



Une présentation des principales méthodes d'évaluation des EIAH en psychologie cognitive

Eric JAMET [CRPCC, Rennes II]

■ **RÉSUMÉ** : L'objectif de cet article est de présenter une série de méthodes d'évaluation des EIAH utilisée en psychologie cognitive. Dans ce champ disciplinaire, l'utilisation de la méthode expérimentale et le recueil d'indices comportementaux variés permettent d'obtenir des indications objectives de l'efficacité d'un EIAH en termes d'apprentissage. Après avoir rappelé quelques éléments théoriques liés à l'apprentissage, on présentera les outils d'évaluation les plus utilisés dans ce champ en distinguant les mesures réalisées pendant ou après l'apprentissage.

■ **MOTS CLÉS** : EIAH, apprentissage, évaluation, Psychologie cognitive

■ **ABSTRACT** : The aim of this article is to present a series of methods used in cognitive psychology to evaluate CSLE. In this field of study, the use of experimental methods and the gathering of various behavioural indices make it possible to obtain objective indicators of the effectiveness for learning of a CEHL. After theoretical elements relative to learning, we will present the evaluation tools that are most frequently used in this field while distinguishing between measurements taken before or after learning.

■ **KEYWORDS** : Computer supported learning environment, evaluation, cognitive psychology.

1. Introduction

2. Quelques rappels théoriques sur l'apprentissage

3. Les méthodes d'évaluations

4. Conclusion

BIBLIOGRAPHIE

1. Introduction

Cet article a pour objectif de présenter quelques méthodes utilisées en Psychologie Cognitive pour évaluer l'apprentissage dans un environnement informatique en les illustrant par des études où elles ont été employées. Cette présentation ne prétend donc à aucun caractère d'exhaustivité. Elle est limitée volontairement à une situation précise d'apprentissage par instruction et à un champ disciplinaire qu'elle ne couvre pas entièrement.

Par environnement d'apprentissage informatique, nous retiendrons ici tout environnement constitué d'un individu (ou plus) en situation d'apprentissage d'un document électronique. Ces environnements peuvent donc être variés : individu isolé ou situation collaborative, document simple (une illustration) ou complexe (un cédérom), apprentissage explicite ou implicite (par exemple via une tâche de recherche d'informations).

Il s'agit donc d'une définition des EIAH restreinte à une situation particulière d'apprentissage de documents. Toutefois, l'influence prépondérante de la qualité des documents utilisés, tant en termes de contenus que de mise en forme matérielle, sur le processus d'apprentissage est suffisante pour que les méthodes présentées dans le cadre de cet article permettent d'apporter des éléments de réponses à de nombreuses questions de recherches dépassant le cadre de la psychologie cognitive.

Concernant le terme d'apprentissage, nous l'utiliserons dans son sens le plus large. Nous considérons donc que ce terme renvoie à la fois à des processus d'acquisition de l'information mais aussi à des processus de construction actifs de savoirs par l'individu. Les processus liés à la compréhension (transfert de connaissances à des situations nouvelles, élaboration d'inférences) sont donc inclus dans le processus d'apprentissage. L'objectif de ce processus n'est donc pas seulement d'ajouter des informations en mémoire mais aussi de construire des représentations mentales cohérentes à l'aide des EIAH (Mayer, 2001).

Concernant maintenant l'évaluation elle-même, nous nous placerons clairement du côté de l'utilisateur. Il s'agira donc d'évaluer l'efficacité d'un environnement en analysant la qualité de l'interaction entre l'utilisateur et le ou les documents électroniques et ses conséquences, ici l'apprentissage. Ces évaluations sont généralement utilisées dans le cadre d'une méthode expérimentale. Dans cette perspective, elles sont menées dans un cadre scientifique rigoureux basé sur un processus de comparaison. Il s'agira par exemple de comparer des versions différentes de documents, des types variés de population, de consigne ou de situation d'apprentissage. Ce type de méthode permet alors d'obtenir des évaluations objectives de la qualité des apprentissages.

Cet article se scindera en trois parties. La première développera succinctement des aspects théoriques liés à l'apprentissage. Les méthodes d'évaluation off-line, qui ont pour caractéristiques d'être utilisées après l'interaction sont exposées dans la seconde partie. Dans un troisième temps, on analysera les méthodes utilisées pendant l'interaction (on-line). Cette présentation sera complétée par des exemples de résultats qu'elles ont permis de mettre en évidence.

2. Quelques rappels théoriques sur l'apprentissage

Sans entrer dans le détail du fonctionnement cognitif impliqué par l'apprentissage, nous en présentons ici quelques principes de base voir Jamet, (Jamet, à paraître), pour une présentation plus développée. Globalement, l'apprentissage d'informations nécessite leur traitement perceptif et sémantique individuel mais aussi leur intégration dans des représentations de plus haut niveau

Les premiers traitements sont donc réalisés dans des mémoires sensorielles et sont dépendants de la modalité d'entrée orale ou visuelle. Ces informations vont être stockées de manière temporaire dans la mémoire de travail. Ce système mnésique à capacité limitée a pour rôle le stockage temporaire d'informations mais aussi la gestion des interactions entre les multiples informations fournies par le document et les connaissances stockées dans la mémoire à long terme de l'utilisateur (Baddeley, 1986).

