

# Une aide à l'évaluation des logiciels multimédias de formation

**Olivier HU, Philippe TRIGANO, Stéphane CROZAT**

UMR CNRS 6599 HEUDIASYC - Université de Technologie de Compiègne  
BP 20529 - 60206 COMPIEGNE Cedex - FRANCE  
Tel (33) 03 44 23 45 02 Fax (33) 03 44 23 45 02  
e-mail : {Olivier.Hu}{Philippe.Trigano}{Stephane.Crozat}@utc.fr

---

*RESUME : De plus en plus de logiciels éducatifs sont proposés dans la grande distribution comme dans le milieu académique, mais il existe peu de méthodes permettant d'en évaluer la qualité. Nous présentons dans cet article une méthode d'aide à l'évaluation intitulée EMPI (Evaluation du Multimédia Pédagogique Interactif), ainsi que le prototype qui l'instrumente. Notre logiciel est destiné aux prescripteurs afin de les aider à évaluer, choisir ou caractériser les logiciels. L'évaluation est donc considérée comme un processus interactif entre l'évaluateur et notre méthode. EMPI se présente sous la forme d'un questionnaire dynamique structuré autour d'une critériologie transdisciplinaire prenant en compte six thèmes complémentaires : les Impressions Générales, la Qualité Technique, l'Ergonomie, les Documents Multimédias, la Scénarisation et les Outils Pédagogiques. Différents outils permettent d'adapter l'évaluation à chaque contexte particulier : un parcours dynamique, la prise en compte de la subjectivité et l'application de calculs originaux. Pour finir, nous rendons compte des validations effectuées et des expériences que nous menons pour s'assurer de l'utilisabilité et la fiabilité d'évaluation du logiciel qui met en œuvre notre méthode.*

*ABSTRACT. Authors and users of multimedia educational software are in front of the lack of experience in this recent field. We submit a method (EMPI: Evaluation of Multimedia, Pedagogical and Interactive software) to evaluate multimedia software used in educational context. Rather to develop an automatic evaluation, we prefer making a method to help human to evaluate and aim to assist the users to make a choice between the large range of software actually available. Our process is to divide the software analysis into six main themes: the general feeling, the technical quality, the usability, the multimedia documents, the scenario and pedagogical tools. Each of these themes is sub-divided into criteria, sub-criteria and questions. The whole forms a hierarchical questionnaire that allows marking software through various aspects, in order to compare it to other software or with a determined pedagogical context. Various tools enables a tuning of the evaluation to each particular context : a dynamic navigation in the questionnaire, the use of the evaluator's subjectivity ... At the end of the paper, we will describe several validations and experiments done to test the fiability of this method.*

**MOTS-CLES :** EIAO, EVALUATION, DIDACTICIELS, QUESTIONNAIRE, MULTIMEDIA

**KEY WORDS :** EVALUATION, TEACHWARE, QUESTIONNAIRE, MULTIMÉDIA

---

## 1. Introduction

Les récents développements de l'informatique ont des répercussions dans de nombreux domaines d'activité, et le milieu éducatif ne fait pas exception. Ainsi, la vulgarisation du CD-Rom, la banalisation des aspects multimédias et le développement d'Internet ont, entre autres, amené de nouveaux débouchés en EIAO. De plus en plus de logiciels sont développés tant dans le milieu académique que dans la grande distribution, mais les acteurs manquent d'outils d'évaluation adaptés à ces nouveaux produits. Certes des méthodes d'évaluation existent, que ce soit dans le domaine des manuels scolaires ou dans celui des logiciels et de leurs interfaces. Mais la complexité des logiciels actuels nous oblige à considérer comme un tout, de nombreux domaines de recherche, comme l'ergonomie, le génie logiciel, l'édition des documents multimédias, la didactique ou les sciences de l'éducation. De ce fait, l'évaluation complète d'un logiciel de formation nous semble une tâche trop complexe pour être automatisée et effectuée par une machine seule. Notre approche consiste donc à proposer un outil capable d'aider un évaluateur humain dans sa démarche d'évaluation. L'évaluation est envisagée comme un processus interactif entre l'évaluateur et notre méthode, dont le rôle est de systématiser la démarche, d'aider et de confronter l'avis de l'évaluateur avec une évaluation objective.

Cet article est décomposé en trois parties. Premièrement, nous décrivons, au travers de divers exemples, quelques-unes des possibilités actuelles des logiciels de formation et par là même, la gamme de logiciels que nous comptons évaluer. Nous présentons ensuite le contexte d'évaluation, le public à qui s'adresse notre méthode, ainsi que les contraintes qui en résultent sur notre approche.

Dans un second temps, nous développons la méthode EMPI (Evaluation des logiciels Multimedia Pédagogiques Interactifs) proprement dite. Nous définissons les critères évalués, en accentuant ceux propres au contexte éducatif. La méthode se présente sous la forme d'un questionnaire dynamique dont nous présentons les techniques de calcul et de manipulation. Plusieurs aspects sont fondamentaux dans notre approche :

- En premier lieu, la méthode EMPI propose des outils permettant la prise en compte de la subjectivité de l'évaluateur : il capitalise ses impressions sur le logiciel, et la méthode lui permet de confronter son opinion avec une évaluation objective.

- Cette évaluation objective est obtenue grâce à des règles de calcul originales : pondération des questions et application de notations particulières afin d'accentuer la détection des défauts majeurs.

- Pour de nombreux critères, un jugement de valeur a priori ne peut être rendu sans une prise en compte du contexte précis d'utilisation. La méthode se borne donc, dans ce cas, à caractériser le logiciel, le jugement étant le fait de l'évaluateur.

- Nous présenterons ensuite l'outil logiciel que nous avons développé et qui instrumente la méthode.

Enfin, nous exposons les principaux résultats déjà obtenus ou en cours d'analyse, lors de la mise au point de la méthode et du logiciel avec des professeurs

mais aussi avec des concepteurs. Nous concluons par une discussion sur la validité de notre approche et sur les perspectives à venir.

## 2. Les caractéristiques de notre évaluation

### 2.1 *Un outil d'aide à l'évaluation*

Notre méthode s'adresse principalement à des enseignants désirant évaluer, analyser ou choisir un logiciel pour l'intégrer dans les situations d'apprentissage qu'ils mettent en place. Nous insistons sur le fait que la méthode présentée dans cet article est une méthode d'aide à l'évaluation de logiciels de formation. La place de l'humain est primordiale et la méthode n'est là que pour le guider et l'assister dans sa démarche d'évaluation.

De ce fait nous envisageons l'évaluation comme une interaction entre l'outil et l'évaluateur. Cette interaction doit, à notre sens, posséder plusieurs caractéristiques essentielles. Premièrement, la méthode doit permettre une évaluation systématique de tous les aspects jugés pertinents dans le cadre d'une future utilisation. L'outil doit donc jouer le rôle de pense-bête ou de *check-list* afin d'éviter toute focalisation ou tout oubli de la part de l'évaluateur. Deuxièmement, lorsque les caractéristiques du logiciel ne peuvent être jugées a priori, sans une connaissance de la situation d'apprentissage, le logiciel doit se contenter de caractériser au mieux le logiciel et d'aider l'évaluateur à juger par lui-même. Enfin, tout jugement d'un logiciel aussi complexe qu'un logiciel multimédia de formation présente une composante subjective qu'il s'agit de capitaliser et de prendre en compte. Nous verrons dans les chapitres 4 et 5 comment nous avons pris en compte ces différents aspects.

### 2.2 *Le public et les logiciels cibles*

L'évaluateur correspond dans notre approche au prescripteur du logiciel : un professeur, un gérant de CDthèque, un particulier désirant s'autoformer, etc. Par exemple, prenons un professeur de mathématique de collège qui désire intégrer un outil logiciel dans son cours. Entre les logiciels disponibles dans le commerce et ceux proposés par l'éducation nationale, via les CRDP<sup>1</sup> par exemple, notre professeur va devoir faire un choix. Notre approche consiste à lui fournir un outil lui permettant de mieux juger ces logiciels.

De ce fait nous voulons évaluer directement le logiciel fini : notre méthode est plus orientée vers une évaluation finale avec les futurs enseignants que d'un outil intégré dans la boucle classique de conception/évaluation de prototype. A l'inverse notre évaluation ne prend pas en compte l'impact sur les apprenants (qui n'est éva-

---

<sup>1</sup> CRDP : Centre Départemental de Documentation Pédagogique

luable qu'a posteriori). Il s'agit donc de fournir une méthode capable de prévoir les défauts et les qualités qui se révéleront lors de l'utilisation par les apprenants.

L'introduction de l'informatique dans les écoles remonte aux années 1960-1970 alors essentiellement tournée vers un enseignement béhavioriste (enseignement programmé, stimuli/réponse) et vers les QCM. Les logiciels ont depuis largement évolués. Les capacités de calcul et de mémorisation des ordinateurs actuels permettent notamment l'intégration d'images, de sons et d'animations. L'interactivité a été grandement développée grâce aux interfaces graphiques et aux simulations. Afin d'illustrer les possibilités actuelles des logiciels de formation, nous allons rapidement présenter quatre familles types : les logiciels de diffusion, les logiciels de type scolaire, les simulations et les systèmes d'enseignement à distance. Ce découpage est arbitraire et ne sert qu'à illustrer les logiciels que nous comptons évaluer au sein de ceux actuellement disponibles.

Grâce à la numérisation et aux outils de compression, les logiciels de diffusions peuvent maintenant s'appuyer sur une grande quantité et une grande diversité de documents. Ainsi, il n'est pas rare de voir des CD-Rom contenir quelques centaines de documents avec plusieurs heures d'enregistrements sonores ou vidéos. Deux aspects sont prépondérants dans ces logiciels : la recherche d'information et la navigation (outils de recherche par mots clés, utilisation d'hyperliens, cartes, indicateurs du parcours, interfaces de navigation intuitives, outils de marquage des documents etc.). A noter que certains de ces logiciels proposent des évaluations de compétences souvent sous la forme de QCM. Ces évaluations sont trop pauvres et trop arbitraires pour être d'un réel intérêt pédagogique. En situation de cours, ces logiciels peuvent donc servir de ressources au même titre que les livres ou les encyclopédies, ils nécessitent donc une intégration de la part de l'enseignant.

Essentiellement basés sur une approche académique de l'enseignement, le contenu des logiciels de type scolaire est découpé en module chacun accompagné d'une série d'exercices. Ces outils sont souvent proches des programmes scolaires avec des titres comme «Les Mathématiques de 1<sup>er</sup> S» ou «Perfect Your English 3<sup>ème</sup>». Contrairement aux précédents, ces didacticiels proposent de réelles applications du contenu étudié : les simples QCM sont maintenant remplacés par des activités où l'apprenant pointe, déplace, écrit ou dessine, bref, interagit directement avec le logiciel via son interface. Cette interaction stimule la motivation de l'apprenant en lui proposant ses activités sous une nouvelle forme (certains logiciels allant jusqu'à immerger l'apprenant dans une fiction) ou en lui permettant quelques fois des manipulations difficiles à mettre en œuvre sous une autre forme. L'intérêt de ces produits réside donc dans l'interactivité suscitée par les applications et les exercices, mais aussi dans la gestion plus ou moins sophistiquée du parcours de l'apprenant au sein des cours présentés. Dans certains logiciels, l'apprenant peut utiliser des parcours prédéfinis ou construire ses propres cheminements. Dans d'autres, le parcours est adapté en fonction des résultats ou du comportement de l'apprenant. Dans le cadre du milieu éducatif, ces logiciels sont utilisés soit directement lors des sessions de cours, soit dans des cours de soutien ou de renforcement en parallèle avec le cours traditionnel. Ces logiciels ont également massivement employés dans le milieu pa-

rascolaire : cours de soutien, camp de vacances, remplacement de cours particuliers etc.

