HAL
open science

Leçons tirées de la conception de AMBRE-add : Comment combiner les objectifs et méthodes d'évaluation pour la conception itérative des EIAH ?

Sandra Nogry, Stéphanie Jean-Daubias, Nathalie Guin-Duclosson

To cite this version:

Sandra Nogry, Stéphanie Jean-Daubias, Nathalie Guin-Duclosson. Leçons tirées de la conception de AMBRE-add : Comment combiner les objectifs et méthodes d'évaluation pour la conception itérative des EIAH ?. STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation), 2006, 13, 28 p. hal-00696321

HAL Id: hal-00696321

https://hal.science/hal-00696321

Submitted on 11 May 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sciences et Technologies de l´Information et de la Communication pour l´Éducation et la Formation

Volume 13, 2006
Article de recherche

Leçons tirées de la conception de AMBRE-add Comment combiner les objectifs et méthodes d'évaluation pour la conception itérative des EIAH ?

Sandra Nogry, Stéphanie Jean-Daubias, Nathalie Guin-Duclosson [LIRIS, Lyon]

■ RÉSUMÉ : Cet article discute des évaluations à réaliser dans le cadre de la conception itérative d'un EIAH. Au cours de la conception, différents aspects de l'interface doivent être évalués : son utilisabilité, son utilité, son acceptabilité ; différentes techniques existent pour chacun de ces aspects. Lors de la conception de l'EIAH AMBRE-add, nous avons combiné diverses techniques, successivement ou simultanément, afin de faire progresser la conception du logiciel. Dans cet article, nous analysons la démarche d'évaluation adoptée dans ce projet en précisant les apports et les limites des différentes évaluations réalisées. Nous montrons en particulier l'intérêt de combiner des techniques permettant de mesurer l'impact du logiciel sur l'apprentissage avec une analyse de l'activité des apprenants. Sur cette base, nous proposons une démarche d'évaluation plus générale pouvant être appliquée dans le cadre de la conception itérative d'EIAH support aux situations d'apprentissage.

■ MOTS CLÉS : conception itérative, utilisabilité, utilité, acceptabilité, usages, démarche d'évaluation.

■ ABSTRACT : This paper deals with the evaluation to conduct during the iterative design of an ILE. During the design of an ILE a set of aspects of the system must be evaluated: usability, utility and acceptance; in this aim, a set of evaluation techniques already exists. For the design of the ITS AMBRE-add, several techniques were combined in order to evaluate and refine the system. In this paper, the approach we adopted to evaluate AMBRE-add is described and analysed in order to specify its contributions and its limits. Finally we propose a general approach of evaluation to be implemented during iterative design of an ILE.

■ KEYWORDS : iterative design, usability, utility, acceptance, usage analysis, evaluation process.

1. Introduction

2. L'évaluation des EIAH

3. Comment conjuguer les différents objets d'évaluation et quelles méthodes utiliser au cours de la conception d'un EIAH ?

4. Quelle démarche d'évaluation mettre en œuvre dans le cadre d'une conception itérative ?

Conclusion

Remerciements

BIBLIOGRAPHIE

1. Introduction

Depuis longtemps déjà, de nombreux auteurs militent pour l'intégration de l'évaluation dès la conception des EIAH (voir par exemple (Nanard et Nanard, 1998)) afin de corriger rapidement des problèmes qu'il serait coûteux voir impossible de corriger en fin de conception. Mais quelle démarche d'évaluation suivre au cours de la conception d'un EIAH ?

Différents objets d'évaluation ont été mis en évidence, différentes techniques ont été décrites et proposées, mais comme le soulignent (Bastien et Scapin, 2004), p. 461, ces techniques sont pour la plupart indépendantes les unes des autres, ce qui pose le problème de leur articulation. Dans cet article, à la lumière de l'évaluation que nous avons conduite dans le cadre du projet AMBRE, nous proposons une démarche d'évaluation des EIAH dans un contexte de conception itérative. Cette démarche est plus particulièrement destinée aux EIAH supports aux situations d'apprentissage, c'est-à-dire aux environnements qui créent des situations d'apprentissage en contraignant l'activité individuelle afin de favoriser l'acquisition de compétences ou de concepts précis pour lesquels les interactions et les processus d'apprentissage sont considérés comme des facteurs clés (Tchounikine, 2002).

Comme l'impact de ces EIAH sur l'apprentissage dépend de la manière dont ils orientent l'activité des apprenants, il nous paraît nécessaire de combiner l'analyse de l'activité des apprenants utilisant le logiciel et son impact sur l'apprentissage. La démarche que nous proposons vise donc à conjuguer différentes méthodes d'évaluation afin d'intégrer au cycle de conception à la fois l'évaluation de l'utilisabilité du logiciel, son impact sur l'apprentissage et l'analyse de l'activité individuelle réalisée par les apprenants dans l'environnement.

Après avoir présenté les différents objets d'évaluation habituellement pris en compte lors de la conception de systèmes informatiques, ainsi que quelques-unes des méthodes associées, nous mettrons en évidence les spécificités de l'évaluation des EIAH en discutant des méthodes adaptées à cette évaluation. Ensuite, nous présenterons les évaluations que nous avons réalisées dans le cadre de la conception de AMBRE-add, un EIAH traitant le domaine des problèmes additifs. L'évaluation de ce système, qui s'est déroulée en plusieurs étapes et a impliqué une centaine d'élèves, présente la caractéristique d'être particulièrement complète. Cette partie sera suivie d'une critique de la démarche suivie : les avantages et les limites des différentes méthodes utilisées seront présentées, puis nous montrerons l'intérêt de combiner l'analyse de l'activité réalisée dans l'environnement à l'évaluation de l'utilisabilité et de l'impact du système sur l'apprentissage. Enfin, nous décrirons la démarche d'évaluation que nous proposons pour la conception itérative d'un EIAH.

2. L'évaluation des EIAH

De nombreuses méthodes destinées à évaluer les différents aspects de l'interface des systèmes informatiques ont été proposées (Kolski, 2001). Trois aspects sont particulièrement importants à évaluer : l'utilisabilité d'un logiciel, son utilité et son acceptabilité.

L'utilisabilité concerne la facilité d'utilisation d'un système. Elle est définie comme étant l'adéquation entre la manière dont une tâche est réalisée par un utilisateur et les capacités cognitives de cet utilisateur (Farenc, 1997). 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). L'utilisabilité est souvent le seul critère ergonomique pris en compte par les concepteurs (Burkhardt et Sperandio, 2004). Cependant Senach distingue l'utilité de l'utilisabilité d'un système, où 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 (Senach, 1993). Quant à l'acceptabilité, elle concerne la perception qu'a l'utilisateur de l'utilisabilité et de

l'utilité du système, et influe sur sa décision d'utiliser cet outil. Elle est dépendante des normes, valeurs, motivations et affects des utilisateurs (Davis, 1986), (Davis, 1989), (Nielsen, 1993), (Dillon et Morris, 1996).

Dans cette section, après avoir précisé les spécificités des EIAH par rapport aux autres environnements informatiques, nous présentons une revue, qui ne se veut pas exhaustive, des différentes méthodes permettant d'évaluer l'utilisabilité, l'utilité et l'acceptabilité d'un EIAH support aux situations d'apprentissage individuel.

Spécificités des EIAH

Les EIAH ont certaines spécificités qu'il est nécessaire de prendre en compte pour évaluer de manière adéquate l'utilisabilité, l'utilité et l'acceptabilité des EIAH.

La première spécificité concerne l'objectif de haut niveau des EIAH : favoriser l'apprentissage. Cet objectif peut différer de la tâche que doit accomplir l'apprenant (l'objectif à court terme). Par exemple, pour favoriser l'apprentissage d'une notion particulière, un premier objectif à court terme peut être de rechercher des documents traitant de cette notion dans un hypermédia, conférant ainsi à l'EIAH deux objectifs de natures différentes.

La seconde spécificité est liée aux utilisateurs des EIAH. Il faut en effet distinguer deux types d'utilisateurs : les apprenants et les enseignants. 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 par les apprenants d'un EIAH donné, mais ils peuvent également en être utilisateurs secondaires (Dubourg et Teutsch, 1997), (Jean-Daubias, 2003) du fait qu'ils peuvent parfois préparer le travail de leurs apprenants en paramétrant le système. Dans ce cas, on doit prendre en compte les deux types d'utilisateurs pour l'évaluation du système.

Par ailleurs, les situations d'utilisation des EIAH peuvent être très variées. 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 les autres 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). Or la situation d'utilisation d'un EIAH a un impact sur l'apprentissage. Il est donc nécessaire de prendre en compte cette situation lors de l'évaluation de l'apprentissage.

Suivant ces spécificités, les méthodes d'évaluation de l'utilisabilité doivent être adaptées aux EIAH (Hû et Trigano, 1998), (Jean, 2003) et les méthodes d'évaluation de l'utilité doivent être spécifiques à l'objectif de haut niveau de ces systèmes : l'apprentissage.

Utilisabilité et EIAH

« L'utilisabilité des EIAH se joue au niveau de l'interface (sa cohérence, sa lisibilité, la façon dont elle représente les actions possibles, etc.), de sa navigation (la cohérence, la simplicité, l'exhaustivité des déplacements possibles), et de sa cohérence avec l'objectif » (Tricot et al., 2003).

Classiquement on oppose deux types de méthodes d'évaluation de l'utilisabilité : l'évaluation analytique et l'évaluation empirique.

L'évaluation analytique de l'utilisabilité consiste à étudier l'interface selon un ensemble de référents afin de contrôler qu'elle possède bien certaines qualités et de détecter les problèmes qu'elle peut poser. L'évaluation analytique est souvent faite sous forme d'une évaluation par inspection, conduite par le concepteur ou par des experts qui peuvent s'appuyer sur des listes de critères conçues par des ergonomes (voir par exemple (Bastien et Scapin, 1993), (Nielsen, 1993), (Lewis et al., 1990), (Schneiderman, 1992). Ces critères et attributs peuvent être utilisés pour évaluer l'utilisabilité des EIAH (Squires et Preece,

1999), mais doivent parfois être adaptés en fonction de l'objectif pédagogique du logiciel, des utilisateurs des EIAH, de la tâche et de la théorie de l'apprentissage sous-jacente à l'environnement (Jean-Daubias, 2003), (Hû et Trigano, 1998).

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. L'utilisation de check-lists semble toutefois être peu adaptée aux EIAH (Squires et Preece, 1999) du fait qu'elles ne prennent pas en compte le contexte d'utilisation réel du logiciel, contexte particulièrement important en EIAH.

L'évaluation empirique consiste à recueillir des données relatives au comportement de l'utilisateur lors de l'utilisation du système. L'utilisation qui est faite du système par ses 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 (Gagné et al., 1988). Ainsi, pour évaluer empiriquement l'utilisabilité d'un système, Nielsen propose d'observer des utilisateurs durant l'utilisation du logiciel suivant des critères déterminés et d'interroger ensuite les utilisateurs (Nielsen, 1993). Ses études montrent qu'un panel de cinq utilisateurs « représentatifs » de chaque type de population visée permet d'identifier 80% des problèmes (Nielsen et Landauer, 1993). Ces techniques d'observation peuvent être complétées par un questionnaire afin d'avoir des indications sur la compréhension ou la satisfaction de l'utilisateur.

Ces méthodes d'évaluation empiriques sont plus flexibles et semblent plus appropriées aux EIAH que les méthodes analytiques, dans la mesure où les utilisateurs observés sont représentatifs des apprenants qui utiliseront ensuite le système. Ces observations, conduites dans le contexte d'utilisation envisagé pour le logiciel, peuvent permettre de déterminer si les apprenants ont le comportement et les résultats escomptés dans des conditions proches de la réalité (Hoecker et Elias, 1986).

Méthodes d'évaluation de l'utilité des EIAH

S'il est important qu'un EIAH soit utilisable, il est également nécessaire de vérifier qu'il est utile, c'est-à-dire qu'il permet d'atteindre l'objectif fixé. 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 des tâches proposées par le système dans le but de permettre cet apprentissage (résolution de problèmes, recherche d'informations, simulations...) (Vivet, 1996), (Jean-Daubias, 2003). Même si ces niveaux sont connectés, il n'y a pas toujours de lien direct entre la réalisation de la tâche et l'apprentissage effectif, un échec dans la réalisation de la tâche pouvant, dans certaines conditions, être bénéfique pour l'apprentissage. Évaluer l'utilité d'un EIAH ne consiste donc pas uniquement à vérifier que l'utilisateur peut réaliser la tâche qu'il souhaite accomplir, ce qui peut être fait avec des méthodes classiques, cela consiste également à évaluer si l'objectif d'apprentissage tel qu'il est défini, souvent conjointement par le concepteur, le prescripteur et éventuellement l'enseignant qui prépare la session d'apprentissage, est atteint par l'apprenant.

