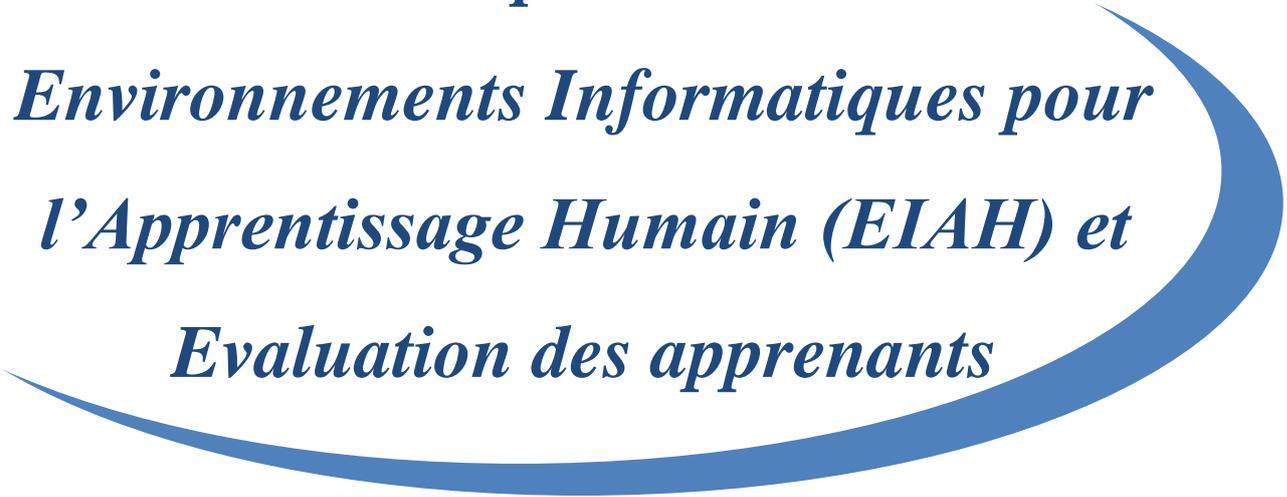


*Chapitre II :*  
*Environnements Informatiques pour*  
*l'Apprentissage Humain (EIAH) et*  
*Evaluation des apprenants*



## **I. Introduction**

Historiquement, les outils informatiques utilisés dans le cadre scolaire ou de la formation ont toujours intégré des moyens d'évaluer les apprenants. Les Environnements Informatiques pour l'Apprentissage Humain (EIAH) (Tchounikine, 2002) ne dérogent pas à cette règle.

L'évaluation joue un rôle particulier en EIAH. Selon Charles (Juwah, 2003), chercheur dans le domaine des EIAH, l'évaluation doit :

- Etre motivante pour l'apprenant ;
- Encourager une activité d'apprentissage soutenue ;
- Contribuer à la progression de l'apprenant ;
- Etre faible en coût humain et facilement maintenable.

Son point de vue est que l'évaluation joue un rôle primordial dans l'activité pédagogique, non seulement comme le moyen de vérifier les acquisitions mais aussi comme le moyen de motiver ces apprentissages et d'inciter les élèves à progresser. C'est la volonté d'obtenir de bons résultats qui doit inciter l'élève à apprendre et c'est la satisfaction qu'il éprouve à réussir qui doit lui donner l'envie de persévérer et d'aller plus loin. L'enseignant n'est plus seulement confronté à la nécessité de comprendre ce qui fait obstacle à la réussite de ses élèves mais il peut aussi, lorsqu'il utilise certains outils informatiques, se servir de l'évaluation comme un "moteur" pour les apprentissages (Dintilhac & Rac, 2005), l'évaluation devenant un moyen pédagogique. Comme dit Papert, "It was hard but it was fun" (Jaillet, 2003). La dimension ludique est exploitée, au risque du consumérisme et de l'abandon des objectifs pédagogiques. Les études réalisées par exemple sur l'outil TDmaths (Dinet et Rouet, 2002), (Dinet et Rouet, 2003) montrent que les apprenants apprécient l'aspect ludique, le fait par exemple de gagner des récompenses lorsqu'ils réussissent une tâche (les bons points de notre enfance). Pourtant, les enseignants utilisent assez peu les scores réalisés par les élèves pour noter leurs compétences, comme s'ils ne prenaient pas ces évaluations au sérieux (Dinet et Rouet, 2002), (Dinet et Rouet, 2003). Les méthodes de calcul des scores sont bien souvent opaques et pas toujours adaptées aux besoins de l'enseignant.

