mettons en évidence à travers deux exemples l'intérêt de combiner les méthodes.

Évaluer l'utilisabilité

En premier lieu, il est important d'évaluer l'utilisabilité d'un système. L'utilisabilité concerne l'adéquation entre la manière dont une tâche est réalisée par un utilisateur et les capacités cognitives de cet utilisateur [7]. 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).

Méthodes d'évaluation de l'utilisabilité

Il existe de nombreuses méthodes permettant de mesurer l'utilisabilité. Pour présenter ces méthodes, nous reprenons la distinction proposée par Senach [22] entre évaluation analytique et évaluation empirique, avant de présenter les spécificités liées à l'évaluation de l'utilisabilité d'un EIAH.

Évaluation analytique

L'évaluation analytique consiste à étudier les interfaces selon un ensemble de référents afin de contrôler qu'elles possèdent bien certaines qualités et de détecter les problèmes qu'elles peuvent poser. Il existe plusieurs manières de conduire une évaluation analytique. Cette évaluation peut être faite par des experts s'appuyant sur des listes de critères conçues par des ergonomes [2] [18] [21]. Ainsi Nielsen [18] et Schneiderman [21] distinguent cinq attributs de l'utilisabilité : l'apprentissage (de l'utilisation du système), l'efficacité (la facilité à réaliser la tâche souhaitée), la mémorisation (la capacité à reprendre en main rapidement le système lors d'utilisations espacées), les erreurs (liées à l'utilisation du logiciel) et la satisfaction subjective de l'utilisateur. Bastien et Scapin [2] proposent, eux, huit critères ergonomiques pour l'évaluation des interfaces (qui parfois recourent les attributs présentés précédemment). Citons pour exemple le critère de gestion des erreurs : « 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. Les erreurs sont ici considérées comme des saisies de données incorrectes, des saisies dans des formats inadéquats, des saisies de commandes avec syntaxe incorrecte, etc. Trois sous-critères participent à la gestion des erreurs : protection contre les erreurs, qualité des messages d'erreurs et correction des erreurs. » [2].

Pour faciliter l'évaluation analytique, il existe également des check-lists qui permettent de vérifier point par point que le logiciel est conforme à un certain nombre de critères.

Une autre méthode d'évaluation analytique consiste à spécifier des tâches et des séquences d'actions dans le logiciel et à imaginer ce que ferait l'utilisateur dans ces situations. Ces inspections cognitives (cognitive walkthroughs) [14] permettent ainsi de s'assurer que le système peut réaliser les actions souhaitées et identifier quels problèmes peuvent se poser.

Évaluation empirique

L'évaluation empirique consiste à recueillir des données relatives au comportement de l'utilisateur lors de l'utilisation du système. Ce type d'évaluation nécessite l'existence d'un système réel (maquette, prototype ou système final) et la présence d'utilisateurs. L'utilisation

qui est faite du système par ces utilisateurs est observée et analysée. Une observation individuelle détaillée de l'interaction entre l'utilisateur et le système permet d'identifier les capacités de l'utilisateur, de détecter les difficultés éventuelles, ou encore de noter les caractéristiques inattendues de la situation [9]. Nielsen indique qu'un panel de cinq utilisateurs « représentatifs » permet d'identifier 80% des problèmes [17]. Ces techniques d'observation peuvent être complétées par des techniques d'entretien ou par un questionnaire afin d'avoir des indications sur la compréhension ou la satisfaction de l'utilisateur.

Utilisabilité et EIAH

Ces différentes méthodes doivent être adaptées aux EIAH. En effet, par rapport aux systèmes informatiques « classiques », ils sont spécifiques par différents points. D'une part, leur objectif de haut niveau : favoriser l'apprentissage chez leur utilisateur, est particulier. Cette spécificité a des conséquences sur la définition de l'utilisabilité des EIAH. Par exemple, dans un EIAH ayant pour objectif à court terme la résolution de problèmes, on peut privilégier un parcours tortueux du point de vue de la résolution de problèmes, plus bénéfique pour l'apprentissage, même s'il semble moins efficace qu'un parcours linéaire. D'autre part, la situation d'utilisation des EIAH est elle aussi particulière. Ces systèmes peuvent être utilisés seuls ou en binôme, à la maison, en salle informatique avec l'enseignant ou en salle de classe en autonomie pendant que le reste des élèves effectuent d'autres activités. L'utilisation du système peut être extrêmement ponctuelle (utilisation unique) ou régulière pendant une période de l'année (par exemple une utilisation par semaine pendant trois mois).