L'apprentissage et la compréhension d'un document complexe passe par l'élaboration et le stockage en mémoire à long terme de représentations symboliques de différentes natures. Certaines, généralement éphémères, sont constituées des éléments linguistiques du document, d'autres concernent son contenu sémantique. Ces dernières sont plus ou moins élaborées en fonction des processus mis en place par le lecteur (auditeur) : hiérarchisation des informations, condensations d'informations proches, oubli d'informations non pertinentes, etc. Dans les modèles plus récents de la compréhension, il est généralement admis qu'un dernier niveau de représentation est élaboré. Ce niveau, nommé modèle mental ou modèle de situation, a une structure analogique au monde réel (Johnson-Laird, 1983), (Fayol et Gaonac'h, 2003) pour une présentation récente en français. Il peut par exemple prendre la forme d'images mentales sans pour autant s'y réduire. C'est à ce niveau que l'interaction entre les connaissances fournies par le texte et celles du lecteur vont permettre l'élaboration d'inférences. La qualité de ce type de processus dépend donc des

connaissances elles mêmes.

Ces représentations mentales élaborées pendant la compréhension du texte diffèrent notamment en termes d'investissement attentionnel. Si la construction de la représentation propositionnelle est automatique et rapide, l'élaboration du modèle mental impose une charge importante sur la mémoire de travail (Johnson-Laird, 1983). De plus, si le degré de connaissances préalables d'un individu influence peu la qualité de la représentation propositionnelle élaborée par un lecteur, l'effet du degré d'expertise dans le domaine du texte influence la qualité du modèle mental (Tardieu et al., 1992). Des sujets novices se distinguent donc généralement des experts quant à leur niveau de réponses à des inférences et non à des questions faisant référence à ce qu'ils ont retenu du texte.

Toute évaluation d'un apprentissage doit donc tenir compte de ces caractéristiques. Ainsi, il faut garder à l'esprit la relative diversité des représentations élaborées et proposer des évaluations adaptées. De même, l'élaboration de ces représentations est dynamique et évolue considérablement avec le temps (que ce soit au cours d'un épisode précis d'apprentissage ou à plus long terme), il est donc nécessaire de proposer des évaluations qui prennent en compte ce caractère dynamique en analysant en temps réel les comportements. Nous nous proposons donc d'exposer successivement deux types de méthodes. Les premières, dites off-line, sont réalisées après l'apprentissage. Les méthodes on-line sont, elles, utilisées au cours de l'apprentissage.

3. Les méthodes d'évaluations

Evaluer un apprentissage peut se faire à l'aide de deux grands types de méthodes caractérisées par leur moment d'utilisation :

- Les premières consistent à relever des indices de la qualité d'un apprentissage *a posteriori* par des techniques classiques de questionnaire de rappel ou de compréhension (méthode off-line).
- Les secondes consistent à recueillir en temps réel des indices de la qualité de cette interaction, par exemple en mesurant le temps d'apprentissage, en analysant les stratégies de consultation du ou des documents, en relevant les commentaires de l'apprenant (méthodes on-line)

Ces méthodes sont nombreuses mais il est nécessaire de reconnaître qu'elles n'ont pas pour la majorité d'entre elles été créées pour évaluer les EIAH mais plutôt pour analyser les phénomènes mnésiques ou la compréhension de texte. Elles sont donc majoritairement issues des travaux de psychologie cognitive ou de psycholinguistique. C'est à notre avis ce qui fait une partie de leur force. En effet, les connaissances accumulées sur l'apprentissage dans ces disciplines permettent aujourd'hui de proposer des méthodes d'évaluations appuyées sur des modèles théoriques de la mémoire ou la compréhension de texte tels que ceux de (Kintsch, 1998) ou (Johnson-Laird, 1983). Par exemple, il est aujourd'hui clairement démontré que certains aspects des documents vont agir uniquement sur certaines des représentations mentales élaborées pendant l'apprentissage. A titre d'exemple, Mayer (Mayer, 1989) a montré que la présence d'une illustration avait un effet positif sur l'apprentissage. Cette constatation relativement vague, dans notre façon de la présenter tout au moins, ne peut être considérée comme vraie que si elle est précisée. Ainsi dans cette expérience, l'effet de l'illustration n'existe pas dans une tâche de reconnaissance des phrases du texte. Il est présent pour une tâche de rappel libre. Il est aussi très fort pour des questions de transfert qui ne relèvent pas directement d'un simple stockage en mémoire puisqu'elles consistent à utiliser les connaissances acquises pour les transférer à des éléments non présentés dans le texte. L'utilisation d'un simple QCM de reconnaissance des phrases aurait abouti à une conclusion du type « pas d'effet de l'illustration sur l'apprentissage... ». C'est la connaissance des différentes représentations mentales qui va permettre de les évaluer spécifiquement. L'évaluation des EIAH doit donc s'appuyer sur un minimum de connaissance du processus d'apprentissage.

3.1 Les méthodes off-line

Ce type d'évaluation a déjà fait l'objet de nombreuses présentations dans le domaine des apprentissages, beaucoup plus rarement dans celui plus spécifique des EIAH (voir néanmoins Tricot ([Tricot, 2003](#))). La caractéristique de ces méthodes est d'être utilisée pour évaluer *a posteriori* l'apprentissage. Elles prennent donc généralement la forme de questionnaires créés pour estimer différentes dimensions de l'interaction : mémorisation, compréhension, charge mentale ou intérêt par exemple.

3.1.1. *Technique de rappel et de reconnaissance*

Ces techniques d'évaluation de la mémorisation sont utilisées dans de nombreuses situations et ne sont pas spécifiques aux EIAH. Elles ont pour objectif de mesurer ce qui a été retenu du document. Ces méthodes sont très variées.