Les logiciels de simulation correspondent généralement à l'enseignement de savoir-faire. Les exemples sont nombreux : l'enseignement des langues (savoir parler) avec une immersion dans un monde parlant la langue étudiée ; en médecine, en management, en économie (savoir décider) avec les études de cas et les simulations ; en manutention, en contrôle de processus (savoir agir) avec les simulateurs de panes ou de comportements. Lorsqu'une formation sur le terrain est trop coûteuse ou trop difficile à mettre en place, ces logiciels offrent une alternative intéressante, notamment en formation continu.

Enfin, les technologies de la communication et des réseaux ouvrent aujourd'hui de nouvelles possibilités en terme de dispositifs de formation à distance. Les réseaux permettent toutes les combinaisons possibles entre les enseignants et les apprenants, entre les apprenants eux-mêmes ou avec des ressources extérieures. La visioconférence, le "chat" commencent à atteindre la maturité suffisante pour un emploi en situation d'apprentissage, mais les outils de communication asynchrone restent prédominants (courrier électronique et forum par exemple). Un enseignant pourra, par exemple, mettre l'ensemble de son cours sur un site Web, l'accès et le parcours au sein des différents chapitres étant contrôlés par mot de passe. Le courrier électronique servira à transmettre les sujets de TD ou de TP et pour les requêtes directes entre les acteurs de la formation. Le travail de groupe pourra être animé par la mise en place d'un forum ou d'une FAQ modérés par les enseignants.

Ainsi les possibilités logicielles offertes sont nombreuses et le choix par un enseignant du meilleur outil pour son cas personnel n'est pas évident. D'où la nécessité de posséder un outil, une méthode d'évaluation et de caractérisation prenant en compte l'ensemble de ces aspects.

### **2.3 Choix de la méthode : un questionnaire dynamique**

Il existe de nombreuses méthodes d'évaluation que ce soit dans le domaine des manuels scolaires [RIC 80] ou dans celui des logiciels [KOL 97]. Outre les caractéristiques que nous avons énoncées précédemment, nous avons posé un certain nombre de contraintes supplémentaires pour choisir la méthode la plus pertinente dans notre cas. Nous ne voulons pas de l'intervention d'un spécialiste de la méthode lors de l'évaluation et nous ne voulons pas nous limiter à un environnement logiciel ou matériel particulier. En confrontant les diverses méthodes existantes avec nos hypothèses de travail [HU 98a], nous sommes arrivés à la conclusion que l'utilisation d'un questionnaire informatisé était la méthode la plus appropriée.

Tout d'abord, l'emploi d'un questionnaire ne nécessite aucune compétence de la part de l'évaluateur. Ensuite, il est relativement facile de gérer dynamiquement un questionnaire structuré à partir d'un corpus de questions. Enfin les résultats d'un questionnaire à questions fermées sont aisément exploitables.

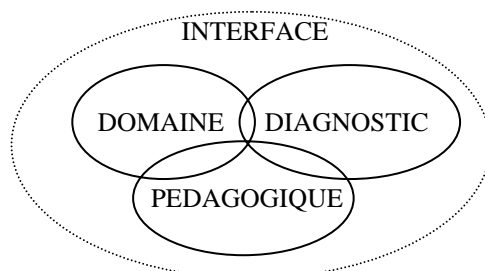
Ce questionnaire doit s'articuler sur une critériologie propre à notre contexte et prenant en compte l'ensemble des caractéristiques des logiciels pédagogiques. Nous allons présenter dans les sections suivantes quels sont les fondements de notre critériologie.

### 3. Les thèmes d'évaluation

Les modèles de didacticiel sont nombreux et la modélisation de l'acte d'apprendre fournit aussi une littérature abondante en EIAO. Nous allons en présenter quelques-uns. Cette liste ne se veut pas exhaustive, elle démontre juste la diversité des approches existantes et les différents modèles qui nous ont inspirés.

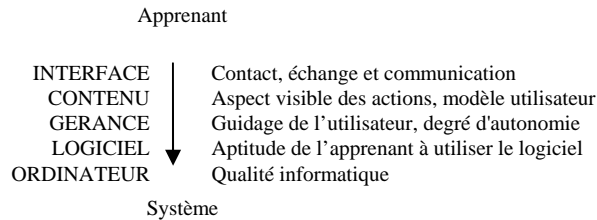
#### 3.1 Quelques modèles de logiciels pédagogiques

Un des modèles classiques issu de l'IA découpe les didacticiels en trois modules : un module du domaine enseigné, un module pédagogique et un module de diagnostic, le tout accessible via une interface (Figure 1).



**Figure 1.** Structure d'un didacticiel (modèle IA)

L'interdépendance et le recouvrement des modules posent un certain nombre de problèmes. La place de l'apprenant et son interaction avec le système par exemple, ne sont pas représentées. Jacques J. Rhéaume propose une structure en couches [RHE 94] qui donne une vision linéaire du didacticiel en partant de l'apprenant jusqu'au système (Figure 2). Si cette structure couvre a priori l'ensemble du domaine, il est tourné vers une analyse quasi exclusivement ergonomique et technique du didacticiel. La pédagogie sous-jacente, l'utilisation de scénarios originaux, le multimédia, la navigation, sont des aspects difficiles à intégrer dans ce modèle.

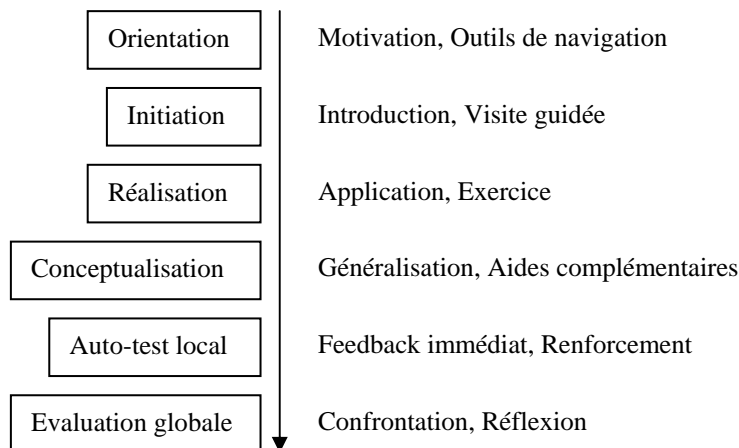


**Figure 2.** Structure en couche [RHE 94]

Or comme le souligne Bruillard et Vivet, l'évaluation technique doit être complétée. Ils identifient trois types d'évaluation [BRU 94]: la pertinence éducative, la qualité technique et l'évaluation ergonomique. La distinction et la mise en exergue de ces trois aspects nous apparaissent comme pertinente, nous l'avons conservée dans notre critériologie.

En ce qui concerne la pertinence pédagogique, nous retrouvons cette séparation chez Philippe Dessus et Pascal Marquet [DES 91] par exemple, qui transposent les travaux réalisés sur l'évaluation des manuels scolaires en y ajoutant la composante interaction homme/machine. L'évaluation qu'ils proposent est divisée en deux parties : d'une part une caractérisation du processus d'apprentissage du didacticiel [BER 88] et d'autre part une évaluation chiffrée de chaque module, à savoir : le contenu, la méthode, la présentation des messages et l'interaction homme-machine. Pour différentes raisons que nous exposerons par la suite (section 4.2), nous avons conservé et étendu cette séparation entre la caractérisation de l'apprentissage et l'évaluation proprement dite.

Ces modèles nous permettent de définir une liste des différents thèmes à évaluer dans un didacticiel (la qualité technique, l'analyse ergonomique et donc l'interface, la scénarisation, la présentation des informations, la pertinence pédagogique, le



**Figure 3.** Ecrans types d'un didacticiel (extrait de [LIN 98b])

contenu etc.). La prise en compte de la tâche de l'apprenant ne s'effectue qu'à travers l'analyse ergonomique qui reste, du fait de notre évaluation a priori, à un haut niveau de détail. Ainsi les caractéristiques propres à l'utilisateur apprenant (par opposition à l'utilisateur opérateur issue de la vision ergonomique de haut niveau) ne sont pas ou peu prises en compte. Il s'agit donc de prendre en considération les résultats issus de la théorie de l'activité appliquée à l'apprentissage. Cette compréhension de l'activité est primordiale si l'on veut correctement instrumenter l'apprenant dans son processus d'apprentissage, et par la même, définir ou évaluer un didacticiel.

Le modèle Helices [LIN 98a] est un modèle de l'action prenant en compte les aspects hiérarchiques (Motifs, Buts, Actions) et les aspects récursifs (Intention, Réalisation, Validation) de l'activité. A partir de cette description de l'activité Monique Linard propose des recommandations en terme de conception [LIN 98b] et par la même, un agencement type des modules d'un didacticiel (Figure 3).

Nous retrouvons plusieurs de ces aspects dans notre approche de l'évaluation :

- Tout d'abord, les aspects Orientation et Initiation nous confortent dans notre volonté de considérer la scénarisation du logiciel avec une attention particulière. Il s'agit à ce niveau de motiver la démarche, d'explicitier les tenants et les aboutissants de la tâche et de présenter les outils de navigation, au sens large, qui permettront d'explorer le logiciel.

- Ensuite les aspects Réalisation et Conceptualisation illustrent et justifient la nécessaire diversité des activités à proposer aux apprenants. Ce point doit donc être identifiable dans toutes démarches d'évaluation. Ces éléments permettent aussi d'introduire la notion d'assistance notamment dans le cadre de la conceptualisation où la présence d'une aide extérieure est souvent nécessaire, que cette aide soit matérielle (base de données, aide contextuelle etc.) ou humaine (tutorat en présentiel, requête en ligne, forum etc.).

- Enfin, les aspects Auto-test local et Evaluation globale confirment la place importante de l'aspect évaluation dans tout didacticiel.

### ***3.2 Proposition d'une structure globale d'évaluation d'un didacticiel***

Notre critériologie reprend donc un certain nombre de points présentés précédemment : l'évaluation technique, l'ergonomie, la nécessité de séparer la caractérisation de la stratégie pédagogique de son évaluation proprement dite, la volonté de mettre en avant la scénarisation et de clairement identifier les aspects liés à l'activité de l'apprenant, à l'aide qui lui est proposée et aux outils d'évaluation existants.

Voici donc la liste des six thèmes d'évaluation à partir desquels nous allons développer notre évaluation d'un logiciel de formation :

- Le thème **Qualité Technique** qui évalue la mise au point informatique du logiciel : rapidité, installation, dysfonctionnement etc.

- Le thème **Ergonomie** qui correspond à l'évaluation de la facilité d'utilisation de l'interface (nous utiliserons par la suite le néologisme utilisabilité).

- Le thème **Documents Multimédias** qui mesure la qualité et la pertinence de la présentation des documents.
- Le thème **Scénarisation** qui évalue la qualité de présentation et d'agencement des informations : navigation, fiction, métaphore etc.
- Le thème **Outils Pédagogiques** qui concerne les outils mis en place entre autre dans le cadre de l'évaluation des apprenants, du suivi pédagogique, de l'assistance etc.

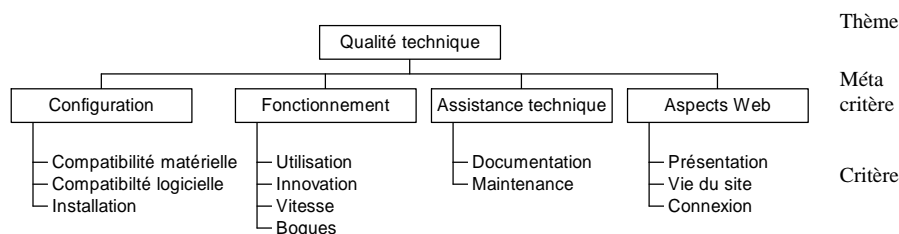
- Le thème **Impressions Générales** qui capitalise les impressions de l'évaluateur.