Il existe de nombreuses méthodes pour évaluer l'apprentissage, pour la plupart empiriques. La **méthode comparative**, développée par la psychologie cognitive, est souvent utilisée pour évaluer les EIAH (voir (Shute et Regian, 1993) pour une description). Elle consiste à comparer l'effet de plusieurs situations sur l'apprentissage qui en résulte à l'aide de tâches à réaliser avant et après la situation d'apprentissage proposée (voir (Tricot, 2002) pour une revue critique des tâches utilisées pour l'évaluation). La difficulté réside dans le choix de la condition contrôle : faut-il comparer l'EIAH 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'utilisation du dispositif de formation. Des **méthodes « on-line »** (Rouet et Passerault, 1999) apportent des indications sur l'activité de l'apprenant au cours de l'utilisation du logiciel. Certaines méthodes permettent d'identifier sur quels éléments l'apprenant a focalisé son attention. D'autres

favorisent la compréhension des processus d'apprentissage mis en œuvre. Ainsi, le **recueil des verbalisations** (Caverni, 1988), qui consiste à demander à l'apprenant de penser à haute voix durant l'activité proposée, permet d'identifier les raisonnements qu'il peut mettre en œuvre pour la réaliser. **La collecte de documents et de traces d'interactions** peut compléter les informations verbales. En complément de ces méthodes, les **méthodes ethnographiques** permettent de prendre en compte la situation dans laquelle l'apprentissage se déroule (Barfurth et al., 1994), (Fasse et Kolodner, 2000).

Acceptabilité des EIAH

Enfin, une troisième dimension à évaluer concerne l'acceptabilité de l'EIAH. Celle-ci peut se définir comme « la valeur de la représentation mentale (attitudes, opinions, etc., plus ou moins positives) à propos de l'EIAH, de son utilisabilité et de son utilité » (Tricot et al., 2003), p. 396. Cette représentation mentale, individuelle ou collective, conditionnerait la décision d'utilisation de l'EIAH, c'est-à-dire d'une part la décision des prescripteurs de proposer ou non le logiciel, et d'autre part la motivation des apprenants à l'utiliser. Comme le précisent (Tricot et al., 2003) « l'acceptabilité peut être sensible à des facteurs tels que la culture et les valeurs des utilisateurs, leurs affects, leur motivation, l'organisation sociale et les pratiques dans lesquelles s'insère plus ou moins bien l'EIAH ». L'acceptabilité peut faire l'objet d'une évaluation par inspection en étudiant l'acceptabilité en termes d'adéquation aux besoins de l'institution, aux attentes des apprenants, aux caractéristiques des apprenants. L'évaluation peut également être envisagée en termes de compatibilité avec l'organisation du temps, l'organisation des lieux, la présence du matériel nécessaire, la visibilité des résultats. Enfin, l'acceptabilité peut faire l'objet d'une évaluation empirique par observations, entretiens ou questionnaires en prenant en compte la motivation, les affects, les cultures et les valeurs des utilisateurs et prescripteurs de l'EIAH -voir (Tricot et al., 2003), page 194.

Selon le modèle proposé par Nielsen, l'acceptabilité d'un système comporte une dimension pratique et une dimension sociale (Nielsen, 1993). L'utilisabilité et l'utilité d'un système font partie des critères qui déterminent son acceptabilité pratique. Elles apparaissent comme des conditions nécessaires à une bonne acceptabilité, mais elles ne sont pas suffisantes. Des études plus récentes ont proposé d'autres modèles et montrent que les relations entre ces trois dimensions ne sont pas faciles à évaluer et sont plus complexes que ce que laisse supposer le modèle de Nielsen, chacune de ces trois dimensions ayant une influence sur les autres -voir (Tricot et al., 2003) pour une revue.

Cette revue n'est pas exhaustive, la question de l'évaluation des EIAH a été traitée plus largement, notamment dans (Delozanne, 2006), (Nogry et al., 2004), (Tricot et al., 2003).

3. Comment conjuguer les différents objets d'évaluation et quelles méthodes utiliser au cours de la conception d'un EIAH ?

Comme nous l'avons souligné dans la partie précédente, pour évaluer un EIAH, il est particulièrement important de vérifier qu'il est utilisable et acceptable par les prescripteurs et les apprenants, qu'il permet de réaliser la tâche que l'apprenant souhaite accomplir, et qu'il a bien un impact sur l'acquisition des connaissances ou des compétences visées. Cette partie a pour but de montrer comment prendre en compte ces différentes facettes précocement au cours de la conception.

Le génie logiciel a développé des méthodes de conception qui permettent d'intégrer l'évaluation au cycle de conception. Ainsi, la conception itérative consiste à affiner progressivement les spécifications en évaluant les solutions retenues, puis en intégrant les modifications choisies jusqu'à obtention d'un produit satisfaisant (Van Eylen et Hiraclidès, 1996 cité par (Jean-Daubias, 2000) ; voir également (Delozanne, 2006)). Cette démarche de conception s'appuie sur la réalisation de maquettes et de prototypes qui sont

évalués puis éventuellement modifiés. Les méthodes d'évaluation de l'utilisabilité ont été conçues pour informer la conception, et peuvent donc facilement être utilisées dans un tel cycle de conception. En revanche, les méthodes d'évaluation de l'impact d'un EIAH sur l'apprentissage ont souvent été utilisées pour évaluer la qualité d'un EIAH en fin de conception ou *a posteriori*. Or, en situation de conception, pour aboutir à une nouvelle version du logiciel, les méthodes d'évaluation doivent non seulement informer sur la qualité du logiciel, mais aussi permettre d'identifier les déficits éventuels du logiciel ainsi que les causes de ces déficits : relèvent-ils d'un problème de contenu, d'une inadéquation entre le dispositif et le format de la connaissance à acquérir (pour une synthèse, voir (De Vries, 2001)), d'un manque de correspondance entre l'objectif d'apprentissage et les besoins de l'apprenant (Tricot et al., 2003) ?

Quelles méthodes d'évaluation utiliser pour éclairer la conception sur chacune de ces facettes ? Certaines méthodes étant plus contraignantes que d'autres à mettre en œuvre, comment les conjuguer pour faire évoluer la conception sans l'alourdir ? S'il existe des descriptions de chacune de ces méthodes, de leur intérêt et de leurs limites, il n'existe pas à notre connaissance de guide qui permette d'orienter la démarche d'évaluation en conception. Aussi, sur la base des travaux existants, nous avons mis en place une démarche d'évaluation pour orienter la conception de l'EIAH AMBRE-add.

Dans cette section, après avoir présenté le principe de cet EIAH et son domaine d'application, nous décrirons les différentes évaluations réalisées dans le cadre de la conception de cet environnement en en présentant les intérêts et les limites, puis nous procéderons à une réflexion critique sur la démarche d'évaluation ainsi mise en place.

L'EIAH AMBRE-add

L'EIAH AMBRE-add, conçu dans le cadre du projet AMBRE (Guin-Duclosson et al., 2002), propose de guider la résolution de problèmes additifs afin de conduire les apprenants à acquérir des classes de problèmes (Vergnaud, 1982), (Riley et al., 1983), (Guin, 1991) et des techniques associées à ces classes de problèmes. Suivant le principe du projet AMBRE, l'EIAH présente à l'apprenant des problèmes résolus puis assiste l'apprenant dans la résolution d'un nouveau problème en le guidant à travers les différentes étapes du cycle du Raisonnement à Partir de Cas, RàPC par la suite (cf. Figure 1) :

- Après avoir lu l'énoncé du problème, l'apprenant doit reformuler le problème afin d'identifier les éléments pertinents pour la résolution.
- Ensuite il choisit, parmi les problèmes-types, un problème proche du problème à résoudre.
- Puis l'apprenant adapte la solution du problème type au problème à résoudre.
- Enfin il range le nouveau problème dans la base de problèmes déjà rencontrés.

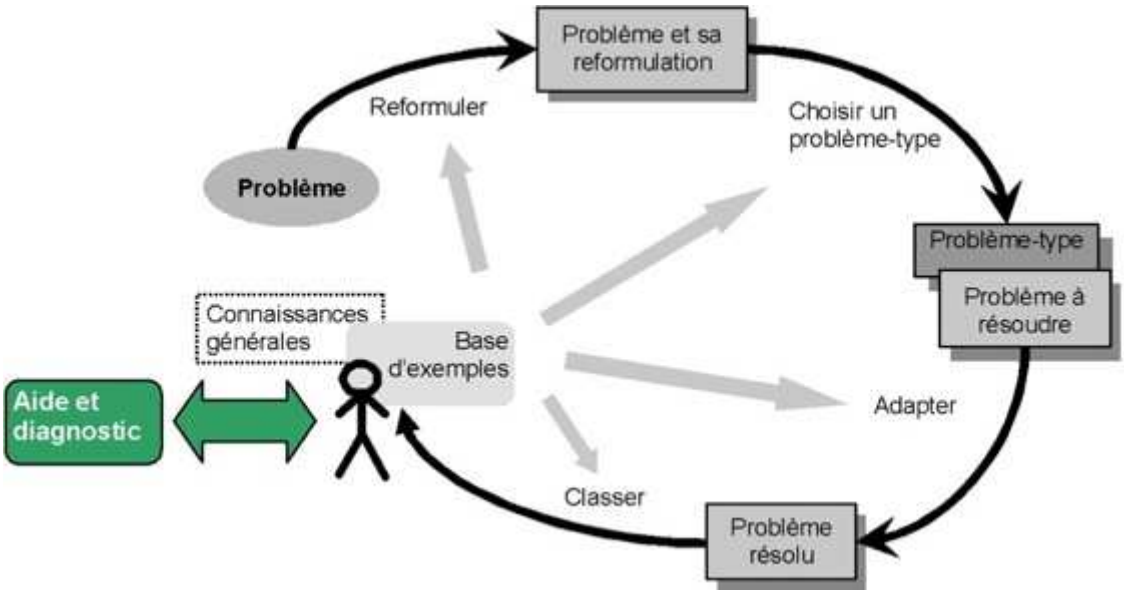

Figure 1 : Le cycle du RàPC adapté au projet AMBRE (« cycle AMBRE ») ; ce cycle décrit les activités de l'apprenant lors de l'utilisation du logiciel.

Ces étapes sont donc guidées par le système mais effectuées par l'apprenant. Dans chacune de ces étapes, un diagnostic est réalisé par le système. Celui-ci évalue la production de l'apprenant et lui propose si nécessaire des explications pour l'amener à comprendre ses erreurs et à corriger ses réponses. Nous considérons que ce diagnostic remplace l'étape de révision présente dans le cycle du RàPC ; ainsi, la révision est distribuée entre les différentes étapes.

L'objectif du projet AMBRE est de montrer que des EIAH conçus suivant ce principe peuvent faciliter l'acquisition de classes de problèmes et des techniques de résolution associées, dans le domaine des problèmes additifs comme dans d'autres domaines d'application.

Les problèmes proposés par AMBRE-add aux apprenants sont des problèmes additifs à une étape étudiés principalement en cycle 2 de l'école primaire française (par des élèves de 7 - 8 ans).

Les problèmes additifs décrivent une situation concrète, par exemple un jeu de billes : « Karim avait 32 billes. À la fin de la récréation, il en a 45. Combien a-t-il gagné de billes pendant la récréation ? ». Pour un problème de ce type, la résolution proposée dans AMBRE-add est décomposée en quatre étapes :

- Décrire le problème à l'aide d'une "opération à trou" : 32 + ? = 45 (équation représentant le problème)
- Écrire comment on effectue le calcul : 45 - 32 = ?
- Effectuer le calcul : 13
- Écrire la réponse à la question : Karim a gagné 13 billes.

Seule l'équation représentant la résolution est considérée comme étant la technique de résolution associée à la classe, le reste de la résolution mobilise d'autres compétences relevant du calcul numérique.

Certains problèmes additifs sont résolus dès la maternelle, alors que d'autres posent encore des difficultés en fin de troisième (Marthe, 1982). Greeno et Riley montrent que les difficultés que rencontrent les jeunes enfants en résolvant des problèmes additifs viennent essentiellement du fait qu'ils n'arrivent pas à se représenter correctement la situation décrite dans l'énoncé (Greeno et Riley, 1987). De nombreuses études montrent que la catégorie du problème et la nature de l'inconnue interviennent dans la difficulté du problème. Le projet AMBRE postule que le fait de demander à l'apprenant de reformuler le nouveau

problème puis de rechercher un problème proche afin d'adapter sa solution pour résoudre le nouveau problème peut lui apprendre à modéliser le problème et à acquérir les classes de problèmes et les techniques de résolution associées.

Cycle de conception et d'évaluation de l'EIAH AMBRE-add

L'EIAH AMBRE-add a été développé au sein d'une équipe pluridisciplinaire en suivant un cycle de conception itérative. Cette démarche de conception itérative a été choisie afin de valider les choix de conception pluridisciplinaires et de détecter précocement les difficultés. Suivant cette démarche, durant la conception de AMBRE-add, des phases de spécification, d'implémentation de prototypes, d'évaluation, et de production de nouvelles spécifications se sont alternées (Figure 2).

