



**HAL**  
open science

## EMPI, un guide logiciel d'aide à l'évaluation du multimédia pédagogique

Stéphane Crozat, Olivier Hû, Philippe Trigano

► **To cite this version:**

Stéphane Crozat, Olivier Hû, Philippe Trigano. EMPI, un guide logiciel d'aide à l'évaluation du multimédia pédagogique. AIPU'99, May 1999, Montréal, Canada. edutice-00000400

**HAL Id: edutice-00000400**

**<https://edutice.hal.science/edutice-00000400>**

Submitted on 10 Mar 2004

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

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

# EMPI, un guide logiciel d'aide à l'évaluation du multimédia pédagogique

Stéphane CROZAT, Olivier HU, Philippe TRIGANO

*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*

## Abstract

Aujourd'hui la question ne se pose pas de savoir s'il faut ou non favoriser l'apprentissage assisté par ordinateur. De fait 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'utilisabilité, les documents multimédia, la scénarisation et la didactique. Chacun de ces six thèmes est divisé en méta-critères, critères et questions. L'ensemble forme 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.

Ce papier présente succinctement la structure du questionnaire, puis s'attache à décrire le fonctionnement de la méthode informatisée ainsi que les validations menées.

## Mots-clés

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

## 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 cd-rom 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 qu'il représente. Des défauts de conception, des contenus trop pauvres, des interfaces inefficaces, des mauvais usages du multimédia, des stratégies pédagogiques inadaptées, sont autant d'exemple d'erreurs courantes. L'alternative et la complémentarité représentées par les didacticiels ne peuvent être négligées pour autant, 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 caractériser et d'évaluer les didacticiels.

Historiquement, la nécessité de mesurer la qualité des didacticiels est issue de deux préoccupations plus anciennes : 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 sera 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 rappelé la structure de notre grille d'évaluation, nous présenterons 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

La méthode EMPI repose essentiellement sur un questionnaire structuré. 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. La structure globale de notre évaluation se base sur six thèmes complémentaires :

- Les **impressions générales**<sup>1</sup> : Les expériences que nous avons menées, ainsi que des études bibliographiques, telles que la théorie des affordances [Gibson 79], nous ont permis de proposer une liste de critères propres à caractériser l'image que le didacticiel offre à l'utilisateur
- La **qualité informatique**<sup>2</sup> : La mise au point technique du logiciel (installation, rapidité, robustesse, ...) est une première discrimination significative. Nos recherches se sont largement inspirées de l'existant dans le domaine du génie logiciel.
- L'**utilisabilité**<sup>3</sup> : Il existe de nombreux critères d'évaluation de l'interface (des sept règles d'or de J. Coutaz aux 3000 règles de J. Vanderdonckt). Nous proposons un compromis inspiré des travaux de l'INRIA [Scapin 86], [Senach 90], [Bastien, Scapin 94], [Scapin, Bastien 97].
- Les **documents multimédia**<sup>4</sup> : Les textes, sons et images constituent la matière première du didacticiel multimédia, ils sont vecteurs de l'information. On envisage dans ce thème l'évaluation de la forme et de la composition des contenus (présentation, lisibilité, adéquation, sémantique, ...).
- La **scénarisation**<sup>5</sup> : La scénarisation est le procédé particulier qui consiste à structurer les documents de façon à en préparer la lecture. On évalue notamment l'intégration des spécificités liées au contexte pédagogique et au support numérique dans l'écriture proposée.

<sup>1</sup> Travaux de recherche et résultats décrits dans [Crozet 98a], [Crozet & al 99]

<sup>2</sup> Recherches en cours de finalisation

<sup>3</sup> Travaux décrits dans [Hû 97], [Hû, Trigano 98], [Hû & al 98]

<sup>4</sup> Travaux décrits dans [Crozet 98a], [Crozet & al 99]

- La **didactique**<sup>6</sup> : A partir de recommandations pédagogiques [Dessus, Marquet 91], [Marton 94], [Park & al 93] et des travaux du collectif européen MEDA [MEDA 90], nous proposons un ensemble de critères permettant de confronter les potentialités et les outils proposés dans le cadre du didacticiel avec le contexte pédagogique particulier dans lequel il doit être utilisé.