Par ailleurs, les utilisateurs d'un EIAH sont généralement de deux types : enseignants et apprenants. En effet, si les apprenants sont clairement les utilisateurs finaux principaux de ces systèmes, les enseignants en sont prescripteurs, au sens où ce sont généralement eux qui provoquent l'utilisation d'un EIAH donné par les apprenants, mais également utilisateurs secondaires [5] [13] du fait qu'ils peuvent parfois préparer le travail de leurs apprenants en paramétrant le système.

Du fait de ces différentes spécificités, les méthodes d'évaluation de l'utilisabilité ne peuvent pas être appliquées telles quelles aux EIAH et doivent être adaptées.

En ce qui concerne l'évaluation analytique, il est possible d'utiliser les critères ergonomiques pour évaluer l'utilisabilité des EIAH [25], mais ces critères doivent parfois être adaptés en fonction de l'objectif pédagogique du logiciel, des utilisateurs des EIAH et de la tâche. Par exemple, l'application du critère de gestion des erreurs [2] aux EIAH nécessite de faire la distinction entre deux types d'erreurs : les erreurs dans l'utilisation du logiciel et les erreurs au

sens de réponses erronées (erreurs conceptuelles) [13]. Si les erreurs d'utilisation du logiciel doivent être empêchées ou corrigées, suivant ainsi la recommandation de Bastien et Scapin, la gestion des erreurs conceptuelles doit être traitée indépendamment en fonction de la théorie cognitive sous-jacente à l'EIAH. Ainsi, dans une perspective behavioriste on choisira de corriger immédiatement chaque erreur, tandis que dans d'autres perspectives on pourra laisser l'apprenant se rendre compte lui-même de ses erreurs.

Les autres méthodes d'évaluation analytique présentées au début de cette partie sont plus difficiles à adapter. L'utilisation de check-lists semble être peu adaptée aux EIAH [25] du fait qu'elles ne prennent pas en compte le contexte d'utilisation réel du logiciel, contexte particulièrement important en EIAH. L'utilisation des « inspections cognitives » peut être envisagées, mais seulement pour certains types d'EIAH. En effet, cette méthode demande souvent une modélisation très détaillée de la tâche, ce qui est réalisable pour des environnements d'apprentissage qui favorisent le développement de compétences très précises, mais pas pour des environnements qui permettent une certaine créativité ou qui peuvent être employés de manières différentes par différents apprenants.

Les méthodes d'évaluation empiriques sont plus flexibles et semblent plus appropriées aux EIAH dans la mesure où les utilisateurs observés sont représentatifs des apprenants qui utiliseront ensuite le système. Les méthodes classiques d'observations et d'entretiens peuvent donc être utilisées. Par ailleurs, il peut être intéressant de conduire ces observations dans le contexte d'utilisation envisagé pour le logiciel afin de déterminer si les apprenants ont le comportement et les résultats escomptés dans des conditions proches de la réalité [11].

Évaluer l'utilité

Après l'évaluation de l'utilisabilité, il est nécessaire d'évaluer l'utilité du système. Rappelons que l'utilité est l'adéquation entre les fonctions fournies par le système et celles nécessaires à l'utilisateur pour atteindre les objectifs de haut niveau du client.

Classiquement, l'évaluation de l'utilité consiste à observer si l'utilisateur est capable d'accomplir sa tâche à partir des fonctionnalités du système. Pour cela, il est nécessaire de bien formaliser la tâche que l'utilisateur doit accomplir.

Dans le cadre des EIAH, l'objectif à atteindre comporte deux niveaux : l'apprentissage (de la discipline enseignée et non de la manipulation du système) et la réalisation de tâches proposées par le système (résolution de problème, recherche d'informations, simulations...) [13]. Même si ces niveaux sont connectés, il n'y a pas de lien direct entre la réalisation de la tâche et l'apprentissage effectif, un échec dans la réalisation de la tâche peut, dans certaines

conditions, être bénéfique pour l'apprentissage. Évaluer l'utilité d'un EIAH ne consiste donc pas seulement à vérifier si l'utilisateur peut réaliser la tâche qu'il souhaite faire (ce qui peut se faire selon les méthodes classiques de mesure de l'utilité), mais aussi à évaluer l'objectif de plus haut niveau qu'est l'apprentissage. Ceci nécessite des techniques spécifiques issues de différentes disciplines dont l'objet d'étude est l'apprentissage.