Les pratiques d'évaluation en EIAH sont cependant plus variées qu'il n'y paraît. Dans ce chapitre, nous tentons de présenter et commentons une vue exhaustive de ces pratiques.

## **II. Les pratiques d'évaluation**

Pour mener les nécessaires évaluations (Dintilhac et Rac, 2005) liées à ses enseignements, l'enseignant dispose de plusieurs types d'évaluation :

- l'évaluation pronostique ;
- l'évaluation formative ;
- l'évaluation formatrice ;
- l'évaluation diagnostique ;
- l'évaluation sommative.

Sont absentes de cette typologie les évaluations normatives (Hadji, 1990) et critériées (Hadji, 1997). L'évaluation normative consiste à comparer les résultats d'un apprenant par rapport à ceux d'autres apprenants. En outre, nous pensons que tout type d'évaluation repose sur des critères, explicites ou non. Les évaluations normatives et critériées sont donc transversales aux cinq autres présentées ici.

### **II.1. L'évaluation pronostique**

L'évaluation pronostique "fonde des décisions de sélection ou d'orientation en fonction de l'aptitude présumée à suivre un nouveau cursus, par exemple telle filière du secondaire ; elle se situe en amont d'un cursus et sous-tend un choix" (Perrnoud, 2001). L'évaluation pronostique permet d'évaluer la capacité d'un apprenant à commencer un apprentissage, un cycle d'étude ou à exercer une profession. C'est une évaluation en amont d'une réalisation ou d'un apprentissage.

### **II.2. L'évaluation formative**

L'évaluation formative est une "évaluation dont l'ambition est de contribuer à la formation" (Hadji, 1990). Dans ce contexte, elle a pour but de réguler l'enseignement. L'évaluation fournit des informations permettant à l'enseignant d'adapter son enseignement aux particularités de l'apprenant, elle entre dans le cadre d'un enseignement différencié (Allal, Cardinet et Perrnoud, 1979). Comme l'explique P. Perrenoud (1997), "toute différenciation appelle une évaluation formative". Pédagogie différenciée et évaluation formative sont intimement liées. L'évaluation formative joue le rôle d'une "discrimination positive" qui vise à emmener chaque apprenant, en tenant compte de ses différences, à un

niveau de connaissances. La pédagogie différenciée s'inscrit dans un processus égalitaire d'acquisition d'un savoir. En ce sens, cette pédagogie et l'évaluation formative qui l'accompagne ont du mal à faire leur place dans l'enseignement scolaire encore habitué à une pédagogie élitiste qui vise à sélectionner les élèves. Contrairement aux autres évaluations, l'évaluation formative ne se contente pas d'évaluer les productions des élèves mais aussi les situations actives permettant de comprendre la démarche des apprenants, leurs rapports aux savoirs, leurs capacités de métacognition, etc. (Perrnoud, 1997).

### **II.3. L'évaluation formatrice**

C'est une forme particulière d'évaluation formative. Selon (Bonniol, 1986), l'évaluation formative s'inscrit dans une visée de régulation de l'apprentissage par l'enseignant tandis que dans l'évaluation formatrice, la régulation est assurée par l'apprenant. En ce sens, l'activité d'auto-évaluation, qu'elle soit individuelle, mutuelle ou collective, est une évaluation formatrice. L'auto-évaluation est une "évaluation interne conduite par le sujet de sa propre action et de ce qu'elle produit. C'est un processus d'altération de son référentiel d'action au cours de confrontations entre son propre référentiel et celui ou ceux d'autrui" (Campanale, 1997). L'auto-évaluation ne peut donc pas être contrainte, elle est tributaire du bon vouloir de l'évalué.

### **II.4. L'évaluation diagnostique**

L'évaluation formative s'appuie en partie sur l'évaluation diagnostique. L'évaluation diagnostique permet d'évaluer un niveau de compétence bien souvent juste avant une nouvelle phase d'apprentissage. Dans le cadre d'une évaluation formative, ce diagnostic permet la remédiation et la mise en œuvre d'une pédagogie différenciée.

La notion de compétence abordée ici est récurrente dans la littérature sur l'évaluation. Selon Chomsky (1965), la compétence en linguistique désigne le système de règles intériorisées qui permet de comprendre et de produire un nombre infini de phrases inédites. Tandis que la performance désigne la manifestation de la compétence des locuteurs et réfère à la diversité des actes de langage et des contextes d'énonciation et de communication. Ainsi, la performance peut être considérée comme la mise en application, en pratique d'une compétence. C'est donc par la performance qu'on évalue la compétence de l'apprenant.