Certaines sont rarement utilisées dans le domaine des EIAH. C'est le cas notamment des tâches de reconnaissance des éléments du document (texte ou illustration) dans une épreuve où ils sont mélangés à des pièges – voir cependant ([Mayer et Gallini, 1990](#)) pour l'utilisation d'une telle épreuve. Les performances dans ce type d'épreuve sont très susceptibles d'être influencés par le hasard et présentent souvent des scores très élevés (notamment pour les épreuves de reconnaissances de matériaux illustrés). Il faut donc être conscient du risque lié parfois à leur faible pouvoir discriminant.

Les tâches de rappel sont elles nettement plus souvent utilisées par exemple sous la forme d'épreuves de complétion de texte ou d'illustrations, de tâche de rappel libre ou de production de résumés. On trouvera des exemples de rappel libre chez Mayer ([Mayer, 2001](#)) ou ([Bétrancourt et Bisseret, 1998](#)).

Une autre forme de rappel consiste à demander après la consultation de rappeler la structure du document. Ce type d'épreuve a été particulièrement utilisé dans des tâches de consultation d'hypertexte – voir par exemple ([Britt et al., 1996](#))

Certaines études cumulent ces différentes formes de rappel. Ainsi dans la recherche publiée par Dee-Lucas et Larkin ([Dee-Lucas et Larkin, 1995](#)) sur les effets des types de plans fournis avant la consultation d'un hypertexte, les sujets doivent à l'issue de la consultation réaliser un résumé, puis un rappel libre du texte, un rappel libre des titres et enfin un rappel indicé du texte à partir de ces titres.

L'une des difficultés liées à l'utilisation de ces épreuves de rappel libre ou de résumé est leur correction, notamment sur des choix liés à l'utilisation de paraphrases plus ou moins proches du texte original ou de la structuration du rappel par l'apprenant. En psycholinguistique textuelle, de trop nombreuses études ont été réalisées sur la base de textes de quelques lignes facilement analysables notamment par des techniques d'analyse propositionnelle. Les documents pédagogiques sont souvent plus complexes et l'analyse du rappel pose alors de sérieuses difficultés. Une solution intéressante réside dans l'utilisation de logiciels d'analyse sémantique latente (LSA) qui permettent de comparer un texte original aux productions des apprenants et de produire une note exprimant le degré de proximité entre les productions et le texte original ([Lemaire et Dessus, 2003](#)).

Un autre problème récurrent de ces épreuves de mémorisation est l'utilisation quasi-systématique de rappel immédiat et non de rappel différé (souvent pour de simples raisons de disponibilités des participants de l'expérience). Ceci limite considérablement le degré de généralisation de ces résultats à des situations où l'information est apprise pour un objectif à plus long terme et le moins qu'on puisse dire est qu'elles sont nombreuses.....

3.1.2. *Le raisonnement inférentiel et les problèmes de transfert de connaissances*

L'évolution des modèles de la compréhension dans les années 1980 a amené nombre de chercheurs à se détacher dans leurs évaluations des simples informations littérales du texte pour s'intéresser à des

évaluations basées sur le raisonnement inférentiel.

Ce type de question, caractérisé par une réponse non explicite dans le texte et nécessitant un processus d'intégration entre plusieurs informations du texte ou entre une information du texte et des connaissances du lecteur, est sensé être révélateur du niveau de qualité du modèle mental élaboré. Les questions de type « paraphrases » testent en revanche la qualité de la représentation propositionnelle construite en mémoire (Kinstch et al., 1990). Mayer (Mayer, 2001) utilise le terme de problème de transfert pour désigner des évaluations similaires dans le cadre des documents multimédias pédagogiques. Ces problèmes sont aussi caractérisés par l'absence de réponse explicite dans le document appris et la nécessité de transférer les connaissances fournies par le texte à d'autres éléments ou situations.

Dans une étude réalisée sur un cours en ligne sur la perception visuelle et auditive (Jamet, 2000), nous avons obtenu des effets spécifiques sur ce type de variable. Le but de cette expérience était d'analyser les effets de l'intégration spatiale du texte dans un schéma explicatif sur les différentes représentations élaborées pendant la compréhension. Après la présentation des documents, l'apprentissage était évalué par un questionnaire de 10 paraphrases et 10 inférences. Des effets différents du format en fonction du type de questions étaient attendus puisque l'effet de l'intégration semble apparaître de manière prioritaire lors de l'élaboration de représentations mentales nécessitant un investissement attentionnel élevé. Conformément à cette hypothèse, les bénéfices de l'intégration ne sont apparus que pour les inférences. Seul l'emploi des questions de type inférence a donc permis de mettre en évidence l'effet recherché.

De manière identique, dans l'étude de (Moreno et Mayer, 2000), le fait de personnaliser les messages par l'emploi de la première et seconde personne dans un document expliquant la formation des éclairs n'améliore pas le rappel mais permet des scores plus élevés pour les problèmes de transfert. Ce type d'épreuves renvoie donc pour partie à d'autres représentations mentales que celles évaluées par une tâche de rappel, il a donc un intérêt certain à les utiliser dans toute évaluation d'un apprentissage.

