

# EMPI : Une méthode informatisée pour l'évaluation des didacticiels multimédias<sup>1</sup>

Stéphane CROZAT, Philippe TRIGANO, Olivier HU

*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*

*Email : Stephane.Crozat@utc.fr, Olivier.Hu@utc.fr, Philippe.Trigano@utc.fr*

## Résumé

Aujourd'hui la question ne se pose pas de savoir s'il faut ou non favoriser l'apprentissage assisté par ordinateur. L'informatique s'insère dans de nombreux processus d'apprentissage, et de fait, pas toujours avec des résultats positifs. Ce constat nous encourage à orienter nos recherches de façon à mieux comprendre la relation entre l'apprentissage et le support numérique, à mieux maîtriser l'outil et ses spécificités, afin d'en orienter au mieux les usages.

Nous proposons une méthode d'aide à l'évaluation des didacticiels multimédia, dont la vocation est d'assister les utilisateurs (principalement des apprenants et des tuteurs) à choisir entre la large palette de logiciels disponibles. Notre approche est de diviser l'analyse du logiciel en six thèmes principaux : les impressions générales, la qualité technique, l'ergonomie de l'IHM, les documents multimédia, la scénarisation et la didactique. Chacun de ces six thèmes est divisé en méta-critères, sous-critères et questions. L'ensemble constitue une base de connaissances, sous forme d'un questionnaire hiérarchique, qui permet de noter le didacticiel selon différentes dimensions, dans l'optique de le comparer avec d'autres logiciels ou avec un contexte pédagogique donné.

Cet article présente la structure détaillée du questionnaire, avec quelques exemples de questions, puis s'attache à décrire le fonctionnement de la méthode informatisée ainsi que les évaluations menées.

## Mots-clés

Evaluation de logiciel – Interfaces homme-machine – Contexte pédagogique – Multimédia – Ergonomie

---

<sup>1</sup> Cette méthode est développée dans le cadre du pôle régional *Nouvelles Technologies Educatives*, soutenu financièrement par le Conseil Régional de Picardie.

## Introduction

On constate aujourd'hui un certain engouement de la part des institutions françaises et européennes [Selys 98] et du grand public vers les Nouvelles Technologies pour l'Education. Il faut les exploiter, il faut les intégrer aux centres de formation quels qu'ils soient, il faut en avoir chez soi, il faut que les enfants les utilisent, il faut que les adultes les utilisent, ... Qui n'entend pas régulièrement ces affirmations ? Mais à y regarder de plus près, force est de constater qu'en pratique l'engouement n'est que *virtuel* et que les didacticiels sont fréquemment ignorés, oubliés, sous utilisés voire même rejetés. Nous ne pensons pas que le principe même soit en cause, il n'y a pas de raisons pour que le support numérique ne trouve pas sa place à côté du livre, de l'enseignement scolaire traditionnel ou de la formation en entreprise (et il n'y a d'ailleurs pas de raisons non plus pour qu'il se substitue à eux comme on l'entend parfois). L'échec relatif du cédérom aujourd'hui peut en grande partie être attribué à leur faible qualité devant un public très exigeant, et qui a toutes les raisons de l'être devant le potentiel inexploité que représente ce support. Des défauts de conception, des contenus trop pauvres, des interfaces inefficaces, de mauvais usages du multimédia, des stratégies pédagogiques inadaptées, sont autant d'exemples d'erreurs courantes. Néanmoins les didacticiels représentent une alternative et une complémentarité dans le domaine de l'apprentissage, dont les potentiels ne peuvent être négligés, notamment en ce qui concerne la formation continue, la formation à distance, la mise à niveau, l'auto-formation, l'éveil scolaire, ...

Notre proposition face à ce constat est d'aider les utilisateurs à appréhender les points forts et les points faibles de ces nouveaux logiciels à vocation éducative. La méthode EMPI (Evaluation des logiciels Multimédia Pédagogiques Interactifs) constitue un outil permettant d'analyser de façon systématique et normalisée un didacticiel. Une évaluation EMPI permet de souligner de façon structurée les caractéristiques du logiciel. Les résultats sont ensuite utilisables par un formateur pour comparer des logiciels entre eux ou bien comparer un logiciel avec une situation d'enseignement et des objectifs d'apprentissage déterminés.

Historiquement, la nécessité de mesurer la qualité des didacticiels est issue de deux thèmes de recherche plus anciens : d'une part l'évaluation des supports pédagogiques, notamment celle des manuels scolaires [Richaudeau 80], et d'autre part l'évaluation des logiciels et des interfaces homme/machine principalement dans le contexte industriel [Kolsky 97]. Nous essayons d'adapter ces deux approches à notre contexte spécifique. Quelle que soit la forme adoptée (enquête, prototypage, analyse de performance, étude en situation, ...) toute évaluation nécessite au moins la mise à plat de trois questions [Depover 94] : « Qui évalue ? Qu'évalue-t-on ? Quand évalue-t-on ? ». Dans notre cas l'évaluateur est l'utilisateur ou le prescripteur du didacticiel ; l'évaluation se veut généraliste et se base sur des dimensions informatiques, ergonomiques, multimédia et didactiques ; dans l'optique d'aider les utilisateurs à choisir, l'évaluation est appliquée essentiellement à des produits finis et porte directement sur les logiciels (et non sur leur impact a posteriori).

Après avoir présenté la structure de notre grille d'évaluation et les critères qui la composent, nous insisterons sur la méthode d'évaluation proprement dite, sa mise en œuvre informatique ainsi que les validations que nous avons menées.

## 1. Proposition de dimensions d'évaluation

### 1.1 Une structure a trois niveaux

La méthode EMPI repose essentiellement sur un questionnaire structuré, à profondeur variable. En nous inspirant d'un certain nombre de modèles portant sur des thèmes proches, [Rhéaume 94], [Weidenfeld & al. 96], [Dessus, Marquet 91], [Berbaum 88], nous avons mis en place une hiérarchie de dimensions d'évaluation. Notre approche a consisté en une sélection et une réorganisation de critères existants dans la littérature afférente à notre problématique<sup>1</sup>, en fonction du cadre d'application particulier qui était le nôtre, à savoir l'évaluation et les didacticiels multimédia. La structure globale (voir Figure 1) de notre évaluation se base sur six thèmes complémentaires (eux-mêmes divisés en méta-critères) :

- Le thème **impressions générales** rend compte de l'image que le didacticiel offre à l'utilisateur.
- Le thème **qualité informatique** permet d'évaluer la mise au point technique du logiciel.
- Le thème **utilisabilité** correspond à l'évaluation ergonomique de l'interface.
- Le thème **documents multimédia** envisage la présentation et la forme des contenus.
- Le thème **scénarisation** s'intéresse à l'ensemble des techniques d'écriture utilisées pour agencer les informations.
- Le thème **didactique** s'intéresse enfin aux ressources pédagogiques mobilisées en fonction du contexte d'apprentissage.

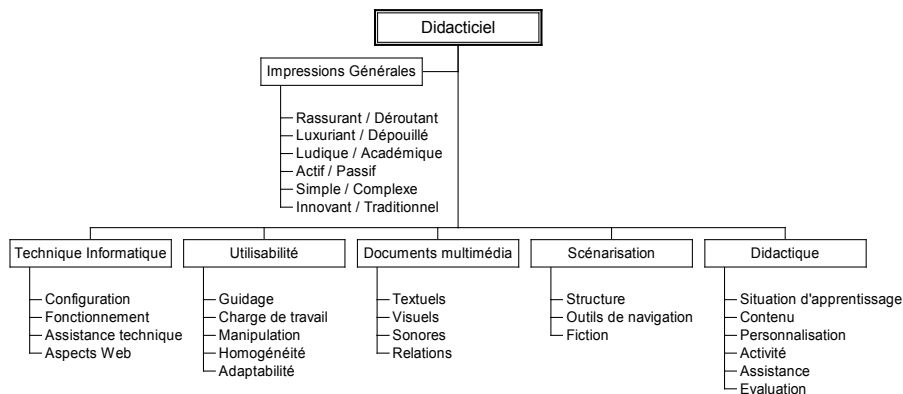


Figure 1 : Hiérarchie des thèmes et méta-critères

<sup>1</sup> Littérature référencée thème par thème ci-dessous

Les méta-critères se divisent encore en sous-critères. C'est finalement au niveau de chaque sous-critère que sont placées les questions qui composent effectivement le questionnaire. Dans les paragraphes suivants, après avoir traité le cas particulier des *impressions générales*, nous nous attacherons à décrire plus avant chaque thème d'évaluation en définissant les méta-critères, énonçant<sup>1</sup> les sous-critères et donnant des exemples de questions.