En sciences de l'éducation, une compétence est souvent vue sous l'angle d'un ensemble de savoir-faire conceptualisés (Malglaive, 1990), (vergnaud, 1995).

## **II.5. L'évaluation sommative**

C'est l'"évaluation par laquelle on fait un inventaire des compétences acquises, ou un bilan, après une séquence de formation d'une durée plus ou moins longue" (Hadjji, 1990).L'évaluation met donc l'accent sur les performances, elle contrôle les connaissances. Elle est en opposition avec l'évaluation formative, elle ne régule pas l'apprentissage, elle le contrôle. L'évaluation sommative peut prendre la forme d'examens périodiques qui valident un apprentissage. Elle conduit à l'obtention d'une note qui sanctionne une activité d'apprentissage afin d'établir un classement, sélectionner les apprenants ou certifier leur niveau. En mettant l'accent sur les performances, l'évaluation sommative s'intéresse essentiellement aux productions réalisées par les apprenants (Campale, 2001).

## **III. Pratiques d'évaluation en EIAH**

L'évaluation sommative est la plus pratiquée, compte tenu de la facilité de sa mise en œuvre. Elle repose bien souvent sur des questionnaires à choix multiples ou fermés relativement faciles à évaluer automatiquement. Les systèmes de tests adaptatifs (CAT : Computer Adaptive Test) (Green, Bock, Humphreys et al, 1984) supportent ce type d'évaluation et permettent d'envisager leur usage dans le contexte d'une évaluation pronostique.

L'évaluation diagnostique a fait l'objet de recherches bien connues en France, avec entre autres le logiciel Pépite (Delozanne et Grugeon, 2004). Contrairement à l'évaluation sommative ou pronostique rencontrée dans les CAT, l'évaluation diagnostique de Pépite permet d'obtenir un bilan de compétences de l'apprenant, et non un simple score.

L'évaluation sommative qui a pour mission de certifier les connaissances de l'apprenant et l'évaluation diagnostique peuvent être associée. Dans la classe, l'enseignant fait souvent appel aux deux, l'évaluation diagnostic permettant à l'enseignant de réguler l'apprentissage. Certains EIAH, tels que TDmaths3 essaient d'intégrer ces deux types d'évaluation.

Toujours dans un souci de régulation des apprentissages, l'évaluation formatrice est aussi rencontrée en EIAH. L'auto-évaluation avec le logiciel GenEval (David, 2003) et l'évaluation par pairs ou co-évaluation avec des systèmes tels que OASYS (Bhalerao et Ward, 2001), constituent les deux formes les plus connues. L'évaluation formatrice est souvent utilisée lorsque l'évaluation n'est pas automatisable (productions libres des apprenants) (Maor, 1998), (Juwah, 2003).

L'évaluation formative est présente sous l'angle de l'assistance à la régulation des activités par l'enseignant. L'objectif est alors de fournir des indicateurs à l'enseignant lui permettant de mener des évaluations formatives en relation avec une activité d'apprentissage. Ces dernières années, de nombreux systèmes de "monitoring" de l'activité pédagogique sont apparus, tels que FORMID SUIVI (Gueraud, Adam, Perin et al, 2005), GISMO (Mazza et Milani, 2005), mais aussi Aplusix (Sander, Nicaud, chachoua et Croset, 2005) qui enregistre la totalité des actions des apprenants.

D'autres évaluations connexes pouvant être utilisées dans une démarche formative ou diagnostique existent en EIAH. C'est le cas de l'évaluation de la participation et de l'évaluation par portfolio.

L'évaluation de la participation s'appuie généralement sur des outils de communications qu'il s'agisse de forums (S.Fujitania, T.Mochiduki ,Y.Isshiki et al, 2003), (Mochizuki, Fujitani, Hisamatsu et al, 2004) (Bratitsis et Dimitracopoulou, 2005) ou de chats (Georgr, 2001).

L'évaluation par portfolio permet d'évaluer la progression de l'apprenant dans ses apprentissages (Ministère de l'Education du Québec, 2002), (Eyssautier et Bavay, 2004).

Dans la suite, nous allons détailler chaque type d'évaluation en EIAH en l'illustrant par un exemple.

### **III.1. L'évaluation sommative et pronostique**

L'évaluation sommative est souvent menée dans des environnements d'apprentissage basés sur des banques de questionnaires à choix multiples (QCM). Chaque questionnaire est organisé en parcours et sanctionne l'apprenant par une note finale. Le test du TOEFL qui permet d'évaluer le niveau d'anglais d'une personne étrangère est certainement le plus connu.