Pour chacun de ces thèmes nous avons cherché à recenser des critères pertinents afin de structurer notre questionnaire [CRO 99]. Avant de développer chacun de ces thèmes, nous pouvons déjà apporter quelques précisions. Si certains thèmes peuvent être appliqués a priori à tous les logiciels de formation (le thème Ergonomie par exemple), d'autres ne s'appliquent qu'à certains types de logiciels (le thème Scénarisation sera moins pertinent pour un logiciel de diffusion). Il ne s'agit donc pas de prétendre qu'un logiciel de formation doit obligatoirement développer les modules correspondant à chacun des thèmes. Ces caractérisations ne sont que les réceptacles de certaines des possibilités logicielles que nous avons développées précédemment.

Nous allons maintenant détailler chacun des thèmes en accentuant sur le thème Outils Pédagogiques propres à notre contexte éducatif. Le thème Impressions Générales sera développé dans le chapitre 4.

### 3.3 Le thème Qualité Technique

Le thème Qualité Technique reflète des problématiques issues du génie logiciel. Sa prise en compte à ce niveau tient dans notre volonté de voir rejeté du milieu éducatif tout logiciel non fiable ou demandant des compétences informatiques inacceptables pour le corps enseignant. La problématique de ce thème n'étant pas propre au contexte éducatif, nous ne ferons que présenter les critères et les méta-critères qui le composent (Figure 4).



**Figure 4.:** Développement du thème Qualité Technique

Dans le tableau ci-dessous (Tableau 1) sont données les définitions des méta-critères du thème Qualité technique, accompagnées de quelques exemples de questions associées.

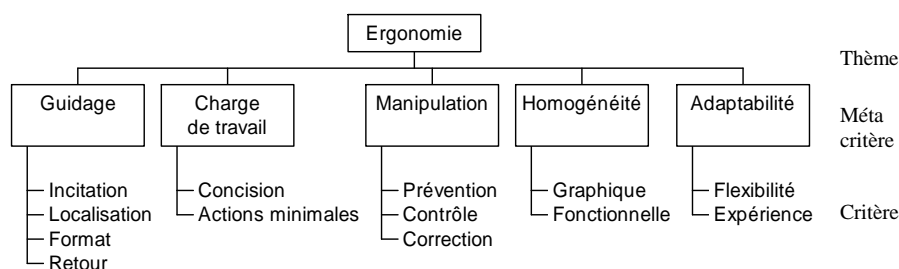
<b>Méta-critère</b>	<b>Définition et exemples de questions associées</b>
Configuration	Mesure la facilité d'installation et le degré de compatibilité du logiciel.
	<i>Quelle est l'importance des ressources mémoires demandées par le logiciel ? Caractériser le nombre des logiciels complémentaires nécessaire au bon fonctionnement ?</i>
Fonctionnement	Mesure la qualité du déroulement du programme.
	<i>Le logiciel a-t-il bloqué ou fermé l'application de façon inattendue ? Caractériser les temps de réponse du logiciel.</i>
Assistance technique	Concerne la qualité des moyens d'assistance technique extérieurs mis à disposition de l'utilisateur.
	<i>Caractériser le degré d'utilisation de la documentation lors de l'exploitation du logiciel. Quelle est la richesse des aides proposées ?</i>
Aspects Web	Mesure la qualité des aspects techniques propres au Web et à Internet.
	<i>Les pages sont-elles optimisées pour le chargement et la présentation ? Les informations sont-elles mises à jour périodiquement ?</i>

**Tableau 1.** Définition et exemples de questions des méta-critères du thème Qualité technique

### 3.4 Le thème Ergonomie

Nos critères en matière d'ergonomie, sont principalement issus des travaux de D. Scapin [SCA 97] [BAS 95] pour les critères proprement dit, et de J. Vanderdonckt [Vanderdonckt 94] pour l'utilisation d'une base de règles ergonomiques. Ces critères ont été adaptés et complétés pour intégrer le contexte multimédia pédagogique. Ces travaux ont déjà fait l'objet de publications tant en ce qui concerne la définition des critères que sur la prise en compte du milieu éducatif dans l'évaluation ergonomique [HU 00a] [HU 99] [HU 98b]. Le contexte éducatif n'apporte, ni ne supprime, aucun critère ergonomique. Certaines considérations doivent simplement être mises en perspective avec le contexte d'utilisation ou la volonté pédagogique de l'enseignant. Que dire du manque de contrôle de l'interface si celui-ci résulte de la prise de contrôle à distance du didacticiel par le professeur ? Comment qualifier la possibilité laissée aux élèves de modifier les paramètres d'affichage, le volume sonore etc., alors que le logiciel doit être utilisé sur trente postes en parallèle ? Les exemples sont nombreux et en aucun cas un jugement de valeur ne peut être rendu, sans prise en compte du contexte précis d'utilisation. La méthode se retrouve donc dans un rôle de caractérisation de certains aspects ergonomiques plus que d'évaluation proprement dite. Cette dernière sera en dernier recours le fait de l'évaluateur sur la base d'une caractérisation objective. Nous verrons dans la description de notre méthode que cette dualité caractérisation / évaluation est un point fondamental de notre approche.

La figure et le tableau suivants (Figure 5 et Tableau 2) présentent les méta-critères et les critères composant le thème Ergonomie, ainsi que la définition des méta-critères accompagnée de quelques exemples de questions.



**Figure 5.** Développement du thème Ergonomie

<b>Méta-critère</b>	<b>Définition et exemples de questions associées</b>
Guidage	Mesure la qualité des caractéristiques de l'interface permettant d'identifier son fonctionnement. <i>Lorsqu'il faut réaliser une action spécifique, est-ce que le système l'indique à l'utilisateur ?</i> <i>Vous est-il arrivé de devoir chercher des composants à l'écran, parce qu'il n'avait pas l'aspect ou le format attendu ?</i>
Charge de travail	Evalue la pertinence des efforts cognitifs que l'utilisateur doit mettre en œuvre pour accomplir sa tâche. <i>Trouvez-vous que l'écran soit surchargé du point de vue des informations et des possibilités d'actions ?</i> <i>Avez-vous trouvé que certaines suites d'actions, pour atteindre un but, étaient trop longues ou trop compliquées ?</i>
Manipulation	Concerne la qualité de la maîtrise par l'utilisateur du déroulement du logiciel. <i>L'utilisateur a-t-il la possibilité d'interrompre un traitement parce qu'il est trop long par exemple ?</i> <i>Est-il arrivé que le logiciel entreprenne une action que vous n'avez pas explicitement demandée ?</i>
Homogénéité	Evalue la cohérence et le maintien tout au long du logiciel d'une charte graphique et fonctionnelle. <i>Est-ce que les caractéristiques de l'interface sont conservées pour des contextes d'utilisation identiques ?</i> <i>Un même intitulé ou icône conduit-il toujours au même comportement ?</i>
Adaptabilité	Concerne l'ensemble des possibilités disponibles pour modifier l'aspect ou les modes d'interaction de l'interface afin d'en améliorer l'utilisation. <i>Le format des textes affichés est-il modifiable ?</i> <i>L'utilisateur, en fonction de son expérience, peut-il accéder à des fonctions de manipulation avancées ?</i>

**Tableau 2.** Définition et exemples de questions des méta-critères du thème Ergonomie

### 3.5 Le thème Documents Multimédias

Les textes, les images et les sons constituent la matière première du logiciel multimédia, ils sont les vecteurs essentiels de l'information. Précisons que notre étude, dans cette partie, ne porte pas sur le contenu, mais sur la pertinence de l'utilisation et de la présentation des documents. L'étude du contenu proprement dit est traitée par ailleurs dans le thème Outils Pédagogiques.

Contrairement au thème Ergonomie qui est découpé en dimensions d'évaluation, le thème Documents Multimédias est principalement découpé en objets. Afin de caractériser les documents du didacticiel, nous avons établi une liste hiérarchisée des objets multimédias existants, accompagnée d'une dimension correspondant à l'agencement de ces documents. Nous retrouvons ainsi dans nos méta-critères (Figure 6) la séparation classique en documents textuels, sonores et visuels, accompagné du méta-critère Interdépendance qui concerne les relations entre les différents documents. La définition de ces méta-critères est donnée dans le tableau ci-dessous (Tableau 3), les exemples de questions seront donnés dans les chapitres correspondant à chaque méta-critère.

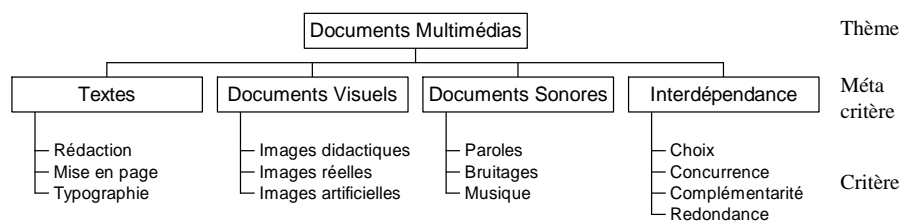


Figure 6. Développement du thème Documents Multimédias

Méta-critère	Définition
Textes	Ce méta-critère mesure la qualité des choix et des formats des documents basés sur le texte écrit.
	<i>Le vocabulaire employé est-il familier pour l'utilisateur ? Les choix de couleurs pour le texte et le fond optimisent-ils la lisibilité ?</i>
Documents visuels	Ce méta-critère évalue la qualité et la pertinence des documents visuels (ensemble des documents basés sur les images et les représentations graphiques, fixes ou animées).
	<i>Est-ce que le contenu des images peut entraîner des ambiguïtés ? Quelle est la qualité de la définition des photographies et vidéos ?</i>
Documents sonores	Ce méta-critère évalue la qualité et pertinence des voix, des musiques et des bruitages.

	<i>Le didacticiel a-t-il recours à l'utilisation de la musique dans un but pertinent (détente, transition, ...) ? Les bruitages sont-ils mis à profit pour remplir des fonctions de signalisation ou de localisation d'actions ou d'événements ?</i>
Interdépendance	Ce méta-critère concerne la qualité d'agencement des différents médias. <i>L'utilisation simultanée de plusieurs médias gêne-t-elle la compréhension ? Des légendes sont-elles utilisées pour résumer le sens des images ?</i>

**Tableau 3.** Définition et exemples de question des méta-critères du thème Documents Multimédias

Les documents textuels sont souvent centraux dans un logiciel de formation. Et si nos habitudes de lecture sont issues de plusieurs siècles d'utilisation du livre et de ses dérivés, le contexte de lecture à l'écran est sensiblement différent : le support est lumineux, fixe et vertical. De plus certains outils de manipulations des documents à l'écran, comme les ascenseurs ou les hyperliens, tout en permettant des manipulations évoluées, peuvent perturber le lecteur s'ils sont mal utilisés [BAL 96]. Les textes et leurs présentations doivent donc être les plus clairs et les plus concis possible afin de compenser la fatigue accrue par une lecture à l'écran. Notre propos n'est pas ici de donner les critères d'écriture des textes en vue de leur affichage sur un écran. Nous ne faisons qu'adopter, au travers de nos critères d'évaluation, les règles d'édition en vigueur pour une bonne appropriation des textes écrits [PAR 91] tout en tenant compte d'études spécifiques aux documents textuels visualisés sur écran [CAR 98] [VAN 98].

Les images didactiques contiennent une véritable information en elles-mêmes. Il s'agit des tableaux, des graphiques, des cartes, etc. Le niveau d'abstraction des images didactiques est central dans la mesure où l'interprétation plus ou moins aisée par l'utilisateur détermine le niveau de langage iconique : des images trop simples ne valorisent pas l'utilisateur et des représentations trop complexes peuvent le rebuter au risque de le voir ne pas utiliser ce support.