## 1.2 Impressions générales

Les expériences que nous avons menées nous ont permis d'observer que le multimédia intègre un élément essentiel, qui tient fondamentalement de la forme, et qui se manifeste par des impressions générales qui s'imposent à l'utilisateur. Nous avons en particulier pu effectuer le constat suivant : lorsqu'un utilisateur est placé en situation de découverte d'un didacticiel, il se forge une première impression sur une très courte durée d'utilisation (de l'ordre de quelques minutes). Pour prendre l'analogie avec un professeur, lorsque ses élèves entrent pour la première fois dans sa classe, ils le *cataloguent* immédiatement. Il a l'air sympa, dur, beau, intelligent, ennuyeux, ... Ils n'ont pas même écouté son premier cours, ils ont déjà de nombreux qualificatifs à son égard. De la même façon un utilisateur qui découvre un logiciel didactique sera capable de le qualifier avant même de vraiment l'utiliser. Cela correspond à l'attrait immédiat qu'exerce le didacticiel sur l'utilisateur.

Ce constat est principalement basé sur plusieurs approches empiriques (études de corpus notamment), dont voici le bref rappel :

**Première expérience : 50 évaluateurs, 1 logiciel** : Une cinquantaine d'évaluateurs, élèves de second cycle à l'UTC, ont été mis en situation d'utilisation du même logiciel, *Prisméo*, réalisé par le Groupe des Ecoles de Télécommunication en 1994. Les utilisateurs ont essayé le logiciel pendant une heure et ont ensuite procédé à une évaluation libre. Ils devaient exprimer leurs impressions et leurs commentaires par écrit. Le corpus est donc composé d'une cinquantaine de textes d'environ une page.

**Deuxième expérience : 100 logiciels** : Le second corpus a été prélevé sur les évaluations d'une centaine de didacticiels différents<sup>2</sup>. Notons que nous avons effectué une analyse manuelle de ce corpus, ainsi qu'un traitement automatique par recherche de lemmes, à l'aide d'un logiciel d'indexation automatique de textes, développé au sein du Laboratoire Heudiasyc [Lamrous, Trigano 97].

**Troisième expérience : 4 évaluateurs, 4 logiciels, 1 heure** : Quatre utilisateurs de culture similaire (informatique) ont évalué chacun quatre logiciels multimédia qu'il ne connaissaient pas, et qui concernaient l'enseignement d'une matière scientifique (*Silicapolis*, *Galilée*, *Dicit*, *La cellule*). Ils disposaient d'une heure environ pour tester l'ensemble des quatre logiciels et noter leurs impressions générales.

---

<sup>1</sup> Nous ne rappellerons pas la définition des sous-critères dans cet article qui veut donner une vision plus générale de la méthode EMPI, mais renverrons chaque fois vers d'autres travaux portant plus précisément sur chaque thème d'évaluation.

<sup>2</sup> Ce corpus est disponible sur le site Internet « <http://www.cyberscol.qc.ca/BD/MEQ> » et a été réalisé par des évaluateurs canadiens.

**Autres expériences :** Des approches critiques menées au sein du laboratoire, des études comparatives de méthodes (voir « MEDA versus EMPI », p.19), ou l'analyse d'autre corpus (notamment liés à l'évaluation de documents hors contexte informatique, tels que les corpus d'analyse de photographies [Debray 94], les corpus de styles graphiques [Poynor 91], ...) ont également alimenté notre réflexion.

A l'issu de ces expériences nous avons pu observer que les impressions générales ressenties se manifestent comme étant :

- **instinctives** : elles ne se fondent pas sur une réflexion de l'utilisateur, mais sur un sentiment immédiat (attirait ou rejet);
- **descriptibles** : les utilisateurs explicitent aisément leurs impressions et le vocabulaire utilisé pour les décrire est relativement restreint (même dans le cas d'une centaine d'évaluation d'une centaine de logiciels) ;
- **homogènes** : les impressions générales ressenties par des utilisateurs distincts en face d'un même logiciel concordent bien ;
- **persistantes** : les impressions initiales se maintiennent dans le temps et sont difficiles à dépasser.

Afin de mieux maîtriser ces observations, nous nous sommes appuyés sur des théories de la perception visuelle, et en particulier celle des *affordances* [Gibson 79].

Dans un second temps, nous avons cherché comment intégrer ces réflexions dans le cadre de notre aide à l'évaluation. Notre objectif est d'aider l'utilisateur à caractériser les impressions générales suscitées par le logiciel, et non bien entendu de porter un jugement direct. Nous avons sélectionné des critères empiriques issus de nos expériences, mais aussi des critères issus de la littérature, de la sémantique de l'image [Cossette 83], à la photographie [Alekan 84] et au cinéma [Vanoye, Goliot-Lété 92], en passant par la musicologie [Kivy 89] ou la publicité [Cornu 90]. Ces études nous ont permis d'aboutir à six paires antagonistes de critères (Figure 2), globaux (afin de simplifier la démarche de caractérisation) et neutres (afin de favoriser la caractérisation et non l'évaluation directe). Notons que la structure de ce thème se distingue de celles que nous retrouverons dans tous les autres modules.

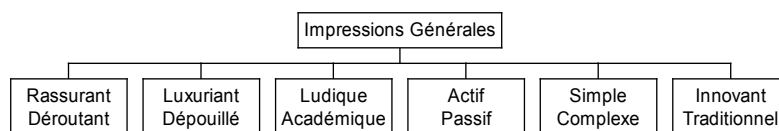


Figure 2 : Critères du thème « impressions générales »

Un exemple de caractérisation à l'aide de ces critères serait : *Le logiciel est « plutôt rassurant » et « très sobre » et « ludique » et « extrêmement actif » et « moyennement complexe » et « classique ».* Quitte à l'évaluateur ensuite de déterminer si le logiciel ainsi décrit peut correspondre ou non au contexte pédagogique dans lequel il devra être utilisé.

### 1.3 Qualité technique

Un bon logiciel est d'abord un logiciel qui fonctionne correctement. Ainsi avant toute autre évaluation, il est primordial de vérifier la bonne qualité des aspects inhé-

rents à tous logiciels : facilité d'installation, vitesse d'exécution, absence de bogues, configuration, ... Face à une population non-informaticienne, les aspects liés à la *qualité technique* sont déjà sources de nombreux rejets. La littérature étant abondante et la recherche stabilisée quant à ce genre de problématique nous avons intégré les critères disponibles sans que cela fasse l'objet d'une véritable recherche de notre part. Seule la partie liée à l'usage du Web, plus innovante à fait l'objet de travaux empirique (analyse de sites) et de confrontations avec d'autres travaux sur le sujet [Vanderdonck 98].

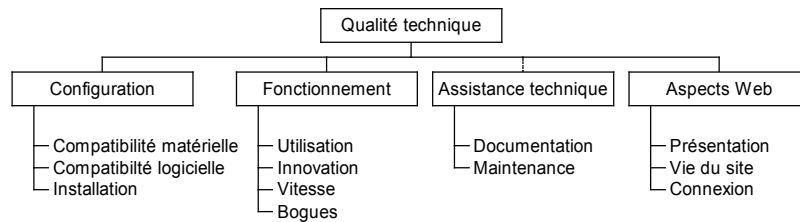


Figure 3 : Critères du thème « qualité technique »

*Définition<sup>1</sup> des méta-critères :*

- Configuration : Gestion des spécificités du logiciel en fonction des environnements matériels sur lesquels il doit être utilisé.
- Fonctionnement : Qualité du déroulement du programme.
- Assistance technique : Ensemble des moyens mis à disposition de l'utilisateur pour la maîtrise de la technique du logiciel.
- Aspects Web : Ensemble des particularités techniques liées à l'utilisation d'Internet au sein du logiciel.

*Exemple de questions associées à ces critères :*

- Y a-t-il une fiche signalétique décrivant les caractéristiques du logiciel (environnement informatique requis, équipement et périphériques nécessaires, ...) ?
- Avez-vous eu des problèmes pour installer correctement le logiciel ?
- Les temps de réponse vous ont-ils parus anormalement longs ?

#### 1.4 Utilisabilité

Il existe de nombreuses listes de critères d'évaluation de l'**interface**, des recommandations relativement générales, comme les sept règles d'or de J. Coutaz ou les huit critères de Ravden et Johnson [Ravden & al 89], jusqu'au guide de J. Vanderdonck [Vanderdonck 94] qui contient plus de 3000 règles. Le problème reste de trouver un compromis entre les critères d'ordre général, qui restent souvent des évidences (bien que pas toujours appliquées...), et les recommandations trop précises qui ne s'appliquent que dans des cas bien particuliers. Pour mettre au point nos critères ergonomiques, nous nous sommes largement inspirés de la liste proposée par l'INRIA [Scapin 86], [Senach 90], [Bastien, Scapin 94], [Scapin, Bastien 97],

<sup>1</sup> Les définitions que nous proposons correspondent à des choix propres et fondent le lexique de la méthode.

complétée et adaptée pour les contextes multimédia et pédagogiques. Notons que le thème utilisabilité fut à l'origine de nos recherches, on renverra à [Hû, Trigano 98] [Hû & al 98] pour de plus amples précisions.