De par la nature pluridisciplinaire du projet AMBRE, des partenaires aux profils variés sont intervenus dans chacune de ces étapes suivant une méthode de conception différenciée (Jean-Daubias, 2004) : des chercheurs en informatique, mais aussi en psychologie cognitive et en didactique des mathématiques, une conseillère pédagogique, des enseignants et des apprenants.

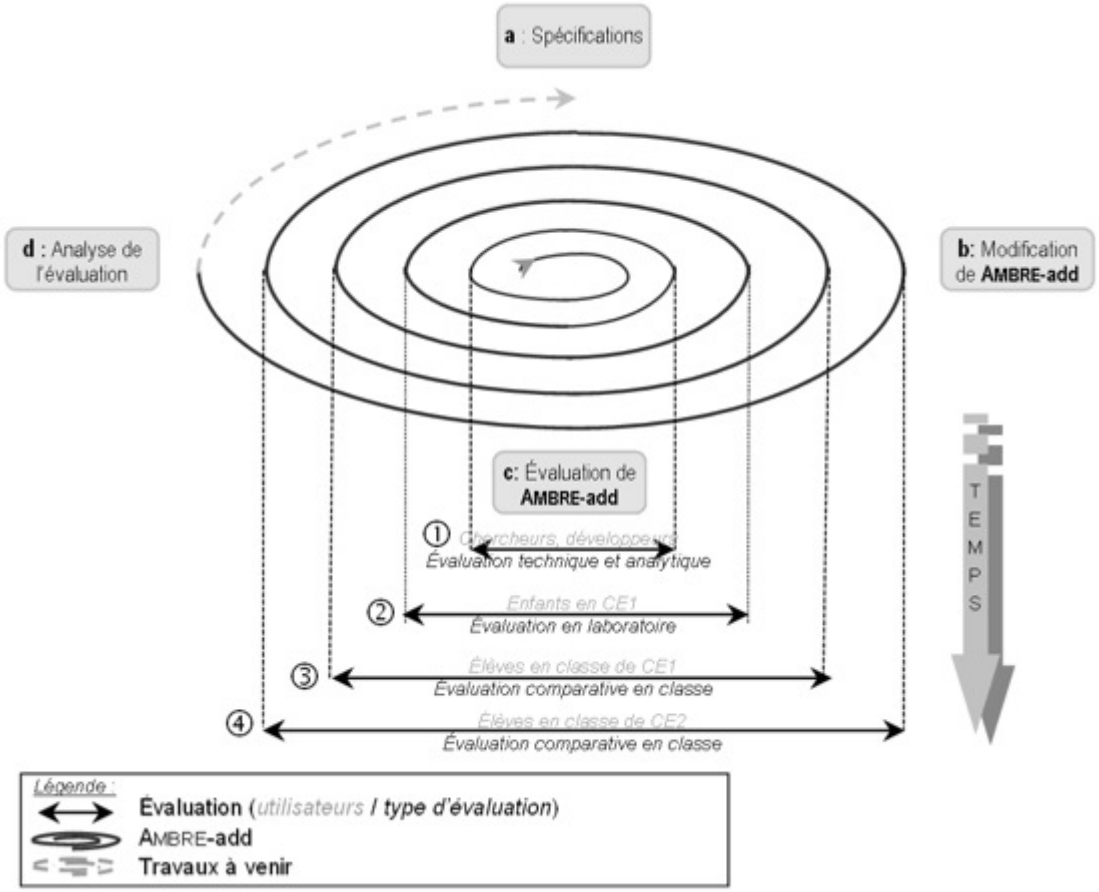

Figure 2 : Le cycle de conception itérative de l'EIAH AMBRE-add qui montre l'élargissement progressif des utilisateurs, des tests et des validations logicielles (inspiré du schéma proposé par Jean, 2000).

Des premières spécifications générales définissant le principe du projet AMBRE ont d'abord été proposées (cf. Figure 2) par l'équipe de conception. Ensuite, ces spécifications ont été appliquées au domaine des problèmes additifs afin de concevoir et d'implémenter l'EIAH AMBRE-add. Dans le cadre d'une conception itérative, ce système a fait l'objet d'une évaluation technique par les développeurs. Ensuite, différentes évaluations analytiques, simples et rapides à mettre en œuvre, ont été réalisées, certaines par les concepteurs afin d'identifier des défauts d'utilisabilité, d'autres avec une conseillère

pédagogique et des enseignants afin d'adapter l'environnement aux élèves de primaire. Chaque évaluation était suivie de nouvelles spécifications et de modifications du système. Une fois que la conception a abouti à une version de l'EIAH en adéquation avec les objectifs initiaux du projet et pouvant être intégrée, *a priori*, dans les dispositifs pédagogiques existants, quatre évaluations empiriques ont été réalisées. Celles-ci avaient pour but d'évaluer l'utilisabilité et l'utilité du système en observant le comportement réel des apprenants utilisant le logiciel ; ces évaluations ont toutes été suivies de nouvelles spécifications et de modifications du logiciel.

Dans la section suivante, nous présentons les méthodes utilisées dans chacune de ces évaluations, ainsi que les spécifications auxquelles elles ont abouties.

Les évaluations menées durant la conception de AMBRE-add

Pour évaluer l'utilisabilité de AMBRE-add et son impact sur l'apprentissage, nous avons conduit plusieurs études en combinant différentes méthodes (présentées dans la partie 2). Nous présupposions implicitement qu'il était nécessaire d'évaluer d'abord l'utilisabilité du système puis son utilité ; l'acceptabilité étant à prendre en compte dans un troisième temps. Aussi, après des évaluations analytiques par inspection, une première évaluation a été conduite en laboratoire avec cinq élèves de CE1 dans le but d'identifier les plus grosses difficultés d'utilisabilité du système. Ensuite une évaluation comparative a été réalisée à l'école auprès de trois classes d'élèves de CE1 pendant plusieurs semaines, afin d'observer l'utilisation du logiciel en situation réelle et d'évaluer l'impact du principe du logiciel (le cycle AMBRE) sur l'apprentissage (voir (Nogry, 2005) pour une description détaillée). Comme nous allons le voir, cette évaluation s'est révélée trop ambitieuse pour différentes raisons que nous exposerons. Elle nous a toutefois fourni des indications relatives à l'utilisabilité et à l'acceptabilité du logiciel par les élèves de CE1, ce qui nous a amenées à réaliser d'autres évaluations, plus légères, auprès d'enfants en CE2. La troisième évaluation a donc consisté à observer 21 élèves de CE2 utiliser le logiciel pendant quatre séances afin d'analyser leur activité dans l'environnement informatique. Une quatrième étude, une évaluation comparative en classe de CE2, a été réalisée afin de confirmer les observations réalisées au cours de l'étude précédente et d'évaluer l'impact du logiciel lui-même sur l'apprentissage.

Évaluations par inspection

L'équipe de conception a d'abord procédé à différentes évaluations analytiques concernant plusieurs aspects du logiciel. Outre l'évaluation technique, nous avons réalisé une évaluation de l'utilisabilité en inspectant les différents aspects de l'interface suivant les critères proposés par Bastien et Scapin en accordant une attention particulière à la cohérence de l'interface, au guidage et à la gestion des erreurs (Bastien et Scapin, 1993), (Jean-Daubias, 2003).

Concernant la cohérence, nous nous sommes assurées que les codes, les dénominations et les interactions proposées soient bien homogènes. Nous avons évalué le guidage en nous assurant que l'utilisateur puisse savoir à tout moment où il se situe dans la séquence d'actions et quelles actions lui sont permises. De plus, nous avons vérifié que les groupements proposés facilitent bien la compréhension et la mise en relation des différents éléments d'interface. En outre, nous avons essayé de réduire le nombre d'erreurs d'utilisation possibles, et nous avons vérifié qu'un message d'aide était systématiquement prévu en cas d'erreur d'utilisation.

AMBRE-add a également fait l'objet d'une évaluation par inspection par des enseignants et une conseillère pédagogique afin de valider l'interface sur le plan pédagogique. Cette inspection a permis d'évaluer différents aspects du logiciel. Tout d'abord, une inspection globale de l'EIAH a été réalisée afin que les enseignants vérifient que les différentes activités proposées par le logiciel dans ses différentes étapes étaient adaptées à des élèves de cycle 2. Cette inspection avait également pour objectif de vérifier la qualité des consignes et des messages d'erreur et de diagnostic, afin d'ajuster le vocabulaire utilisé aux élèves de CE1. Cette inspection a par ailleurs porté sur le contenu proposé par le logiciel : une analyse

systématique des problèmes présentés a permis de vérifier que les thèmes abordés étaient adaptés et que les valeurs utilisées dans les problèmes correspondaient bien au niveau des apprenants.

Ainsi, cette inspection pédagogique nous a permis de compléter l'inspection ergonomique de l'utilisabilité en apportant un avis d'experts sur l'adéquation des consignes et des messages au public visé et sur l'adéquation du contenu proposé au niveau des apprenants.

Évaluation empirique conduite en laboratoire

Une première évaluation empirique a été conduite en laboratoire afin de mettre en évidence les plus grosses difficultés rencontrées par des enfants en classe de CE1 lors de l'utilisation de l'EIAH AMBRE-add. Comme le préconisent Nielsen et Landauer (Nielsen et Landauer, 1993), cinq enfants ont été observés individuellement lors d'une séance d'utilisation du système d'une durée de 45 minutes. Ils étaient ensuite interrogés sur différents aspects de l'utilisation du logiciel. Cette observation était guidée par des critères ergonomiques sélectionnés parmi ceux proposés par Bastien et Scapin, Nielsen, ainsi que Schneiderman, tels que la prise en main du logiciel, la compréhension générale, l'efficacité, la gestion des erreurs, la surcharge cognitive et la satisfaction (Bastien et Scapin, 1993), (Nielsen, 1993), (Schneiderman, 1992). Pour chaque critère, un ensemble d'observables avaient été préalablement défini.

Apports de cette évaluation

Cette première observation et l'analyse des réponses des enfants au questionnaire nous ont permis de décrire l'utilisation du logiciel suivant les critères choisis et de mettre en évidence les problèmes d'utilisabilité rencontrés par les enfants lors de l'utilisation de l'interface. Nous avons par exemple observé certaines difficultés pour comprendre le principe général du logiciel ; une séance d'utilisation ne semblait pas suffisante pour prendre en main le logiciel. Des problèmes d'interaction ont également été mis en évidence : dans la phase d'analyse des exemples, certains éléments d'interface évoquaient des cases à remplir mais n'étaient pas associés à une action, ce qui induisait les enfants en erreur. Enfin, nous avons mis en évidence une difficulté d'ordre mathématique récurrente chez tous les utilisateurs. Dans l'étape d'adaptation, après avoir construit l'addition à trou (ex : 32+ ?=43), les apprenants éprouvaient de très grosses difficultés à trouver l'opération correspondante (ex : 43-32= ?) (cf. Figure 3). Les enfants ne disposent en effet pas des outils mathématiques pour effectuer cette transformation en classe de CE1. Ce problème n'avait pourtant été évoqué ni par la conseillère pédagogique ni par les enseignants durant les évaluations analytiques.

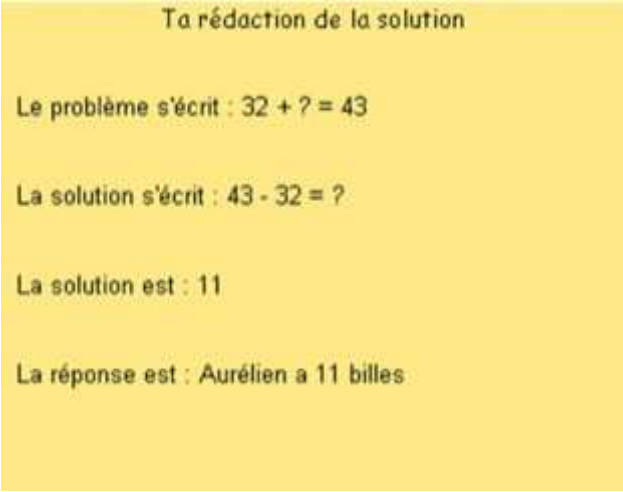

Figure 3 : Plan de rédaction de la solution dans AMBRE-add.

Ces différentes observations nous ont amenées à proposer des recommandations qui ont conduit à modifier certains éléments d'interface, à revoir la présentation du logiciel afin de faciliter sa prise en main

et sa compréhension, et à modifier l'étape d'adaptation afin d'éliminer la difficulté liée au calcul.

Limites

Cette évaluation comporte cependant certaines limites. Durant l'observation, si nous avons remarqué qu'une séance d'utilisation n'était pas suffisante pour prendre en main le logiciel, nous n'avons pas eu la possibilité de déterminer le nombre de séances nécessaires pour prendre en main le logiciel, les enfants n'étant venus au laboratoire que pour une seule séance.