Les autres documents visuels, images réelles ou images artificielles, sont souvent accompagnés ou accompagnent une information disponible sous une autre forme. On retrouvera ces notions d'accompagnement dans le méta-critère Interdépendance. Ces deux critères comprennent des images fixes et des images animées, les notions de fluidité se retrouvent donc dans les deux. Néanmoins la construction des images répond à des critères différents selon leurs sources. Ainsi le critère Images réelles doit prendre en compte les notions de définitions, de cadrage et d'éclairage, alors que le critère Images artificielles, considérera l'originalité et l'attrait du graphisme ou la qualité des textures et de facétisation des images virtuelles.

La fonction du son est clairement définie dans le cas du cinéma [SOR 92] : la parole condense le récit, les bruitages permettent de renforcer les objets et les personnages, la musique crée une atmosphère. Malgré le coté caricatural de cette définition nous retiendrons ce découpage (Figure 6) tout en prenant en compte le fait que le son est une dimension à part entière et non simplement un appui à l'image pour préciser ou augmenter son sens. Ainsi la musique a un impact sur l'émotion ressen-

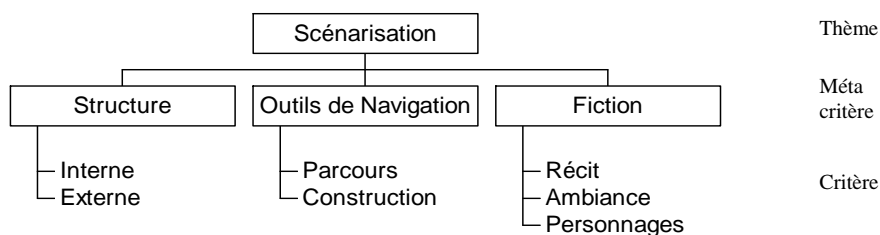
tie par l'utilisateur [CHI 94], et donc sur la façon dont il aborde sa session de travail. La parole est caractérisée par une évanescence qui n'en fait pas un support idéal pour transmettre des informations trop complexes, par contre elle est particulièrement adaptée pour faire passer un message, des consignes, pour fixer la concentration de l'utilisateur ou accompagner une information textuelle. Enfin les bruitages sont particulièrement utiles pour interpeller l'attention de l'utilisateur sur les actions que lui ou le logiciel manifeste.

Cette « nomenclature » des documents nous permet de caractériser les médias utilisés par le didacticiel. Chaque média présente un certain nombre de caractéristiques qu'il convient de mettre à profit pour en optimiser l'usage. Le méta-critère Interdépendance évalue la pertinence de leurs utilisations en terme de choix de média, de compétition entre les médias etc. (Figure 6).

### 3.6 Le thème Scénarisation

Dans le cadre d'un didacticiel multimédia, nous définissons la Scénarisation comme le procédé particulier qui consiste à structurer les documents de façon à en préparer au mieux la lecture, au sens de l'appropriation, par l'utilisateur. Le thème Scénarisation ne prend pas directement en compte les informations à transmettre, mais la façon dont elles sont agencées. Dans le cas d'un logiciel de diffusion, il s'agira donc principalement des outils de navigation au sein des différents documents. Dans celui des sites Web, il faudra prendre en compte la structuration interne au site et la navigation entre sites. Mais la scénarisation peut être beaucoup plus contextualisée. En effet, certains logiciels utilisent des métaphores ou des fictions pour agencer leurs contenus et il convient de considérer cette autre forme de parcours du logiciel.

Nous abordons donc ce thème selon trois approches complémentaires : l'organisation des données avec le méta-critère Structure, les moyens permettant de parcourir cette structure avec le méta-critère Outils de navigation, et enfin le cadre dans lequel cette navigation a lieu avec le méta-critère Fiction (Figure 7, Tableau 4).



**Figure 7.** Développement du thème Scénarisation

<b>Méta-critère</b>	<b>Définition</b>
Structure	Ce méta-critère concerne l'organisation des informations dans le didacticiel.
	<i>La structure et des liens qui la composent suit-elle une logique propre ? Le didacticiel offre-t-il des liens vers Internet ?</i>
Outils de navigation	Ce méta-critère évalue la qualité des outils permettant à l'utilisateur d'appréhender la structure.
	<i>Le logiciel propose-t-il à l'utilisateur des outils lui permettant de savoir où il en est et ce qui lui reste à faire ?</i>
	<i>L'utilisateur a-t-il la possibilité d'annoter les documents auxquels il accède ?</i>
Fiction	Ce méta-critère mesure la qualité et la pertinence de l'utilisation d'une dimension imaginaire dans le parcours du didacticiel ou dans la présentation des contenus.
	<i>Le récit permet-il de structurer l'apprentissage ?</i>
	<i>L'utilisation du didacticiel est-elle génératrice d'émotion ?</i>

**Tableau 4.** Définition et exemples de questions des méta-critères du thème Scénarisation

Le support numérique offre la possibilité de structurer librement les documents. Dans les logiciels de diffusion, les documents sont principalement présentés sous forme d'arborescences. Par contre, dans les sites Web par exemple, la structure peut devenir beaucoup plus complexe au risque de générer une perte de cohérence. Outre la structure interne au logiciel, notre évaluation portera aussi sur l'intégration du logiciel dans une structure plus vaste : site à l'intérieur d'Internet, organisation de plusieurs logiciels etc.

Nous avons estimé nécessaire d'évaluer les outils de navigation permettant à l'utilisateur de gérer cette structure. Ces outils concernent notamment la gestion des parcours (index, résumés, cartes dynamiques de ce qui a été parcouru, navigation contextuelle, etc.) et de réécriture des contenus (carnets de notes, hyperliens personnalisés, annotations, bilans, etc.). A notre sens dès lors que l'utilisateur dispose de tels outils, la structure peut retrouver sa cohérence et son exploitation devient réellement possible.

Le méta-critère Fiction concerne la dimension imaginaire proposée dans certains logiciels. Dans les logiciels de type parascolaire par exemple, les produits destinés aux jeunes sont systématiquement construits autour d'une fiction, d'une mascotte ou d'une ambiance particulière : un accompagnateur qui guide l'apprenant, une île dont il faut s'échapper en résolvant des problèmes ou des expériences proposées dans le laboratoire d'un savant fou par exemple. Sans prétendre que ce genre de métaphores convient à tout le monde et permet de tout apprendre, nous admettons que la fiction est un facteur d'implication et d'autonomie (l'utilisateur participe à l'histoire), de cohérence et d'appropriation (elle fixe un cadre général qui canalise la lecture). L'utilisation d'une fiction adaptée est donc un moyen de proposer l'utilisation du didacticiel comme un loisir, d'où une forte présence dans le parascolaire. Notre

évaluation porte principalement sur la cohérence et la pertinence du récit, des personnages et des ambiances mis en place dans le logiciel.

### 3.7 Le thème Outils Pédagogiques

Le contexte pédagogique des logiciels que nous évaluons est pris en compte dans l'ensemble des thèmes que nous avons retenus, mais le thème Outils Pédagogiques traite spécifiquement de ces aspects. Ce thème est généralement absent des méthodes d'évaluation de logiciels, notamment à cause de la difficulté à le traiter hors contexte. Nous le considérons donc comme un apport fort de notre méthode par puissance de caractérisation. C'est pourquoi nous allons le développer plus que ceux cités précédemment.

Dans ce thème, la dualité caractérisation / évaluation est encore plus présente que dans les autres thèmes. En effet, il est très difficile de juger de la qualité d'un logiciel sans une connaissance préalable du contexte dans lequel il va être utilisé. Ainsi les méta-critères et les critères de ce thème sont orientés vers une caractérisation du logiciel, la question de la pertinence de ces caractéristiques étant le fait de l'évaluateur.

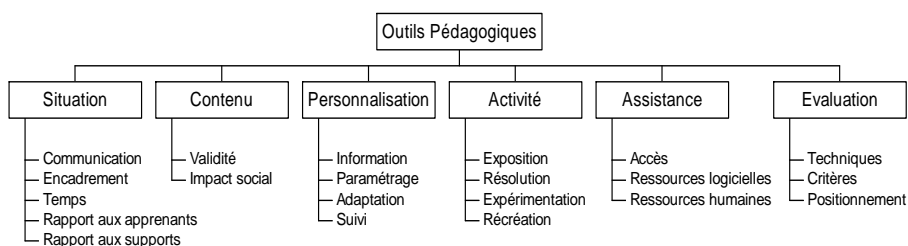


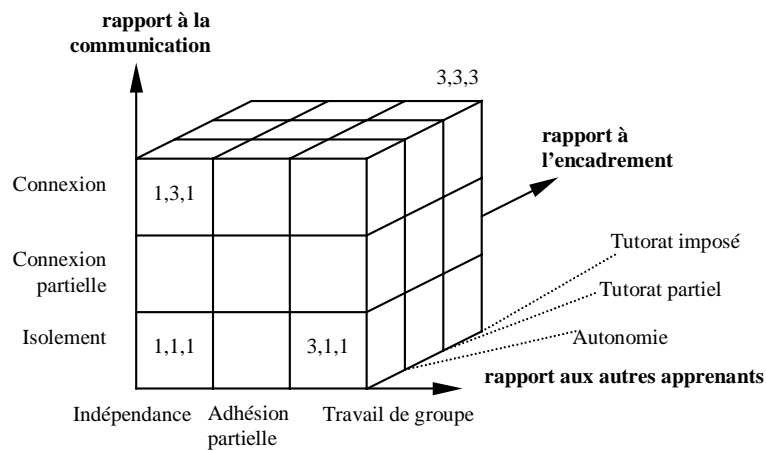
Figure 8. Développement du Thème Outils Pédagogiques

Méta-critère	Définition
Situation	Ce méta-critère correspond à la description et à la mesure de pertinence des situations techniques (communication, travail de groupe etc.) prévues par le didacticiel.
Contenu	Mesure la validité et la qualité du contenu.
Personnalisation	Evalue la qualité et la richesse des outils permettant une personnalisation de l'apprentissage.
Activité	Evalue la pertinence et la richesse des activités proposées aux apprenants.
Assistance	Caractérise et évalue la qualité et la richesse des outils d'aide pédagogique proposés à l'apprenant.
Evaluation	Caractérise et évalue la qualité et la bonne adéquation des outils et des dimensions d'évaluation.

Tableau 5. Définition des méta-critères du thème Outils Pédagogiques

### 3.7.1 Le méta-critère Situation

Ce méta-critère permet à l'évaluateur de décrire une partie de l'organisation technique du dispositif d'apprentissage. L'utilisation des réseaux en situation d'apprentissage a multiplié les possibilités organisationnelles. On peut donc être confronté aux relations les plus simples (un apprenant devant un ordinateur), comme aux situations les plus complexes (notions de groupes, problèmes de communication, tutorat à distance, etc.). Pour faciliter l'identification d'une situation type, nous utilisons une matrice (Figure 9) décrivant certains de ces aspects.

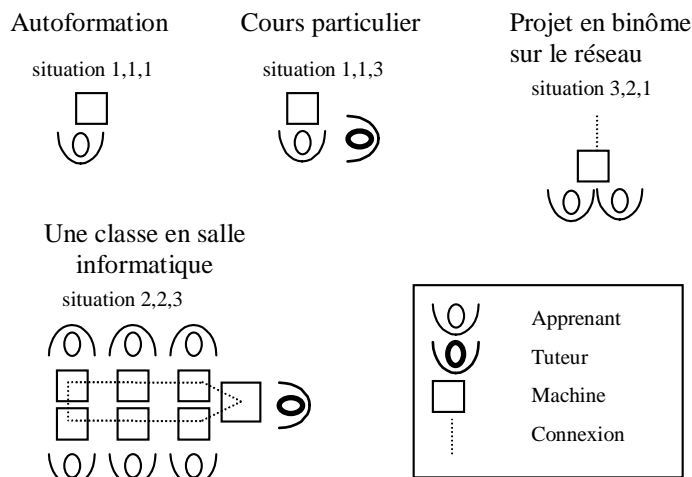


**Figure 9.** Matrice des situations d'apprentissage

Cette matrice permet de définir la situation d'apprentissage selon trois axes qui correspondent à trois critères (Tableau 6) :

- Le rapport à la communication qui permet d'identifier la situation en terme de communication et d'interaction entre l'apprenant et le système informatique.
- Le rapport aux autres apprenants qui caractérise l'intégration ou non de l'apprenant dans un groupe.
- Le rapport à l'encadrement qui mesure le suivi pédagogique de l'apprenant.