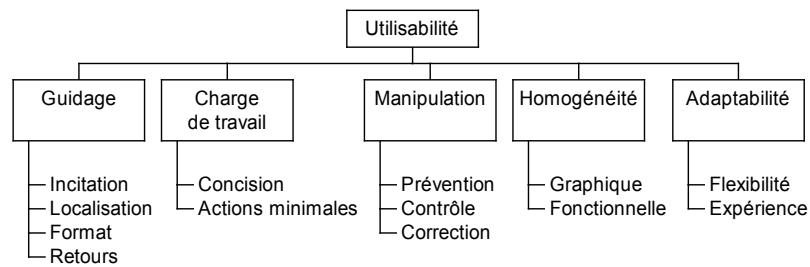


Figure 4 : Critères du thème « utilisabilité »

*Définition des méta-critères :*

- Guidage : Ensemble des techniques mises en œuvre pour aider l'utilisateur à accomplir sa tâche.
- Charge de travail : Efforts cognitifs que l'utilisateur doit mobiliser pour accomplir sa tâche.
- Manipulation : Maîtrise par l'utilisateur du déroulement du logiciel.
- Homogénéité : Maintien tout au long du logiciel d'une charte graphique et fonctionnelle assurant l'union et la cohérence.
- Adaptabilité : Ensemble des possibilités disponibles pour modifier l'interface du logiciel afin d'en améliorer l'utilisation.

*Exemple de questions associées à ces critères :*

- Vous est-il arrivé de devoir chercher des composants à l'écran, parce qu'ils n'étaient pas à l'endroit attendu ?
- Vous est-il arrivé de ne pas savoir si une de vos actions avait été prise en compte ?
- Vous est-il arrivé de ne plus savoir où vous vous trouviez ou comment vous déplacer dans l'arborescence ou le scénario du logiciel ?

### 1.5 Document

Les textes, les images et les sons constituent la matière première du logiciel multimédia, ils sont les vecteurs essentiels de **l'information** dans le didacticiel. Précisons que notre étude, à travers ce thème, ne porte pas sur le contenu des documents, mais sur la pertinence de l'utilisation qui en est faite afin de transmettre leur contenu. Plus précisément l'objet de notre étude, s'inspirant des théories du support [Goody 79] a été de déterminer en quoi chaque type de document était spécifique et quel *supplément* il apportait à l'information véhiculée.

Les documents textuels sont centraux dans la perspective de l'apprentissage, puisqu'ils combinent les avantages du langage et de la permanence de l'écrit. Néanmoins leur utilisation dans les didacticiels est délicate, en ce sens que l'écran, lumineux, immobile et vertical, est plutôt mal adapté à la lecture de textes. Il permet une capacité de concentration moindre que sur un support papier et les hyperliens

tendent à rompre la continuité de la lecture [Balpe et al 96]. Nous sommes attachés dans ce cadre aux critères utilisés actuellement par les professionnels de l'édition [Parker, Thérien 91] majoritairement transposables à notre cas, ainsi qu'aux critères pensés pour le support numérique [Baticle 85] ou la lecture sur écran [Type, Frommer 85]. L'image est largement utilisée dans les didacticiels, de part les facilités offertes par le multimédia. Néanmoins son caractère hautement polysémique en fait un moyen peu propice à la communication précise. Les images didactiques [Costa, Moles 91] construites pour l'enseignement et associées au texte peuvent être plus efficaces qu'un texte pour la représentation de certaines informations, à condition de respecter les règles qui leurs sont propres. Les illustrations trouvent néanmoins une place d'importance pour accompagner les informations, et nous proposons de construire des critères d'évaluation de ces illustrations, à partir des domaines qui leurs sont liés, tels que la photographie [Barthes 80], la sémantique de l'image [Cossette 83], l'audiovisuel [Vanoye, Goliot-Lété 92], ... Les fonctions du son sont clairement identifiées dans le cadre du cinéma [Sorlin 92], et les particularités de la musique, liées à l'émotion notamment, nous permettent également de dégager des critères de caractérisation du son. Enfin il est nécessaire de prendre en compte dans notre évaluation le fait que, dans un cadre multimédia, en plus de leur usage indépendant, les documents prennent leur signification accompagnés par d'autres documents co-présents. Et l'ensemble tisse une toile de **relations** aussi essentielles dans la compréhension du didacticiel que la compréhension de chaque document [Bachimont 99a].

Notons que les considérations liées aux documents multimédia et à leur scénarisation (traitée dans le paragraphe suivant) ont été largement développées dans [Crozat 98].

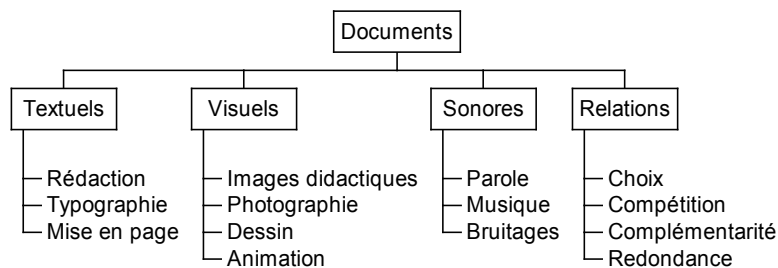


Figure 5 : Critères du thème « documents »

*Définition des méta-critères :*

- Documents textuels : Forme d'appropriation basée sur le texte écrit.
- Documents visuels : Formes d'appropriation basées sur les images et les représentations graphiques, fixes ou animées.
- Documents sonores : Formes d'appropriation basées sur le texte oral et sur les autres sollicitations de l'ouïe.
- Relations : Etude de la redondance et de la complémentarité des différents documents entre eux.

*Exemples de questions associées :*

- Relève-t-on dans la rédaction des textes des erreurs de frappe ou de français ?

- Le degré d'iconicité des images didactiques est-il adapté aux utilisateurs ?
- L'utilisation simultanée de documents visuels et textuels favorise-t-elle la compréhension ?

### 1.6 Scénarisation

Dans le cadre du didacticiel multimédia, on définit la scénarisation comme le procédé particulier qui consiste à structurer les documents de façon à en préparer au mieux la lecture par l'utilisateur. La scénarisation ne prend donc pas directement en compte les informations à transmettre, mais la façon dont elles sont agencées. Cela impose une forme d'écriture **originale**, dans la mesure où le numérique propose de traiter avec des concepts nouveaux. Le support numérique offre une conception nouvelle de la structure du document et son étude fait actuellement l'objet de recherches dans le domaine du multimédia [Pognant, Scholl 96], [Milon, Cormerais 99]. Cette structure particulière a pour conséquence la délinéarisation de la lecture, qui implique également une perte de sens [Bachimont 99b], qu'il est nécessaire de prendre en compte à travers des outils permettant notamment la gestion de l'espace de lecture, la réécriture des contenus ou la perception de la cohérence globale de l'hyperdocument. Notons que selon nous la possibilité d'intégrer l'apprentissage au travers une fiction répond également à ces préoccupations. En effet elle est facteur d'implication et de cohérence pour l'utilisateur. La narratologie s'intéresse aujourd'hui au problème des logiciels multimédia interactifs, en adaptant les connaissances dont elle dispose dans le domaine de la littérature [Carrière 93] ou du cinéma [Gaudreault, Jost 90]. Le schéma actantiel proposé par [Pajon, Polloni 97] paraît particulièrement adapté aux structures de fiction dans les didacticiels puisqu'une analogie peut-être faite entre le but à atteindre dans le récit et le but pédagogique réel.

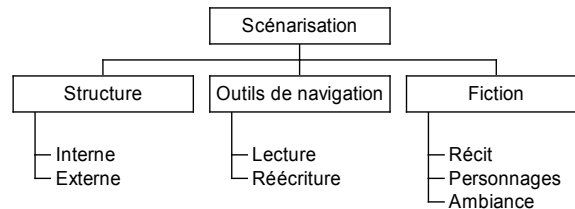


Figure 6 : Critères du thème « scénarisation »

#### Définition des méta-critères :

- Structure : Représentation de l'ensemble des parcours possibles au sein de l'ensemble des nœuds d'information du logiciel.
- Outils de navigation : Ensemble des moyens mis à la disposition de l'utilisateur pour s'appropriier la structure.
- Fiction : Intégration d'une dimension imaginaire et non-directement liée à l'apprentissage dans la façon de parcourir les documents.

#### Exemple de questions :

- Le logiciel propose-t-il à l'utilisateur des outils lui permettant de réécrire les informations transmises au cours de la navigation ?

- Le parcours du didacticiel s'appuie-t-il sur un récit ?
- Le logiciel a-t-il recours à des personnages pour incarner l'utilisateur et/ou le tuteur ?