Dans la suite de l'article, nous présentons des méthodes qui peuvent être utilisées pour évaluer l'apprentissage et nous montrons comment elles peuvent être associées pour évaluer un EIAH donné.

Méthodes d'évaluation de l'apprentissage

Il existe de nombreuses méthodes pour évaluer l'apprentissage. Ces méthodes, issues de techniques non spécifiques aux EIAH, ont été adaptées pour évaluer de tels systèmes. Nous les présentons en distinguant les méthodes qui ont pour but de quantifier l'apprentissage, des méthodes qualitatives plus exploratoires.

Méthodes quantitatives

Les méthodes quantitatives visent à mesurer de manière objective l'impact d'un dispositif sur l'apprentissage.

Méthode comparative

La méthode comparative, développée par la psychologie cognitive, est souvent utilisée pour évaluer les EIAH. Elle consiste à comparer l'effet de plusieurs situations sur l'apprentissage qui en résulte. Cette méthode nécessite d'avoir des hypothèses précisément définies à tester. Tricot et Lafontaine [26] en rappellent le principe général : prescrire une tâche en rapport avec l'apprentissage visé pour évaluer l'état des connaissances du participant ; faire ensuite apprendre à l'apprenant ce que l'on veut lui faire apprendre (dans notre cas, au travers de l'utilisation de l'EIAH) ; prescrire enfin une seconde tâche, analogue à la première. S'il y a une différence positive de performance entre les deux tâches, on considère alors que le sujet a appris quelque chose. Pour vérifier que la différence obtenue est bien due à la phase d'apprentissage (à l'utilisation de l'EIAH), on utilise généralement une condition contrôle dans laquelle les participants font " exactement la même chose ", sauf ce qui est évalué. Pour que cette évaluation soit valide, on prend la précaution de présenter dans la condition contrôle et dans la condition testée les mêmes contenus, la même démarche, pendant le même temps, avec le même environnement, la même consigne...

Une difficulté de cette méthode (pour plus d'informations sur cette méthode, voir [23]) réside dans le choix de la condition contrôle : faut-il comparer l'EIAH à évaluer à un enseignement

oral ? à un autre système ? à une version tronquée de l'EIAH testé ? Malgré cette difficulté, cette méthode permet d'observer le résultat d'un changement dû au système et d'inférer les connaissances acquises par l'apprenant avec un certain degré de généralité. Toutefois, elle ne permet pas de comprendre ce qui se passe au cours de l'apprentissage ou au cours de l'utilisation du dispositif de formation.

Méthode micro-génétique

L'analyse micro-génétique [24] a été développée et utilisée en psychologie développementale pour favoriser une compréhension de l'évolution des connaissances de l'apprenant en relevant de nombreux échantillons de « la pensée » des apprenants au fur et à mesure qu'elle évolue.

Cette méthode consiste à faire passer des tests de manière intensive au cours de la période d'apprentissage et, pour chaque apprenant, à analyser très finement les réponses, essai par essai.

Les différentes stratégies utilisables pour répondre ayant été identifiées au préalable, il est possible d'identifier celles qui sont utilisées successivement par l'apprenant, l'ordre d'apparition de ces stratégies, leur fréquence d'utilisation ainsi que la variabilité interindividuelle. L'analyse du changement est effectuée à la fois par individu et pour le groupe. Cette méthode demande beaucoup de temps mais elle permet d'obtenir des résultats fins qui peuvent faciliter la compréhension de l'évolution des connaissances.

Méthodes « on-line »

D'autres méthodes, dites « on-line » [20] permettent de savoir ce que fait l'apprenant au cours de l'activité proposée (l'utilisation du logiciel). Ces méthodes sont pour la plupart issues de recherches en compréhension de texte. Elles permettent essentiellement d'identifier sur quels éléments l'apprenant a focalisé son attention. Ces méthodes sont donc tout à fait intéressantes pour l'étude des processus attentionnels de l'apprenant au cours de la tâche (particulièrement dans les hypermédias), mais offrent peu d'informations sur les processus d'apprentissage mis en œuvre. D'autres méthodes de recueil, s'apparentant plus à des méthodes qualitatives (comme les verbalisations) peuvent apporter des informations complémentaires. Nous les présentons dans la section suivante.