3.1.3. Tâche de réalisation

Certains savoirs, notamment les savoirs procéduraux, peuvent difficilement être évalués par des questionnaires qui, par essence, renvoient à des connaissances déclaratives. Ceci est particulièrement important pour certains apprentissages notamment dans le domaine technique ou des apprentissages impliquant l'utilisation de matériels. Certains documents ont pour objet de transmettre ce type de savoirs procéduraux. L'utilisation de tâches visant à faire agir les sujets après leur apprentissage peut donc s'avérer nécessaire. Les études utilisant ce type d'évaluation dans le cadre des EIAH sont très largement minoritaires en psychologie cognitive. On trouvera néanmoins dans les travaux de Sweller plusieurs épreuves de ce type. Tindall-Ford, Chandler et Sweller (Tindall-Ford et al., 1997) démontrent par exemple que la réalisation de schémas électriques peut être améliorée si le schéma explicatif utilisé au cours de l'apprentissage est accompagné d'une explication orale plutôt que d'une explication écrite. La réalisation de procédures plus ou moins complexes dans un tableur informatique (Cerpa et al., 1996) peut être aussi facilitée si les instructions verbales du document explicatif sont intégrées au schéma plutôt que présentées séparément.

3.1.4. Les questionnaires subjectifs

Cette catégorie d'outils d'évaluation a pour caractéristique d'être proposée à l'apprenant sous forme de questionnaires sollicitant un jugement sur des notions aussi variées que l'intérêt, la valeur éducative ou la charge mentale perçue. Il s'agit donc d'évaluer les impressions subjectives de l'utilisateur d'un EIAH. Ce type d'évaluation permet donc d'avoir un retour sur le degré de satisfaction après l'utilisation d'un logiciel mais il ne garantit absolument pas son efficacité pédagogique. Un EIAH peut être jugé par des élèves ou par ses concepteurs comme très satisfaisant et n'entraîner aucun progrès en terme d'apprentissage. Ces questionnaires peuvent néanmoins, s'ils sont utilisés avec prudence et en complément de méthodes plus objectives, permettre de recueillir des indices variés qui permettront de faire des choix adaptés aux objectifs du concepteur (par exemple, privilégier l'efficacité pédagogique ou au contraire l'impression de satisfaction).

Certains de ces questionnaires ont pour objectif d'évaluer l'effort mental lié à la tâche d'apprentissage par des techniques de questionnaires où les sujets jugent après coup de l'effort qu'ils ont du fournir. Les modèles théoriques de l'apprentissage multimédia font souvent référence à la notion de capacité limitée du système cognitif, notamment de la mémoire de travail (Mayer, 2001), (Sweller, 1999) par exemple, et à la notion de charge mentale. Evaluer cette charge en mémoire peut donc fournir des indices sur la qualité d'un document ou d'une situation d'apprentissage. La notion de charge mentale est floue et souvent critiquée à juste titre dans la littérature – voir le numéro spécial de Psychologie Française coordonnée par Tricot et Chanquoy, (Tricot et Chanquoy, 1996). L'évaluation est donc difficile du point de vue conceptuel mais aussi technique.

La question de la mesure de la charge mentale a fait l'objet de nombreux travaux tant sur les documents techniques que pédagogiques. La première catégorie de mesures concerne des indices psychophysiologiques (diamètre de la pupille, variation du rythme cardiaque, potentiels évoqués). Le second type de mesure consiste à utiliser des échelles subjectives d'évaluation. Il s'agit donc dans ce cas de demander à l'utilisateur de produire un jugement sur l'effort mental qu'il a investi dans la tâche. Ce type de mesure peut donc paraître très critiquable notamment parce qu'il sous-entend que la personne comprenne précisément la notion et soit capable de s'auto-évaluer. Néanmoins cette technique est généralement corrélée à des mesures plus invasives comme les techniques de double tâche que l'on décrira ultérieurement (Paas et Van Merriënboer, 1993). Kalyuga, Chandler et Sweller (Kalyuga et al., 2000) utilisent par exemple des échelles de type Lickert en 7 points pour que les apprentis évalués jugent de leur difficulté ou de l'effort mental investi après l'apprentissage d'un document électronique. Enfin, il faut rappeler que des échelles d'évaluation de la charge mentale très générales sont utilisées régulièrement en ergonomie et peuvent être mises en œuvre assez aisément. Il s'agit par exemple du SWAT (Reid et al., 1988), du NASA-TLX (Hart et al., 1988) ou du Workload Profile (Tsang et Velazquez, 1996).

Cette méthode d'échelle peut aussi être utilisée pour juger de l'intérêt, de la satisfaction ou de la valeur éducative du document. Sur ce dernier point, Bétrancourt, Dillenbourg et Montarnal (Bétrancourt et al., 2003) ont observé que contrairement à ce qui est parfois postulé la possibilité d'exercer un contrôle sur le déroulement du document (interactivité) n'entraîne pas un jugement plus positif. Nous avons pu aussi observer que ce type de jugement peut parfois permettre de départager deux documents aussi efficaces ou que les documents les plus efficaces ne sont pas nécessairement ceux qui entraînent le degré de satisfaction le plus élevé (Le Bohec et Jamet, à paraître). Moreno, Mayer, Spires et Lester (Moreno et al., 2001) observent quant à eux des effets positifs à la fois sur l'intérêt et les tests de transferts lorsqu'un agent pédagogique animé est ajouté à l'interface dans un logiciel de sciences de la vie.

3.1.5. Conclusion

Ces méthodes off-line sont probablement les plus utilisées mais cela ne signifie pas qu'elles soient les plus pertinentes. Leur principal défaut tient dans leur nature, elles sont réalisées après la phase d'acquisition des connaissances. Nous avons souligné le caractère dynamique des processus engagés par les individus au cours d'un apprentissage. Evaluer ces caractéristiques *a posteriori* est quasiment impossible. Il est donc primordial de se donner quelques moyens pour appréhender ce dynamisme en évaluant en temps réel les processus engagés par l'apprenant.