### 1.7 Didactique

La littérature propose bon nombre de critères ou de recommandations pédagogiques [Park & al 93] : le processus d'apprentissage peut-être évalué à partir des critères de présentation, contenu, méthodes et interactivité [Dessus, Marquet 91]. De nombreux facteurs pédagogiques sont également proposés tels que la motivation, le rythme, la participation, la structuration, les méthodes pédagogiques, le guidage de l'apprenant, la répétition et la variété d'activités de l'apprenant, les exercices, l'application et l'évaluation de connaissances, ... [Marton 94]. En ce qui concerne plus particulièrement le contrôle du déroulement de l'apprentissage on recommande la hiérarchisation et la segmentation des contenus, le repérage dans le temps et l'espace des informations, le suivi de la progression de l'apprenant, ... [Hannafin, Peck 88]. [Gagné, Medsker 96] proposent quant à eux un ensemble de conditions préalablement nécessaire à un apprentissage donné. On note aussi les références à des réflexions plus spécifiquement liées aux contextes d'usage de didacticiels, tel la problématique de l'interactivité posée par [Vivet 96] ou de l'assistance [Piché & al 98]. Enfin, une attention particulière a également été portée sur les critères utilisés par le collectif européen Joanna MEDA [MEDA 90] (voir §3.3).

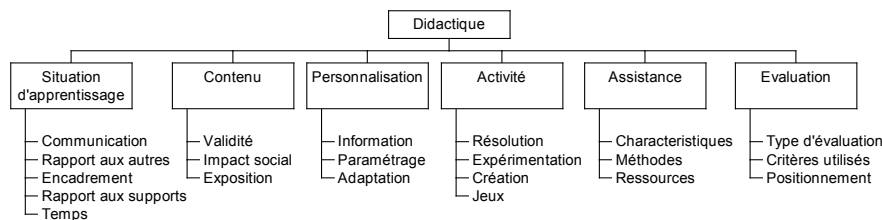


Figure 7 : Critères du thème « didactique »

#### Définition des méta-critères :

- Situation d'apprentissage : Description du dispositif matériel et humain accompagnant l'utilisation du didacticiel.
- Contenu : Relatif à la nature même des informations véhiculées dans le didacticiel.
- Personnalisation : Ensemble des techniques permettant l'ajustement entre le didacticiel et l'utilisateur dans le but de rendre plus performant l'apprentissage.
- Activité : Description des activités disponibles dans le didacticiel pour permettre à l'apprenant la manipulation du contenu exposé.
- Assistance : Ensemble des outils et méthodes disponibles pour aider de façon formative l'apprenant dans ces tâches d'apprentissage.
- Evaluation : Relatif aux évaluations, formatives ou sommatives, des connaissances construites par l'apprenant dans le cadre de son apprentissage.

#### Exemple de questions associées:

- Est-ce que l'apprenant a accès au réseau local ? à Internet ? Est-il isolé ?

- Quelle sont la véracité du contenu, la pertinence des pré-requis et des objectifs en fonction du public visé ?
- Quelle est la qualité de l'évaluation effectuée avant l'utilisation (calibrage), en cours d'utilisation (progression) et en fin de parcours (test final) ?

## 2. La méthode EMPI

Nous venons de rappeler brièvement la structure du questionnaire que nous proposons. Un logiciel (à l'état de prototype) permet actuellement d'appliquer et d'exploiter ce questionnaire. Nous nous proposons dans cette partie d'en exposer les fonctionnalités principales afin de cerner les spécificités de la méthode EMPI.

### 2.1 Profondeur variable

La méthode EMPI ambitionne d'être une méthode généraliste, et en ce sens elle se veut adaptée à des évaluateurs, des situations d'évaluation et des logiciels différents. Nos expériences ont montré que cela posait des problèmes, notamment du fait que certains aspects du questionnaire peuvent ne pas être pertinents dans un contexte donné (si un évaluateur est déjà compétent dans ce domaine, si le logiciel à évaluer n'est pas concerné, si l'évaluation est volontairement restreinte du fait de contrainte de temps par exemple). Afin de gérer ces diversités, et après avoir essayé plusieurs solutions, telle que des questionnaires court, moyen et long, nous avons opté pour un questionnaire unique, mais à **profondeur variable**. La Figure 8 met en évidence l'interface permettant la navigation entre les critères. L'évaluateur peut approfondir chaque notion selon le degré de précision qu'il veut atteindre, l'intérêt qu'il lui porte ou les compétences éventuelles qui le dispensent d'examiner précisément tel ou tel point. Au niveau le plus haut du questionnaire on retrouve la liste des thèmes, et au niveau le plus bas les questions liées aux sous-critères.

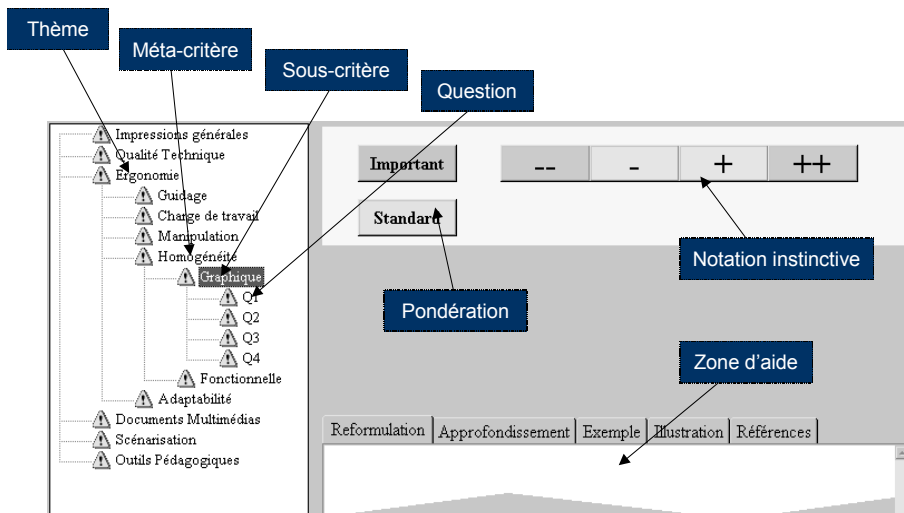


Figure 8 : Navigation entre les critères

## 2.2 Système de notation

**Notes instinctives et notes calculées** : Le processus d'évaluation sur lequel se base la méthode EMPI passe par la notation des critères précédemment définis. Notre système de notation est mixte dans la mesure où chaque critère est noté par deux biais différents :

- La notation instinctive (NI) est l'appréciation directe des thèmes, méta-critères et critères par l'évaluateur. Cette évaluation se fait sur une échelle fixe [-- ; - ; + ; ++] (voir Figure 8).
- La notation calculée (NC) est obtenue à partir des réponses aux questions (voir Figure 12). Une note entre -10 et +10 est ainsi attribuée à chaque question, et la moyenne permet de faire remonter des notes calculées au niveau de chaque sous-critère, puis méta-critère, puis thème.

Notons que les réponses instinctives sont optionnelles. De plus, puisque le questionnaire est à profondeur variable, certaines questions peuvent ne pas avoir été utilisées. Chaque critère est donc dans l'un des cas de figure suivant :

- Il dispose d'une note instinctive ou d'une note calculée : cette note fait loi.
- Il dispose d'une note instinctive et d'une note calculée : l'évaluateur devra prendre une décision, à la fin de son évaluation, et déterminer une note finale.
- Il ne dispose d'aucune note : le critère est ignoré pour cette évaluation.

**Notation finale** : Une fois l'ensemble du questionnaire parcouru, l'évaluateur doit donc revenir sur les critères pour lesquels une note instinctive et une note calculée existent simultanément. La méthode lui permet de confronter ces deux notes, afin qu'il prenne une décision. Pour chaque critère la NI et la NC sont mises en correspondance et agrémentées de deux indices permettant de gérer les écarts éventuels entre ces deux notes :

- L'indice de **cohérence** indique à l'évaluateur s'il a été logique dans ses notations instinctives. Par exemple si l'évaluateur a indiqué que la *navigation* était très bonne, puis que ses sous-critères (la *structure*, la *lecture* et la *réécriture*) étaient mauvais, on détecte une confusion de sa part, et l'indice de cohérence sera faible.
- L'indice de **corrélation** met en relief, pour chaque critère, la correspondance entre la note instinctive et la note calculée des questions qui le composent. Par exemple si les réponses aux questions du critère *structure* donnent une moyenne de notes calculées de 8 et que la note instinctive de *structure* est 6, on considérera qu'il y a correspondance, et donc l'indice de corrélation sera élevé.

Cette phase permet à l'évaluateur d'effectuer une forme d'auto-validation de son évaluation, puisqu'il compare son évaluation instinctive des critères avec les réponses qu'il a données aux questions.

Notons que le logiciel fournit également à l'évaluateur une proposition de note finale basée sur 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. Actuellement notre prototype fournit un rapport imprimable qui permet la saisie des notes finales. La Figure 9 permet d'illustrer un extrait d'un tel rapport.