Méthodes qualitatives

Si les méthodes décrites précédemment permettent de quantifier l'impact de l'EIAH sur l'apprentissage (dans une situation contrôlée), les résultats obtenus avec ces méthodes sont souvent difficiles à généraliser à une situation complexe telle que l'apprentissage en classe. De plus, ces méthodes quantitatives ne prennent que rarement en compte les composantes

affectives de l'apprentissage pourtant essentielles pour la bonne acceptation et l'utilisation du système [16].

Les méthodes qualitatives développées en sciences humaines offrent des outils pour appréhender la globalité du phénomène étudié. Elles ne sont pas spécifiques à l'étude de l'apprentissage mais peuvent permettre de prendre en compte l'apprentissage en situation ainsi que les composantes affectives de l'apprentissage.

Recueil des verbalisations

Le recueil des verbalisations [4] consiste à demander à l'apprenant de penser à haute voix durant la tâche proposée afin d'identifier les raisonnements qu'il met en œuvre pour réaliser la tâche demandée. Cette méthode est très utilisée dans de nombreuses études anglo-saxonnes sur l'apprentissage.

Adaptation des méthodes ethnographiques

Les méthodes ethnographiques consistent à observer une situation «de l'intérieur». L'observateur fait partie de la situation qu'il observe, il est pleinement conscient de la subjectivité de sa situation et adopte donc une position réflexive sur sa pratique d'observation. Pour l'évaluation des EIAH, l'observation peut porter sur un apprenant, un groupe d'apprenant ou une classe entière [3]. L'observation individuelle apporte des informations intéressantes et diverses permettant de comprendre les interactions entre l'apprenant et le système, par exemple sur les fonctionnalités utilisées, les difficultés rencontrées, les stratégies adoptées ou encore la motivation de l'apprenant. L'observation d'apprenants qui travaillent en groupes, peut donner des informations sur leurs interactions avec le système ainsi que sur les connaissances qu'ils élaborent à travers leurs interactions [10]. L'observation d'une classe dans son ensemble peut permettre d'identifier son fonctionnement et peut aider à mettre en évidence les caractéristiques de l'enseignement [8]. L'observation peut être facilitée par le recours à des instruments d'observation (grilles d'observation [8] [3], eye-tracking) et des instruments d'enregistrement (vidéo) afin de faire a posteriori une analyse plus fine de la situation.

En complément des observations, il peut être pertinent de tenir un journal de bord afin d'augmenter la fiabilité des observations en intégrant le point de vue subjectif et réflexif du chercheur [3]. Ce journal de bord peut documenter l'évaluation mise en place en décrivant le déroulement du projet ou en relevant les éventuels changements d'attitude intervenus au cours du projet. Il peut également contenir des données réflexives (perception de l'évolution des processus d'apprentissage, questions émergent de l'observation).

Entretiens

Les données issues de l'observation peuvent être complétées par des entrevues avec les apprenants individuelles ou collectives. Ces entretiens peuvent être ouverts (les thèmes abordés varient en fonction du participant), structurés (les questions sont prédéfinies) ou semi-structurés (les questions à poser sont prédéfinies, mais une place est laissée pour des réponses plus individualisées). Les réponses et les remarques des apprenants récoltées durant l'entretien peuvent éclairer l'interprétation des résultats obtenus par d'autres méthodes, permettre de comprendre le cheminement interne de l'apprenant ou donner des indications sur le degré de conscience de l'apprenant face son apprentissage.

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 [3].