3.2 Les méthodes on-line

Ces méthodes consistent, comme il a déjà été précisé, à tenter de recueillir des indices de l'activité au cours d'un épisode d'apprentissage. Les temps de traitements, l'enregistrement du parcours dans le document ou des mouvements oculaires sont autant d'exemples susceptibles de rentrer dans cette catégorie (Rouet et Passerault, 1999). Si les méthodes off-line ont souvent été adaptées aux EIAH, les méthodes on-line ont en revanche souvent été créées dans des situations de consultation de documents électroniques, tout simplement parce que leur complexité ou leur degré de précision impliquait l'utilisation d'un ordinateur...

3.2.1 Les techniques de fenêtres mobiles

Ces techniques de fenêtres mobiles sont employées en psycholinguistique textuelle. Dans ce type de paradigme, c'est l'action de l'utilisateur sur le clavier ou la souris qui permet de découvrir des segments de taille variable du texte (Auto présentation segmentée, APS). Cette technique permet de mesurer de manière relativement précise les temps de traitement. Elle a permis par exemple de mettre en évidence une phase d'intégration en fin de phrase lorsque la consigne pousse les lecteurs à un traitement plus approfondi du texte (Haberlandt et al., 1989). Dans la version RSVP (Rapid Visual String Presentation), c'est l'ordinateur qui impose le rythme de lecture à des cadences variées – voir (Coirier et al., 1996) pour une présentation de ces techniques.

Les segments découverts à chaque appui sont de tailles variées. Par exemple, dans l'étude de Rouet, Vidal-Abarca, Bert Erboult et Millogo (Rouet et al., 2001), ils correspondent à un paragraphe. Cette technique permet ici d'étudier non seulement les temps de lecture des paragraphes, mais aussi les stratégies de consultation d'un document électronique en fonction du type de tâche de recherche d'information. Ces techniques sont donc particulièrement utiles pour l'étude des processus de l'information pendant la lecture, notamment lorsque des techniques plus complexes et plus coûteuses telles que les mouvements oculaires ne sont pas disponibles.

3.2.2 *L'enregistrement informatique de l'activité utilisateur*

Pendant des tâches de consultation de documents électroniques, il est possible de recueillir des indices variés de l'activité de l'utilisateur notamment la nature des pages consultées ou leur temps de consultation. Classiquement, les pages consultées sont stockées de manière séquentielle dans un fichier de données, généralement à l'insu de l'utilisateur. Ces variables ont été très utilisées notamment dans les tâches de recherche d'informations en comparant par exemple le degré de redondance des pages ouvertes dans la consultation ou le nombre de pages consultées dans la tâche par rapport au chemin minimal pour trouver l'information (de Vries et Tricot, 1998).

Ces données permettent aussi de reconstruire *a posteriori* le parcours dans le document – voir par exemple (Britt et al., 1996) – ou d'évaluer le temps nécessaire à la réalisation des différentes étapes de la tâche ou le taux de consultation des pages du document. Par exemple, dans leur étude sur l'effet des types de plan dans l'apprentissage d'un hypertexte, (Dee-Lucas et Larkin, 1995) recueillent des mesures concernant le temps passé à lire le texte, le temps passé à lire le plan ou la proportion des pages consultées.

La difficulté consiste évidemment à traiter ces données qui peuvent être très volumineuses mais aussi à faire des choix sur la précision des mesures qui peuvent aller de l'enregistrement d'un temps global de consultation au stockage de chaque action de l'utilisateur. Le lecteur intéressé trouvera chez (Rouet et Passerault, 1999) une très intéressante discussion de la notion de grain dans l'évaluation et du degré de précision à adopter en fonction des objectifs de l'évaluation de l'apprentissage.

3.2.3 *L'enregistrement des mouvements oculaires*

Le développement des systèmes d'enregistrement des mouvements oculaires permet aujourd'hui de recueillir de manière plus aisée et plus précise les déplacements du regard dans des documents électroniques (Baccino et Colombi, 2001). Les techniques de reflets cornéens permettent ainsi d'enregistrer les mouvements oculaires en utilisant le reflet d'une lumière infrarouge sur la cornée. L'avantage de cette technique est son caractère relativement écologique car ces systèmes dits « tête libre » ne sont pas contraignants dans la tâche d'apprentissage. Ces mesures comprennent les différentes fixations ou saccades (mouvement), mais aussi le parcours temporel précis de l'utilisateur dans l'interface. Elles permettent de recueillir par exemple des indices sur la difficulté de recherches d'une information ou la quantité de fixations sur des zones précises du document. Toutefois, il est nécessaire de rappeler qu'il n'y a pas nécessairement une correspondance stricte entre la zone fixée et ce qui est traité mentalement.

Hannus et Hyönä, (Hannus et Hyönä, 1999) ont ainsi étudié ces mouvements oculaires pendant l'apprentissage d'un document de biologie composé de textes et d'illustrations chez des enfants de 10 ans.

Le temps de fixation de l'illustration correspondait à 6% du temps total, ce qui souligne des stratégies d'utilisation privilégiée du texte dans ce contexte. De plus, les fixations sur les parties pertinentes de l'illustration étaient liées à certaines aptitudes intellectuelles, les enfants les plus faibles de ce point de vue fixant plus souvent des zones non pertinentes du document. De même, Hyönä, Lorch et Kaakinen ([Hyönä et al., 2002](#)) ont pu mettre en évidence des stratégies différenciées de consultation de documents électroniques (textes scientifiques) chez des adultes dans une tâche d'élaboration d'un résumé.