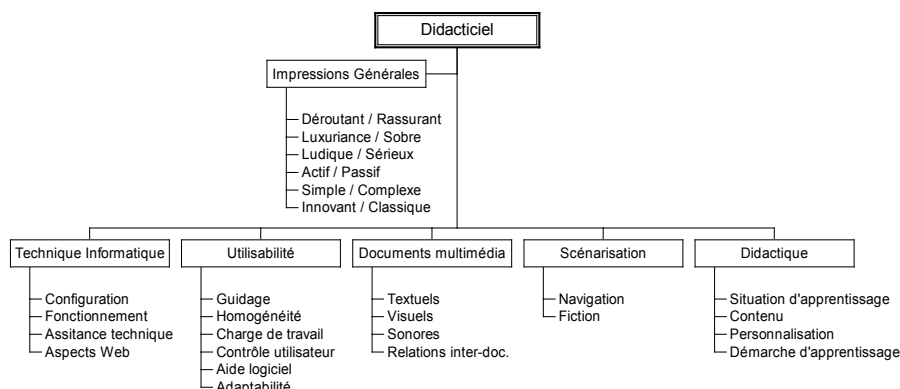


Figure 1 : Hiérarchie des thèmes et méta-critères

Les six thèmes se divisent en méta-critères (voir Figure 1), eux même divisés en critères, afin de former une structure à trois niveaux. C'est finalement au niveau de chaque critère que sont placées les questions qui composent effectivement le questionnaire.

## 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 très généraliste, et en ce sens 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 2 met en évidence un exemple d'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 critères.

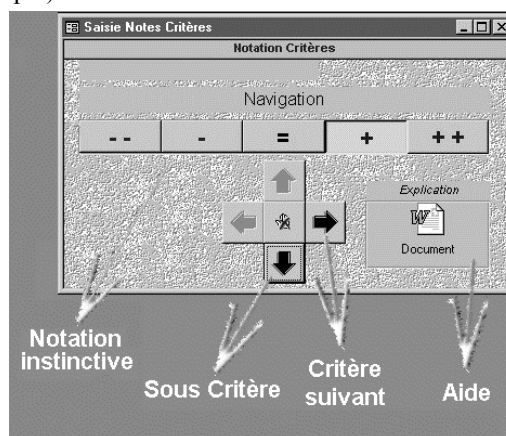


Figure 2 : Navigation entre les critères

### 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 :

<sup>5</sup> Travaux décrits dans [Crozat 98a], [Crozat & al 99]

<sup>6</sup> Travaux décrits dans [Salesse 97], [Trigano & al 98]

- La notation instinctive (NI) est l'appréciation directe des thèmes, méta-critères et critères par l'évaluateur. Cette évaluation se fait sur une échelle [– – ; – ; = ; + ; ++], voir Figure 2.
- La notation calculée (NC) est obtenue à partir des réponses aux questions (voir Figure 5). 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 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 les 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 évaluations 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, il y a un problème de cohérence de sa part.
- L'indice de corrélation met en relief la correspondance entre la note instinctive et la note calculée.

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 ne permet pas une confrontation *on-line* des notations. Le logiciel fournit un rapport imprimable qui permet la saisie des notes finales. La Figure 3 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 3 : Extrait de rapport d'évaluation

**Notation exponentielle** : La notation calculée repose sur un principe permettant de mettre en valeur les informations 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 4.

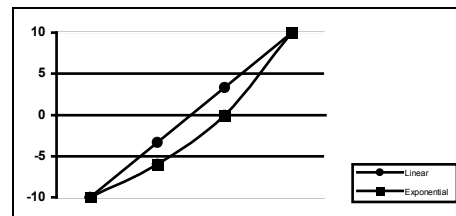


Figure 4 : Notation exponentielle vs notation linéaire

**Pondération des questions** : A l'intérieur de chaque critère on distingue une ou deux questions pour lesquelles la réponse est particulièrement signifiante et déterminante. Un système de pondération est donc appliqué de façon à distinguer les questions standards des fondamentales.