Collectes de documents

Lors d'une évaluation, un grand nombre de documents sont produits. Il peut être intéressant de les analyser, par exemple pour confirmer une interprétation. C'est pourquoi il est important de récolter tous ces documents, en particulier les productions des apprenants (productions informatisées des élèves, productions écrites comme les brouillons), les documents contextuels (photos, plan de l'environnement) et les traces des interactions entre l'apprenant et le système. Ces traces peuvent se résumer aux productions de l'élève mais peuvent être aussi beaucoup plus riches. Les traces recueillies peuvent être très diverses (frappe du clavier, clic, déplacement de la souris, utilisation des menus et fonctionnalités, dialogues...). Toutefois, il n'est pas pertinent de tracer tous ces éléments. De ce fait, les concepteurs choisissent quelles traces ils souhaitent enregistrer, mais même choisies rigoureusement, ces traces sont difficilement utilisables à l'état brut. C'est pourquoi des modèles d'interprétation de ces traces ont été proposés. Ainsi, Dubourg et al. [6] considèrent les événements systèmes tels que les clics et les frappes au clavier comme des observables de trop bas niveau pour étudier l'interaction. Ils définissent donc les événements-logiciel (menu, fonction activée) comme des séquences d'événements-système correspondant à une action significative pour la situation d'interaction. Selon Balacheff [1] la succession des événements-logiciel recueillis au cours d'une session définit le modèle comportemental de l'apprenant. Une interprétation de ce modèle permet de construire le modèle épistémique dont le rôle est d'attribuer une signification aux comportements de l'apprenant en identifiant ses buts, stratégies et connaissances.

Interprétées convenablement, ces traces d'interaction peuvent donc être une source d'information importante sur les processus d'apprentissage de l'apprenant.

En résumé, pour évaluer l'utilité d'un EIAH, c'est-à-dire identifier son impact sur l'apprentissage, nous avons présenté deux types de méthodes : les méthodes quantitatives, objectives, qui permettent de quantifier l'apprentissage et donnent des résultats généralisables, mais qui ne permettent pas de prendre en compte des situations complexes, et des méthodes qualitatives qui permettent de prendre en compte la complexité de la situation, mais dont les résultats n'ont pas vocation à être généralisés, leur validité étant restreinte à des contextes proches de la situation étudiée.

Notons que si nous utilisons cette distinction entre méthodes qualitatives et quantitatives pour faciliter notre exposé, les méthodes présentées forment plutôt un continuum entre ces deux types d'approches.

Quelle méthode pour quel environnement ?

Parmi les différentes méthodes d'évaluation de l'apprentissage, l'évaluateur doit choisir la (ou les) méthode appropriée pour évaluer un EIAH donné. Ce choix peut être guidé par la théorie de l'apprentissage sous-jacente à l'EIAH à évaluer. En effet, les environnements d'apprentissage sont toujours conçus en référence à une théorie de l'apprentissage (par exemple : cognitivisme, constructivisme, socio-constructivisme), même si cette référence n'est pas explicite. Ces différentes manières d'appréhender l'apprentissage peuvent être mises en parallèle avec les différents cadres théoriques dans lesquels les méthodes d'évaluation ont été développées.

La théorie cognitiviste fait l'analogie entre la cognition humaine et un système de traitement de l'information. Dans cette théorie, l'apprentissage est le processus qui conduit à l'intégration d'informations nouvelles en mémoire (cette théorie accorde donc une grande importance aux différents systèmes de mémoire), le comportement observé (les performances) reflète les représentations mentales formées par l'apprenant. Les méthodes qui mesurent les performances individuelles (méthodes comparative ou microgénétique) sont cohérentes avec une vision cognitiviste de l'apprentissage. Elles peuvent de ce fait apporter des informations pertinentes pour évaluer des EIAH dont le cadre théorique est le cognitivisme.

Dans la théorie constructiviste, l'apprentissage est vu comme le résultat d'une interaction entre le sujet et son environnement. Selon cette théorie, confronté à une situation nouvelle, l'individu va mettre en oeuvre des idées et démarches en fonction des représentations qu'il

possède déjà. Les situations de conflit entre les informations nouvelles et les conceptions propres à l'apprenant conduisent alors celui-ci à produire un nouveau sens pour interpréter la situation. L'approche socio-constructiviste prend en compte le contexte social dans lequel l'individu évolue. L'apprentissage se fait dans l'interaction entre le sujet et son environnement social et culturel. Dans ces deux théories, la situation est essentielle pour l'apprentissage. Les méthodes qualitatives dont l'objectif est de considérer l'homme dans le milieu où il évolue semblent alors bien adaptées aux EIAH sous-tendus par cette théorie.