Cette technique est très prometteuse. Les principaux freins à son développement sont la relative complexité technique de l'utilisation de ces appareils et du traitement des résultats mais aussi et surtout le prix élevé des matériels. De plus, comme le soulignent ([Baccino et Colombi, 2001](#)) cette méthode n'a d'intérêt que si elle est inscrite dans un cadre expérimental rigoureux et qu'elle n'est pas utilisée, comme c'est encore trop souvent le cas notamment dans le domaine commercial, pour donner un air scientifique à une étude qui ne l'est pas.

3.2.4 Les techniques de doubles tâches.

Concernant l'évaluation de la charge mentale, le premier type de mesure évoqué consistait à utiliser des questionnaires (voir méthodes off-line). Une autre méthode consiste à utiliser des techniques de double tâche. Le principe de base est que les capacités de la mémoire de travail sont limitées mais peuvent être affectées de manière flexible à des tâches concurrentes. Si deux tâches exigeant les mêmes ressources sont effectuées simultanément, ces ressources doivent être partagées. Autrement dit, les ressources affectées à une tâche secondaire seront dépendantes de la quantité de ressources affectées à la tâche principale. Des indices de réalisation de cette tâche secondaire sont donc révélateurs du niveau de charge induit par la tâche principale.

A titre d'exemple, Brunken, Steinbacher et Leutner ([Brunken et al., 2002](#)) ont montré à l'aide de cette technique que des documents multimédias composés d'illustrations et de textes écrits induisent une charge cognitive plus élevée que si ce texte est présenté à l'oral. Dans cette expérience, des étudiants apprennent un document d'anatomie (tâche principale) composée d'une illustration et d'un texte oral ou écrit. La tâche secondaire consiste à appuyer sur la barre espace lorsqu'une lettre A inscrite en haut de l'écran devient rouge. Les temps de réaction sont mesurés pour cette tâche.

Ces temps de réaction sont plus longs en situation de double tâche qu'une situation de tâche simple, notamment quand le document est uniquement visuel. La conclusion des auteurs est donc que le document audiovisuel implique une charge en mémoire moins importante que la version écrite. Ceci vient donc confirmer les explications d'une série d'études sur l'effet de modalité préalablement publiées et où il était montré une supériorité des documents audiovisuelles en termes d'apprentissage ([Jamet, 2002](#)).

Néanmoins, il convient de souligner que la nature de la tâche concurrente ne doit pas être ignorée. Dans cette étude, elle est visuelle, ce qui peut avoir évidemment une influence considérable sur la nature des résultats observés. Gyselinck, Cornoldi, Dubois, De Beni et Erlich ([Gyselinck et al., 2002](#)) ont ainsi mis en évidence des effets différenciés du type de tâche concurrente dans une situation d'apprentissage d'un document multimédia illustré ou non.

3.2.5 Les protocoles verbaux

Cette technique consiste à demander aux sujets de l'étude de « penser à voix haute » pendant l'étude, ici la situation d'apprentissage. Ces verbalisations peuvent être concurrentes à la tâche mais elles ralentissent le traitement. Elles peuvent être différées après la tâche, par exemple en diffusant la vidéo du protocole au sujet et en lui demandant de la commenter ([Hoc et Leplat, 1983](#)), voir ([Caverni, 1988](#)) pour une revue.

Cette technique a été peu utilisée dans le domaine des EIAH en psychologie cognitive. Dans leur revue de questions consacrées à ces méthodes dans le domaine des hypermédias, Rouet et Passerault citent les travaux de Gray ([Gray, 1990](#)) dans une tâche de recherche d'informations. Ces auteurs soulignent le

caractère critiquable de la méthode à 3 niveaux :

- Seuls les processus conscients peuvent être verbalisés. Ils ne constituent pourtant qu'une partie des traitements réalisés.
- La verbalisation peut interférer avec l'activité de compréhension.
- Enfin, les verbalisations peuvent avoir trait aux représentations qu'a l'apprenant de son activité plutôt qu'à l'activité elle-même.

Il faut néanmoins garder à l'esprit la distinction entre protocoles verbaux et analyse de productions verbales. Les remarques que nous avons formulées tiennent surtout dans des situations où l'on cherche à analyser des processus cognitifs à travers la verbalisation. Toutes les techniques d'analyse de productions verbales n'ont pas cet objectif. Cette méthode peut par exemple être particulièrement utile pour analyser les échanges entre élèves dans des situations d'apprentissage collaboratif.

4. Conclusion

La question du choix des méthodes d'évaluation d'apprentissage est cruciale mais complexe. Les méthodes exposées présentent des niveaux variés de précision, de complexité de mise en œuvre, de pertinence par rapport à la tâche d'apprentissage évaluée.

En tenant compte de ces remarques, il nous paraît évident que c'est l'utilisation conjointe de certaines de ces mesures qui est le plus susceptible de permettre une évaluation satisfaisante d'une situation d'apprentissage. A titre d'exemple, (Dee-Lucas et Larkin, 1995) utilisent dans leur étude des mesures d'enregistrement de l'activité, des épreuves de rappel et des questionnaires d'évaluation subjective du document. Un protocole similaire est utilisé par (Wenger et Payne, 1996). Dans leur étude, l'apprentissage est évalué dans des situations de consultations de textes électroniques par des tâches de rappel libre ou dans l'ordre du texte, par une épreuve de compréhension et par des enregistrements des temps de lecture et des caractéristiques de la navigation des apprenants dans le document.