Afin d'illustrer cette matrice, nous donnons quelques exemples de situations typiques avec leurs indices correspondant dans la matrice (Figure 10).



**Figure 10.** Exemple de situations d'apprentissage types avec les indices correspondants

Cette description technique du dispositif ne se veut pas exhaustive. Nous la complétons avec deux autres notions : l'aspect temporel de l'utilisation (critère Temps) et le rapport entre le support informatique et les autres supports d'apprentissage plus traditionnels comme les livres ou les vidéos (critère Rapport aux supports).

En ce qui concerne le rapport au temps, deux aspects sont complémentaires. D'une part l'usage, un logiciel peut être utilisé sur des périodes courtes (exercice par exemple), à l'échelle de l'heure (un module de cours) ou dans un dispositif complet de formation (didacticiel utilisé toute l'année). D'autre part la communication, le dispositif pédagogique peut entraîner une utilisation discrète, synchrone ou asynchrone, et cela sur les trois axes décrits par la matrice précédente.

Critère	Définition et exemples de questions associées
Communication	Caractérise et mesure la qualité et la richesse des outils de communications utilisés ou intégrés par le didacticiel.
	Quelles sont les possibilités de communication entre les apprenants ? Caractériser la synchronisation des outils de communication proposés aux apprenants.
Encadrement	Caractérise et évalue la qualité des outils d'encadrement dont dispose le tuteur au sein du didacticiel.
	Quel est le type d'encadrement privilégié par le didacticiel ? Le tuteur peut-il diriger l'utilisation du logiciel à distance ?
Temps	Caractérise et mesure la qualité de prise en compte des contraintes temporelles des apprenants.
	Quel est le temps standard d'une session du didacticiel ? Le didacticiel permet-il de gérer le temps écoulé entre les sessions ?

Rapport aux apprenants	Caractérise les possibilités d'utilisation du logiciel en terme de <b>manipulation de groupe</b> .
	Le produit est-il prévu pour être exploité par plusieurs personnes sur une même interface ? Le logiciel est-il approprié à l'utilisation coopérative par plusieurs utilisateurs sur des machines différentes ?
Rapport aux supports	Mesure la souplesse d'utilisation et la complémentarité du didacticiel avec d'autres supports.
	Le didacticiel peut-il être utilisé de façon autonome, sans autre support de formation ? Le didacticiel peut-il s'articuler avec d'autres supports de formation ?

**Tableau 6.** Définition des critères et exemples de questions du méta-critère Situation

### 3.7.2 Contenu

Notre approche ne nous permet pas de prendre en considération la discipline concernée par le didacticiel. Nous nous contentons donc d'attirer l'attention de l'évaluateur sur la validité des informations d'une part (mise à jour du contenu, variété, granularité etc.) et sur le fait que ces informations doivent respecter un certain nombre de valeurs sociétales (Tableau 7). Outre les grands sujets des droits de l'homme (racisme, sexisme, travail des enfants, esclavagisme etc.), il convient d'interpeller les évaluateurs sur la prise en compte et le respect des critères socioculturels de la population cible [NEN 99].

Critère	Définition et exemples de questions associées
Validité	Mesure la qualité a priori et la pertinence du contenu vis-à-vis des objectifs affichés par le didacticiel.
	Relève-t-on des erreurs dans le contenu des informations présentées à l'apprenant ? Le degré d'approfondissement du contenu est-il pertinent par rapport au niveau des apprenants ?
Impact social	Concerne le respect des critères socioculturels.
	Le didacticiel respecte-t-il le statut des individus évoqués (âge, sexe, etc.) ? Le didacticiel respecte-t-il les cultures et les coutumes de la population cible ?

**Tableau 7.** Définition des critères et exemples de questions du méta-critère Contenu

### 3.7.3 Personnalisation

Le support informatique a cela de différent, comparé entre autre au livre ou à la cassette vidéo, que l'aspect calculatoire des traitements qu'il autorise lui permet potentiellement de s'adapter de façon dynamique à l'utilisateur. Dans un contexte pédagogique où la personnalisation de la démarche est primordiale, nous attirons l'attention de l'évaluateur sur les outils disponibles en ce sens dans le logiciel évalué. Nous avons regroupé ces outils de personnalisation selon quatre approches complémentaires (Tableau 8) : le critère Cadrage, le critère Paramétrage, le critère Adaptation et enfin le critère Suivi.

<b>Critère</b>	<b>Définition et exemples de questions associées</b>
Cadrage	Concerne la description de la population cible du didacticiel.
	Des informations concernant les pré-requis et le niveau des apprenants sont-elles disponibles ?
	Des informations concernant les objectifs de formation sont-elles disponibles ?
Paramétrage	Mesure la richesse et la pertinence des moyens mis à la disposition des utilisateurs pour adapter le didacticiel à son contexte d'utilisation.
	Le tuteur dispose-t-il de paramètres lui permettant d'adapter le contenu à son enseignement ?
	Le didacticiel permet-il de respecter le rythme de chaque apprenant à travers la gestion des sessions ?
Adaptation	Concerne la capacité d'adaptation dynamique du didacticiel en fonction du déroulement de l'apprentissage.
	Le didacticiel s'adapte-t-il en fonction du parcours au sein des connaissances ?
	Le didacticiel s'adapte-t-il en fonction des résultats d'évaluation ?
Suivi	Caractérise et évalue les outils permettant un traitement dans le temps des informations.
	Le didacticiel gère-t-il un historique de l'apprentissage ?
	Le didacticiel permet-il d'analyser les données de suivi d'un groupe d'apprenants ?

**Tableau 8.** *Définition des critères et exemples de questions du méta-critère Personnalisation*

#### 3.7.4 Activité

Notre méthode s'applique à définir les différents types d'activités proposées à l'apprenant ainsi que les stratégies mises en œuvre au sens large du terme (structure traditionnelle cours/TD/TP ; formation à base de cas ; apprentissage par découverte, etc.). Le fait de permettre à un apprenant d'agir est un atout potentiel important dans la mesure où cela permet de confronter des informations avec une réalité, ce qui est intéressant du point de vue pédagogique (Vygotsky, Piaget, ...). Deux questions fondamentales à se poser par rapport à cette problématique sont proposées par Martial Vivet : « Le scénario crée-t-il de l'interactivité ? L'activité générée est-elle utile à l'apprentissage ? » [VIV 96]. Les approches actuelles visent à accroître les possibilités d'action de l'utilisateur, en passant du niveau d'exploration basique, à un niveau d'expérimentation [PER 98] voire de création [LEG 98].

Nous avons regroupé les activités de l'apprenant selon quatre critères (Tableau 16) : les activités de lecture au sens large du terme (assimilation de documents), les activités de résolution (raisonnements), les activités de type expérimentation (application de savoir-faire) et les activités ludo-éducatives dans le critère Récréation.

<b>Critère</b>	<b>Définition et exemples de questions associées</b>
Exposition	Qualité de la présentation des contenus : multiplication des points de vue, approfondissement, mise en scène etc.

	Est-ce qu'un même contenu est proposé selon des points de vue variés ? Est-ce que les contenus peuvent être plus ou moins approfondis ?
Résolution	Richesse et pertinence des activités de résolution (exercice, problème etc.) proposées : niveau de difficulté, variété etc.
	Est-ce que le didacticiel propose des exercices de natures et de formes différentes ? Le didacticiel propose-t-il des exercices de difficultés variées ?
Expérimentation	Richesse et pertinence des activités de manipulation (TP, expérimentation, création etc.) proposées : degrés de liberté, variété etc.
	Le didacticiel stimule-t-il la créativité et l'activité des utilisateurs ? Le didacticiel permet-il d'assister à des expériences ou à des cas réels ?
Récréation	Qualité des activités ludiques ou ludo-éducatives proposées.
	Le didacticiel propose-t-il des activités récréatives ou ludiques aux apprenants ? Quelle est la qualité des activités récréatives en termes d'apprentissage par le jeu ?

**Tableau 15.** Définition des critères et exemples de questions du méta-critère Personnalisation

### 3.7.5 Assistance

Après les outils d'assistance "automatique" (tuteurs intelligents, outils de résolution automatique etc.) dont l'efficacité reste discutable, les aides logiciels s'orientent vers des outils de médiation entre l'apprenant et les ressources matérielles ou humaines. Car en dehors des capacités d'aide internes au didacticiel, il convient de vérifier aussi si des outils de médiation vers une assistance extérieure sont proposés. Dans le cadre de certains dispositifs de formation (formation à distance ou groupe de projet par exemple) ces possibilités sont même primordiales : demande d'aide par courrier électronique, demande d'intervention du tuteur par le didacticiel, forum, etc. Ici encore, le but de notre questionnaire sera donc de passer en revue les différents types d'outils d'assistance existant à travers trois critères (Tableau 10). Le critère Accès caractérise les types d'aide et les moyens mis à disposition de l'apprenant pour y accéder : aide contextuelle, agent, requête en langage naturel par exemple. Le critère Ressources matérielles concerne la richesse et les types de ressources disponibles : banque de données, support méthodologique, logiciels annexes, assistants etc. Enfin, le critère Ressources humaines caractérise et évalue la pertinence des dispositifs mis en place pour requérir de l'aide extérieure ou pour permettre à l'enseignant d'intervenir dans la formation : courrier électronique, forum de questions, gestion de la présence du tuteur etc.

Critère	Définition et exemples de questions associées
Accès	Caractérise et mesure la pertinence des moyens d'accès à l'aide mis en place dans le logiciel.
	Le didacticiel propose-t-il une aide spontanément ? Le didacticiel permet-il de fournir une aide correspondant à une requête de l'utilisateur ?
Ressources matérielles	Concerne la richesse et la qualité des aides méthodologiques apportées : approfondissement pas à pas, description de méthodes etc.

	Le didacticiel propose-t-il des renvois vers le contenu théorique depuis les exercices ou expérimentations ? Le didacticiel propose-t-il une ou plusieurs aides méthodologiques pour aider l'apprenant à traiter ses problèmes ?
Ressources humaines	Caractérise et évalue la pertinence des moyens d'accès à des personnes ressources. Le didacticiel prévoit-il de faire appel à des compétences extérieures ? Le didacticiel propose-t-il à l'enseignant d'intégrer ses interventions dans le parcours du logiciel ?

**Tableau 10.** Définition des critères et exemples de questions du méta-critère Assistance

### 3.7.6 Evaluation

L'évaluation est l'un des aspects majeurs de l'apprentissage. Mais cette évaluation peut revêtir plusieurs formes selon sa position temporelle dans le cursus (calibrage, contrôle continu, évaluation finale), selon le public auquel s'adressent les résultats (auto-contrôle, exercices corrigés, examen) et selon les indicateurs utilisés (taux de réponse, taux d'erreur, temps de réponse etc.). Les techniques d'évaluation sont donc très diverses : QCM, questionnaires indications de confiance, actions réflexes, désignation, analyse syntaxique, construction, contrôle de simulation, etc.

Le choix de ces dimensions et de ces techniques dépendra principalement du type de savoir que l'on désire enseigner (savoir-faire, procédure, conceptualisation) mais aussi de la définition même donnée à la notion de compétence : Considère-t-on, par exemple, qu'il est pire de donner une réponse fautive que pas de réponse du tout ? Quelle est la place de l'inaction dans le contrôle d'une simulation ? Le méta-critère Evaluation est donc divisé en trois critères (Tableau 11) : le critère Technique qui caractérise les types d'évaluation présents dans le logiciel ; le critère Critère qui décrit les dimensions utilisées par les outils d'évaluation ; et le critère Positionnement qui mesure la pertinence de l'évaluation par rapport aux objectifs (bilan de compétences etc.) ou aux références institutionnelles (niveau scolaire, certificat etc.).