	NI	NC	Cohérence	Corrélation	Proposition
<b>Scénarisation</b>	<b>-0,67</b>	<b>-1,25</b>	<b>80%</b>	<b>95%</b>	<b>-1,0 +/- 0,0</b>
<b>Navigation</b>	<b>-1,09</b>	<b>-1,5</b>	<b>85%</b>	<b>96%</b>	<b>-1,3 +/- 0,0</b>
<i>Structure</i>	-5	-10		75%	-8,7 +/- 0,5
<i>Lecture</i>	-2	0		83%	-0,5 +/- 0,0
<i>Réécriture</i>	4	2		86%	2,5 +/- 0,1
<b>Fiction</b>	<b>-1,88</b>	<b>-1</b>	<b>83%</b>	<b>93%</b>	<b>-1,4 +/- 0,1</b>
<i>Récit</i>	-8	-10		90%	-9,0 +/- 0,3
<i>Ambiance</i>	2	0		83%	0,5 +/- 0,0
<i>Personnages</i>	0	0		100%	0,0 +/- 0,0
<i>Emotion</i>	3	6		81%	5,2 +/- 0,4

Figure 9 : Extrait de rapport d'évaluation

On retrouve : dans la partie gauche les notes instinctives (NI) et les notes calculées (NC) ; au centre les indices de cohérence et de corrélation ; et à droite la proposition de note finale. Cette représentation actuelle étant jugée trop complexe, nous travaillons à une représentation graphique plus rapide à appréhender. La Figure 10 permet de récapituler l'ensemble des étapes de la notation, pour aboutir finalement à la visualisation de résultats, Figure 13 .

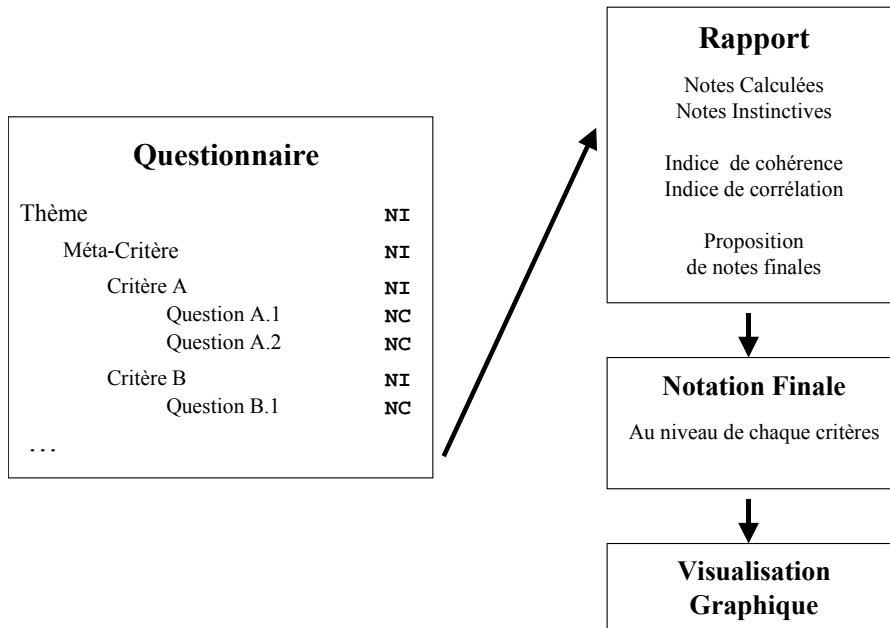


Figure 10 : Etapes de la notation avec la méthode EMPI

**Notation exponentielle** : Rappelons que les notes calculées varient entre -10 et +10. La notation calculée repose sur un principe permettant de mettre en valeur les in-

formations atypiques (défauts importants notamment) sans les noyer sous les informations standards.

Par exemple à des réponses du type [Très bien, Bien, Mauvais, Très mauvais], on associera une notation exponentielle [10, 0, -6, -10] plutôt qu'une notation linéaire [10, 4, -4, -10], comme représenté sur la Figure 11.

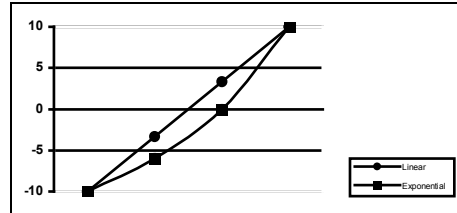


Figure 11 : Notation exponentielle vs notation linéaire

**Pondération** : A l'intérieur de chaque critère on distingue souvent une ou deux questions pour lesquelles la réponse est particulièrement significative et déterminante. De même certains critères d'un thème peuvent se révéler plus importants que d'autre au regard d'un contexte spécifique. Un système de pondération est donc proposé de façon à distinguer les points standards et fondamentaux (voir Figure 8).

### 2.3 Différents types de questions

**Questions DD** : La méthode a originellement débuté sur l'évaluation des didacticiels selon le seul thème de l'utilisabilité (ergonomie). La technique de notation privilégiée était donc basée sur la détection de défaut (DD). C'est à dire qu'à chaque réponse à une question pouvait être associée directement une note. Ce système, renforcé par la notation exponentielle a fait ses preuves (voir 3.2), notamment grâce au caractère relativement peu contextuel de nos critères ergonomiques.

**Questions CE** : Lorsque nous avons abordé des domaines moins objectifs et plus contextuels, comme la scénarisation ou la didactique, les questions DD n'étaient plus suffisantes. Nous avons alors mis en place un second type de question, destiné à coexister avec le premier, les questions de caractérisation / évaluation (CE). Elles permettent de diviser un problème en deux phases : une question de caractérisation qui permet de décrire un contexte et une question d'évaluation proprement dite visant à juger cette situation (voir Figure 12). L'intérêt d'une telle démarche est de permettre une prise de conscience chez l'évaluateur, qui dans un premier temps mène une réflexion sur la situation existante, puis prend une décision et porte un jugement quant à cette situation.

Figure 12 : Question « caractérisation / évaluation »

## 2.4 Présentation de résultats

Une fois que l'évaluateur a déterminé les notes finales, la méthode lui propose des représentations graphiques de ses résultats sous forme d'histogrammes ou de *Pareto*<sup>1</sup>. La représentation est à précision variable, à l'instar du questionnaire, et peut concerner les différents niveaux du questionnaire (thèmes, méta-critères ou sous-critères). Notons que les histogrammes classiques sont plus adaptés pour des comparaisons entre logiciels, tandis que les *Pareto* mettent mieux en évidence les défauts et qualités du didacticiel.

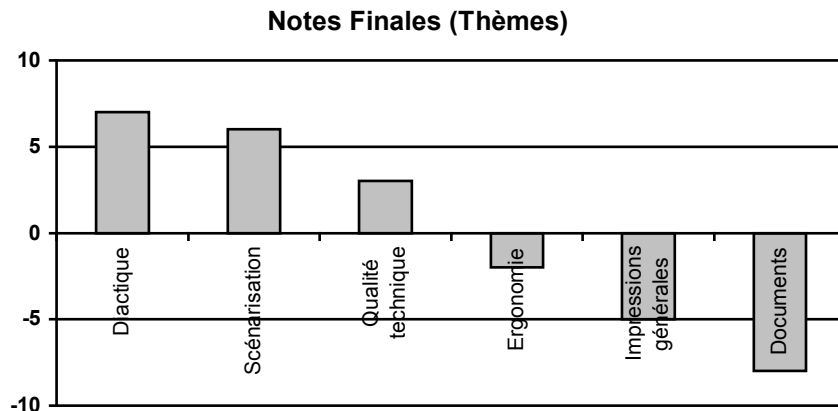


Figure 13 : Représentation des résultats au niveau des Thèmes sous forme de Pareto

<sup>1</sup> Histogramme particulier trié par ordre décroissant.

## 2.5 Aide

Une aide contextuelle est proposée à l'évaluateur au cours de son parcours du questionnaire (voir Figure 8). Insistons sur le fait que cette aide est indispensable dans la mesure où elle offre à l'évaluateur des outils pour manipuler des concepts qu'il ne maîtrise pas forcément. Associée à chaque question, critère, méta-critère et thème, cette aide constitue une véritable base de connaissances, et peut être vue comme un outil didactique sur l'évaluation. Elle est structurée sous trois formes principales, remplissant des objectifs différents :

- Des indications générales et des reformulations afin d'explicitier la question, d'aider à trouver la réponse la plus adaptée, de sécuriser l'évaluateur (voir Figure 14).
- Une aide approfondie, plus théorique, afin d'aider à mieux comprendre la problématique sous-jacente, à gérer des cas particuliers, à objectiver le regard de l'évaluateur.
- Des exemples caractéristiques permettant d'illustrer les concepts abordés.