Cette multiplication (raisonnée...) des mesures a de multiples intérêts. Elle permettra d'éviter des erreurs méthodologiques dites « de seconde espèce ». En effet, le choix d'une seule mesure peut conduire à ne pas mettre en évidence de différence tout simplement parce que cette mesure est inadaptée, pas assez précise, pas assez discriminante ou qu'elle n'évalue les « bonnes » performances dans cette situation précise. Les relations entre performances mesurées peuvent être complexes. Une absence d'effet sur l'efficacité (i.e. la performance à une épreuve) peut être simplement expliquée par un effet sur l'efficience (i.e. la vitesse d'apprentissage). En d'autres termes, des difficultés liées à une situation d'apprentissage peuvent avoir été compensées en ralentissant le traitement, les effets ne seront pas observés sur des épreuves de mémoire ou de compréhension mais sur les mesures on-line de temps d'apprentissage.

Il ne s'agit évidemment pas de multiplier à outrance des mesures (notamment les plus invasives) en dénaturant ainsi totalement la situation d'apprentissage, mais de se donner un maximum d'objectivité dans une situation la plus proche possible de la situation « naturelle » visée par l'évaluation et la tâche est évidemment ardue.

En respectant ces précautions, l'usage de ces méthodes peut avoir des retombées très importantes, notamment en termes de conception de documents pédagogiques. Ainsi, on connaît désormais le rôle bénéfique des illustrations sur la mémorisation, mais aussi sur la compréhension (Gyselinck, 1996). D'autres travaux ont permis de montrer les effets parfois positifs, parfois négatifs de la duplication de l'oral et l'écrit (redondance) dans les documents multimédias (Le Bohec et Jamet, 2005). On connaît aussi mieux

les caractéristiques de la lecture d'hypertextes, les problèmes cognitifs qui en découlent et les solutions de conception susceptibles de les limiter –voir par exemple (Rouet et Potelle, 2005). Dans des domaines proches, tels que la conception de diaporamas, des recommandations ergonomiques peuvent désormais être proposées à partir des résultats du type d'études exposées ici (Jamet, à paraître). Les travaux de Richard Mayer, cités à plusieurs reprises ici, sont désormais exploités dans des ouvrages directement destinés à des concepteurs de cours en ligne (Clark et Meyer, 2003). Toutefois, l'exploitation des résultats de ces travaux ne pourra se développer que si les méthodes sous-jacentes sont connues, non seulement des autres disciplines du champ des EIAH, mais aussi des concepteurs. C'est à ce type de diffusion des connaissances que tente de contribuer cet article.

BIBLIOGRAPHIE

BACCINO, T. & COLOMBI, T. (2001). L'analyse des mouvements des yeux sur le Web, In A.Vom Hofe (Ed.), *Revue d'intelligence artificielle*, p. 127-148. Paris : Hermès.

BADDELEY, A. (1986). *Working memory*. Oxford : Oxford University Press.

BÉTRANCOURT, M. & BISSERET, A. (1998). Integrating textual and pictorial information via pop-windows: an experimental study. *Behaviour and Information Technology*, 17 (5), p. 263-273.

BÉTRANCOURT, M., DILLENBOURG, P., & MONTARNAL, C. (2003). Computer Technologies in Powerful learning environments: The case of using animated and interactive graphics for teaching financial concepts. In L. V. E. De Corte, N. Entwistle & J. van Merriënboer (Ed.), *Unravelling basic components and dimensions of powerful learning environments*. (p. 143 - 157): Advances in Learning and Instruction Series, Pergamon: Elsevier Science Ltd

BRITT, M.A., ROUET, J.-F. & PERFETTI, C.A. (1996). Using hypertext to study and reason about historical evidence. In J.-F. Rouet, J.J., Levonen, A.P. Dillon, and R.J. Spiro (Eds.). *Hypertext and Cognition* (p. 43-72). Mahwah, NJ: Lawrence Erlbaum Associates.

BRUNKEN, R., PLASS, J.L. & LEUTNER, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38, p.53-61.

CAVERNI, J.-P. (1988). La verbalisation comme source d'observables pour l'étude du fonctionnement cognitif. In Caverni, J.P., Bastien, C., Mendelsohn, P., Tiberghien, G., *Psychologie cognitive, modèles et méthodes*. Presse Universitaire de Grenoble: p. 253- 73.

CERPA, N., CHANDLER, P., & SWELLER, J. (1996). Some conditions under which integrated computer based training software can facilitate learning. *Journal of Educational Computing Research*, 15, p. 345 – 367

CLARK, R. C., & MAYER, R.E., (2003). *E-Learning and the Science of Instruction: Proven Guide-lines for Consumers and Designers of Multimedia Learning*. San Francisco: Jossey-Bass Pfeiffer.

COIRIER, P., GAONAC'H, D. ET PASSERAULT, J.M. (1996) *Psycholinguistique textuelle, approche cognitive de la compréhension et de la production des textes*. Armand Collin. Paris.

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

DEE-LUCAS, D. & LARKIN, J. (1995). Learning from electronic texts: Effects of interactive overviews

for information access. *Cognition and Instruction*, 13(3), p. 431-468.

FAYOL, M., GAONACH, D., 2003. *Aider les élèves à comprendre, du texte au multimédia*. Editions Hachette Education.

GRAY, S. H. (1990). Using protocol analyses and drawings to study mental model construction during hypertext navigation. *International Journal of Human-Computer Interaction*, 2(4), p. 359-378.

GYSELINCK, V. (1996). Illustrations et modèles mentaux dans la compréhension de textes. *L'année psychologique*, 96, p. 495-516.

GYSELINCK, V., CORNOLDI, C., EHRLICH, M.-F., DUBOIS, V., & DE BENI, R. (2002) Visuospatial memory and phonological loop in processing texts and illustrations. *Applied Cognitive Psychology*, 16, p. 665-685.