## **Chapitre II : Environnements Informatiques pour l'Apprentissage Humain (EIAH) et Evaluation des apprenants**

---

Ce test fait partie de ce qu'on appelle les tests adaptatifs par ordinateurs ou CAT pour "Computer Adaptive Test". Ces tests permettent de certifier des connaissances de base (Biggs, 1999). Outre une évaluation sommative, un CAT peut aussi offrir une évaluation pronostique comme le test du GRE qui conditionne l'accès aux formations universitaires par les bacheliers aux États-Unis.

Les tests du TOEFL et du GRE reposent sur ce système. Ils sont administrés par le service des tests éducatifs, qui est une organisation privée à but non lucratif travaillant en étroite collaboration avec les services d'éducation américains.

Le principe de base d'un CAT est simple. Un élève se place seul devant un ordinateur et répond à une liste de questionnaires à choix multiples. Pour chaque question, si la réponse est correcte, la question suivante sera un peu plus difficile. Dans le cas contraire, la question suivante sera un peu plus facile (Auger et Séguin, 1992). Il en va ainsi tout au long du test, ce qui permet au logiciel de déterminer le niveau du candidat et de lui attribuer un score. Les systèmes de CAT s'appuient sur le modèle de la théorie de réponse aux items (Item Response Theory) (Lord, 1980). Ce modèle guide le concepteur dans la construction d'un système de CAT. Le but de ces modèles est d'optimiser la cohérence entre les scores obtenus et les réelles compétences de l'évalué.

### **III.1.1. L'évaluation dans les CAT**

Dans un CAT, on évalue des réponses à un QCM. A chaque question, la réponse est juste ou fausse, c'est une notation binaire qui conduit à une note finale après évaluation de la dernière réponse. Plus le nombre de questions est important plus la note finale est précise. L'évaluation globale menée par un CAT est sommative et individuelle. Elle appartient au registre du "testing". Un CAT permet de certifier un niveau de compétences, que ce soit en anglais pour le TOEFL ou sur les connaissances pré-universitaires pour le GRE.