Par ailleurs, les conditions d'observation ont pu influencer le comportement des enfants : la présence de l'observateur a pu les pousser à être plus concentrés et plus attentifs qu'ils ne l'auraient été en utilisant le logiciel en classe. Par ailleurs, les enfants ont souvent eu la tentation de demander de l'aide à l'observateur plutôt que d'utiliser l'aide du logiciel (même si l'observateur minimisait ses interventions), ce qui a sans doute limité la qualité des observations portant sur l'utilisation du système d'aide et de diagnostic.

Évaluation de l'impact du logiciel sur l'apprentissage

Après avoir identifié les principales difficultés rencontrées lors de l'utilisation du logiciel et apporté des modifications à AMBRE-add, nous avons conduit une deuxième évaluation, dans une situation proche d'une situation réelle d'utilisation scolaire. L'objectif principal de cette évaluation était de mesurer l'impact du cycle AMBRE sur l'apprentissage en utilisant une méthode comparative consistant à utiliser AMBRE-add ou un logiciel contrôle et comportant différents tests permettant de comparer les progrès accomplis par les élèves suivant le système utilisé.

Durant cette étude, le logiciel était utilisé par demi-classe, en salle informatique, pendant six séances d'une demi-heure. Le fait de réaliser cette évaluation en classe permet d'étudier l'utilisation du logiciel dans une situation écologique, familière aux élèves. Tous les paramètres de cette situation d'utilisation ne pouvant pas être contrôlés, diverses méthodes qualitatives ont été utilisées afin d'observer l'activité des apprenants durant les séances d'une part et de prendre en compte la situation dans laquelle le système était utilisé, ainsi que les interactions entre apprenants et encadrants d'autre part.

La section suivante présente les techniques mises en place pour évaluer chacun de ces aspects, puis les résultats obtenus. Nous discuterons ensuite l'intérêt et les limites de ces techniques au vu des objectifs visés.

Impact de AMBRE-add sur l'apprentissage

Pour tester l'impact du cycle AMBRE sur l'apprentissage de classes de problèmes et des techniques associées, nous avons comparé l'utilisation de AMBRE-add à celle de deux logiciels contrôles. L'objectif de cette comparaison était de montrer que guider la résolution d'un problème à travers les étapes du cycle AMBRE (cf. Figure 4), a un impact plus important sur l'apprentissage que d'autres formes de résolutions de problèmes arithmétiques sur ordinateur.

Le premier logiciel contrôle (« résolution simple », cf. Figure 4) présentait d'abord des problèmes résolus puis proposait de résoudre un nouveau problème directement, sans le reformuler ou identifier un problème proche. Après la résolution, un écran présentait à l'apprenant un bilan du problème. Comme, dans cette condition, le nombre d'étapes pour la résolution était beaucoup moins important que dans AMBRE-add, une tâche complémentaire était proposée afin d'équilibrer le temps passé entre les deux systèmes, elle consistait à rechercher des informations dans un énoncé.

Figure 4 : Logiciels utilisés pour évaluer l'impact de AMBRE-add dans une approche comparative.

Le deuxième logiciel contrôle (« reformulation et résolution », cf. Figure 4) présentait des problèmes résolus puis demandait à l'apprenant de résoudre un nouveau problème en le reformulant avant de chercher sa solution. Un bilan de la résolution était ensuite proposé. Contrairement à AMBRE-add, ce système ne proposait pas de choisir et d'adapter un problème-type. La comparaison de l'impact sur l'apprentissage de ce logiciel avec l'impact de AMBRE-add devait permettre d'évaluer si le choix d'un problème type et son adaptation a une influence réelle sur l'apprentissage, ou si l'étape de reformulation peut expliquer à elle-seule l'impact du système. Il est à noter que ces deux logiciels contrôle ont été conçus sur la base de AMBRE-add, la charte graphique était identique, les messages d'aide et d'explications ont été adaptés à partir de ceux proposés par AMBRE-add, quant aux problèmes proposés dans ces trois logiciels, ils étaient identiques.

Nous avons choisi de conduire cette expérience auprès de trois classes d'élèves de CE1, effectif qui nous paraissait suffisant pour pouvoir mettre en évidence des différences significatives entre groupes.

Dans chacune des trois classes de CE1 participant à l'évaluation, la moitié des élèves utilisait le logiciel AMBRE-add tandis que l'autre moitié utilisait l'un des deux logiciels contrôle implémentés pour cette expérience.

L'impact des différents logiciels sur l'apprentissage des classes de problèmes et des techniques associées était mesuré à travers différents tests (voir (Nogry et al., 2004) pour une présentation détaillée). Parmi ceux-ci, un test papier-crayon consistant à résoudre six problèmes était proposé dans le but de mesurer les progrès réalisés par les élèves suivant le logiciel contrôle utilisé. Celui-ci était présenté avant l'utilisation des logiciels (test 1), après la quatrième séance (test 2), immédiatement après la dernière séance (test 3) et un mois après (test 4).

Nous avons complété cette approche expérimentale par l'enregistrement des traces d'interaction, un recueil d'observations, un questionnaire et des entretiens.

Les traces d'interaction comportaient des informations telles que le nombre de problèmes résolus, le temps passé sur chaque étape ou encore le type d'erreurs rencontrées. Les observations devaient permettre de caractériser l'utilisation effective du logiciel, ainsi que les interactions entre apprenants et observateurs. Pour cela, des fiches individuelles d'observation ont été construites pour étudier l'activité des élèves lors de la réalisation des différentes étapes du cycle AMBRE. Les différentes questions posées par les apprenants, les difficultés qui les amenaient à faire appel aux encadrants et le type d'aide qui leur était apporté étaient également notés. Par ailleurs, après la dernière séance un questionnaire et un entretien collectif étaient proposés afin de connaître l'avis subjectif des apprenants sur les logiciels, telles que les difficultés ressenties ou leur satisfaction, et afin de recueillir des données personnelles, telle que la

fréquence d'utilisation d'un ordinateur.

Principaux résultats

L'analyse des tests de résolution de problèmes (cf. Figure 5) montre une amélioration significative des performances après l'utilisation des logiciels ($F(3,192)=18.1$, $p<0.001$), on observe une amélioration des performances quel que soit le logiciel utilisé. Par ailleurs, l'analyse ne met pas en évidence d'effet du type de logiciel utilisé ($F(2,64)<1$, ns) ni d'interaction entre les variables logiciel et test ($F(6,192)=1.15$, $p=0.33$), ce qui signifie que les élèves qui ont utilisé AMBRE-add ne progressent pas plus que les élèves qui ont utilisé un logiciel contrôle.

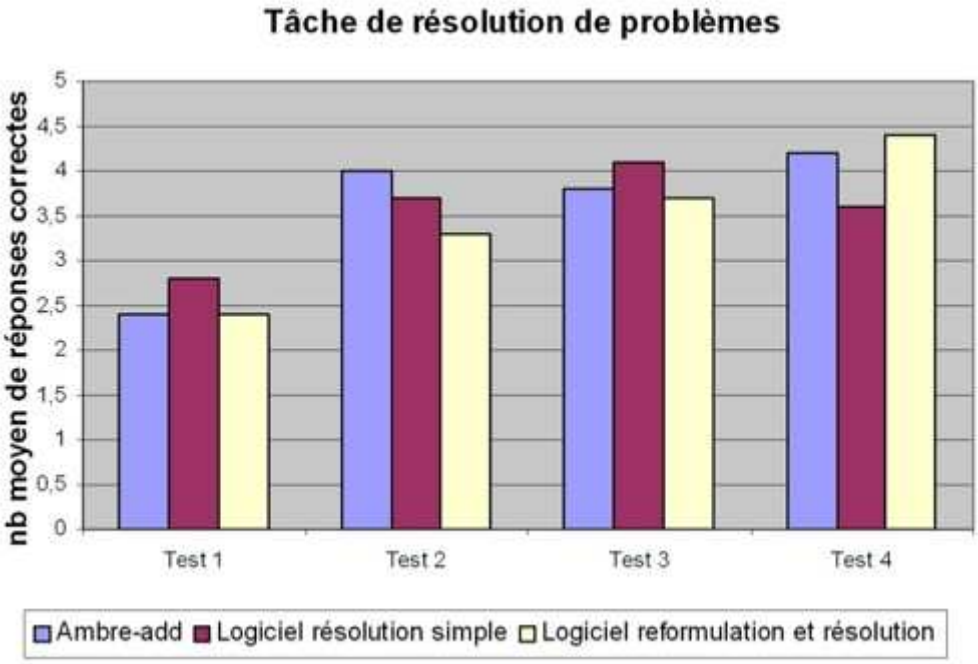

Figure 5 : Nombre moyen de problèmes correctement résolus dans chaque test en fonction du logiciel utilisé.

D'après ces résultats, le cycle AMBRE semble avoir un impact équivalent aux autres méthodes de résolutions (voir (Nogry, 2005) pour une présentation et une discussion plus détaillées des résultats).

Quelles conclusions tirer de ces résultats ? Faut-il remettre en cause l'impact du cycle AMBRE sur l'apprentissage ? Comment tirer partie de cette expérience pour la conception de l'EIAH ? Être capable de concevoir une nouvelle version du logiciel ayant un impact sur l'apprentissage plus important que d'autres systèmes passe par une analyse des causes de l'absence de différence entre les groupes ayant utilisé les trois systèmes. L'analyse des traces d'interaction et des observations réalisées durant l'utilisation des logiciels peut apporter diverses explications à ces résultats quantitatifs.

D'abord, les traces d'interaction récoltées par les différents logiciels montrent que les enfants résolvent moins de problèmes avec AMBRE-add qu'avec le logiciel résolution simple (neuf problèmes versus quatorze problèmes résolus en moyenne sur six séances) ; le nombre de problèmes résolus avec AMBRE-add durant chaque séance est relativement faible (moins de deux problèmes résolus par séance). L'observation de l'utilisation et l'analyse des questions posées au cours des séances montrent que les enfants qui utilisent AMBRE-add éprouvent beaucoup de difficultés à lire les instructions, à lire et à comprendre les messages d'aide et d'explication, ainsi qu'à identifier les raisons de leurs erreurs. De plus, certains enfants adoptaient un comportement passif face au logiciel (ils n'essayaient pas de corriger leurs erreurs), ou une recherche de la solution par essai-erreur. Les difficultés rencontrées semblent avoir pesé sur la motivation des élèves, ces différents facteurs ont pu de surcroît limiter la progression de ces

derniers. En outre, les questions posées durant l'utilisation des trois logiciels n'étaient pas de même nature et les réponses apportées par les encadrants dans les conditions contrôle ont pu biaiser la comparaison entre logiciels. Les utilisateurs de AMBRE-add étaient parfois bloqués dans une étape précise de la résolution ; les encadrants pouvaient alors répondre à leurs questions en reformulant les messages d'explication fournis par le système. En revanche, lorsque les utilisateurs du logiciel « résolution simple » proposaient une solution qui s'avérait fausse, le système leur demandait de recommencer, ce qui les conduisait souvent à une impasse. Ils demandaient alors aux encadrants de les aider, ce que ces derniers faisaient en leur expliquant comment résoudre le problème. Par ailleurs, l'utilisation des logiciels ayant lieu à l'école durant une séance de mathématiques, les encadrants apportaient aux apprenants les explications nécessaires, ce qui a pu constituer un biais pour l'expérience.

Ces résultats ont permis de mettre en évidence des problèmes d'utilisabilité et d'acceptabilité du système par des élèves de CE1 : AMBRE-add semble ne pas être tout à fait adapté aux caractéristiques des enfants de ce niveau, faibles lecteurs, pas toujours très autonomes ; le logiciel est pour certains aspects en contradiction avec les pratiques auxquelles les enfants étaient habitués (recherche collective, utilisation de dessins ou d'une barre de calcul pour trouver la solution). Cette évaluation nous a conduites à apporter différentes modifications à AMBRE-add. Les messages d'aide et de diagnostic ont été modifiés pour les rendre plus compréhensibles, un compagnon et un synthétiseur vocal ont été ajoutés au système afin d'assister la lecture sous une forme ludique. Par ailleurs, des activités plus simples sont développées dans l'EIAH dans le but de préparer les enfants à l'utilisation du logiciel ou de leur proposer une remédiation adaptée à leurs difficultés. Nous avons par ailleurs reconsidéré le public auquel AMBRE-add est destiné : il semble en effet être plus adapté à des enfants en classe de CE2.

Intérêts et limites des méthodes utilisées durant cette évaluation

Cette section propose une revue critique des apports et limites des différentes méthodes d'évaluation utilisées.

Notons d'abord que cette expérience s'est révélée très couteuse en temps de préparation (conception et implémentation des logiciels contrôles), de passation et d'analyse. Deux membres de l'équipe de conception ont été mobilisées une journée par semaine pendant six semaines pour encadrer les séances d'utilisation, une personne a également passé plusieurs journées à l'école pour les différents tests.