Une fois la méthode choisie, il est important de préciser le type d'apprentissage visé par l'EIAH (connaissances précises, savoir-faire, savoir-être, compétences transversales) afin d'identifier les comportements à observer, de choisir les tâches à prescrire pour tester l'apprentissage, ou encore de définir les hypothèses à tester. Si l'on souhaite faire acquérir des savoir-faire dans une perspective cognitiviste, on peut par exemple tester l'apprentissage avec une tâche de résolution de problème. Si l'on souhaite faire acquérir des connaissances déclaratives avec une plateforme d'apprentissage coopératif dans une perspective socio-constructiviste, on pourra observer et analyser la référence aux connaissances à apprendre dans les dialogues entre apprenants.

Une nécessaire pluralité des méthodes utilisées

Si la théorie de l'apprentissage conduit à appliquer préférentiellement un type de méthode, le fait que l'EIAH est utilisé dans une situation complexe conduit souvent à combiner différentes méthodes pour obtenir à la fois des données quantitatives sur l'apprentissage et des données sur le contexte d'utilisation du système. Pour aborder cette question de la combinaison de plusieurs méthodes, nous présentons deux exemples : l'évaluation de l'EIAH Ambre-PA [19] et celle de la plateforme collaborative « Learning By Design » [12].

L'EIAH Ambre-PA (problèmes additifs) est destiné à faire apprendre une méthode pour la résolution de problèmes additifs. Pour ce faire, Ambre-PA propose à l'apprenant de résoudre des problèmes additifs en étant guidé à travers les étapes du raisonnement à partir de cas (reformulation du problème, choix d'un problème proche, adaptation du problème proche pour résoudre le nouveau problème). Nous faisons l'hypothèse que l'utilisation de Ambre-PA va aider l'apprenant à acquérir une méthode de résolution, c'est-à-dire à apprendre à identifier la classe d'un problème et à appliquer la technique de résolution adaptée. Cet EIAH, destiné à être utilisé individuellement en classe, se place dans une perspective cognitiviste. Pour évaluer cet environnement, nous avons comparé les résultats obtenus avec Ambre-PA avec les résultats obtenus avec un logiciel qui proposait une résolution simple des mêmes problèmes

(sans guidage). Cette expérience s'est déroulée dans une école primaire (dans la salle informatique), où des élèves de 3 classes de CE1 (enfants de 8 ans) ont utilisé l'un des logiciels une fois par semaine pendant six semaines. Pour compléter les résultats obtenus à différents tests quantitatifs proposés avant, pendant et après l'utilisation du logiciel, nous avons utilisé des méthodes qualitatives (analyse des traces d'interaction, grille d'observation de l'utilisation du logiciel, questionnaire et entrevue collective semi-structurée lors de la dernière séance d'utilisation). Ces méthodes complémentaires nous ont permis d'une part de mieux comprendre ce qui se passe pendant l'utilisation du logiciel et d'autre part de prendre en compte la situation d'apprentissage. Nous avons complété cette expérience par l'observation individuelle et le recueil de verbalisations d'enfants en dehors du cadre de la classe afin de vérifier nos interprétations de ces données.

Dans cette expérience, nous avons à la fois associé différentes méthodes au sein d'une même expérience et multiplié les expériences pour compléter nos données. Les méthodes qualitatives viennent étayer et renforcer les analyses quantitatives et permettent d'envisager des pistes explicatives aux résultats recueillis.

Une autre manière de combiner différentes méthodes est proposée dans l'évaluation de la plateforme collaborative « Learning By Design » (LBD) [12] destinée à développer l'esprit scientifique des apprenants. Ce projet est sous-tendu par une vision plus socio-constructiviste de l'apprentissage. L'évaluation associe des méthodes qualitatives ethnographiques à la méthode comparative afin d'évaluer l'effet de la plateforme sur le développement de l'esprit scientifique en fonction de la situation d'utilisation de LBD. Des ethnologues ont donc visité régulièrement les classes qui participaient au projet pour identifier les caractéristiques des enseignants et les activités réalisées à l'aide de LBD dans ces classes. Ensuite, la compréhension individuelle (évaluée par questionnaires et entretiens individuels semi-structurés) des élèves utilisant LBD a été comparée à la compréhension d'élèves suivant un enseignement traditionnel. La confrontation des résultats obtenus aux tests de compréhension avec les données issues des observations ethnographiques a permis d'identifier les compétences acquises grâce à l'utilisation de LBD en fonction des caractéristiques de l'enseignement et de la classe. Ainsi, cette évaluation a mis en évidence les situations dans lesquelles LBD était le plus bénéfique.