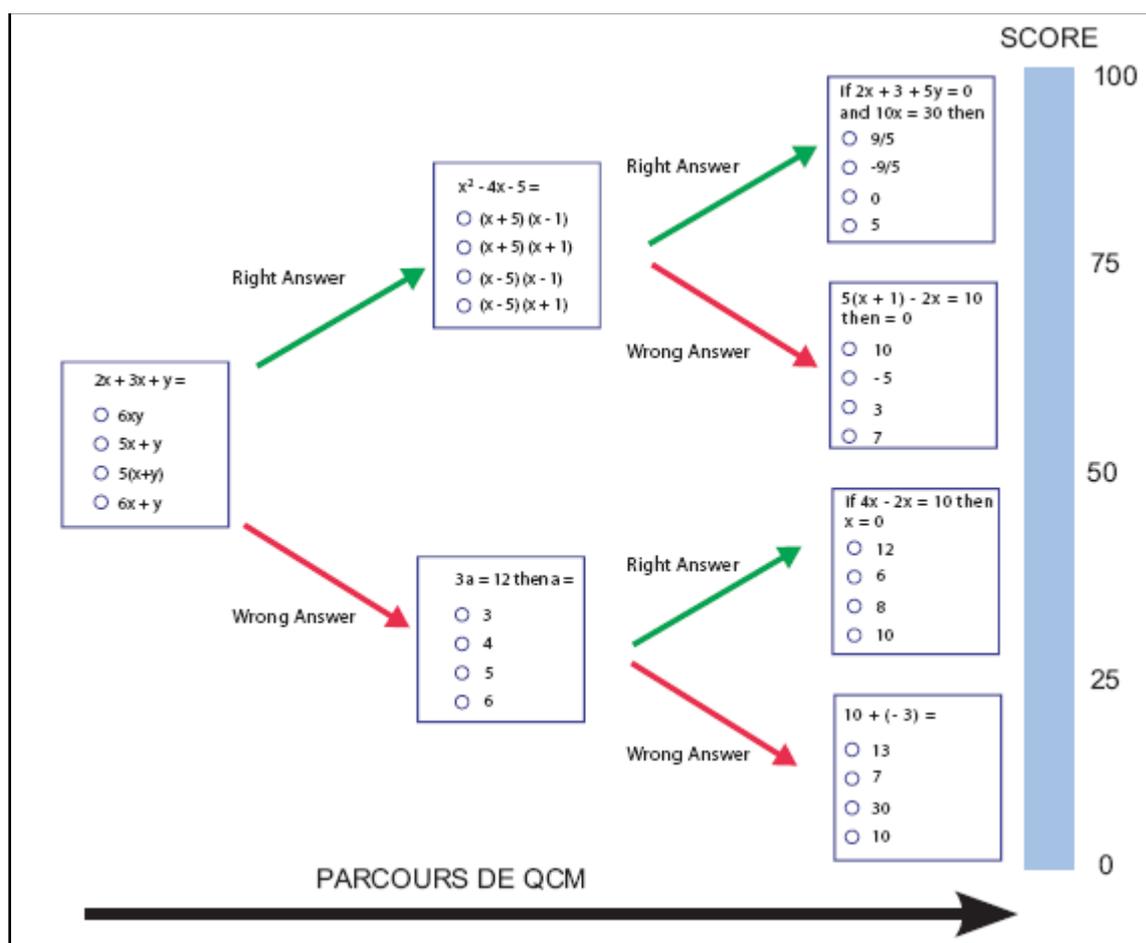


Figure N° 18: Principe

### III.2. L'évaluation diagnostique

Contrairement aux évaluations sommatives et pronostiques des CAT, les systèmes d'évaluation diagnostique ne se contentent pas de corriger les productions des apprenants et de comptabiliser des points, mais proposent des diagnostics de compétences de l'apprenant. Chaque réponse de l'apprenant est alors analysée pour essayer de l'associer à une erreur envisagée par les concepteurs de l'exercice et à son origine potentielle. Le logiciel Pépité (Delozanne et Grugeon, 2004) et ses dérivés illustrent la mise en œuvre de cette évaluation diagnostique en EIAH. Pépité (Jean-Daubias, 2000), (Delozanne et Grugeon, 2004) est un logiciel qui a pour objectif d'aider les enseignants à diagnostiquer les compétences de leurs élèves en algèbre élémentaire. Pépité a été réalisé dans le cadre du projet pluridisciplinaire Lingot (Delozanne, B.Rugeon, Artigue et Rozalski, 2003), dont l'objectif est de concevoir des logiciels qui aident les enseignants à réguler les apprentissages en algèbre. Pépité est l'axe diagnostique du projet Lingot. C'est un outil logiciel composé de trois parties :

## Chapitre II : Environnements Informatiques pour l'Apprentissage Humain (EIAH) et Evaluation des apprenants

- PepiTest qui est destiné aux élèves. Il permet à ces derniers de résoudre une liste de 22 exercices constitués de questions ouvertes et fermées en algèbre ;
- PepiDiag qui est le module d'analyse des réponses. Il interprète les réponses des élèves à partir des analyses didactiques élaborées en amont du projet ;