Méthode comparative. Comme nous l'avons présenté précédemment, les différents tests effectués montrent que AMBRE-add semble avoir un impact équivalent à celui des logiciels contrôle sur l'apprentissage visé. Cependant, le choix de la condition contrôle parait ici poser problème : les utilisateurs de AMBRE-add ont résolu moins de problèmes que les utilisateurs de la condition résolution simple. Par ailleurs, les questions posées et les aides fournies par les encadrants n'étaient pas de même nature dans les deux cas. Afin de préserver l'équivalence entre systèmes nous avions proposé des messages d'aide et d'explication proches dans les différents systèmes, mais cette expérience montre que ces messages n'étaient pas suffisants dans le cadre d'une résolution simple ; ils devaient être non seulement reformulés par les encadrants, mais aussi complétés par des informations proches de celles fournies dans l'étape de reformulation de AMBRE-add.

Compte tenu des différences observées concernant l'utilisation des trois logiciels, la portée des tests quantitatifs permettant de comparer les progrès des apprenants suivant le système utilisé est limitée.

En revanche, la caractérisation des différences d'activité réalisée avec chaque environnement grâce à des méthodes « on line » telles que les observations et les traces d'interactions se sont révélées très utiles pour informer la conception.

Observation de l'activité réalisée avec les logiciels. Comme nous l'avons vu, les informations récoltées durant les séances sur l'utilisation des différents logiciels et les questions posées par les apprenants se sont révélées d'une grande importance pour comprendre les résultats des analyses quantitatives.

Toutefois, contrairement à ce que nous avions prévu, il n'a pas été possible d'observer systématiquement et pendant une durée prolongée l'utilisation des logiciels par les élèves. En effet, durant les premières séances, les encadrants ont été très sollicités par les apprenants. Lors de l'analyse nous avons néanmoins pu récolter des informations sur l'utilisation du logiciel à travers la description des difficultés observées par les encadrants lorsqu'ils intervenaient. Certaines erreurs et certaines difficultés récurrentes qui faisaient obstacle à l'utilisation du logiciel par les enfants, ainsi que certains comportements d'essai-erreur, ont été identifiés.

Notons cependant que ces informations n'ont pas été notées systématiquement, et dépendaient de la subjectivité des observateurs. Aussi les phénomènes observés ne sont pas quantifiables. Les **traces d'interaction** récoltées lors de l'utilisation des trois systèmes sont un bon complément des observations. Dans cette étude, elles ont été utilisées pour quantifier le nombre de problèmes résolus par séance ou encore le temps passé sur chaque problème et sur chaque étape. Faciles à acquérir, ces données ont apporté des indications relatives à la portée générale des observations.

Observation en classe de CE2

Suite à l'expérience précédente qui a mis en évidence un défaut d'acceptabilité de AMBRE-add en classe de CE1, nous avons réalisé une évaluation auprès d'enfants en classe de CE2. À ce niveau (Nogry, 2005), les problèmes additifs à une étape tels que ceux présentés par AMBRE-add continuent d'être traités, les enfants de ce niveau ayant encore des difficultés à modéliser ces problèmes ; cependant, ils ont moins de difficultés de lecture que les élèves de CE1 et sont plus autonomes. L'objectif de cette étude était de vérifier l'utilisabilité et l'acceptabilité de AMBRE-add auprès d'élèves de CE2. Pour cela nous avons observé la manière dont ils l'utilisent afin de comparer l'utilisation effective du logiciel (l'usage réel) à l'utilisation prescrite par les concepteurs et d'identifier les difficultés rencontrées par ce public.

Méthode

Dans cette expérience, 21 élèves en classe de CE2 ont utilisé AMBRE-add individuellement dans la salle informatique de l'école, durant quatre séances de 45 minutes. Durant ces séances, nous avons observé l'utilisation effective du logiciel. Au cours de chaque séance, l'un des encadrants observait tour à tour les enfants ; il notait sur une fiche d'observation des informations relatives à la réalisation de chaque étape, la réaction de l'enfant par rapport à l'aide et au diagnostic, ainsi que les difficultés rencontrées ; au cours de l'ensemble des séances, tous les enfants étaient observés au moins une fois. À l'issue de la quatrième séance, chaque élève était interrogé sur la stratégie qu'il employait dans l'étape d'adaptation (étape centrale dans le cycle AMBRE). De plus, nous avons recueilli des traces d'interaction. Enfin, à l'issue de la quatrième séance, nous avons fait remplir aux enfants un questionnaire ; les questions portaient sur les difficultés perçues par les apprenants, la compréhension du vocabulaire utilisé dans l'interface et la satisfaction des utilisateurs.

Résultats

Conformément à nos attentes, les élèves de CE2 ont été très autonomes et n'ont pas eu de difficulté à lire les consignes et les messages d'aide et de diagnostic ou à mettre en œuvre des techniques opératoires. Dès la seconde séance, ils ne posaient quasiment plus de questions. De plus, ils ont pris en main le logiciel en une seule séance. L'analyse des traces montre en outre qu'ils résolvaient les problèmes plus rapidement que les élèves de CE1.

Nous avons par ailleurs observé que l'utilisation effective du logiciel par les CE2 était plus proche de l'utilisation prescrite que ne l'était l'utilisation du logiciel par les CE1. Nous avons par exemple noté que plus de la moitié des élèves disent utiliser la solution du problème-type choisi pour trouver la solution du problème à résoudre lorsqu'ils éprouvent des difficultés, ce qui est conforme au principe proposé par AMBRE-add. Les élèves de CE2 semblaient également moins souvent faire appel à des stratégies d'essai-

erreur. Cependant, cette observation a montré que certaines difficultés relatives à la navigation ou à la compréhension des messages dans l'étape de reformulation persistaient chez certains élèves. L'analyse de ces difficultés nous a conduites à concevoir un tutoriel mieux adapté aux apprenants, qui présente davantage le fonctionnement de la navigation dans le logiciel et facilite la compréhension du principe général du logiciel. Cette évaluation nous a par ailleurs encouragées à évaluer l'impact du logiciel sur l'apprentissage auprès d'élèves de CE2, AMBRE-add semblant être bien adapté à ce public.

Intérêts et limites de cette évaluation

Dans cette évaluation, nous avons limité le nombre de méthodes employées en nous centrant sur l'observation individuelle, l'analyse de traces d'interaction et de questionnaires. Ce dispositif a l'avantage de pouvoir être mis en place facilement, auprès d'un nombre de classes restreint, tout en apportant des résultats intéressants pour la conception : il permet de décrire la manière dont les élèves utilisent le logiciel en situation scolaire et les difficultés rencontrées par les apprenants. Il permet aussi de prendre en compte des variations individuelles. Cette forme d'évaluation donne des indications qui permettent de juger de l'utilisabilité et de l'acceptabilité du système auprès de ce public et dans ce contexte. Néanmoins, ce dispositif ne suffit pas pour quantifier finement les différents types de comportements observés, ou pour observer l'évolution de l'utilisation du logiciel par un apprenant au cours du temps. En outre, il n'apporte pas non plus d'information sur les progrès réalisés par l'apprenant grâce à l'utilisation du logiciel.

Évaluation comparative en classe de CE2

Les observations encourageantes réalisées avec des élèves de CE2 et l'intérêt que les enseignants portent à la démarche proposée par le logiciel et aux problèmes traités nous ont encouragées à tester l'impact de AMBRE-add sur l'apprentissage auprès de ce public. Pour cela nous avons réalisé une étude comparative conjuguée à des méthodes d'observation plus systématiques que dans les évaluations précédentes, et à une analyse des traces d'interaction. Suite aux limites mises en évidence dans la première évaluation comparative, nous avons choisi une seule condition contrôle qui consistait non pas à utiliser un logiciel présentant les mêmes problèmes, mais à réaliser une autre activité en classe avec l'enseignant ; différents tests étaient réalisés pour mesurer les progrès des apprenants. Cette condition contrôle avait simplement pour but de vérifier qu'AMBRE-add est bien source de progrès, et que la progression n'est pas due à un simple effet de répétition des tests présentés.

Méthode

23 élèves en classe de CE2 ont utilisé AMBRE-add en salle informatique durant quatre séances de 45 minutes. Durant cette évaluation, les enfants passaient d'abord un premier test (T1, cf. Figure 6) consistant à résoudre six problèmes additifs. Ensuite, dans chaque classe, la moitié des enfants utilisait AMBRE-add en salle informatique durant quatre séances, tandis que l'autre moitié restait en salle de classe avec l'enseignant pour réaliser une autre activité sans rapport avec la résolution de problèmes. À l'issue de ces quatre séances, les élèves passaient un second test (T2), constitué de six problèmes de même nature que dans le premier test. Ensuite, les deux groupes étaient inversés : les enfants n'ayant pas encore utilisé le logiciel utilisaient le logiciel pendant quatre séances, tandis que les autres restaient en classe (cf. Figure 6). À l'issue des ces quatre séances, un post-test était de nouveau proposé aux élèves (T3). Par ailleurs, durant chaque session, plusieurs apprenants étaient observés in(Baker et al., 2004)grille (proposée par (Baker *et al.*, 20(Lloyd & Loper, 1986) une méthode de (Lloyd & Loper, 1986)) permettant de quantifier différents comportements prédéfinis était également remplie en observant chaque élève trois fois par séance. Enfin, un questionnaire était proposé aux élèves après la dernière séance d'utilisation.

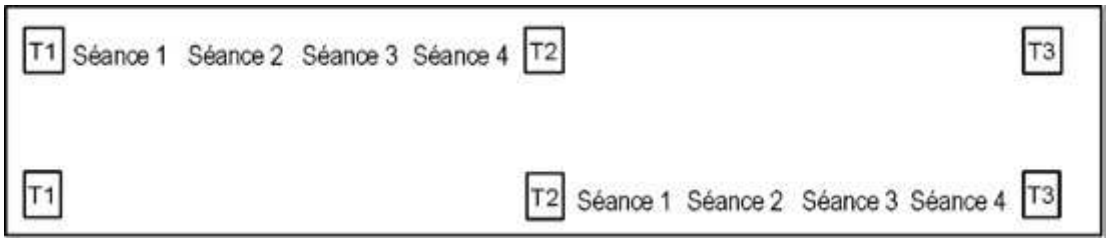

Figure 6 : Déroulement de l'expérience d'évaluation comparative en classe de CE2.

Résultats

Notons tout d'abord que l'analyse de l'activité des apprenants durant l'utilisation du logiciel a confirmé les observations faites dans l'expérience précédente, la quantification des comportements a par exemple montré que le nombre d'élèves adoptant un comportement d'essai-erreur était limité.

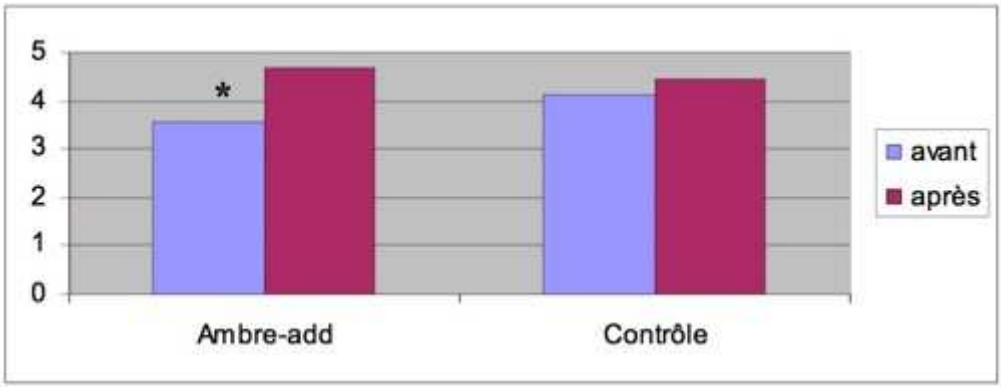

Figure 7 : Nombre moyen de problèmes correctement résolus par condition et par test.

Par ailleurs, la comparaison des progrès réalisés par chacun des groupes dans les deux parties du test montre que l'utilisation de AMBRE-add est associée à un progrès plus important que la réalisation de l'activité contrôle (cf. Figure 7). Nous avons regroupé les résultats obtenus dans chaque condition (AMBRE-add vs contrôle)[1] puis réalisé une analyse de variance sur les réponses correctes en considérant les variables *logiciel* (AMBRE-add vs condition contrôle) et *test* (avant vs après) comme des variables intra-sujets. Cette analyse met en évidence un effet du *test* ($F(1,20)=10.4$; $p<.005$) et une tendance à un effet d'interaction *logiciel*test* ($F(1,20)=3.7$; $p=.070$). Cette interaction permet de conclure que les progrès observés sont plus importants lorsqu'AMBRE-add est utilisé que dans la condition contrôle. AMBRE-add est donc source de progrès chez les enfants en classe de CE2. Cependant cette conclusion doit être nuancée car la valeur de l'effet d'interaction ($p=.070$) est légèrement supérieure au seuil de significativité admis ($p=.05$)[2]. À cela deux explications peuvent être avancées. D'une part le nombre de participants à l'expérience est assez faible, ce qui limite la puissance du test statistique. D'autre part, une analyse des résultats individuels montre que quelques élèves, en grande difficulté, ne progressent pas.