Ces associations de méthodes ont certaines limites. Elles nécessitent beaucoup de temps et une bonne maîtrise des techniques employées (ce qui peut nécessiter une équipe importante d'évaluateurs). L'association de méthodes pose également la question de la validité des

résultats. En effet, comme les méthodes qualitatives et quantitatives n'ont pas les mêmes critères de validation, quels critères adopter pour les méthodes mixtes ?

Malgré ces difficultés, la combinaison de méthodes apporte des informations beaucoup plus riches en permettant à la fois de quantifier les acquisitions de connaissances (ou de compétences) et de prendre en compte le contexte dans lequel a lieu l'apprentissage. Ainsi, ce type d'évaluation peut permettre de définir la situation dans laquelle un environnement conduit aux meilleurs résultats [27].

Conclusion

Dans cet article, nous nous sommes intéressées à la question de l'évaluation des EIAH. S'il ne fait plus de doute que l'évaluation doit faire partie du cycle de conception d'un Environnement Informatique pour l'Apprentissage Humain, les questions relatives à ce que l'on souhaite évaluer et comment l'évaluer restent ouvertes. Pour aborder ces questions, nous nous sommes appuyées sur la distinction entre utilisabilité, utilité et usage d'un système informatique. En effet, au cours de la conception de tout système, il est recommandé d'évaluer à la fois l'utilisabilité du système et son utilité. Pour l'évaluation de l'utilisabilité des EIAH, nous avons présenté quelques-unes des méthodes développées en IHM et proposé des adaptations au cas particulier des EIAH. Il nous semble que l'évaluation de l'utilité des EIAH requière une approche différente. L'utilité des EIAH correspondant à l'impact du système sur l'apprentissage, nous avons présenté différentes méthodes, quantitatives et qualitatives, permettant d'évaluer l'impact d'un EIAH sur l'apprentissage. Nous avons ensuite abordé la question du choix de la méthode en fonction de l'EIAH à évaluer, il nous semble en effet important de prendre en compte les théories de l'apprentissage sous-jacentes à l'EIAH à tester. Souvent, les EIAH sont appelés à être utilisés dans des situations complexes. Pour prendre en compte à la fois l'effet de l'EIAH sur chaque apprenant et une telle complexité, il semble parfois préférable de combiner plusieurs méthodes d'évaluation.

Les méthodes présentées ici nous semblent les plus directement utilisables pour évaluer les EIAH, la liste n'est donc pas exhaustive. Par ailleurs, nous avons choisi de présenter uniquement les méthodes de recueil et non les méthodes d'analyse. Enfin, nous nous sommes limitées à l'évaluation au cours de la conception, nous n'avons donc pas abordé la question de l'évaluation des usages qui intervient plutôt en fin du cycle de conception-

Remerciements

Nous tenons à remercier vivement Magali Ollagnier-Beldame pour sa participation à la réflexion qui a conduit à la rédaction de cet article.

