

DES-TEF 2001-2002

Évaluer des logiciels éducatifs

**Bernadette Charlier
Cellule d'Ingénierie Pédagogique
Département Education et Technologie
FUNDP**

Evaluer des logiciels éducatifs hyper et multimédias

Cadre de référence¹

1. Pourquoi un cadre de référence ?

L'évaluation des logiciels éducatifs est un domaine vaste et complexe. Il existe de nombreuses approches méthodologiques et plusieurs acteurs sont impliqués. Un cadre de référence est nécessaire pour situer les principales approches méthodologiques.

Ce cadre a été élaboré au cours d'une recherche commanditée en 1997 et 1998 par la DG. XXII des Communautés Européennes (programme SOCRATES) à laquelle la Cellule d'Ingénierie Pédagogique du Département Education et Technologie a participé.

➤ Quel critère de référence ?

Il n'est pas possible de parler d'évaluation sans évoquer un critère de référence. Ainsi, dans cette étude, on considère qu'un des premiers buts d'un logiciel éducatif est d'améliorer l'apprentissage. Partant de cette intention, on peut considérer que toute méthode d'évaluation se fonde sur une théorie de l'apprentissage (explicite ou non). Il faut cependant souligner que bien que cette proposition apparaisse comme une lapalissade, beaucoup de méthodes d'évaluation apparaissent comme « a-théoriques ».

Pour constituer le cadre de référence présenté de manière synthétique dans ces notes, les auteurs ont privilégié la théorie socio-constructiviste.

Un modèle à trois dimensions prenant en compte les buts de l'évaluation, les contextes (y compris les acteurs concernés) et les méthodologies utilisées permettant de représenter la variété des pratiques a été développé.

Dans la suite, nous construisons ce modèle progressivement en détaillant d'abord chacune de ses dimensions. Nous illustrerons ensuite certaines pratiques par quelques exemples caractéristiques.

2. Les buts de l'évaluation

On peut distinguer quatre buts principaux pour l'évaluation de logiciels éducatifs :

➤ Améliorer l'enseignement et l'apprentissage ;

C'est-à-dire rechercher des usages des logiciels pouvant améliorer l'apprentissage et l'enseignement.

Exemple : compte-rendu EPI (N°65) (en annexe)

➤ Susciter la mise en place d'une politique régionale ou nationale ;

C'est-à-dire recommander voire imposer des critères pour l'élaboration des cahiers des charges ou pour la sélection des logiciels utilisés

¹ Inspiré de Squires et al. 1999

Exemple : recommandation issue de cette étude (en annexe) voir aussi RIP français

➤ Contribuer au développement d'un logiciel éducatif

C'est-à-dire fournir aux développeurs les informations nécessaires à la qualité du développement (depuis l'élaboration du cahier des charges, en passant par l'évaluation des prototypes, et l'évaluation en contexte)

Exemple : évaluation participative de NESTOR (en annexe)

➤ Contribuer au développement du marché du logiciel éducatif

C'est-à-dire favoriser le marché du logiciel en réalisant des études de marché, en définissant des normes ou des labels communs.

Exemple : <http://www.educnet.education.fr/res/bliste.htm> label RIP accordé par le CNDP

3. Les méthodes d'évaluation

Les méthodes habituellement utilisées peuvent être classées en six grandes catégories² :

➤ Méthodes catégorielles

Il s'agit d'attribuer un label, une étiquette au logiciel. Il y a plusieurs exemples de typologies de ce type. Par exemple, celle proposée par l'OCDE en 1989 : drill, tutoriel, tutoriel intelligent, simulation et modélisation, résolution de problèmes, jeux, bases de données, traitement de texte, gestion de l'apprentissage, aide à l'expérimentation de laboratoire, exploration et découverte.

➤ Check-list (ou méthode par jugement)

Il s'agit de fournir un large ensemble de questions ayant trait à plusieurs ensembles de critères différents : par exemple l'ergonomie, les qualités pédagogiques, ... L'évaluateur est appelé à répondre oui ou non pour chacun des critères. Le profil ainsi constitué fournit le jugement de l'évaluateur sur la qualité du logiciel. Un bon exemple est la méthode MEDA.

➤ Evaluation qualitative

Il s'agit de rassembler et d'analyser un ensemble de données qualitatives issues d'interviews, d'observations, de questionnaires. Un bon exemple est la démarche d' « évaluation en contexte » du logiciel MEDA réalisée au cours de ce module spécialisé.

➤ Evaluation quantitative

²Ces catégories ne sont pas exclusives. Elles aident cependant à envisager un ensemble de possibles.

Ce type d'évaluation est basée sur des traitements statistiques appliqués à des prétests et post-tests.

➤ Evaluation éclectique

Correspond à l'application cohérente de plusieurs méthodes intégrant à la fois des méthodes d'évaluation qualitative et quantitative.

➤ Evaluation critique

Est basée sur l'avis subjectif d'experts. Jusqu'il y a peu le CNDP fonctionnait par commissions d'inspecteurs qui donnaient leur avis sans explicitation de critères particuliers. La Commission Européenne organise ses évaluations également par commissions d'experts.

➤ Evaluation synthétique

Correspond à la tentative d'élaborer un cadre théorique synthétique pour réaliser l'évaluation. Un bon exemple est la grille proposée par Reeves (1994) en annexe.

4. Les acteurs de l'évaluation

Trois types d'acteurs peuvent être identifiés : les évaluateurs, le public et les évalués.