### 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. 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 5).

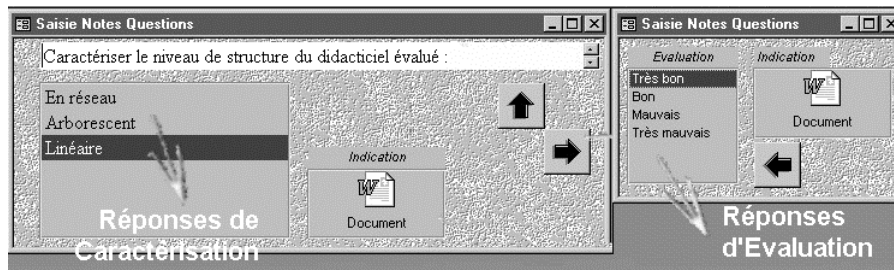


Figure 5 : Question " caractérisation / évaluation "

## 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*. La représentation est à précision variable, à l'instar du questionnaire, et peut concerner les différents niveaux du questionnaire (thèmes, méta-critères ou critères). Notons que les histogrammes classiques sont plus adaptés pour des comparaisons entre logiciels, tandis que les Paretos mettent mieux en évidence les défauts et qualités du didacticiel.

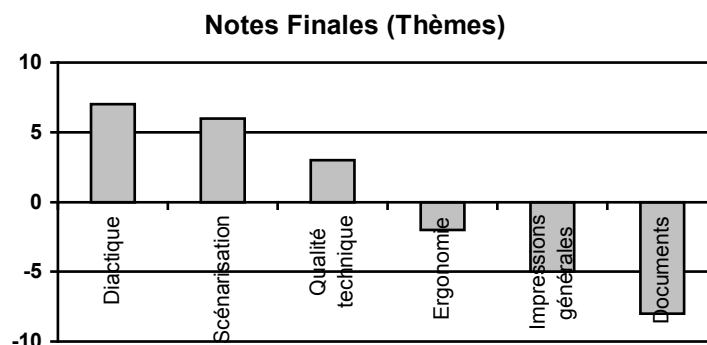


Figure 6 : Représentation des résultats au niveau des Thèmes sous forme de Pareto

## 2.5 Aide

Une aide contextuelle est proposée à l'évaluateur au cours de son parcours du questionnaire (voir Figure 2). Cette aide est structurée sous trois formes principales, remplissant des objectifs différents :

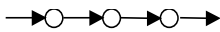
- Des indications générales et des reformulations afin d'explicitier la question, d'aider à trouver la réponse la plus adaptée, de sécuriser l'évaluateur (voir Figure 7).
- Une aide approfondie, plus théorique, afin d'aider à mieux comprendre la problématique sous-jacente, à gérer des cas particuliers, à objectiver le regard de l'évaluateur.
- Des exemples caractéristiques permettant d'illustrer les concepts abordés.

## Question

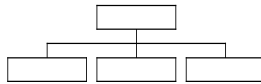
Caractériser le niveau de structure du didacticiel évalué : Linéaire / Arborescente / Réseau

## Indications

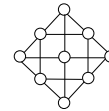
Une structure linéaire est séquentielle et l'utilisateur ne peut que contrôler le flux d'information. Il s'agit des cas classiques du livre ou de la bande magnétique. Cette structure ne profite pas des avantages du support numérique, mais elle présente l'avantage d'être plus naturelle à appréhender par 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 spécifique 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.



Structure linéaire



Structure arborescente



Structure réseau

Figure 7 : Exemple d'aide contextuelle du 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 approche 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 en 1996. 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 fait l'objet de nouvelles validations dont nous aurons les conclusions en janvier 1999, tandis qu'une nouvelle expérience à grande échelle sera menée en mars 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 8), détaillé dans [HU 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 8 : 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 importante. Par exemple la même différence, de 4 unités, entre -2 et 2 (valeurs opposées) est jugée de plus de conséquences qu'entre 4 et 8 (qui expriment tous deux une note positive, mais avec des intensités différentes).