HABERLANDT, K., GRAESSER, A. C., & SCHNEIDER, N. J. (1989). Reading strategies in fast and slow readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, p. 815-823.

HANNUS, M., & HYÖNÄ, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, 24, p. 95-123.

HART, S. G., STAVELAND, L. E., HANCOCK, P. A., & MESHKATI, N. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human mental workload*. (p. 139-183): North-Holland.

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

HYÖNÄ, J., LORCH, R.F., JR., & KAAKINEN, J.K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, 94, p. 44-55.

JAMET, E. (2000). « L'intégration spatiale d'éléments textuels et illustratifs améliore-t-elle la performance ? ». *Revue d'intelligence artificielle*, 14, p. 167-188.

JAMET, E. (à paraître). Peut-on concevoir des documents électroniques plus efficaces? L'exemple des diaporamas. *European review of Applied Psychology*.

JOHNSON-LAIRD, P. N. (1983) *Mental models : Toward a cognitive science of language, inference, and consciousness*. Cambridge, MA : Harvard University Press.

KALYUGA, S., CHANDLER, P., & SWELLER, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, p. 126-136.

KINTSCH, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge, MA: Cambridge University Press.

KINTSCH, W., WELSCH, D., SCHMALHOFER, F., & ZIMNY, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29, p. 133-159.

LE BOHEC, O., & JAMET, E. (2005). Les effets de redondance dans l'apprentissage à partir de documents

multimédia. *Le Travail Humain*, 68(2), p. 97-124.

LE BOHEC O., & JAMET E. (A paraître). Level of redundancy, note-taking and multimedia learning. In J-F. Rouet, R. K. Lowe, & W. Schnotz (Eds). *Multimedia comprehension*. Cambridge University Press.

LEMAIRE B., DESSUS P. (2003) Modèles cognitifs issus de l'Analyse de la sémantique latente. *Cahiers Romains de Sciences Cognitives*.1(1), p. 55-74

MAYER R.E., (1989) Systematic thinking fostered by illustration in scientific text. *Journal of Educational Psychology*, vol.8, p. 240-246.

MAYER, R.E. (2001). *Multimedia learning*. Cambridge University Press.

MAYER, R.E., & GALLINI, J.K. (1990). When Is an Illustration Worth Ten Thousand Words? *Journal of Educational Psychology*, 82(4), p. 715-726.

MORENO, R. & MAYER, R.E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology*, 92, p. 724-733.

MORENO, R., MAYER, R. E., SPIRES, H., & LESTER, J. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19, p. 177-213.

PAAS, F., & VAN MERRIENBOER, J. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors*, 35, p. 737-743.

REID, G. B., NYGREN, T. E., HANCOCK, P. A., & MESHKATI, N. (1988). The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. In *Human mental workload*. (p. 185-218): North-Holland.

ROUET, J-F. & PASSERAULT, J.-M. (1999). Analyzing Learner-hypermedia interaction: An overview of online methods. *Instructional Science*, 27(3/4), p. 201-219.

ROUET, J-F., & POTELLE, H. (2005). Navigational Principles in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. (p. 297-312): Cambridge University Press.

ROUET, J-F., VIDAL-ABARCA, E., BERT-ERBOUL, A. & MILLOGO, V. (2001) Effects of information search tasks on the comprehension of instructional text. *Discourse Processes*, 31, p. 163-186

SWELLER, J., (1999). Instructional design in technical areas. *Australian Education Review*, 43.

TARDIEU, H., EHRLICH, M.-F. & GYSELINCK, V. (1992). Levels of representation and domain-specific knowledge in comprehension of scientific texts. *Language and Cognitive Processes*, 7, p. 335-352.

TINDALL-FORD, S., CHANDLER, P., & SWELLER, J. (1997). When Two Sensory Modes Are Better Than One. *Journal of Experimental Psychology: Applied*, 3(4), p. 257-287.

TRICOT, A. (2003). IHM, cognition et environnements d'apprentissage. In G. Boy (Ed.), *L'ingénierie cognitive : IHM et cognition*. (pp. 411-447). Paris : Hermès Science.

TRICOT, A., CHANQUOY, L. (Eds) (1996). La charge mentale. *Psychologie française*, 41(4).

TSANG, P. S., & VELAZQUEZ, V.L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3), p. 358-381.

WENGER, M. J., & PAYNE, D.G. (1996). Comprehension and retention of nonlinear text: Considerations of working memory and material-appropriate processing. *American Journal of Psychology*, 109, p. 93-130.

■ A propos de l'auteur

Eric Jamet est enseignant-chercheur à l'université Rennes II en psychologie cognitive et ergonomique. Il est directeur adjoint du Centre de Recherche en Psychologie, Cognition et Communication (CRPCC) et responsable d'un programme de recherches consacré aux nouvelles technologies. Ses travaux portent sur les traitements cognitifs réalisés pendant l'apprentissage à partir de documents pédagogiques ou techniques (diaporamas, cédéroms ou sites web) ainsi que sur les difficultés d'utilisation de systèmes complexes d'interaction homme - machine.

Adresse : LPE - CRPCC - Université Rennes II, Place du recteur H. Le Moal 35043 Rennes CEDEX

Courriel : eric.jamet@uhb.fr

Toile : http://www.uhb.fr/sc_humaines/psycho_expe/index.php?rub=fiche&iduser=30

Référence de l'article :

Eric JAMET , Une présentation des principales méthodes d'évaluation des EIAH en psychologie cognitive, *Revue STICEF*, Volume 13, 2006, ISSN : 1764-7223, mis en ligne le 03/02/2007, <http://sticef.org>

© Revue Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation, 2007