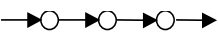
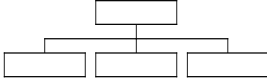
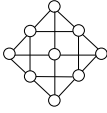
<b>Question</b>		
Caractériser la structure du didacticiel évalué : Linéaire / Arborescente / Réseau		
<b>Indications</b>		
<p>Une structure linéaire est séquentielle et l'utilisateur ne peut que contrôler le flux d'information. Il s'agit des cas classiques du livre ou de la bande magnétique. Cette structure ne profite pas des spécificités du support numérique, mais elle présente l'avantage d'être plus naturelle à appréhender par l'utilisateur. Une structure arborescente est hiérarchique et typiquement basée sur les menus et sous-menus. Il s'agit d'un compromis entre structure linéaire et réseau. La structure réseau est particulière au support numérique (comme les structure Internet). Elles permettent potentiellement des lectures plus riches et plus adaptées puisque chacun peut trouver son chemin particulier dans le réseau. Mais ces structures peuvent également désorienter l'utilisateur, si celui-ci n'est pas assez préparé et guidé dans sa lecture.</p>		
		
Structure linéaire	Structure arborescente	Structure réseau

Figure 14 : Exemple d'aide contextuelle de premier niveau

## 3. Validation

### 3.1 Boucle conception / évaluation

La méthode que nous proposons est issue de recherches dans des domaines variés, tels que l'informatique, l'ergonomie, mais également les sciences cognitives, sociales ou artistiques. Devant la complexité de notre entreprise nous avons adopté une ap-

proche progressive, qui débuta par l'étude de l'aspect ergonomie, pour se porter ensuite sur les angles didactiques et enfin multimédia. A chaque étape notre approche est d'extraire des critères de la littérature concernée, de tester ces critères, de les intégrer à un prototype et de les évaluer en situation réelle. Après chaque évaluation nous entamons un nouveau cycle.

Des travaux critiques ont donc été régulièrement menés depuis la première version de la méthode proposée par [Reithmuller 96]. Ces travaux ont permis de faire progresser itérativement la méthode EMPI jusqu'à sa version actuelle. Nous présentons dans la partie suivante des expériences en situation, ainsi qu'une étude comparative avec MEDA. Notons que ces travaux de validation, dont les résultats ont déjà été exploités, ont principalement porté sur le thème *utilisabilité*, à travers une version simplifiée de EMPI, correspondant à un premier prototype réalisé sous Excel. La version actuelle de la méthode a récemment fait l'objet de nouvelles validations, une nouvelle expérience à grande échelle ayant été menée en mai 1999.

### 3.2 Expériences en situation

**Indice de divergence** : Afin de pouvoir effectuer des comparaisons sur un grand nombre d'évaluations, et ainsi de vérifier la pertinence de nos critères, nous avons mis au point un indice de divergence (Figure 15), variant entre 0 et 1, détaillé dans [Hù 97]. Cet indicateur permet de mesurer les écarts entre les différentes notes attribuées par différents évaluateurs sur un même critère et sur un même didacticiel.

$$I_{i,j} = \frac{|E_i - E_j|}{10 + \max(|E_i|, |E_j|)}$$

avec  $E_i$  : résultat de l'évaluation  $i$   
 $I_{i,j}$  : Indice de divergence entre l'évaluation  $i$  et l'évaluation  $j$

Figure 15 : Calcul de l'indice de divergence

Cet indice a la particularité de tenir compte du fait qu'une même différence de note peut avoir une importance plus ou moins grande. Par exemple la même différence, de 4 unités, entre -2 et 2 (valeurs opposées) est jugée plus conséquente qu'entre 4 et 8 (qui expriment tous deux une note positive, mais avec des intensités différentes). Nous proposons les appréciations suivantes pour interpréter les valeurs numériques des divergences observées :

**Mode expérimental** : Diverses expériences ont été menées (une dizaine d'évaluateurs<sup>1</sup> ont appliqué la méthode à une trentaine de logiciels) ainsi qu'une expérience à grande échelle (quarante-cinq évaluateurs<sup>2</sup> ont été comparés sur un même didacticiel<sup>3</sup>). Le mode expérimental que nous avons adopté est basé sur une

<sup>1</sup> Des étudiants de premier et second cycle.

<sup>2</sup> Des étudiants de second cycle.

<sup>3</sup> Prisméo de l'ENST

courte période de découverte du logiciel à évaluer (environ une heure), puis par l'évaluation proprement dite en trois étapes (durée totale d'environ quatre heures) :

1. L'évaluateur critique le logiciel à partir de ses propres critères
2. L'évaluateur critique le logiciel à partir des critères que nous avons identifiés
3. L'évaluateur applique la méthode EMPI

**Hypothèses** : Nos travaux de validation se sont pour le moment attachés à vérifier deux hypothèses (résumées sur la Figure 16) :

- La stabilité des évaluations d'un même logiciel par des évaluateurs différents. Cette stabilité nous est indispensable pour pouvoir proposer des évaluations standardisées et donc interprétables indépendamment des conditions d'évaluation (dans une certaine mesure, cette remarque sera discutée en conclusion)
- L'inclusion des informations fournies instinctivement par les évaluateurs dans les informations fournies par EMPI. Cette relation est également indispensable dans la mesure où notre méthode est orientée vers les utilisateurs et doit donc s'appuyer sur leur propre représentation des logiciels évalués.

(1) $\forall \text{logiciel, } \text{évaluateur1, } \text{évaluateur2,}$ $EMPI(\text{logiciel, } \text{évaluateur1}) \cong EMPI(\text{logiciel, } \text{évaluateur2})$
(2) $\forall \text{logiciel, } \text{évaluateur,}$ $EMPI(\text{logiciel, } \text{évaluateur}) \subset EVAL\_INSTINCTIVE(\text{logiciel, } \text{évaluateur})$

Figure 16 : Hypothèses à vérifier

**Résultats** : Nous énonçons ci-dessous nos résultats principaux<sup>1</sup> :

- L'utilisation de critères permet une évaluation plus complète, car plus systématique, par rapport à l'évaluation libre, ce qui vérifie la relation d'inclusion (2). Notons que l'extension de cette relation a également été vérifiée, à savoir qu'étant donné un panel de plusieurs évaluateurs, chaque évaluation EMPI comprend l'ensemble des remarques issues de toutes les évaluations instinctives.
- L'utilisation du questionnaire complet permet de réduire notablement la divergence entre les évaluateurs (Figure 17), et donc de normaliser l'évaluation, ce qui va dans le sens de la relation d'équivalence (1).

100% des critères ont une divergence inférieure à 0,33
95% des critères ont une divergence inférieure à 0,28
75% des critères ont une divergence inférieure à 0,22
25% des critères ont une divergence inférieure à 0,17

Figure 17 : Répartition des critères en fonction de la divergence

- L'utilisation de la notation exponentielle permet également de réduire la divergence (environ 10% en moyenne) et est plus proche des évaluations instinctives (les utilisateurs sanctionnant généralement lourdement les logiciels présentant des défauts évidents).

---

<sup>1</sup> Notons que les détails chiffrés de nos résultats d'expériences sont disponibles dans [Hû & al 98]

- Les aspects multimédia, moins normés que les aspects ergonomiques, présentent un taux de divergence plus grand. Notons néanmoins que la version du questionnaire testée n'utilisait pas les questions CE et que de nouvelles recherches ont été menées pour approfondir ces aspects depuis.
- D'une façon générale ces évaluations ont permis de mettre en évidence les points faibles du questionnaire (par exemple la *charge de travail*, divergence 0,31) et de nous diriger vers des approfondissements, des restructurations et des re-formulations.

**Validité d'une évaluation EMPI** : Ces résultats nous ont permis de déterminer qu'une évaluation EMPI même très courte (une heure) apporte plus d'informations à l'évaluateur qu'une évaluation instinctive de même durée, et peut être considérée comme un aperçu fiable du logiciel. Une évaluation standard, d'une durée de trois à quatre heures, fournit un rapport détaillé sur le logiciel et permet de remplir les objectifs de comparaison entre logiciels et de recherche d'adéquation à une situation d'apprentissage donnée. Des évaluation plus longues (huit heures) permettent de détecter en profondeur les défauts des logiciels traités. Cela n'est pas nécessaire dans notre cadre, mais peut le devenir si EMPI est utilisé comme un outil d'expertise du logiciel, dans un objectif d'amélioration par exemple. Ces utilisations *détournées* qu'EMPI suscite seront discutée plus avant en conclusion.

### 3.3 MEDA versus EMPI

Le collectif européen Joanna MEDA a proposé en 1990 une méthodologie permettant d'évaluer des logiciels de formation en entreprise, à partir d'un questionnaire basé sur des critères pédagogiques, ergonomiques, commerciaux, techniques, ... [MEDA 90]. Une étude comparative entre MEDA et EMPI, menée en 1998, nous a permis de mettre en évidence les points communs et les divergences entre les deux méthodes. L'étude a été menée grâce à l'évaluation de plusieurs mêmes didacticiels multimédia par les deux méthodes. Nous avons utilisé MEDA 97 sur cédérom, ne prenant en compte que la partie concernant l'ergonomie pour notre étude, la version de EMPI utilisée ne maîtrisant alors que ce thème. Nous présentons ici les remarques et propositions principales qui ont pu être faites à la suite de cette étude. On renverra à [Crozat 98] pour de plus amples détails sur l'étude.