Critère	Définition et exemples de questions associées
Technique	Ce critère caractérise les types d'évaluation mis en place dans le logiciel : auto-évaluation, contrôle continu etc. Le didacticiel permet-il à l'apprenant de s'auto-évaluer ? Le didacticiel prévoit-il l'intervention du tuteur dans les phases d'évaluation ?
Critère	Ce critère décrit et évalue la pertinence des dimensions évaluées par le logiciel : mémoire, rapidité, confiance en soi etc. Quelle est la diversité des caractéristiques évaluées chez l'apprenant ? Les dimensions évaluées par le didacticiel sont-elles en accord avec les objectifs d'apprentissage ?
Positionnement	Ce critère caractérise les possibilités de positionnement des connaissances de l'apprenant par rapport à lui-même ou par rapport aux niveaux institutionnels.

	Le didacticiel est-il en mesure de fournir un bilan de compétence ? Le didacticiel donne-t-il la possibilité de se positionner par rapport à un niveau académique ?
--	--

**Tableau 11.** *Définition des critères et exemples de questions du méta-critère Evaluation*

Nous venons de présenter la critériologie que nous utilisons dans nos évaluations ainsi que quelques exemples de questions. Nous allons maintenant décrire comment nous utilisons cette structure et quels sont les résultats fournis par notre méthode.

#### **4. Manipulation du questionnaire et calcul des résultats**

##### **4.1 Un questionnaire adaptatif**

L'arborescence de thèmes, de méta-critères et de critères que nous venons de présenter est constituée, tant au niveau des nœuds que des feuilles, par les questions proprement dites. Actuellement, il y a plus de deux cent cinquante questions dans notre corpus et le temps qu'il faudrait pour répondre à toutes est beaucoup trop long pour être acceptable dans certaines situations. Afin d'optimiser ce temps et du fait que notre approche de l'évaluation tend vers une interaction entre l'outil et l'évaluateur, notre questionnaire s'adapte dynamiquement à l'évaluation en cours. Deux caractéristiques de notre méthode permettent cette adaptation : le parcours au sein du questionnaire est libre et la structure de l'arborescence est modifiée dynamiquement.

Tout d'abord, le parcours est libre, ce qui permet à l'évaluateur, d'approfondir un point précis, en développant les sous-niveaux ou de passer directement à la notion suivante, selon ses besoins, ses compétences ou la pertinence de cette notion dans le contexte du logiciel particulier. Ensuite la structure du questionnaire peut-être modifiée dynamiquement. Elle peut l'être par l'évaluateur qui a la possibilité d'indiquer si tel ou tel critère est essentiel, normal ou sans objet, auquel cas les sous-niveaux et les questions correspondantes sont supprimés. La structure peut aussi être modifiée par les réponses données par l'évaluateur. Prenons un exemple : dans le thème Documents Multimédias si à la question "Le logiciel propose-t-il des séquences vidéos ?", la réponse est "Non", alors dans le critère Contrôle la question "Est-ce que des boutons de contrôle des séquences vidéos sont proposées ? (pause, stop, retour etc.)" sera automatiquement supprimée. Ainsi les réponses données au fur et à mesure par l'évaluateur modifient la structure ou l'importance des nœuds de l'arborescence.

## 4.2 Prise en compte de la subjectivité de l'évaluateur

Une dimension est souvent oubliée dans les méthodes d'évaluations existantes : la perception subjective qu'a l'utilisateur du logiciel. Ces impressions nous semblent primordiales dans le contexte éducatif [HU 99] et cela à plusieurs titres.

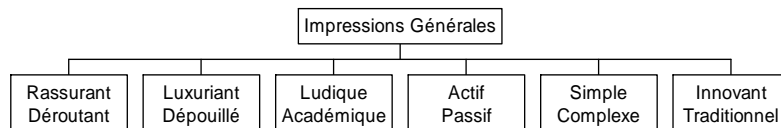
La séduction, le plaisir d'utilisation ont un impact évident sur la motivation des apprenants et donc sur l'apprentissage. Or le caractère subjectif de ces aspects, nous oblige à nous en remettre à l'opinion de l'évaluateur. Deux hypothèses rendent cette opinion valide dans notre approche. D'une part l'évaluateur est considéré comme un utilisateur, ne serait-ce que pendant le cours laps de temps où il teste le logiciel. Il est donc à même de fournir ses impressions sur le logiciel. D'autre part, nos évaluateurs sont les prescripteurs des logiciels évalués, nous posons comme hypothèse qu'ils ont une connaissance suffisante du public d'apprenant pour que leurs opinions soient pertinentes.

Un autre argument en faveur de la prise en compte de l'avis subjectif de l'évaluateur concerne l'hétérogénéité des compétences des évaluateurs. Les prescripteurs, les enseignants ne sont pas tous des spécialistes en ergonomie, en informatique ou en multimédia. Il est pratiquement impossible de leur fournir un questionnaire détaillant toutes les caractéristiques de tous les logiciels de formation possible. En permettant à l'évaluateur de donner son avis et en l'aidant à le faire, nous augmentons nos chances de détecter un défaut dont l'évaluateur ressent les conséquences sans clairement en identifier les causes avec les questions proposées.

Deux outils permettent, dans notre méthode, de prendre en compte l'opinion de l'évaluateur en cours d'évaluation : le thème Impressions Générales et les Notes Instinctives.

### 4.2.1 Le thème Impressions Générales

Le rôle de ce thème est de capitaliser en début d'évaluation, les impressions ressenties par l'évaluateur lors de l'utilisation du logiciel. Pour ce faire nous utilisons une série d'adjectifs (Figure 11) issus de l'analyse de divers corpus d'évaluation<sup>1</sup> et des théories de la perception visuelle (notamment celle des affordances [GIB 79]). Ces adjectifs sont les plus neutres possibles pour qu'aucun jugement de valeur ne viennent perturber la caractérisation du logiciel à ce niveau.



**Figure 11.** Développement du thème Impressions Générales

<sup>1</sup> La mise au point de cette liste est entre autre détaillée dans [Crozat, Trigano, Hu 99]. Les corpus utilisés sont issus de nos propres expériences et d'une série d'évaluations effectuée au Canada et disponi-

Ce thème n'est pas construit sur le même modèle que les autres puisqu'il ne possède qu'un niveau correspondant à des couples de méta-critères. Les questions correspondantes sont du type : « Comment avez-vous trouvé le logiciel ? » et une réponse possible serait « Très Rassurant », « Plutôt Dépouillé », « Très Académique », « Plutôt actif », « Très Complexe » et « Plutôt Traditionnel ». Cette description fournit donc une image du logiciel dont l'évaluateur pourra, après son évaluation, vérifier la bonne adéquation avec la situation pédagogique envisagée.

#### 4.2.2 Les Notes Instinctives

En plus de son avis global sur le logiciel, l'évaluateur peut, lors de son parcours de l'arborescence, donner une note dite instinctive, à chaque thème, méta-critère ou critère. Cette note n'est pas obligatoire, mais elle permet à l'évaluateur, soit de noter un aspect sans avoir à l'approfondir s'il s'estime compétent pour le faire, soit de capitaliser son opinion sur cet aspect en vue d'une confrontation avec le résultat des questions proprement dites. Cette note instinctive est toujours donnée sur une échelle qualitative ( -- ; - ; + ; ++ ) sans position neutre pour contraindre l'évaluateur à prendre une décision. Nous présentons dans le chapitre suivant comment ces notes instinctives sont utilisées dans les résultats finaux.

#### 4.3 Calcul des notes

Une fois le questionnaire rempli, la méthode fournit à l'évaluateur un rapport lui permettant d'appréhender son évaluation de façon globale.

Tout d'abord une note est calculée pour chaque nœud de l'arborescence. Au niveau des questions la note est donnée soit directement par la réponse indiquée, soit en combinant les réponses d'une question double caractérisation/évaluation (Figure 12). Notons que l'évolution des notes en fonction des réponses n'est pas obligatoirement linéaire. Afin d'accentuer la détection de certains défauts majeurs, une notation dite exponentielle est appliquée aux questions clés.

<b>Exemple de question double</b>				
<b>Caractérisation</b> : « Une aide sur les possibilités de l'interface est-elle proposée ? »				
<b>Evaluation</b> : « Que pensez-vous de l'aide apportée ? »				
	<i>Evaluation</i>			
<i>Caractérisation</i>	Très bonne	Bonne	Mauvaise	Très mauvaise
Toujours	+10	+6	+2	-2
Souvent	+6	+2	-2	-6
Parfois	+2	-2	-4	-10
Jamais	-10	-10	-10	-10

ble sur Internet à l'adresse suivante « <http://www.cyberscol.qc.ca/BD/MEQ> ». En ce qui concerne la méthode d'analyse, on se reportera à [LAM 97].

**Figure 12.** *Exemple de notes associées à une question double caractérisation/évaluation*

Au niveau des critères, des méta-critères et des thèmes, les notes correspondent à la moyenne des notes des sous-niveaux, pondérées par leur poids respectif (essentiel, normal ou sans objet, voir chapitre 4.1).

Ainsi, l'évaluateur dispose des notes calculées et des notes instinctives fournies au fur et à mesure du parcours. Ces notes sont mises en correspondance et accompagnées de deux indices permettant de gérer les écarts éventuels entre notes instinctives et notes calculées :

- L'indice de cohérence qui indique à l'évaluateur s'il a été logique dans ses évaluations instinctives entre deux niveaux (par exemple, entre le méta-critère et les critères qui le composent).

- L'indice de corrélation qui met en valeur la ressemblance entre la note instinctive et la note calculée.

$$I_{cohérence} = Moy_i \left( \frac{|NI - NI_i|}{10 + \max(NI, NI_i)} \right); \quad I_{corrélation} = \frac{Min(NI, NC)}{Max(NI, NC)} \text{ avec}$$

$NI$  : Note instinctive du niveau

$NI_i$  : Note instinctive du  $i$ ème sous niveau

$NC$  : Note calculée du niveau

**Figure 13.** *Calcul de l'Indice de cohérence et de l'Indice de corrélation*

A partir du rapport, l'évaluateur fixe pour chaque critère la note finale qu'il estime la plus juste et qui sera le fruit de la comparaison entre note instinctive et note calculée. Notons que la méthode propose à l'évaluateur une note finale qui effectue une moyenne pondérée entre notes calculées et instinctives, tout en sachant que les notes instinctives sont d'autant plus prises en compte qu'elles sont cohérentes entre elles. Cette note finale n'étant qu'une proposition, elle est modifiable par l'évaluateur. Une fois ces notes fixées, une visualisation graphique des résultats est fournie.

## 5. La version logicielle

Un prototype logiciel mettant en œuvre la méthode a été développé. Il reprend l'ensemble des caractéristiques que nous avons décrites : une arborescence des critères dans laquelle l'évaluateur se déplace librement, la possibilité de donner une note instinctive et celle d'indiquer l'importance du nœud en cours ; la capitalisation des impressions générales de l'évaluateur ; la proposition d'une note finale accompagnée d'un bilan de l'évaluation effectuée sur cet aspect ; et enfin une représentation graphique des résultats.