**Mode expérimental** : Diverses expériences ont été menées (une dizaine d'évaluateurs<sup>7</sup> ont appliqué la méthode à une trentaine de logiciels) ainsi qu'une expérience à grande échelle (quarante-cinq évaluateurs<sup>8</sup> ont été comparés sur un même didacticiel<sup>9</sup>). Le mode expérimental que nous avons adopté est basé sur une courte période de découverte du logiciel à évaluer (environ une heure), puis par l'évaluation proprement dite en trois étapes :

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

**Résultats** : Une description détaillée de ces évaluations et de leurs résultats est disponible dans [Hû & al 98]. On rappellera néanmoins les quelques résultats suivants :

- L'utilisation de critères permet une évaluation plus complète, car plus systématique, par rapport à l'évaluation libre.
- L'utilisation du questionnaire complet permet de réduire notablement la divergence entre les évaluateurs, et donc de normaliser l'évaluation.
- L'utilisation de la notation exponentielle permet également de réduire la divergence et est plus proche des évaluations instinctives (les utilisateurs sanctionnant généralement lourdement les logiciels présentant des défauts évidents).
- Les aspects multimédia, plus subjectifs 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 pointer sur les points faibles du questionnaire et de diriger nos axes de recherches.

### 3.3 MEDA versus EMPI

Le collectif européen Joanna MEDA propose 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 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 cd-rom, 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 98b] 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 aspects 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 6).

---

<sup>7</sup> Des étudiants de premier et second cycle.

<sup>8</sup> Des étudiants de second cycle.

<sup>9</sup> Prisméo de l'ENST



**Représentation des disparités à l'intérieur d'un critère :** La possibilité de rendre compte sur une représentation graphique (par exemple comme sur la Figure 9) 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.

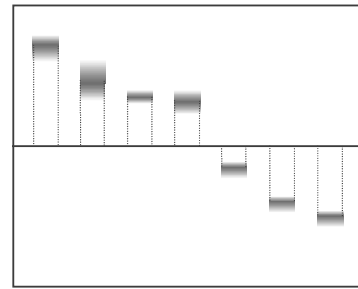


Figure 9 : Pareto avec indice de disparité

**Réutilisation des questions :** Dans MEDA certaines questions affectent plusieurs critères. Cette vision paraît juste dans certains cas (par exemple un menus 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 aider à choisir sans 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).

## Conclusion et Perspectives

Proposer un outil généraliste pour aider à l'évaluation de tous types de didacticiels est une tâche ambitieuse, certains diront impossible, mais 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. Ceci nous conduit 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 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, ...

Nos recherches futures se dirigent également vers l'explicitation de la méthode, afin que l'évaluateur acquière une compétence à l'utilisation de la base de connaissance représentée par l'aide intégrée. 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

- [Alekan 84] Henri Alekan, “ Des lumières et des ombres ”, Le sycomore, 1984.
- [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 98a] Stéphane Crozat, “ Méthode d’évaluation de la composition multimédia des didacticiels ”, Mémoire de DEA, UTC, 1998.
- [Crozat 98b] Stéphane Crozat, “ MEDA versus EMPI ”, Rapport interne, UTC, 1998.
- [Crozat & al 99] Stéphane Crozat, Philippe Trigano, Olivier Hû, “ EMPI : Une méthode informatisée pour l’évaluation des didacticiels multimédias ”, Rapport interne, UTC, 1998.
- [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 ”, HPTM’97, Hermes, 1997.
- [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.
- [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 97] Olivier Hû, “ Méthodologie d’évaluation du multimédia pédagogique ”, Mémoire de DEA, UTC, 1997.
- [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.
- [Kolsky 97] Kolski C., “ Interfaces Homme-machine : application aux systèmes industriels complexes ”, Hermes, 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.
- [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 cd-romculturels ”, Hermès, 1996.
- [Ravden & al. 89] Ravden S.J., Johnson G.I., “ Evaluating usability of Human-Computer Interfaces : a practical method ”, Ellis Horwood, Chichester, 1989.
- [Rhéaume 94] Rheaume 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.
- [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.
- [Trigano & al 98] Phillipe Trigano, Olivier Hû, Stéphane Crozat, “ EMPI : Une méthode pour évaluer les logiciels multimédia de formation ”, Rapport interne, UTC,
- [Vanderdonckt 94] Vanderdonck J., “ Guide ergonomique de la présentation des applications hautement interactives ”, Presses Universitaires Namur, 1994.
- [Vanderdonckt 98] Jean Vanderdonckt, “ 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., Leclet D., Richard D., “ Techniques de base pour le multimédia ”, Ed. Masson, Paris, 1996.