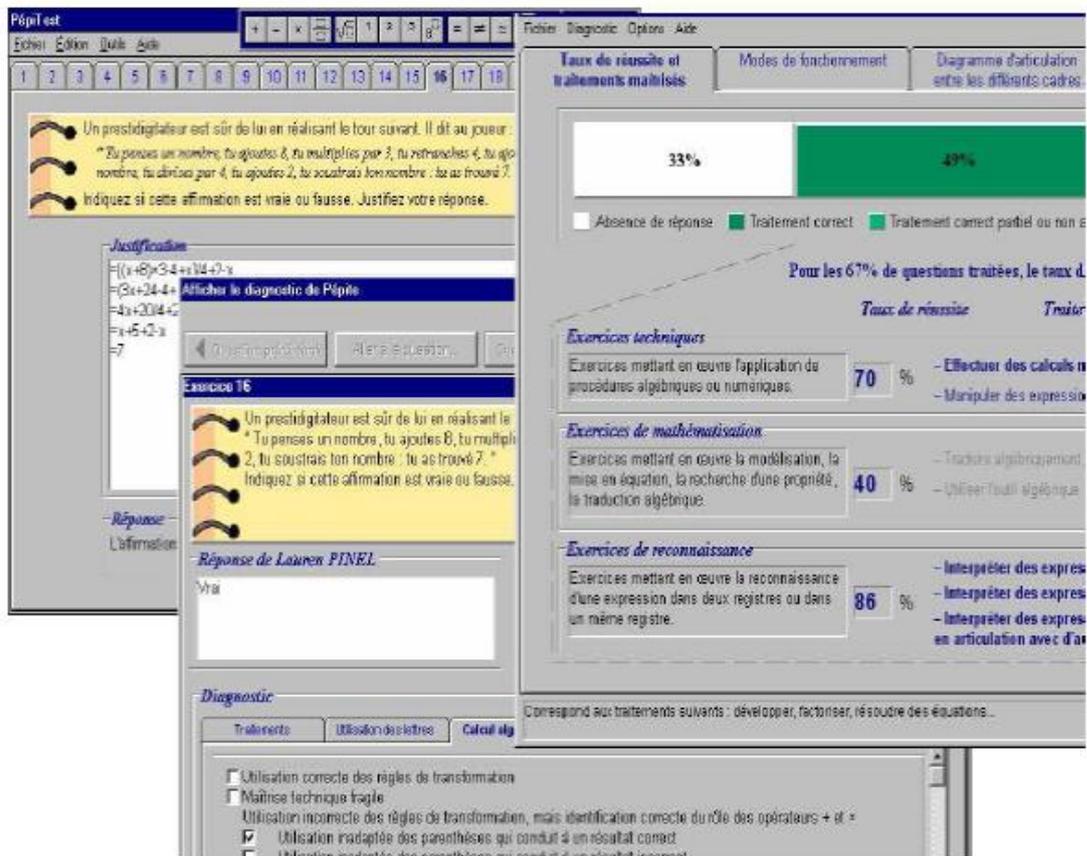


Figure N° 19: Pépite : de gauche à droite, les interfaces de PepiTest, PepiDiag et PepiProf

- PepiProf qui est l'outil enseignant. Il établit le profil de chaque élève et permet de modifier les réponses apportées par les élèves dans le cas où l'enseignant les trouverait contestables.

### III.2.1. L'évaluation dans Pépite

Pépite permet d'évaluer les compétences des apprenants en algèbre. Cette évaluation sommative et individuelle de l'apprenant est automatique. Il fournit grâce à PepiDiag une synthèse en pourcentages d'acquisition des compétences mises en jeu par le questionnaire. L'enseignant obtient pour chaque apprenant un diagnostic composé de quatre parties :

- La partie "Taux de réussite et traitements maîtrisés" indique le taux de réussite par type d'exercice ;

- La partie "Modes de fonctionnement" détaille la manière de justifier les réponses et entre plus en détail dans la résolution des exercices. On s'attache aux compétences transversales de l'apprenant ;
- La partie "Diagramme d'articulation entre les différents cadres" représente graphiquement les modes de fonctionnement de l'apprenant ;
- La partie "Résumé du profil" propose un résumé textuel du diagnostic.
- Voici ci-dessous un extrait d'un résumé de profil PepiDiag.
- Utilisation des lettres pour faire du calcul algébrique avec des règles fausses dans 40% des réponses. La maîtrise du calcul algébrique reste insuffisante (maîtrise correcte dans seulement 45% des questions). Les difficultés sont liées à :
- Règles de transformation non maîtrisées, mais identification correcte du rôle des opérateurs + et x dans 20% des réponses (utilisation inadaptée des parenthèses qui conduit à un résultat correct, utilisation de règles de transformation fausses identifiées).
- Identification incorrecte du rôle des opérateurs + et x dans 5% des réponses (les règles de transformation utilisées "assemblent" les termes).

Conversion correcte dans 72% des réponses. Justification de type scolaire dans 50% des réponses.

Pépité permet aussi de réaliser un profil collectif des apprenants. Il est possible de visualiser le diagnostic "moyen" d'une classe.

Une première version de Pépité a été développée en Delphi en 2000, suivie en 2002 par une nouvelle version Java. Pépité a été très largement expérimentée. Cependant, les résultats de ces expérimentations sont assez mitigés. PepiDiag montre ainsi quelques inconsistances dans la grille d'analyse didactique lorsqu'elle est appliquée par le logiciel, c'est-à-dire d'une manière très rigide et sans discernement. En outre la grille d'évaluation présentée hors contexte induit de nombreuses difficultés : par exemple, certains enseignants ne comprennent pas les items du diagnostic (Jean-Daubias, 2000) [DG04].

### III.3. L'évaluation sommative et diagnostique

Certains exercices commerciaux tels que TDmaths intègrent les deux types d'évaluation, sommative et diagnostique. TDmaths prend aussi en compte l'évaluation transversale évoquée plus haut (cf. II) : l'évaluation normative.

TDmaths est un exerciceur de mathématiques destiné aux élèves de niveau collège. Il équipe près de 400 collèges, soit un minimum de 12000 utilisateurs. TDmaths est une application client-serveur java.