**Mode de réponse** : Alors que MEDA propose une notation en *amplitude* – en qualifiant directement le critère par une appréciation « mauvais », « bon », « très bon », ... – EMPI propose une notation en *fréquence*, c'est à dire qu'elle qualifie indirectement les critères en fonction de la fréquence des problèmes relevés. En pratique, on est souvent amené à vouloir répondre par des formulations du type « souvent bien ET exceptionnellement très mauvais », comme par exemple pour exprimer qu'au sein de l'homogénéité globalement bonne de telle interface, le fait d'avoir inversé sur un écran le bouton « quitter » et le bouton « aide » s'est avéré lourd de conséquences. Ceci nous conduit à penser que sur ce point les deux méthodes sont complémentaires.

**Personnalisation des grilles** : On distingue l'approche *asynchrone* de MEDA qui sépare en deux phases la sélection des questions puis la réponse effective, de

l'approche *synchrone* de EMPI qui offre de répondre ou non au fur et à mesure des questions. L'avantage de la première approche est qu'elle permet de contextualiser le questionnaire, et d'avoir une vision globale sur ses objectifs d'évaluation. La seconde, quant à elle, est plus applicable pratiquement, dans la mesure où c'est souvent en répondant effectivement qu'on se rend le mieux compte de la pertinence des questions par rapport à son logiciel. Cette fois encore les deux approches se complètent, et nous proposons une solution fondée sur une personnalisation guidée dans un premier temps, permettant de proposer un questionnaire type modifiable a priori, puis de laisser à l'utilisateur la possibilité d'ignorer les questions et les critères en fin de compte inutiles ou inapplicables.

**Système de notation et présentation des résultats** : La pondération des critères et des questions telle que proposée dans MEDA est efficace pour mettre en valeur les points importants en fonction du contexte. Certaines spécificités d'EMPI ont également été validées, notamment la bonne exploitabilité de la notation exponentielle, la représentation des critères sous forme de *Pareto* (voir Figure 13).

**Représentation des disparités à l'intérieur d'un critère** : La possibilité de rendre compte sur une représentation graphique (par exemple comme sur la Figure 18) de la disparité des réponses aux questions d'un même critère manque aux deux méthodes. En effet il est important de marquer la différence entre un critère *moyen* dans le sens où toutes les questions le concernant obtiennent une réponse moyenne ; et un critère *moyen* c'est à dire comportant de très bonnes et de très mauvaises choses.

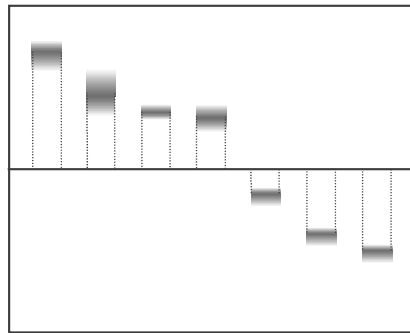


Figure 18 : Pareto avec indice de disparité

**Réutilisation des questions** : Dans MEDA certaines questions affectent plusieurs critères. Cette vision paraît juste dans certains cas (par exemple un menu inadapté gênera à la fois la scénarisation et l'ergonomie d'utilisation). Cette approche nous a conduit à émettre deux hypothèses pour EMPI :

- faire remonter la note calculée d'une question à plusieurs critères lorsque celle-ci les concerne
- définir les critères de telle façon qu'ils ne se recoupent pas et rédiger chaque question pour un seul critère

Nous avons, pour le moment au moins, opté pour la seconde hypothèse, plus simple à mettre en œuvre et plus cohérente pour l'évaluateur.

**Importance de l'aide à l'évaluation** : Un autre des éléments importants de ces méthodes est leur capacité à fournir des explications à l'évaluateur. On peut remarquer trois niveaux d'information différents potentiellement utiles à l'évaluation :

- l'explicitation de la question : « Qu'est ce que ça veut dire ? »
- la théorie sous-jacente : « Pourquoi me pose-t-on cette question ? »

- l'exploitation de la question : « Comment ma réponse est-elle utilisée par la méthode ? »

Il est important que ces aides soient les plus neutres possible, de façon à guider l'évaluateur dans son choix sans l'influencer outre mesure. Enfin il paraît également intéressant d'introduire le multimédia dans l'aide explicative. En effet certaines questions pourraient être clarifiées si elles étaient correctement illustrées (par exemple en opposant exemples et contre-exemples d'interfaces, en proposant des tests, ...).

Cette expérience nous a également permis de constater la bonne cohérence entre les méthodes MEDA et EMPI, qui chacune avec leurs propres questions parviennent à des résultats proches pour des évaluations des mêmes didacticiels par un même évaluateur. Indépendamment de la validation relative qui en ressort, l'étude comparative des résultats nous a permis de souligner certains points faibles à améliorer (comme la difficulté pour EMPI de traiter la charge cognitive sans éléments pédagogiques).

#### **4. Conclusions, discussions, perspectives**

Proposer un outil généraliste pour aider à l'évaluation de tous types de didacticiels est une tâche ambitieuse, certains diront impossible, d'autres indispensable. Devant des difficultés telles que la gestion des aspects subjectifs ou l'évaluation de critères intrinsèquement contextuels, nous avons adopté une attitude humble : la méthode ne permet pas réellement l'évaluation, mais le guidage d'un humain dans son processus d'évaluation. EMPI fournit une description structurée du didacticiel, mais cette description doit ensuite être interprétée par un formateur en fonction d'un contexte, d'objectifs, d'un dispositif, ... Ceci nous conduit également à certaines restrictions, comme par exemple le fait que l'évaluateur doive avoir une certaine connaissance du contexte pédagogique ou encore des capacités de recul pour généraliser ses appréciations.

L'approche modulaire que nous proposons va dans le sens de l'adaptation de chaque évaluation en fonction des évaluateurs et des domaines. Néanmoins la version actuelle ne propose pas d'assistance particulière pour aider l'évaluateur à choisir a priori les critères pertinents dans son cadre. Les validations que nous avons menées et celles à venir nous permettent de cerner les limites de la méthode. Ainsi nous commençons à entrevoir que les modules *qualité technique* ou *utilisabilité* ne nécessiteront aucune compétence particulière de la part de l'évaluateur, alors qu'il en sera autrement pour le module *didactique*. De même la méthode ne s'applique pas identiquement à tous les contextes pédagogiques, et nous commençons à entrevoir des champs privilégiés en terme de matière, de public, de type de didacticiel, ...

La validité des évaluations EMPI peut également être discutée puisqu'elle dépend du contexte, des évaluateurs, des modalités d'évaluation, et de l'interprétation qu'en fera le formateur. Nos travaux de validations prouvent qu'il est possible de réduire la divergence entre les évaluations grâce à des structuration et des formulations des connaissances claires pour l'utilisateur. La validité d'une évaluation sera néanmoins

d'autant plus forte que l'évaluateur aura des connaissances propres et que l'évaluation sera approfondie. Ajoutons qu'une évaluation EMPI effectuée par un évaluateur non expert sera toujours inférieure à une réelle expertise, exhaustive et en situation du logiciel. De même l'interprétation d'une évaluation EMPI ne donne lieu qu'à des prévisions et sera moins certaine que l'évaluation d'un apprentissage réel après utilisation du logiciel. Notre ambition est donc de rendre possible une évaluation dans les nombreux cas où ces méthodes meilleures ne peuvent être appliquées (Quel enseignant a les moyens d'effectuer des évaluations a posteriori sur plusieurs logiciels ?), tout en étant conscient des limites d'une telle approche par rapport à ces méthodes.

La base de connaissances structurée qu'est EMPI ouvre des horizons autre que le cadre stricte qui est le sien aujourd'hui. Nous entrevoyons deux prolongements à nos travaux actuels, la formation et l'utilisation dans un contexte de conception. Nos recherches futures se dirigent donc vers l'explicitation de la méthode, afin que l'évaluateur acquière une compétence, notamment grâce aux informations intégrées dans l'aide. Nous avons en effet remarqué que l'utilisation de la méthode aide à l'assimilation des concepts manipulés. Nous souhaitons effectivement faire en sorte que cet effet secondaire soit le plus constructif possible, de façon à former l'œil des évaluateurs qui utilisent EMPI à la *lecture* des didacticiels. Si l'on peut aider à lire, on peut inévitablement aider à *écrire*, et nous travaillons donc également à l'utilisation des critères dans un contexte de conception, par l'intermédiaire de propositions de chartes ergonomiques, graphiques, pédagogiques.