Bibliographie

1. N. Balacheff. Didactique et intelligence artificielle. In *N. Balacheff et M. Vivet, Didactique et intelligence artificielle*, (La pensée sauvage éditions), 7-42, 1994.
2. C. Bastien et D. Scapin, Critères ergonomiques pour l'évaluation des interfaces utilisateurs, RT n°156, INRIA, juin 1993.
3. M.A. Barfurth, J. Basque, M. Chomienne, et L.R. Winer. Les instruments de collecte de données de recherche qualitative dans des environnements pédagogiques informatisés. In *Apprendre dans des environnements pédagogiques informatisés*. P. Bordeleau (dir.), Montréal, éditions Logiques, 485-548, 1994.
4. J.P. Caverni. La verbalisation comme source d'observables pour l'étude du fonctionnement cognitif. In *Caverni, J.P., Bastien, C., Mendelsohn, P., Tiberghien, G., Psychologie cognitive, modèles et méthodes*. Presse Universitaire de Grenoble, 253- 73, 1988.
5. X. Dubourg et P. Teutsch. Interface Design Issues in Interactive Learning Environments, *IFIP WG 3.3 Working Conference, Human-Computer Interaction and Educational Tools*, Sozopol, mai 1997.
6. X. Dubourg., E. Delozanne et B. Grugeon. Situations d'interaction dans un environnement d'apprentissage : le système Repères. *Actes des Quatrièmes Journées EIAO de Cachan*. Paris, Eyrolles, 223-244, 1995.
7. N. Farenc. ERGOVAL : Une méthode de structuration des règles ergonomiques permettant l'évaluation automatique d'interfaces graphiques. *Thèse en informatique*, Université Toulouse I. 1997.
8. B.B. Fasse, and J.L. Kolodner. Evaluating Classroom Practices Using Qualitative Research Methods: Defining and Refining the Process. *Proceedings of International Conference of the Learning Sciences 2000*, 93-198, 2000.
9. R.M. Gagné, L.J. Briggs and W.W. Wager. *Principles of instructional design*. New York, Holt, Renhart and Winston Eds, 1988.
10. M. Gilly, J.-P. Roux, et A. Trognon (Éds.). *Apprendre dans l'Interaction : analyse des médiations sémiotiques*. Collection « Langage - Cognition - Interaction », dirigé par A. Trognon et M. Charolles. Presses Universitaires de Nancy et Publications de l'Université de Provence, 1999.
11. D. Hoecker and G. Elias. User evaluation of the LISP intelligent tutoring system. *Proceedings of the human factors society*, 32(3), 313-324, 1986.
12. J. Holbrook, J. Gray, B.B. Fasse, P.J. Camp, and J.L. Kolodner. Managing complexity in classroom curriculum implementation sites: Triangulating Multi-Level Assessment and Evaluation. *Paper presentation in Building Sustainable Science Curriculum: Acknowledging and Accomodating Local Adaptation*. American Educational Research Association, Seattle, WA, 2001. (<http://www.cc.gatech.edu/projects/lbd/pubconf.html>)
13. S. Jean. Application de recommandations ergonomiques : spécificités des EIAO dédiés à l'évaluation. *Actes des Rencontres Jeunes Chercheurs en IHM 2000*, 39-42, 2000.

14. C. Lewis, P.G. Polson C. Wharton and J. Rieman. Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In *J.C. Chew & J. Whiteside (eds), CHI'90: Human Factors in Computing Systems*. ACM: New York, 235-242, 1990.
15. D. Litmann and E. Soloway. Evaluating ITSs: The cognitive science perspective. In *M. Polson & J. J. Richardson (Eds.), Foundations of Intelligent Tutoring Systems*. Hillsdale, NJ: LEA, 1988.
16. M.A. Mark, and J.E. Greer. Evaluation methodologies for intelligent tutoring systems. *Journal of Artificial Intelligence in Education*, 4(2/3), 129-153, 1993.
17. J. Nielsen. Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies*, 41 (3), 385-397, 1994.
18. J. Nielsen. Usability Engineering, Academic Press, 1993.
19. S. Nogry, S. Jean-Daubias and N. Duclosson. ITS evaluation in classroom: the case of the Ambre ITS. *ITS 2004 (soumis)*.
20. J.-F. Rouet et J.-M. Passerault. Analyzing learner hypermedia interaction: An overview of online methods. *Instructional Science*, 27, 201–219, 1999.
21. B. Schneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Reading, MA : Addison-Wesley, 1992.
22. B. Senach. L'évaluation ergonomique des interfaces homme – machine. In *J.-C. Sperandio éditeur, L'ergonomie dans la conception des projets informatiques*, Octares éditions, 69-122, 1993.
23. V.J. Shute, and J. W. Regian. Principles for Evaluating intelligent Tutoring Systems. *Journal of Artificial Intelligence and Education*, 4 (2/3), 245-271, 1993.
24. R.S. Siegler, and K. Crowley. The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, 46, 606-620, 1991.
25. D. Squires, and J. Preece. Predicting quality in educational software: Evaluating for learning, usability, and the synergy between them. *Interacting with Computer*, 11 (5), 467-483, 1999.
26. A. Tricot et J. Lafontaine. Une méthode pour évaluer conjointement l'utilisation un outil multimédia et l'apprentissage réalisé avec celui-ci. *Le Français dans le Monde*, Janvier, 41-52, 2002.
27. M. Vivet. Evaluating Educational Technologies: Evaluation of Teaching Material Versus Evaluation of Learning? *CALISCE 1996*, 37-38, 1996.