➤ Les évaluateurs

Un ensemble de personnes peuvent agir comme évaluateurs. Il s'agit des enseignants, des chercheurs, des développeurs, des éditeurs etc.. Evidemment, certaines catégories d'évaluateurs sont plus intéressés par certains buts de l'évaluation. Par exemple, les enseignants seront plus intéressés par l'amélioration de l'enseignement et de l'apprentissage alors que les éditeurs seront plus intéressés par les questions de marketing.

➤ Les évalués

Les évalués peuvent aussi être différents. Il peut s'agir d'apprenants connus, d'apprenants stéréotypés caractérisés par un certain nombre d'indicateurs ou de grandes catégories comme les enfants de 12 à 14 ans.

➤ Le public de l'évaluation³

Il peut s'agir d'enseignants, d'apprenants, de parents, de développeurs, d'éditeurs, d'institutions,...

³Autrement dit: ceux à qui les résultats de l'évaluation sont destinés

5. Les contextes de l'évaluation

En croisant les différents types de publics, on peut définir des contextes pour l'évaluation :

➤ Un contexte personnel

Le public et l'évaluateur sont la même personne interagissant avec l'évalué.
Typiquement, il s'agit de l'enseignant qui évalue l'usage d'un logiciel dans sa classe.

➤ Un contexte institutionnel

Le public dans ce cas est un groupe local : une école, un éditeur.

➤ Un contexte global ou systémique

Le public dans ce cas est par exemple un Ministère de l'Education et l'ensemble de la population qu'il représente.

6. Le modèle

Ce tableau croise les contextes d'évaluation, les méthodes et les buts poursuivis. Pour plusieurs cellules des exemples ont été repérés. Dans la suite, nous reprenons quelques uns de ceux-ci. Le lecteur désireux d'obtenir le rapport complet en anglais incluant une bibliographie très fouillée peut l'obtenir auprès de Bernadette Charlier. En outre, nous l'invitons à rechercher ses propres exemples.



	Personnelle	Locale	Globale
Catégorielle	▲		●
Check-list	▲		●
Qualitative	▲	▲	▼
Quantitative			
Eclectique			▼
Critique	▲		
Synthétique			

▲ Améliorer enseignement et apprentissage

● Favorise les politiques universelles

▼ Soutien au génie logiciel

■ Profite au marché du logiciel éducatif

7. Quelques exemples

➤ L'évaluation personnelle

1. Les catégories ou les typologies

En annexe, vous trouverez 12 exemples de typologies. Leur principal avantage est de faire penser à des usages possibles.

2. Les checklists

Un bon exemple est MEDA.

Plusieurs critiques ont été formulées à propos des check-lists :

- elles ne sont pas toujours basées sur une théorie de l'apprentissage ;
- elles prennent difficilement en compte le contexte d'usage ;
- il est difficile de donner un poids aux critères ;
- l'attention est trop centrée sur les aspects techniques ;
- elle ne sont pas adaptées aux applications innovantes ;
- l'usage en classe n'est pas pris en compte.

Pensez-vous que MEDA 97 souffre des mêmes défauts ?

3. Les évaluations critiques

De plus en plus d'évaluations d'experts sont publiées en ligne. Un site <http://www.techlearning.com/> permet de rechercher les revues off ou on-line existantes.

4. L'évaluation qualitative

On peut parler d'évaluation en contexte, il s'agit d'observer l'interaction entre l'enseignant, l'apprenant et le logiciel.

Par exemple, un auteur espagnol (Marqués, 1995) propose de s'intéresser aux :

- caractéristiques préalables : objectifs pédagogiques, étudiants, ressources ;
- stratégies de l'enseignant : ses explications préalables, les activités prévues ;
- organisation de la session de travail avec le logiciel et les méthodes d'utilisation du logiciel (navigation par exemple) ;
- l'évaluation des résultats obtenus.

➤ **L'évaluation locale**

1. Evaluation qualitative en contexte local.

Il s'agit le plus souvent d'études de cas qualitatives. Un bon exemple est fourni par le travail réalisé par Françoise Docq et Amaury Daele à propos de l'usage du campus virtuel de LEARN-NETT (en annexe).

Les développeurs utilisent ce que l'on appelle la technique du focus-group. On propose un prototype à un petit groupe d'apprenants. On observe l'usage en situation et ensuite on propose une discussion de type brain-storming aux participants pour rechercher les améliorations possibles.

2. Approche éclectique

Un bon exemple est la méthode TILT. Elle inclut :

- un échange préalable avec les enseignants pour définir les objectifs, le projet d'usage, le choix des méthodes
- une évaluation par un évaluateur hors contexte (par exemple avec une check-list)
- l'élaboration d'une épreuve d'évaluation des apprentissages (quizz ou autre)
- une observation en classe
- la rédaction d'un rapport provisoire
- la discussion de ce rapport avec les enseignants et éventuellement les apprenants
- la rédaction du rapport final.

➤ **L'évaluation globale**

Outre les méthodes décrites précédemment, on peut relever les revues en ligne comme celle du CNDP ou les commissions d'experts.

Certaines listes de critères pour la labélisation sont intéressantes, telle celle de la California Clearing House :

- approche constructiviste
- méthode collaborative
- stratégies pour stimuler différentes activités cognitives
- possibilité d'exploration autonome
- questions ouvertes
- stratégies pour stimuler la créativité des étudiants.

Référence

D. Squires et al. A study of evaluation of hyper – and multi – media educational software, Final report, 1999. European Commission. Tender N° DGXXII/26/96