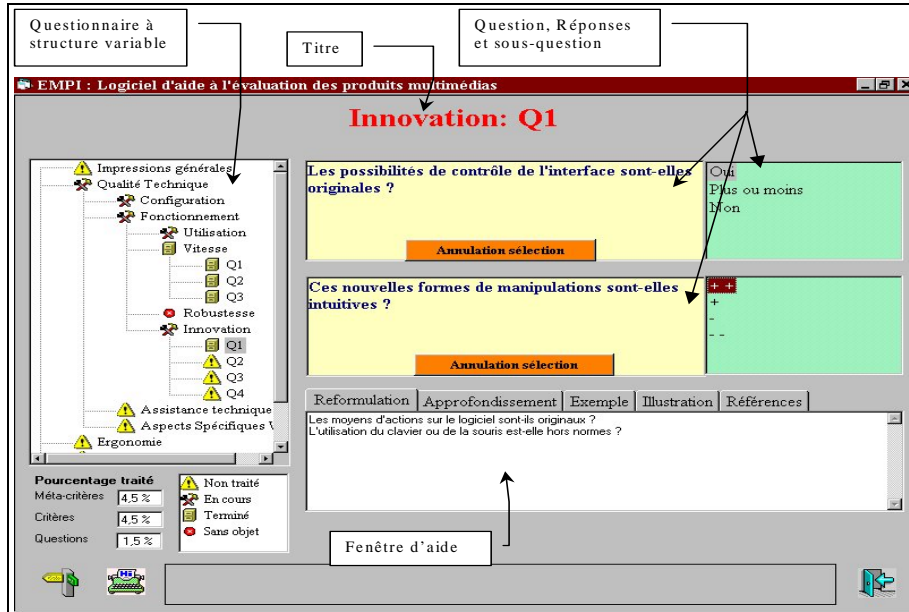


Figure 14. Interface standard de parcours du questionnaire

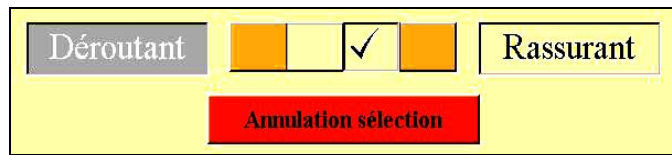


Figure 15. Exemple de fenêtre de description des Impressions Générales

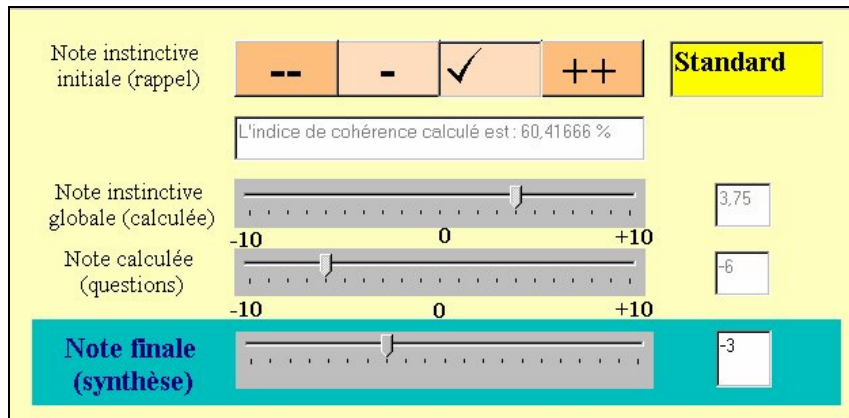


Figure 16. Exemple de fenêtre de proposition de note finale

## 6. Validation

### 6.1 Conception/Evaluation de notre méthode

Plusieurs versions du questionnaire ont été mises au point, [HU 98a] et plusieurs validations ont donc été menées [HU 98b] :

- Une dizaine d'évaluateurs de compétences diverses ont appliqué la méthode sur une trentaine de logiciels. Ces évaluations ont été comparées à des analyses approfondies des logiciels afin de tester la qualité des évaluations.
- Une cinquantaine d'évaluateurs (étudiants du second cycle) ont été comparés sur l'évaluation de deux logiciels pour mesurer les disparités entre évaluations.
- Les méthodes EMPI et MEDA [MED 97] ont été confrontées.
- La critériologie d'EMPI a été comparée aux critères édités par le ministère de l'éducation national (certification RIP) et aux questionnaires de la méthode Méthane développée par l'OVAREP [OVA 92].

Ces expériences correspondent à des validations effectuées dans le cadre d'une conception itérative et non à des études terrains avec des évaluateurs en situation réelle. A ce niveau, nos objectifs étaient donc multiples : valider les techniques de calcul utilisées, tester la stabilité de nos évaluations et vérifier l'utilisabilité de notre questionnaire avant de le soumettre à des évaluateurs réels.

Outre l'amélioration cyclique de notre méthode, plusieurs résultats importants sont apparus à ce stade :

- L'application de la notation exponentielle et des pondérations augmente de façon significative la stabilité d'évaluation (les différences sur les points secondaires sont atténuées) et améliore la détection des principaux défauts du logiciel (la présence d'un défaut n'est plus noyée dans la masse des points évalués).
- L'utilisation de notes chiffrées comme unique forme de résultat ne permet pas de rendre compte de tous les aspects de l'évaluation. Il est par exemple difficile de faire la différence entre une caractéristique mauvaise et l'absence de cette caractéristique dans le logiciel évalué.
- L'utilisation d'un arbre dynamique, dont le parcours est libre, offre une souplesse non négligeable dans la mise en place de notre méthode dans une plate-forme d'évaluation : évaluation selon différents niveaux de détails, temps d'évaluation acceptable (moins de 2h), possibilité de cibler l'évaluation sur certains aspects.
- La confrontation entre l'opinion subjective et l'évaluation proprement dite améliore le réalisme de l'évaluation, à savoir la correspondance entre la description critique issue de l'évaluation et la réalité du logiciel analysé par des experts.

Après cette phase de mise au point, il s'agit de valider notre approche sur le terrain en situation réelle, travail que nous réalisons actuellement.

## 6.2 Mises en situation

### 6.2.1 Tester l'utilisabilité de la méthode par des enseignants

Il s'agit de vérifier si l'approche de l'évaluation que nous proposons correspond à notre public cible, les enseignants. Nous avons pour cela fait appel à des enseignants de différents niveaux (instituteur, professeur de collège, professeur agrégé et enseignant de 3<sup>ème</sup> cycle universitaire) et de différentes disciplines (langues, informatiques, enseignements techniques etc.). Le but de cette expérimentation est double : d'une part vérifier auprès d'utilisateurs potentiels, la pertinence et la facilité d'utilisation de notre méthode ; d'autre part tester les limites de notre méthode sur une plus grande gamme de logiciels et de situations pédagogiques.

Sur cette question des limites de notre méthode, plusieurs résultats apparaissent déjà. D'une part la description des logiciels cibles que nous avons donnée en début d'article est confirmée. En effet l'évaluation de logiciels outils (traitement de textes, manipulation de données etc.) ou de micro-monde ne donne pas de résultats pertinents. Par contre l'application de la méthode à des logiciels culturels ou ludoculturels (jeux de type Versailles ou Egypte Ed. Cryo) semble prometteuse, d'autant plus que ces logiciels sont effectivement utilisés en salle de classe.

### 6.2.2 Vérifier la fiabilité des évaluations sur des cas réels

Plusieurs expériences sont effectuées afin de vérifier si les résultats fournis par notre méthode d'évaluation concordent avec l'utilisation réelle du logiciel dans un processus pédagogique. L'une d'elle correspond à une enquête menée auprès de différents collèges ayant tous acquis un même logiciel, à savoir « Perfect Your English » coédité par le CNDP. Cette contrainte sur l'usage d'un même produit a rendu relativement difficile de regrouper un grand nombre d'établissements. Notre expérience porte au total sur une quinzaine d'établissements.

#### *Objectif*

Le but de cette validation est de confronter les résultats obtenus avec notre méthode et les réalités des usages sur le terrain. Il s'agit donc de vérifier que les usages effectivement constatés font partie de ceux décrits par notre méthode. Mais également si les défauts ou les qualités détectés a priori par notre méthode correspondent à la réalité, et dans quelle mesure ces résultats sont-ils proportionnels à la gêne ou à la satisfaction observée.

#### *Choix des évaluateurs*

Le choix de l'évaluateur ou des évaluateurs qui vont établir l'évaluation référence est problématique. En effet, plusieurs biais peuvent apparaître selon la qualité de l'évaluateur (expert de la méthode ou enseignant) et selon le nombre d'évaluations prises en compte (évaluation unique ou moyenne de résultats) :

- évaluateur expert : mauvaise connaissance des situations pédagogiques.
- évaluateur enseignant : mauvaise connaissance du logiciel (non mesurée ici).
- évaluation unique : proche d'une situation réelle mais subjectivité importante.
- moyenne d'évaluation : meilleure évaluation mais inaccessible en situation réelle.

Nous avons effectué une synthèse entre une évaluation faite par un expert (afin de couvrir l'ensemble des possibilités offertes par la méthode) et différentes évaluations effectuées par des enseignants (pour améliorer la pertinence de l'évaluation).

#### *Protocole*

Les principales étapes de nos enquêtes sont pour chaque établissement :

- 1- observation de la situation d'usage du logiciel : dispositif technique, organisation des sessions etc.
- 2- description par l'enseignant de la mise en place du logiciel : installation, insertion dans le cursus, buts etc.
- 3- description par l'enseignant des points forts et des points faibles du logiciel et de cette forme d'apprentissage.
- 4- si possible observation des élèves, sinon recueil d'informations sur l'usage des élèves via les enseignants ou les surveillants.
- 5- Reclassement de l'ensemble des données.
- 6- Comparaison avec les évaluations EMPI.

#### *Résumé des enquêtes et confrontations avec les évaluations<sup>1</sup>*

Observation terrain	Evaluation EMPI
<b>1- Observation de la situation</b>	
- Deux principaux contextes d'utilisation : soit en parallèle du cours (groupe de soutien et/ou de perfectionnement) ou en libre accès dans les centres de ressources (bibliothèques, club scolaire). Il n'y a qu'un seul cas où le logiciel est employé en cours.	- Les situations d'apprentissage envisageables : pas de travail de groupe, pas de liens avec d'autres supports, pas de prise en compte explicite du tuteur. - Un logiciel plutôt tourné vers l'autoformation.
<b>2- Description de la mise en place</b>	
- Le logiciel a été installé par le service informatique, l'enseignant ou même les élèves.	- Installation facile. - Pas de module de désinstallation. - Bonne documentation relativement complète depuis l'installation jusqu'à la proposition de scénario de mise en place à l'usage des enseignants. - Buts et pré-requis clairement identifiés.
<b>3- Description du logiciel</b>	
<b>4- Observation des usages</b>	
Pour les enseignants : - manque des possibilités de mise à jour. - difficile à intégrer en cours mais nécessite quand même une formation ou un accompa-	- De nombreux éléments actifs ne sont pas explicites. Les feed-back sont souvent pauvres ou absents. - la Charge de travail : des exercices nécessitent des manipulations trop complexes ; Les validations des exercices sont souvent

<sup>1</sup> Nous ne donnons ici qu'une synthèse des résultats, l'ensemble des évaluations et des comptes rendus d'interview sont disponibles en document interne ou sur simple demande aux auteurs [HU 00b].

<p>gnement des élèves. En ce qui concerne les apprenants :</p> <ul style="list-style-type: none"> <li>- de nombreuses options sont non utilisées.</li> <li>- la découverte et l'identification des affichages sont souvent difficiles sans aide.</li> <li>- il y a un temps d'adaptation parfois long pour acquérir les modes opératoires des exercices.</li> <li>- absence de scénario ou d'aide pertinente pour guider l'apprenant dans son exploitation et du logiciel et du contenu.</li> <li>- émulation entre apprenants mais une vision ludo-éducative parfois parasite : usage de "trucs" pour passer les "niveaux".</li> </ul>	<p>« doublées » par des étapes superflues.</p> <ul style="list-style-type: none"> <li>- l'Homogénéité : l'exécution de la commande Retour, présente dans toute l'application, reste difficile à prévoir sans aide complémentaire.</li> <li>- l'Adaptabilité : en dehors de la possibilité de changer la langue des commentaires (français/anglais) aucun paramètre n'est modifiable par l'utilisateur.</li> <li>- la Navigation : il est difficile de marquer, de refaire ou de trier les exercices à faire.</li> <li>- le Contenu : bonne qualité mais absence de mise à jour.</li> <li>- les Activités : exercices de type "scolaire", peu d'activité intégrant l'usage de l'enregistrement.</li> </ul>
---	---

### Conclusion

Cette enquête ne concerne qu'une quinzaine de collège, on ne peut donc pas généraliser. Néanmoins ces observations sont révélatrices des usages et plusieurs points de comparaison nous semblent importants :

- les défauts ergonomiques révélés par notre méthode ont effectivement eu des répercussions sur les usages : les élèves qui utilisent le logiciel seuls n'exploitent pas son potentiel faute de s'être rendu compte du rôle de certains objets de l'interface ; les autres ont eu besoin d'une prise en main plus ou moins longue.