Pour aider ces élèves, une adaptation des activités proposées par le logiciel au niveau des élèves est souhaitable. L'équipe de conception envisageait déjà d'intégrer un modèle de l'apprenant au système. Cette évaluation confirme l'importance d'un tel modèle. En s'appuyant sur ce modèle, le système pourra par exemple adapter les problèmes proposés au niveau des élèves ou proposer des activités complémentaires destinées à remédier à des problèmes plus spécifiques rencontrés par les élèves en difficulté en les faisant travailler sur un point particulier.

Intérêts et limites de cette étude

Les résultats des deux études conduites en classes de CE2 sont assez satisfaisants du point de vue des concepteurs car ils montrent que l'EIAH AMBRE-add semble à la fois utilisable (bonne prise en main,

bonne compréhension), utile (il permet aux élèves de résoudre des problèmes et est source de progrès) et acceptable par des élèves en classe de CE2.

Sur le plan méthodologique, les observations individuelles et la grille d'observation quantitative utilisée ont permis de caractériser l'activité des apprenants durant les séances d'utilisation, ainsi que les difficultés rencontrées. En outre, l'impact du logiciel sur l'apprentissage a pu être mesuré grâce à la méthode comparative utilisée. Notons cependant que les résultats auraient été plus robustes si l'expérience avait été conduite dans un nombre de classes plus important.

La condition contrôle choisie, plus simple à mettre en œuvre, a été suffisante pour tester l'impact du logiciel lui-même sur l'apprentissage. Néanmoins, nous sommes bien conscientes que cette évaluation ne permet pas de savoir si cet impact vient du principe proposé par AMBRE-add, guider la résolution du problème suivant les étapes du cycle AMBRE, ou du fait de l'entrainement répété à la résolution de problèmes. Toutefois, il nous a paru nécessaire de vérifier l'impact du logiciel lui-même avant d'évaluer l'effet du principe sous-jacent à l'EIAH en comparant celui-ci à des systèmes proposant d'autres formes de résolution.

En résumé, au cours du cycle de conception itérative du logiciel AMBRE-add, nous avons réalisé quatre évaluations, une première en laboratoire, puis les trois suivantes à l'école, auprès d'élèves de CE1 et CE2. Celles-ci nous ont permis d'identifier des problèmes de différentes natures ce qui nous a amené à réaliser différentes modifications du logiciel, à y ajouter des fonctionnalités et à reconsidérer le public pour lequel AMBRE-add est le plus adapté. Ces différentes évaluations ont ainsi conduit progressivement à la conception d'une version du logiciel qui semble utilisable, utile et acceptable par une majorité d'élèves des classes de CE2 testées.

4. Quelle démarche d'évaluation mettre en œuvre dans le cadre d'une conception itérative ?

Quelles conclusions tirer des différentes évaluations réalisées dans le cadre de la conception itérative de AMBRE-add ? Que peut-on généraliser à partir de cette expérience ?

Comme nous l'avons vu, différentes techniques peuvent être mises en œuvre au cours de la conception itérative d'un EIAH afin d'évaluer son utilisabilité, son utilité pour l'apprentissage visé et son acceptabilité. Il n'existe pas de consensus sur la manière de combiner ces méthodes. Dans le cadre de la conception itérative de AMBRE-add, nous avons donc établi une démarche d'évaluation.

Comme le soulignent Littman et Soloway « there can be no doubt that evaluating ITS is costly, frustrating and time-consuming. » (Littman et Soloway, 1988). Cependant, toutes les techniques d'évaluation n'ont pas le même coût. La mise en œuvre de cette démarche nous a permis de mieux caractériser leurs intérêts et leurs limites. Nous avons ainsi pu identifier pour chacune des techniques le type de connaissances qui peuvent être acquises, mais aussi des indications sur le « coût » lié à leur mise en œuvre, c'est-à-dire le temps de préparation, de mise en place et d'analyse de chaque méthode et les compétences qu'elles demandent. Sur cette base, nous proposons une démarche d'évaluation en situation de conception itérative d'EIAH supports à des situations d'apprentissages individuels.

Leçons tirées la conception itérative de AMBRE-add

Dans cette section nous tirons les leçons de la démarche de conception et d'évaluation adoptée pour AMBRE-add en étudiant les intérêts, limites et champs d'application des différentes méthodes utilisées.

L'évaluation **par inspection,** réalisée par un expert ou en suivant des critères d'évaluation ergonomiques, est considérée comme l'évaluation la moins couteuse et la plus rapide à mettre en œuvre. De plus, elle peut être utilisée très tôt dans le processus de conception (Bastien et Scapin, 2001), (Bastien et Scapin,

2004). Notre expérience a confirmé les qualités de cette méthode. Toutefois, les critères ergonomiques permettent d'évaluer uniquement les aspects utilisabilité du système. De plus, ces évaluations ne permettent pas d'éliminer tous les problèmes d'utilisabilité, comme l'ont montré pour AMBRE-add les évaluations empiriques qui ont suivi les inspections. Ceci s'explique par le fait que la personne qui fait l'inspection n'est pas un utilisateur ordinaire placé dans une situation d'utilisation courante (Burkhardt et Sperandio, 2004), cette différence étant d'autant plus importante lorsque l'apprenant est jeune. Les critères ergonomiques ont en effet été développés sur la base d'études chez les adultes et il n'existe pas encore de critères largement diffusés pour évaluer l'utilisabilité de logiciels destinés à des enfants. L'observation du public cible utilisant le logiciel est donc nécessaire pour compléter les inspections afin d'améliorer l'utilisabilité du système.

L'observation d'utilisateurs en laboratoire complétée par un entretien a permis de mettre en évidence les plus gros problèmes d'utilisabilité et d'identifier une difficulté d'ordre mathématique non détectée par les experts pédagogues qui avaient inspecté préalablement le logiciel. En revanche, la durée de l'observation, limitée à une séance d'utilisation n'a pas permis de différencier les difficultés de prise en main du logiciel de celles qui persistent au delà de la première utilisation. En outre, les enfants volontaires pour participer n'étaient pas tout à fait représentatifs des enfants de leur niveau. Une sélection plus rigoureuse des utilisateurs observés améliorerait la qualité des conclusions tirées d'une telle étude.

Cette forme d'évaluation, très informative quant à l'utilisabilité du logiciel, ne permet pas de juger de son impact sur l'apprentissage ou de son acceptabilité. Aussi pour l'évaluation de AMBRE-add nous avons complété cette approche en réalisant une seconde évaluation empirique afin de tester l'utilité du logiciel (nous pensions évaluer l'acceptabilité dans une étape suivante). L'évaluation de l'impact du logiciel sur l'apprentissage consiste à mettre en évidence un développement des connaissances. Cet impact étant directement lié à la manière dont le logiciel contraint l'activité de l'apprenant lors de l'utilisation du logiciel, il nous paraît important de caractériser cette activité ou les différentes formes qu'elle peut prendre. Aussi nous avons conjugué une approche comparative à des techniques permettant de caractériser l'activité des apprenants au cours d'une expérience réalisée à l'école auprès de plusieurs classes de CE1.

L'**approche comparative** avait pour but de vérifier que si un progrès est observé, celui-ci est bien lié à l'utilisation du logiciel plutôt qu'à des activités réalisées par ailleurs. Comme nous l'avons décrit précédemment, la comparaison de AMBRE-add avec les logiciels contrôles choisis a été problématique. Les élèves utilisant l'un des logiciels contrôle ont résolu plus de problèmes et ont posé des questions de nature différente. L'activité réalisée avec les trois logiciels était donc différente tant du point de vue de la méthode de résolution de problèmes que des interactions avec les encadrants ; les situations étaient alors peu comparables, ce qui limite la portée des résultats quantitatifs. En outre, les conditions contrôle choisies avaient plus pour objectif de mettre en évidence un effet du principe pédagogique sous-jacent au logiciel que de montrer l'impact du logiciel lui-même sur l'apprentissage. Il était sans doute trop ambitieux de chercher à tester l'effet du cycle AMBRE sur l'apprentissage avant d'avoir testé l'utilité du logiciel. Nous reviendrons sur la question du choix de la condition contrôle dans la section suivante.

Les **techniques permettant de caractériser l'activité des apprenants** durant l'utilisation des logiciels (observations, recueil des questions posées, traces d'interaction) ont été les plus informatives pour la conception ; les analyses faites à partir des observations et des questions posées ont permis de déterminer le temps nécessaire pour que l'EIAH soit pris en main, d'identifier les difficultés persistantes rencontrées par les apprenants ou encore de caractériser certains aspects de l'utilisation effective du logiciel. Dans les études conduites en CE2, des analyses plus fines de l'activité des apprenants ont été réalisées en complétant la prise de note « à la volée » par des observations plus systématiques des apprenants. Celles-ci ont permis de comparer l'utilisation effective du logiciel avec l'utilisation prescrite et d'identifier certaines difficultés. Nous avons également quantifié certains comportements caractéristiques adoptés par les apprenants au cours de chaque séance afin de suivre leur évolution. Notons que si les méthodes

d'observation sont peu couteuses à mettre en place, leur analyse peu être longue et fastidieuse suivant le niveau de détail de l'activité auquel on souhaite accéder. De plus, ces méthodes posent des problèmes méthodologiques : à quel niveau de granularité analyser l'activité ? Quelle action est pertinente et porteuse de sens pour l'apprenant ? Certaines formes d'entretiens avec l'apprenant peuvent aider à interpréter l'observation. L'analyse des traces soulève le même type de questions. Utiliser les traces demande également une réflexion en amont pour déterminer quelles traces enregistrer et un travail d'analyse important pour les exploiter.

Dans la dernière expérience réalisée en CE2, la **combinaison d'une approche comparative et d'une analyse de l'activité** a permis de mettre en correspondance les progrès des apprenants avec leur activité durant les séances et les difficultés qu'ils ont rencontrées. Cette combinaison nous semble très prometteuse pour mieux comprendre les raisons d'un progrès ou d'une absence de progrès et proposer de nouvelles spécifications pour la conception.

L'analyse de l'activité instrumentée des apprenants et des usages est donc complémentaire de la mesure de l'impact d'un logiciel sur l'apprentissage ; de plus elle est très informative au cours de la conception itérative des EIAH. Elle peut faciliter l'identification des causes de certaines difficultés ressenties par les utilisateurs, et ce faisant peut faciliter la proposition de solutions alternatives par l'équipe de conception

L'intérêt de l'analyse de l'activité instrumentée et des usages d'un logiciel au cours de la conception a déjà été mis en évidence en ergonomie cognitive (voir par exemple (Béguin et Rabardel, 2000), (Decortis et al., 2001)). Néanmoins, dans le domaine des EIAH, cette analyse est souvent réservée à des logiciels déjà commercialisés ou à des sites web déjà accessibles aux apprenants (voir par exemple (Trouche, 2002), (De Vries et Tricot, 1998). Le domaine des sciences de l'éducation s'intéresse à cette question (voir par exemple (Bruillard et Baron, 2006)), mais étudie plutôt la manière dont les pratiques s'institutionnalisent sur des échelles de temps beaucoup plus longues (de l'ordre de plusieurs années). L'analyse des activités instrumentées pour la conception des EIAH n'est prise en compte que depuis peu de temps (voir (Hautecouverture et al., 2004), (Cottier et Choquet, 2005)) et est pour l'instant peu répandue.

Ces différentes considérations sont reprises dans la section suivante pour présenter la démarche d'évaluation que nous proposons.

Proposition d'une démarche d'évaluation en situation de conception itérative

Dans cette section nous proposons des recommandations, fondées sur notre expérience d'évaluation et enrichies d'apports théoriques, destinées à orienter le choix d'une démarche d'évaluation à appliquer en conception. Ces recommandations sont suivies par la proposition d'une démarche qui nous paraît bien adaptée pour évaluer, au cours d'une conception itérative, des logiciels supports aux situations d'apprentissage destinés à être utilisés en milieu scolaire.

Rappelons tout d'abord que la conception de dispositifs techniques se caractérise par une problématique de départ large et peu circonscrite, puis par une réduction progressive des possibilités de choix : l'équipe de conception dispose d'abord d'un grand nombre de degrés de liberté, puis ceux-ci se réduisent et les choix deviennent de plus en plus irréversibles (De Terssac, 1996). La conception des EIAH n'échappe pas à cette structuration, aussi le choix de l'ordre des évaluations réalisées et la manière dont les résultats de ces évaluations sont intégrés à la conception ont une réelle influence sur le produit final de la conception.