Les élèves progressent en autonomie, encouragés et guidés par leur enseignant. L'ensemble du programme d'algèbre du collège est présenté sous forme de cours synthétiques, illustrés par des exemples et structurés en chapitres. Les liens entre les chapitres dessinent la carte des mathématiques. Cette carte présente les différents parcours possibles dans le programme, y compris les passages entre les niveaux (6ème, 5ème, 4ème, 3ème).

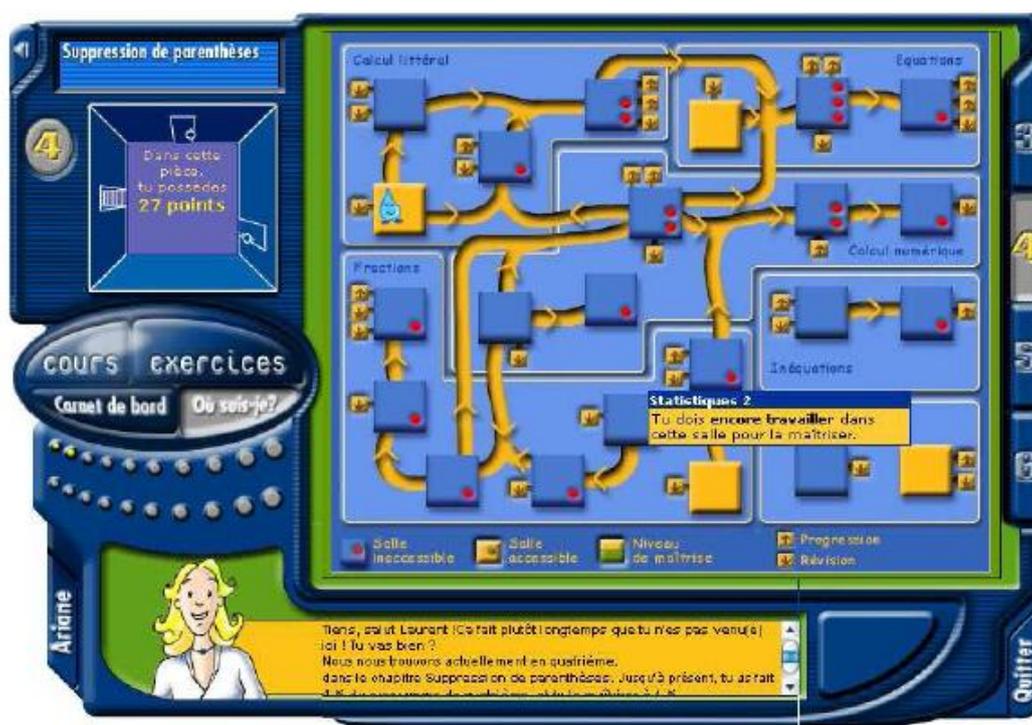


Figure N° 20: TDmaths : parcours de l'apprenant

En suivant ces liens, l'élève progresse au fur et à mesure de sa maîtrise des notions, validée par ses succès aux exercices. Les exercices, de difficulté croissante, ont des énoncés aléatoires. TDmaths offre ainsi plus de 600 exercices, une grande variété de formes (QCM, saisie alphanumérique, calcul mental chronométré, représentation graphique) et une infinité

## Chapitre II : Environnements Informatiques pour l'Apprentissage Humain (EIAH) et Evaluation des apprenants

d'énoncés. A chaque nouvelle connexion, l'élève peut reprendre son travail là où il l'a laissé.

Les caractéristiques de suivi du travail de l'élève par l'enseignant sont les suivantes :

- évaluation permanente des élèves conforme aux objectifs des évaluations nationales;
- analyse précise des connaissances de chaque élève : notions travaillées, maîtrise de ces notions;
- historique des sessions de travail et des résultats obtenus.

### III.3.1. L'évaluation dans TDmaths

Dans TDmaths, trois types d'évaluation sont intégrés :

– Une évaluation sommative individuelle des performances sanctionnée par des scores obtenus dans les exercices de mathématiques ;

– Une évaluation diagnostique des apprenants. Les compétences transversales des apprenants sont évaluées et un diagnostic graphique est ensuite proposé à l'enseignant. Il peut alors envisager des remédiations à partir de cette évaluation. Dans ce cas, l'évaluation s'inscrit alors dans un cadre formatif ;

– Une évaluation normative des apprenants. Que ce soit au niveau des scores ou des compétences transversales, le logiciel propose de positionner les résultats de ces évaluations par rapport à la classe.