- les contradictions qui étaient apparues lors de l'évaluation (usage en autoformation favorisée mais absence d'aide logiciel de qualité ; proposition d'usage du logiciel pour les enseignants mais aucuns moyens de modifier ou d'intégrer le logiciel) se retrouvent dans le discours des enseignants.

Ainsi nous avons de bons résultats en ce qui concerne la qualité de prédiction de notre méthode. Par contre la forme de ces résultats reste discutable puisqu'il est souvent nécessaire de passer par une étape de synthèse pour pouvoir les comparer efficacement avec le discours des enseignants.

### 6.3 Confrontation avec des concepteurs

Rappelons que notre méthode, et le logiciel qui l'instrumente, ont pour but d'aider un enseignant à évaluer un logiciel de formation. Néanmoins un résultat parallèle à nos validations concerne l'exploitation de nos évaluations par les concepteurs de logiciels. Nous avons rencontré plusieurs équipes de concepteurs (enseignants et informaticiens) afin de déterminer en quoi nos résultats leur étaient utiles. Deux aspects semblent particulièrement appréciés ; les Impressions Générales et la distinction entre Notes Instinctives et Notes Calculées.

Le thème Impressions Générales fournit un cadre de description clair et concis du logiciel et permet de mieux définir la population naturelle de ce dernier. Prenons le cas d'un logiciel destiné à une formation pour adultes, si les Impressions Générales du logiciel sont « Déroutant », « Très ludique », « Luxuriant » et « Très Inno-

vant », il y a un risque pour que la population cible ne corresponde pas avec la population réelle du logiciel.

De plus la distinction entre Notes Calculées et Notes Instinctives permet de mettre en évidence les différences entre la réalité et la perception des utilisateurs. Cette information permet, par exemple, de classer les priorités de mise à jour : selon la stratégie de développement appliquée, il sera plus pertinent de s'intéresser aux aspects techniquement au point mais mal jugés par les utilisateurs, plutôt que d'améliorer des aspects de moins bonne qualité mais où les utilisateurs n'ont aucun problème.

## 7. Conclusion et Perspectives

Notre objectif était de fournir un outil d'aide à l'évaluation destiné aux prescripteurs de logiciel. La mise au point de la liste complète des critères nous a contraint à dresser un état de l'art transdisciplinaire : informatique, ergonomie et sciences de l'éducation, mais aussi design, photographie, musicologie, narratologie jusqu'au cinéma et à la production de jeux vidéo. La cohérence entre toutes ces approches n'est pas chose aisée et certaines redondances ou contradiction peuvent apparaître. De plus, la multiplicité des situations et des stratégies pédagogiques rend toute évaluation a priori de produits finis complexe et sujette à discussion.

Il ne s'agit donc plus d'automatiser l'évaluation mais d'avoir une vision plus proche de l'accompagnement. L'utilisation d'une structure dynamique, la prise en compte de la subjectivité de l'évaluateur, la caractérisation plutôt que l'évaluation, et l'application de techniques de calcul originales permettent à EMPI de fournir une évaluation stable et de qualité des logiciels multimédias de formation. Il reste maintenant à valider totalement l'utilisation du logiciel, instrumentant cette méthode, auprès des enseignants. De cette validation nous pourrions extraire des résultats en termes d'utilisabilité de notre propre logiciel d'évaluation, ainsi qu'une mesure de la qualité et de la pertinence de l'évaluation en fonction des attentes et des habitudes spécifiques à notre public cible.

Outre les concepteurs qui comme nous l'avons décrit peuvent utiliser les résultats fournis par EMPI, notre méthode ouvre plusieurs pistes et perspectives de recherche. Si les questions posées par EMPI permettent d'évaluer un logiciel, un corpus de recommandations issues de la même structure devrait apporter une aide à la conception. Certaines recherches s'orientent donc naturellement vers le développement d'une méthode d'aide à la conception de didacticiels.

En tout état de cause nous espérons pouvoir fournir, après la prise en compte des différentes validations en cours sur le terrain, un instrument d'aide à l'évaluation permettant d'appréhender les didacticiels dans leur ensemble, au-delà du « tape à l'œil » de l'emballage et vers un enseignement de qualité.

### Financement

Cette recherche est effectuée dans le cadre du pôle régional Nouvelles Technologies Educatives, soutenu financièrement par le Conseil Régional de Picardie.

### Bibliographie

[BAL 96] BALPE J.-P., LELU A., PAPY F., SALEH I., *Techniques avancées pour l'hypertexte*, Editions Hermès, Paris, 1996.

[BAS 95] BASTIEN C., SCAPIN D., « Evaluating a use interface with ergonomic criteria », *International Journal of Human Computer Interaction* 7, 1995, p105-121.

[BER 88] BERBAUM J., « Un programme d'aide au développement de la capacité d'apprentissage », Grenoble, Université de Grenoble II, Multigraphié, 128p, 1988.

[BRU 94] Bruillard E., Vivet M., « Concevoir des EIAO pour des situations scolaires – Approche Méthodologique », *Didactique et Intelligence Artificielle*, Balacheff & Vivet (coord.), La Pensée Sauvage (Ed.), 1994.

[CAR 98] CAROT S., BETRANCOURT M., « Ergonomie des documents techniques informatisés : expériences et recommandations sur l'utilisation des organisateurs para-linguistiques », *Les Hypermédias*, Tricot A. & Rouet J.-F (coord.), Hermes, 1998.

[CHI 94] CHION M., *Musiques : Médias et Technologies*, Flammarion, 1994.

[CRO 99] CROZAT S., HU O., TRIGANO P., « EMPI : Une méthode informatisée pour l'évaluation des didacticiels multimédias », *RIHM'99* ; vol.1, n°2, novembre 1999.

[DES 91] DESSUS P., MARQUET P., « Outils d'évaluation de logiciels éducatifs. Université de Grenoble ». *Bulletin de l'EPI*. 1991.

[GIB 79] GIBSON J., « The ecological approach to visual perception », LEA, Londres, 1979.

[GRI 96] GRISLIN M., KOLSKI C., « Evaluation des interfaces Homme/Machine lors du développement des systèmes interactifs. », *Technique et Science Informatique*, 1996, vol.15, n°3

[HU 98a] HU O., TRIGANO P., « Proposition de critères d'aide à l'évaluation de l'interface homme-machine des logiciels multimédia pédagogiques », *IHM'98*, Nantes, septembre 1998.

[HU 98b] HU O., TRIGANO P., CROZAT S., « E.M.P.I. : une méthode pour l'Evaluation du Multimédia Pédagogique Interactif », *NTICF'98*, INSA Rouen, novembre 1998.

[HU 99] HÛ O., TRIGANO P., CROZAT S., « Considering subjectivity in software evaluation – Application for teachware evaluation », in *proceedings CADUI II, Computer Aided Design of User Interfaces*, J. Vanderdonck & A. Puerta (Eds), Kluwer Academic Publisher, Louvain, 1999, 331-336.

[HU 00a] HU O. « Quelques réflexions sur l'évaluation ergonomique et le contexte pédagogique » *Rencontres Jeunes Chercheurs IHM*, Ile de Berder, 3-5 mai 2000.

[HU 00b] HU O. « Compte rendu d'enquête sur l'usage d'un logiciel d'apprentissage de l'anglais au collège » Document Interne, Heudiasyc, Université de technologie de Compiègne, 2000.

[KOL 97] KOLSKI C., *Interfaces Homme-machine : application aux systèmes industriels complexes*, Editions Hermès, 1997.

[LAM 97] LAMROUS S., TRIGANO Ph , « L'organisation des espaces documentaires : vers une exploitation optimale », *Document Electronique*, Hermès, Vol 1, n°4, décembre 1997.

[LEG 98] LEGLISE M., « Un logiciel en situation dans un dispositif d'apprentissage de la conception », *CAPS'98*, Université de Caen, juin 1998.

[LIN 95] LINARD M., ZEILIGER R., « Designing an interface in an educational context : first steps to a principle approach », in *Proceedings 5<sup>th</sup> Conference INTER'ACT 95*, IFIP, Lillehammer, Juin 1995, K. Norby, P. Helmersen, D. Gilmorer, S. Arnesen (eds), Chapman&Hall, London, 401-404.

[LIN 98a] LINARD M., BELISLE C., ZEILIGER R., « Reconciling information processing and activity theories : HELICES, a foundation model for creating cognitive artefacts », *ISCRAT98, International Society for Cultural Research and Activity Theory*, 4<sup>th</sup> Congress, Juin 1998, Aarhus, DK.

[LIN 98b] LINARD M., « L'écran de TIC, dispositif d'interaction et d'apprentissage : la conception des interfaces à la lumière des théories de l'action », in : *Dispositifs et médiation des Savoirs, Colloque Grems-Grame*, Université de Louvain-la-Neuve, 1998.

[MED 97] MEDA, *61 critères d'évaluation de logiciels de formation-éducation*. MédiaScreen, cd-rom, Université de Liège, 1997

[NEN 99] NENDJO ELLA A., KOLSKI Ch., WAWAK F., JACQUES C., YIM P., « An approach of computer-aided choice of UI evaluation criteria and methods », in *proceedings CADUI II, Computer Aided Design of User Interfaces*, J. Vanderdonck & A. Puerta (Eds), Kluwer Academic Publisher, Louvain, 1999, 319-329.

[OVA 92] ORAVEP, METANE : méthode d'analyse numérique pour la formation des adultes. Paris : ORAVEP, 1992. (<http://www.oravep.asso.fr/>).

[PAR 91] PARKER R., THERIEN L., *Mise en page et conception graphique*, Reynald Goulet, 1991.

[PER 98] PERRIN H., BONNAIRE R., « Un logiciel pour la visualisation de mécanismes de gestion des processus du système UNIX », *NTICF'98*, INSA Rouen, novembre 1998.

[POG 96] POGNANT P., SCHOLL C. *Les cd-rom culturels*, Editions Hermès, 1996.

[RHE 94] RHEAUME J., « L'évaluation Des Multimédias Pédagogiques : De L'évaluation Des Systèmes A L'évaluation Des Actions », *Educatechnologie*<sup>1</sup>, Vol.1, n°3, Sept. 1994.

[RIC 80] RICHAUDEAU F., *Conception et production des manuels scolaires*, Paris, Retz, 290p, 1980.

---

<sup>1</sup> Educatechnologie : Revue Electronique (<http://www.fse.ulaval.ca/fac/ten/reveduc/html/>)

[SCA 97] SCAPIN D., BASTIEN C., « Ergonomic criteria for evaluating the ergonomic quality of interactive systems », *Behaviour & Information Technology* 16, 1997, p226-231.

[SOR 92] SORLIN P., *Esthétiques de l'audiovisuel*, Nathan, 1992.

[VAN 94] VANDERDONCKT J., « Guide ergonomique de la présentation des applications hautement interactives », Presses Universitaires, Namur, 1994.

[VAN 98] VANDERDONCKT J., « Conception ergonomique de pages WEB », Vesale, 1998.

[VIV 96] VIVET M., « Evaluating educational technologies : Evaluation of teaching material versus evaluation of learning ? », *CALISCE'96*, San Sebastian, juillet 1996.