Trois dimensions doivent être évaluées : l'utilisabilité, l'utilité du système et son acceptabilité. D'après le modèle classique de Nielsen, l'utilisabilité est considérée comme une pré-condition à l'utilité du logiciel, l'utilité étant elle-même une pré-condition à l'acceptabilité du système (Nielsen, 1993). Suivant ce modèle, une démarche qui consiste à évaluer l'utilisabilité, puis l'utilité et enfin l'acceptabilité s'impose. Cependant, supposer que l'utilité est nécessaire pour que le logiciel soit acceptable ne signifie pas qu'elle

soit suffisante pour assurer l'acceptabilité du logiciel (Dillon et Morris, 1996). Que faire alors si, à un stade avancé de la conception, l'évaluation d'un logiciel utilisable et utile met en évidence une faible acceptabilité ? Pour éviter une telle situation, (Tricot et al., 2003) proposent que l'évaluation de l'utilisabilité soit réalisée conjointement à l'évaluation des autres dimensions. L'expérience d'évaluation en conception de l'EIAH AMBRE-add va dans ce sens : s'il n'était pas prévu d'évaluer l'acceptabilité au cours de l'expérience de CE1, nous nous sommes rendues compte que cette dimension était importante à prendre en compte dès ce moment, les observations réalisées lors de l'utilisation en milieu scolaire permettant d'accéder à différentes indications de l'acceptabilité du système.

Le critère du coût de mise en place d'une évaluation peut également être pris en compte dans le choix des méthodes à utiliser. Ce critère doit cependant être rapporté au type de connaissances issues de cette évaluation utilisables pour la conception. Par exemple, même si la mise en place d'une évaluation en classe pendant plusieurs séances est couteuse, elle est indispensable pour évaluer à la fois l'utilisabilité et l'acceptabilité d'un EIAH destiné à être utilisé en situation scolaire.

Sur la base de ces considérations, la démarche que nous proposons combine différentes techniques afin de prendre en compte aussi précocement que possible l'utilisabilité, l'utilité du système et son acceptabilité au cours du cycle de conception itérative. Cette démarche est plus particulièrement destinée à la conception d'EIAH supports aux situations d'apprentissage, qui contraignent l'activité de l'apprenant afin de le conduire à acquérir de nouvelles compétences ou connaissances et notamment à ceux qui sont destinés à être utilisés individuellement, au sein d'une institution à visée éducative.

Après avoir défini les spécifications du logiciel et réalisé une première maquette, une première étape consiste à effectuer une évaluation par inspection. Cette méthode présente l'avantage de pouvoir être mise en œuvre très précocement sans avoir besoin d'implémenter toutes les fonctionnalités du logiciel et peut permettre de réorienter la conception à moindre coût. Cette analyse peut porter sur l'utilisabilité du système en utilisant les nombreux critères existants, sur des aspects pédagogiques ou didactiques en faisant appel à des experts de ces domaines, ou encore sur l'acceptabilité *a priori* du système. Sur cette base, de nouvelles spécifications peuvent être produites jusqu'à l'implémentation d'un système réel techniquement stable.

Des évaluations empiriques de ce système, conduites en laboratoire, sont alors nécessaires, particulièrement si les utilisateurs visés sont des enfants pour qui peu de recommandations ergonomiques existent. L'observation de l'utilisation du logiciel par des utilisateurs représentatifs et un entretien individuel permet de mettre en évidence les principaux défauts d'utilisabilité et éventuellement des difficultés d'ordre didactique, pour aboutir à de nouvelles spécifications et à une modification du prototype. Une utilisation répétée du logiciel pendant plusieurs séances peut s'avérer utile pour dépasser la phase de prise en main et observer la manière dont les utilisateurs utilisent le logiciel à plus long terme. Cette évaluation peut être réitérée plusieurs fois si la première évaluation conduit à des modifications importantes du logiciel.

L'étape suivante consiste à observer et à analyser les activités des apprenants lors de l'utilisation du logiciel en situation scolaire. En effet, les usages d'un dispositif technique tel qu'un EIAH ne dépendent pas uniquement de l'outil lui-même mais émergent également en fonction des schèmes des apprenants (Rabardel, 1995), et en interaction avec les pratiques déjà existantes et les caractéristiques de l'institution dans laquelle elle est intégrée (Hussenot, 2005). L'analyse de l'activité peut apporter des informations sur l'utilisabilité (par exemple le temps nécessaire pour que l'EIAH soit pris en main ou les difficultés persistantes), l'acceptabilité (par exemple l'évolution de la motivation ou la compatibilité avec l'organisation du temps et des lieux) ou sur les différents types d'usage du système et leur évolution au fur et à mesure des séances. Elle peut également permettre de prendre en compte les variations interindividuelles dans l'utilisation du logiciel. L'analyse des activités instrumentées, et plus largement des usages, peut être réalisée à partir d'observations, de l'analyse des traces d'interaction, et d'entretiens permettant de prendre en compte le point de vue subjectif du sujet, sur son activité (ses motivations, le

sens qu'il donne à son activité). Cette analyse peut conduire à des modifications du système et éventuellement à la constitution d'un scénario pédagogique pour faciliter la prise en main du logiciel et orienter son utilisation (Hautecouverture et al., 2004), (Trouche, 2002). Elle pourra être répétée au cours du cycle de conception, afin d'affiner la conception du produit à travers des observations d'usage régulières (Hautecouverture et al., 2004).

De notre point de vue, cette analyse de l'activité doit être combinée dès que possible à une analyse du développement des connaissances ou compétences du sujet au cours de l'activité. En effet, une mise en correspondance du développement des connaissances et compétences individuelles avec les activités que l'apprenant réalise au sein de l'environnement d'apprentissage peut permettre de mieux comprendre quel type d'activité est associé à un développement. Cette analyse peut être basée sur l'évolution des réponses des sujets produites au cours du temps dans l'interaction avec l'EIAH, ou par des tests adaptés au domaine étudié donnés avant, pendant et après l'utilisation du système.

Le type d'analyse de l'activité et le type de tests destinés à mesurer l'apprentissage doivent être choisis en fonction du domaine d'application de l'EIAH ou des connaissances en jeu et adaptés à l'avancement de la conception. Dans les phases initiales, une analyse inductive (non guidée par des hypothèses) faite à partir d'un dispositif d'observation et d'analyse peu contraignant, peut permettre d'observer des phénomènes importants inattendus, et d'anticiper des modifications du logiciel. à ce moment là, les tests destinés à mesurer l'apprentissage peuvent être utilisés à titre indicatif, pour mettre en évidence des tendances plutôt que pour obtenir des mesures très précises et des différences statistiquement valides.

Pour augmenter la validité des méthodes destinées à mesurer le développement des connaissances ou des compétences, une condition contrôle peut être mise en place. Le choix de cette condition pose des difficultés, nous proposons ici quelques pistes pour orienter ce choix. Dans un premier temps, une comparaison de l'EIAH à une situation sans intervention cf. (Ainsworth et al., 1998) peut apporter des informations intéressantes concernant l'utilité du système. En effet, sous réserve que les tests proposés soient adaptés, si un progrès est observé, cette comparaison pourra permettre de vérifier que l'effet obtenu n'est pas du à une activité réalisée par ailleurs (par exemple avec l'enseignant). Une analyse comparative peut également être mise en œuvre pour atteindre d'autres objectifs, par exemple pour tester l'effet d'une fonction particulière du logiciel. L'EIAH est alors comparé à une autre version du logiciel (Aleven et al., 1999), (Aleven et Koedinger, 2002), (Luckin et al., 2001). Dans cette condition, on s'intéresse moins à l'utilité du système lui-même qu'à l'utilité de l'une des fonctionnalités. Mais comme on l'a vu dans notre propre expérience, la définition de l'autre version de l'EIAH est délicate, et il convient d'être attentif à l'activité que cette autre version suscite pour pouvoir interpréter les résultats de la comparaison.

Enfin, notons que les approches comparatives sont souvent utilisées pour comparer le logiciel conçu à un logiciel existant ou à l'efficacité d'un enseignement classique dispensé par un enseignant cf. (Shute et Glaser, 1990), (Koedinger et al., 1997), (Meyer et al., 1999). De notre point de vue, la comparaison à ces conditions contrôle relèvent plus de l'évaluation en fin de conception ou *a posteriori*, que de l'évaluation en conception. Ces évaluations ont alors pour but de mettre en évidence la supériorité de l'EIAH (il conduit à apprendre plus vite, mieux, ou à l'acquisition de connaissances ou compétences complémentaires), par rapport aux produits ou aux méthodes pédagogiques existants pour enseigner des contenus similaires.

Pour finir, précisons que cette démarche doit bien sûre être adaptée en fonction des objectifs et des contraintes de l'équipe de conception. Elle nous semble applicable à différents domaines, à condition d'adapter les méthodes mises en œuvre pour évaluer l'apprentissage, celles-ci étant totalement dépendantes des connaissances ou compétences à acquérir par l'intermédiaire de l'EIAH.

Conclusion

Devant la multiplicité des aspects à évaluer et des méthodes proposées, quelle démarche d'évaluation mettre en œuvre en conception ?

Nous avons été confrontées à cette question lors de la conception de l'EIAH AMBRE-add, et, pour y répondre, nous nous sommes appuyées sur les études existantes afin de mettre en place plusieurs évaluations combinant différentes méthodes afin d'évaluer l'utilisabilité et l'utilité du logiciel. Dans cet article, nous avons décrit les méthodes existantes, puis nous avons procédé à une revue critique des différentes évaluations conduites lors de la conception de AMBRE-add. Nous avons ainsi mis en évidence les apports à la conception et les limites des différentes méthodes utilisées. Cette analyse critique nous a conduites à mettre en évidence l'intérêt d'analyser les activités des apprenants lors de l'utilisation du système très tôt au cours de la conception et à proposer une démarche d'évaluation précoce en situation de conception. Ainsi, nous proposons d'évaluer d'abord l'utilisabilité de l'EIAH en utilisant les méthodes décrites en IHM, puis d'analyser les activités en situation réelle d'utilisation pendant plusieurs séances afin de compléter l'évaluation de l'utilisabilité, d'identifier les différentes formes d'usage du logiciel et de récolter des données sur l'acceptabilité. Cette analyse des usages peut être combinée à une évaluation de l'impact de l'EIAH sur l'apprentissage afin d'identifier les formes d'usages associées à un apprentissage. Chacune de ces évaluations pourra conduire à produire des spécifications pour implémenter une nouvelle version du logiciel dans le cadre d'une conception itérative.

Remerciements

Cette recherche a été soutenue par le programme pluridisciplinaire STIC-SHS « Société de l'Information » du CNRS. Nous remercions également les informaticiens qui ont développé le logiciel, les enseignants qui ont participé à cette étude ainsi que les trois relecteurs de cet article pour leurs remarques constructives.

BIBLIOGRAPHIE

AINSWORTH S. E., WOOD D., O'MALLEY C. (1998). There is more than one way to solve a problem: Evaluating a learning environment that supports the development of children's multiplication skills, *Learning and Instruction*, Vol 8 n°2, 141-157.

ALEVEN V., KOEDINGER K. R. (2002). An Effective Meta-cognitive Strategy: Learning by Doing and Explaining with a Computer-Based Cognitive Tutor. *Cognitive Science*, Vol 26 n°2, 147-179.

ALEVEN V., KOEDINGER K. R., CROSS K. (1999). Tutoring answer explanation fosters learning with understanding. Proceedings of AIED-99, Amsterdam, Netherland, IOS Press, 199-206.

BAKER R.S., CORBETT A.T., KOEDINGER K.R. (2004). Detecting Student Misuse of Intelligent Tutoring Systems. *Proceedings of ITS'2004*, Maceio, Brasil, Springer, 531-540.

BASTIEN C., SCAPIN D. (2004). La conception de logiciels interactifs centrés sur l'utilisateur : étapes et méthodes. *In Ergonomie*, Falzon P. (ed), Paris : PUF, 451-462.

BASTIEN J. M. C., SCAPIN D. L. (2001). Évaluation des systèmes d'information et Critères Ergonomiques. In C. Kolski (Ed.), *Systèmes d'information et interactions homme-machine. Environnement évolués et évaluation de l'IHM. Interaction homme-machine pour les SI*. Paris : Hermes, Vol. 2, p. 53-79.

BASTIEN C., SCAPIN D. (1993). Critères ergonomiques pour l'évaluation des interfaces utilisateurs, *RT*

n°156, INRIA.

BARFURTH M.A., BASQUE J., CHOMIENNE M., WINER L.R. (1994). Les instruments de collecte de données de recherche qualitative dans des environnements pédagogiques informatisés. *In Apprendre dans des environnements pédagogiques informatisés*, Bordeleau P. (ed), Editions Logiques, 485-548.

BéGUIN P., RABARDEL P. (2000). Concevoir des activités instrumentées. *Revue d'Intelligence Artificielle*, Vol 14, 35-54.

BURKHARDT J.-M., SPERANDIO J.-C. (2004). Ergonomie et conception informatique. *In Ergonomie*, Falzon P. (ed), Paris : PUF, 437-450.

BRUILLARD, E., BARON, G.-L. (2006). Usages en milieu scolaire : caractérisation, observation et évaluation. *In Environnements informatiques pour l'apprentissage humain*, Grandbastien M., Labat J.-M. (eds), chapitre 12, Hermès-Lavoisier.

CAVERNI J.-P. (1988). La verbalisation comme source d'observables pour l'étude du fonctionnement cognitif. *In Psychologie cognitive, modèles et méthodes*. Caverni J.P., Bastien C., Mendelsohn P., Tiberghien G. (eds), Presse Universitaire de Grenoble, 253- 273.