---

### Références citées

- [Alekan 84] Henri Alekan, « Des lumières et des ombres », Le sycamore, 1984.
- [Bachimont 99a] Bruno Bachimont, « L'intelligence artificielle comme écriture dynamique : de la raison graphique à la raison computationnelle », in « Au nom du sens », Actes de Cerisy la Salle, Grasset, Ed Jean Petitot, 1999.
- [Bachimont 99b] Bruno Bachimont, « Herméneutique du support et ingénierie des connaissances : de la différence numérique », (à paraître).
- [Balpe et al 96] Jean-Pierre Balpe, Alain Lelu, Fabrice Papy, Imad Saleh, «Techniques avancées pour l'hypertexte», Hermès, Paris, 1996.
- [Barthes 80] Roland Barthes, « La chambre claire : Note sur la photographie », Editions de l'Etoile, Gallimard, Le Seuil, 1980.
- [Bastien, Scapin 94] Bastien C., Scapin D., Evaluating a user interface with ergonomic criteria. Rapport de recherche INRIA n°2326 Rocquencourt, aout 1994.
- [Baticle 85] Yveline Baticle, « Clés et codes de l'image : L'image numérisée, la vidéo, le cinéma », Magnard, Paris, 1985.
- [Berbaum 88] Berbaum J., Un programme d'aide au développement de la capacité d'apprentissage, Grenoble, Université de Grenoble II, Multigraphié, 128p, 1988.
- [Carrière 93] Jean-Claude Carrière, « Raconter une histoire », FEMIS, 1993.
- [Chion 94] Michel Chion, « Musiques : Médias et technologies », Flammarion, 1994.

- [Cornu 90] Geneviève Cornu, « Sémiologie de l'image dans la publicité », Les Editions d'Organisation, 1990.
- [Cossette 83] Claude Cossette, « Les images démaquillées », Riguil Internationales, 2<sup>ème</sup> édition, Québec, 1983.
- [Costa, Moles 91] Joan Costa, Abraham Moles, « La imagen didáctica », Ceac, Barcelone, 1991.
- [Crozat 98] Stéphane Crozat, « Méthode d'évaluation de la composition multimédia des didacticiels », Mémoire de DEA CDS, Université de Technologie de Compiègne, 1998.
- [Crozat & al 99a] Stéphane Crozat, Olivier Hû, Philippe Trigano, « EMPI : a questionnaire based method for the evaluation of multimedia interactive pedagogical software », International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'99), session 'Multimedia Communications, Systems and Applications', USA, Juin 1999.
- [Crozat & al 99b] Stéphane Crozat, Olivier Hû, Philippe Trigano, « A method for evaluating multimedia learning software », IEEE Multimedia ICMCS'99, Florence, Italie, juin 1999.
- [Crozat & al 99c] Stéphane Crozat, Olivier Hû, Philippe Trigano, « EMPI, un guide logiciel d'aide à l'évaluation du multimédia pédagogique », congrès AIPU, Montréal, Canada, mai 1999.
- [Debray 94] Régis Debray, « L'œil naïf », Seuil, 1994.
- [Depover 94] Depover C., Problématique et spécificité de l'évaluation des dispositifs de formation multimédias, *Educatechnologie\**, vol.1, n°3, septembre 1994.
- [Dessus, Marquet 91] Dessus P., Marquet P., Outils d'évaluation de logiciels éducatifs. Université de Grenoble. Bulletin de l'EPI. 1991.
- [Durand & al 97] Alain Durand, Jean-Marc Laubin, Sylvie Leleu-Merviel, « Vers une classification des procédés d'interactivité par niveaux corrélés aux données », H<sup>2</sup>PTM'97, Hermes, 1997.
- [Gagné, Medsker 96] Robert M. Gagné, Karen L. Medsker, « The conditions of learning : Training applications », Harcourt Brace & Compagny, USA, 1996.
- [Gaudreault, Jost 90] André Gaudreault, François Jost, « Le récit cinématographique », Nathan, 1990.
- [Gibson 79] James J. Gibson, « The ecological approach to visual perception », LEA, Londres, 1979.
- [Goody 79] Jack Goody, « La raison graphique : La domestication de la pensée sauvage », Les Editions de Minuit, 1979.
- [Hannafin, Peck 88] Hannafin M.J., Peck K.L., « The Design, Development and Evaluation of Instructional Software », NY, MacMillan Publishing Company, 1988.
- [Hû 97] Olivier Hû, « Méthodologie d'évaluation du multimédia pédagogique », Mémoire de DEA CDS, Université de Technologie de Compiègne, 1997.
- [Hû & al 98] Olivier Hû, Philippe Trigano, Stéphane Crozat, « E.M.P.I. : une méthode pour l'Evaluation du Multimédia Pédagogique Interactif », NTICF'98, INSA Rouen, novembre 1998.

- [Hû, Trigano 98] Olivier Hû, Philippe Trigano, « Proposition de critères d'aide à l'évaluation de l'interface homme-machine des logiciels multimédia pédagogiques », IHM'98, Nantes, septembre 1998.
- [Kivy 89] Peter Kivy, « Sound sentiment : An essay on the musical emotion », Temple University Press, Philadelphia, 1989.
- [Kolsky 97] Kolski C., Interfaces Homme-machine : application aux systèmes industriels complexes, Hermes, 1997.
- [Lamrous, Trigano 97] Sid Lamrous, Phillippe Trigano, « L'organisation des espaces documentaires : vers une exploitation optimale », revue Document Electronique, Hermès, Vol 1, n°4, décembre 1997.
- [Marton 94] Marton Ph., La conception pédagogique de systèmes d'apprentissage multimédia interactif : fondements, méthodologie et problématique, Educatechnologie\*, vol.1, n°3, septembre 1994.
- [MEDA 90] MEDA, « Evaluer les logiciels de formation », Les Editions d'Organisation, 1990.
- [Milon, Cormerais 99] Alain Milon, Franck Cormerais, "Gestion et management de projet multimédia: Du cahier des charges à la commercialisation", L'Harmattan, Paris, 1999.
- [Pajon, Polloni 97] Patick Pajon, Olivier Polloni, « Conception multimédia », cd-rom, CINTE, 1997.
- [Park & al. 93] Park I., Hannafin M.-J., Empiically-based guidelines for the design of interactive media, Educational Technology Research en Development, vol.41, n°3, 1993
- [Parker, Thérien 91] Roger C . Parker, Lise Thérien, « Mise en page et conception graphique », Reynald Goulet, 1991.
- [Piché & al 98] Patrick Piché, Claude Frasson, Esma Aimeur, « Amélioration de la formation au moyen d'un agent perturbateur dans un système tutoriel intelligent », NTICF'98, INSA Rouen, novembre 1998.
- [Pognant, Scholl 96] Patrick Pognant, , Claire Scholl « Les cédéromculturels », Hermès, 1996.
- [Poynor 91] Rick Poynor, , « Typography Now : The next wave », Edward Booth-Clibborn, 1991.
- [Ravden & al. 89] Ravden S.J., Johnson G.I., Evaluating usability of Human-Computer Interfaces : a practical method. Ellis Horwood, Chichester, 1989.
- [Rhéaume 94] Rheau J., L'évaluation Des Multimédias Pédagogiques : De L'évaluation Des Systèmes A L'évaluation Des Actions, Educatechnologie\*, Vol.1, N°3, Sept. 1994.
- [Richaudeau 80] Richaudeau F., Conception et production des manuels scolaires, Paris, Retz, 290p, 1980.
- [Riethmuller 96] Riethmuller V., Méthodologie d'évaluation de l'interface Homme/Machine. Rapport de DEA CDS, Université de Technologie de Compiègne, septembre 1996
- [Scapin 86] Scapin D., Guide ergonomique de conception des interfaces Homme/Machine. Rapport technique INRIA Rocquencourt, n°77, octobre 1986.

- [Scapin, Bastien 97] Dominique Scapin, J.M.Christian Bastien, « Ergonomic criteria for evaluating the ergonomic quality of interactive systems », Behaviour & Information Technology, n°16, 1997.
- [Selys 98] Gérard De Selys, « L'école grand marché du XXI<sup>ème</sup> siècle », Le Monde Diplomatique, juin 1998.
- [Senach 90] Senach B., Evaluation ergonomique des interfaces Homme/Machine : une revue de la littérature. Rapport INRIA, Sophia-Antipolis, n°1180, Rocquencourt, mars 1990.
- [Sorlin 92] Pierre Sorlin, « Esthétiques de l'audiovisuel », Nathan, 1992.
- [Type, Frommer 85] J.H. Mc Type, W.D. Frommer, « Effects on character / background colour combination on CRT character lisibility », HFS'85, proceedings of human factor society, 1985.
- [Vanderdonck 94] Vanderdonck J., Guide ergonomique de la présentation des applications hautement interactives, Presses Universitaires Namur, 1994.
- [Vanderdonck 98] Jean Vanderdonck, « Conception ergonomique de pages WEB », Vesale, 1998.
- [Vanoye, Goliot-Lété 92] Francis Vanoye, Anne Goliot-Lété, « Précis d'analyse filmique », Nathan, 1992.
- [Vivet 96] Martial Vivet, « Evaluating educational technologies : Evaluation of teaching material versus evaluation of learning ? », CALISCE'96, San Sebastian, juillet 1996.
- [Weidenfeld & al. 96] Weidenfeld G., Caillot M., Cochard G.-M., Fluhr C., Guerin J.-L., Lecllet D., Richard D., Techniques de base pour le multimédia, Ed. Masson, Paris, 1996.