COTTIER P., CHOQUET C. (2005). De l'usager construit à l'usager participant. *Actes de EIAH'2005*, Montpellier, France, 449-454.

DAVIS, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: Theory and results. (Doctoral dissertation, Sloan School of Management, Massachusetts Institute of Technology).

DAVIS, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), p. 319-339.

DECORTIS F., DAELE L., POLAZZI L., RIZZO A., SAUDELLI B. (2001). Nouveaux instruments actifs et activités narratives. *Revue d'Interaction Homme-Machine*, Vol 2 n°2, 1-30.

DELOZANNE, E. (2006), Interfaces en EIAH, *In Environnements informatiques pour l'apprentissage humain*, Grandbastien M., Labat J.-M. (eds), chapitre 10, Hermès-Lavoisier, 223-244

DE TERSSAC, G. (1996). Le travail de conception : de quoi parle-t-on ? *In Coopération et conception*. De Terssac G., Friedberg E. (eds), Octares, 1-22.

DE VRIES E. (2001). Les logiciels d'apprentissage, panoplie ou éventail ? *Revue Française de Pédagogie*, Vol 137, 105-116.

DE VRIES E., TRICOT A. (1998). Évaluer l'utilisation d'hypermédias : intérêts et limites des variables de performance. *Hypertextes et Hypermédias*, n° hors série, 175-190.

DILLON A., MORRIS M. (1996). User acceptance of information technology : theories and models. *Annual Review of Information Science and Technology*, p.3-32.

DUBOURG X., TEUTSCH, P. (1997). Interface Design Issues in Interactive Learning Environments. *Proceedings of IFIP WG 3.3 Working Conference: Human-Computer Interaction and Educational Tools*, Sozopol, Bulgarie.

FARENC N. (1997). ERGOVAL : une méthode de structuration des règles ergonomiques permettant l'évaluation automatique d'interfaces graphiques. Thèse de doctorat de l'Université Toulouse I.

FASSE B.B., KOLODNER J.L. (2000). Evaluating Classroom Practices Using Qualitative Research Methods: Defining and Refining the Process. *Proceedings of International Conference of the Learning*

Sciences, 93-198.

GAGNé R.M., BRIGGS L.J., WAGER W.W. (1988). *Principles of instructional design*. New York, Holt, Renhart, Winston (eds).

GREENO J.G., RILEY M.S. (1987). Processes and development of understanding. *In Metacognition, motivation and understanding*, Weinert F.E., Kluwe R.H. (eds), 289-313.

GUIN D. (1991). La notion d'opérateur dans une modélisation cognitive de la compréhension des problèmes additifs. Math. Inf. Sci. Hum. Vol. 113, 5-33.

GUIN-DUCLOSSON N., JEAN-DAUBIAS S., NOGRY S. (2002). The AMBRE ILE: How to Use Case-Based Reasoning to Teach Methods. *Proceedings of ITS'2002*, Biarritz, France, Springer, 782-791.

HAUTECOUVERTURE J.-C., GRéGORI N., PAQUELIN D., CHAROY F., GODART C., PATTEN M., FAUGERAS I. (2004). Analyse d'usage d'une plate-forme de coopération : Usage et développement logiciel. *Cahiers romans de sciences cognitives*, Vol 1 n°3, 45-77.

HOECKER D., ELIAS G. (1986). User evaluation of the LISP intelligent tutoring system. *Proceedings of the human factors society*, Vol 32 n°3, 313-324.

Hû O., TRIGANO P. (1998). Propositions de critères d'évaluation de l'interface homme-machine des logiciels multimédias pédagogiques. *Proceedings of IHM'98*, Nantes.

HUSSENOT, A. (2005). Trajectoires d'usage d'une solution TIC : traduction, "enaction" et appropriation. *3ème GDR TIC et Société*, Paris.

JEAN S. (2000). PÉPITE : un système d'assistance au diagnostic de compétences, Thèse de doctorat de l'Université du Maine.

JEAN-DAUBIAS S. (2003). Vers une définition des spécificités des EIAH dédiés à l'évaluation pour l'application de recommandations ergonomiques. *Revue d'Interaction Homme-Machine*, Vol 4 - n°1, 2003, 535-538.

JEAN-DAUBIAS S. (2004). De l'intégration de chercheurs, d'experts, d'enseignants et d'apprenants à la conception d'EIAH, *Actes de TICE 2004*, Compiègne, 290-297.

KOEDINGER, K. R., ANDERSON, J. R., HADLEY, W. H., & MARK, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, Vol 8, 30-43.

KOLSKI C. (2001), *Analyse et conception de l'IE : interaction homme-machine par les SP1*, Hermès, Paris.

LEWIS C., POLSON P.G., WHARTON C., RIEMAN J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. *Proceedings of CHI'90: Human Factors in Computing Systems*, ACM: New York, 235-242.

LITMANN D., SOLOWAY E., 1988. Evaluating ITSs: The cognitive science perspective. *In Foundations of Intelligent Tutoring Systems*, Polson M., Richardson J. J. (eds), Hillsdale, NJ: LEA.

LLOYD J.W., LOPER A.B. (1986). Measurement and Evaluation of Task-Related Learning Behavior: Attention to Task and Metacognition. *School Psychology Review*, Vol.15 n°3, 336-345.

LUCKIN R., PLOWMAN L., LAURILLARD D., STRATFOLD M., TAYLOR J., (2001). Narrative evolution: learning from students' talk about species variation. *International Journal of AIED*, Vol 12, 100-123.

MARTHE P. (1982). Problèmes de type additif et appropriation par l'élève des groupes additifs Z et D entiers relatifs et décimaux relatifs. Thèse de doctorat, EHESS, Paris.

MEYER T. N., MILLER T. M., STEUCK K., KRETSCHMER M. (1999). A multi-year large-scale field study of a learner controlled intelligent tutoring system. *Proceedings of AIEd'99*, Lajoie S., Vivet M. (eds), 191-198.

NANARD J., NANARD M. (1998). La conception d'hypermédias, *In Les hypermédias, approches cognitives et ergonomiques*, Tricot A., Rouet J.-F. (eds), Paris, Hermès, 15-34.

NIELSEN J. (1993). *Usability Engineering*, Academic Press.

NIELSEN J., LANDAUER T.K. (1993). A mathematical model of the finding of usability problems. *Proceedings of ACM INTERCHI'93 Conference*, Amsterdam, The Netherlands, 206-213.

NOGRY S. (2005). Faciliter l'apprentissage à partir d'exemples en situation de résolution de problèmes - Application au projet AMBRE. Thèse de doctorat de l'Université Lumière Lyon 2.

NOGRY S., JEAN-DAUBIAS S., OLLAGNIER-BELDAME M. (2004). Évaluation des EIAH, une nécessaire diversité des méthodes. *Actes de TICE'2004*, Compiègne, France, 265-271.

NOGRY S., JEAN-DAUBIAS S., DUCLOSSON N. (2004). ITS evaluation in classroom: the case of the AMBRE-awp. *Proceedings of ITS 2004*, Maceio, Brasil, springer, 511-520.

RABARDEL P. (1995). *Les hommes et les technologies, approche cognitive des instruments contemporains*. Paris, Colin.

RILEY M.S., GREENO J.G., HELLER J.I. (1983). Development of children's problem-solving ability in arithmetic. *In The development of mathematical thinking*, Ginsburg H.P. (ed). New-York: Academic Press.

ROUET J.-F., PASSERAULT J.-M. (1999). Analyzing learner hypermedia interaction: An overview of online methods. *Instructional Science*, Vol 27, 201–219.

SCHNEIDERMAN B. (1992). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Reading, MA: Addison-Wesley.

SENACH B. (1993). L'évaluation ergonomique des interfaces homme – machine. *In L'ergonomie dans la conception des projets informatiques.* Sperandio J.-C. (ed). Octares : 69-122.

SHUTE V.J., REGIAN J. W. (1993). Principles for Evaluating Intelligent Tutoring Systems. *Journal of Artificial Intelligence and Education*, Vol 4 n° 2/3, 245-271.

SHUTE V. J., GLASER R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments*, Vol 1, 51-77.

SQUIRES D., PREECE J. (1999). Predicting quality in educational software: Evaluating for learning, usability, and the synergy between them. *Interacting with Computer*, Vol 11 n°5, 467-483.

TCHOUNIKINE P. (2002). Pour une ingénierie des environnement informatiques pour l'apprentissage humain. *Revue Information-Interaction-Intelligence*, Vol 2 n°1, 59-95.

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. *Actes de EIAH 2003*, Strasbourg, France, Desmoulins C., Marquet P., Bouhineau D. (eds), 391-402.

TRICOT A., LAFONTAINE J. (2002). Une méthode pour évaluer conjointement l'utilisation un outil

multimédia et l'apprentissage réalisé avec celui-ci. *Le Français dans le Monde*, 41-52.

TROUCHE L. (2002). Genèses instrumentales, aspects individuels et collectifs. *In Calculatrices symboliques - Transformer un outil en un instrument du travail mathématique : un problème didactique*, Guin N., Trouche L. (eds), la pensée sauvage édition, 243-275.

VERGNAUD G. (1982). A classification of cognitive tasks and operations of the thought involved in addition and substraction problems. *In Addition and substraction: A cognitive perspective*, Carpenter P.T., Moser, J.M., Romberg, T.A. (eds), Hillsdale, Erlbaum, 39-58.

VIVET M. (1996). Evaluating Educational Technologies: Evaluation of Teaching Material Versus Evaluation of Learning? *Proceedings of* CALISCE, 37-38.

---

[1] Le fait de contrebalancer l'ordre de confrontation au logiciel peut être considéré comme une précaution expérimentale plutôt qu'une variable, ce qui autorise à regrouper les résultats obtenus par les deux groupes pour chaque condition. Notons que dans cette analyse, les scores obtenus au second test sont pris en compte deux fois.

[2] Habituellement, on considère qu'une différence entre deux groupes n'est pas due au hasard lorsque la probabilité est inférieure à 5% ($p<.05$).

À propos des auteurs

Sandra NOGRY est post-doctorante, rattachée au LIRIS (Laboratoire d'InfoRmatique en Images et Systèmes d'information, UMR 5205) au sein de l'équipe SILEX (Supporting Interaction and Learning by Experience) et chercheuse associée à l'équipe Compréhension, Raisonnement et Acquisition de Connaissances (laboratoire paragraphe, Université Paris 8). Ses recherches portent sur l'évaluation des EIAH, la conception, l'apprentissage et la résolution de problèmes par analogie.

Adresse : Université de Lyon, Lyon, F-69003, France ; université Lyon 1 ; LIRIS - UMR5205, Villeurbanne, F-69622, France

Courriel : sandra.nogry@gmail.com

Toile : http://liris.cnrs.fr/sandra.nogry/

Stéphanie JEAN-DAUBIAS est maître de conférences en informatique à l'Université Claude Bernard Lyon 1. Elle est rattachée au LIRIS (Laboratoire d'InfoRmatique en Images et Systèmes d'information, UMR 5205) au sein de l'équipe SILEX (Supporting Interaction and Learning by Experience). Ses recherches, mises en œuvre au sein des projets PERLEA et AMBRE, portent sur la personnalisation des apprentissages, en cherchant à proposer aux enseignants des outils les aidant à effectuer cette personnalisation.

Adresse : Université de Lyon, Lyon, F-69003, France ; université Lyon 1, Lyon, F-69003, France ; LIRIS - UMR5205, Villeurbanne, F-69622, France

Courriel : Stephanie.Jean-Daubias@liris.univ-lyon1.fr

Toile : http://liris.cnrs.fr/stephanie.jean-daubias/

Nathalie Guin-Duclosson est maître de conférences en informatique à l'Université Lyon 1. Elle est rattachée au LIRIS (Laboratoire d'InfoRmatique en Images et Systèmes d'information, UMR 5205) au sein de l'équipe SILEX (Supporting Interaction and Learning by Experience). Ses recherches portent sur

les systèmes à base de connaissances pour les EIAH. Ses thèmes d'intérêt actuels sont liés à la personnalisation des EIAH : interprétation des traces d'activités, profils d'apprenants, génération d'activités adaptées. L'ensemble des ces thématiques de recherches sont mises en œuvre au sein du projet AMBRE.

Adresse : Université de Lyon, Lyon, F-69003, France ; université Lyon 1, Lyon, F-69003, France ; LIRIS - UMR5205, Villeurbanne, F-69622, France

Courriel : Nathalie.Guin@liris.univ-lyon1.fr

Toile : http://liris.cnrs.fr/nathalie.guin

Référence de l´article :

Sandra Nogry, Stéphanie Jean-Daubias, Nathalie Guin-Duclosson, Leçons tirées de la conception de AMBRE-add. Comment combiner les objectifs et méthodes d'évaluation pour la conception itérative des EIAH ?, *Revue STICEF*, Volume 13, 2006, ISSN : 1764-7223, mis en ligne le 14/03/2007, http://sticef.org

© Revue Sciences et Technologies de l´Information et de la Communication pour l´Éducation et la Formation, 2006