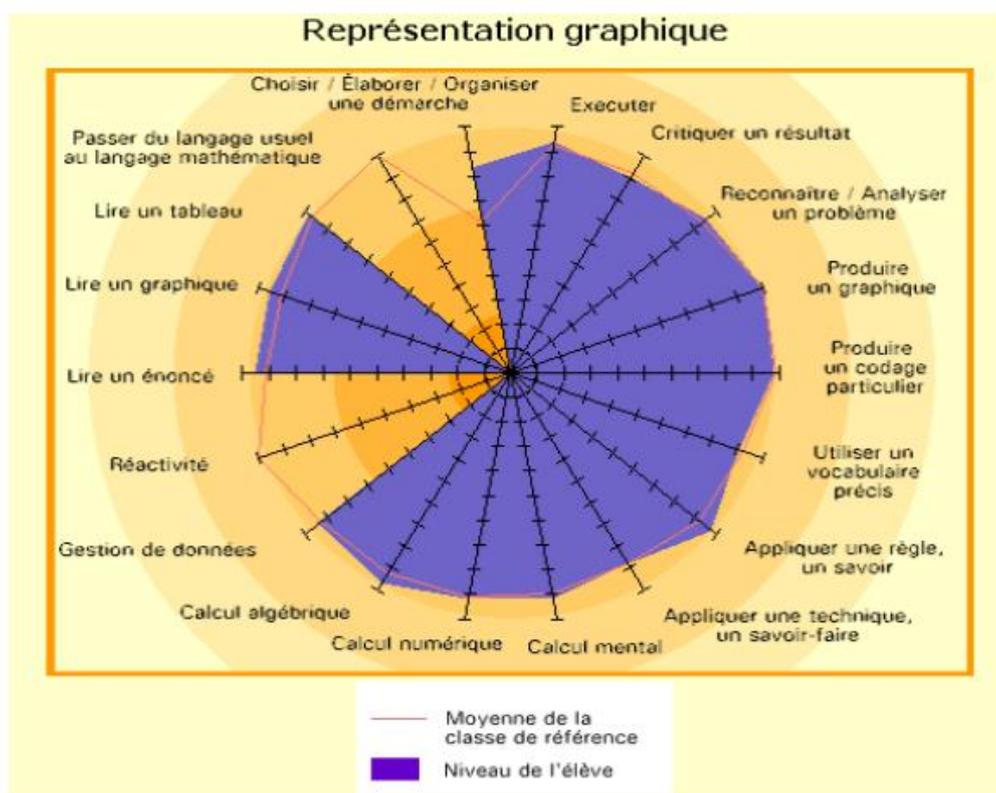


Figure N° 21: TDmaths : interface enseignant, carte des compétences transversales

## **Chapitre II : Environnements Informatiques pour l'Apprentissage Humain (EIAH) et Evaluation des apprenants**

---

Les compétences transversales évaluées dans TDmaths sont des savoir-faire tels que "lire un énoncé", "critiquer un résultat", etc.

Dans le sens où TDmaths est un outil commercial largement diffusé, il est représentatif des pratiques en EIAH dans les collèges français. Du point de vue de l'évaluation, TDmaths est dual. D'une part, il offre une évaluation de type "testing" partagée par le plus grand nombre d'EIAH, d'autre part il introduit l'évaluation diagnostique des compétences. Celle-ci est plus souvent rencontrée dans des outils de recherche tels que Pépité (Delozanne et Grugeon, 2004). Il confirme donc l'attente des enseignants d'une évaluation fine des compétences des apprenants.

### **L'auto-évaluation**

Dans certains cas, il peut être utile de proposer à l'apprenant de s'auto-évaluer afin de favoriser l'auto-régulation de ses apprentissages en s'insérant dans le cadre d'une évaluation formative. Le logiciel GenEval (David, 2003) propose une liste de questions, dans laquelle l'apprenant peut naviguer. Pour chaque question, l'apprenant peut bénéficier ou non d'indications pour répondre. Une fois la réponse donnée et corrigée, l'apprenant évalue sa maîtrise des compétences mises en jeu dans les questions en se donnant une note.

GenEval (David, 2003) est un logiciel qui permet de générer des exercices d'auto-évaluation. Ce logiciel a été développé par le Centre d'Auto-Formation et d'Innovation Multimédia (CAFIM) dans le cadre du projet européen ARIADNE (Alliance of RemoteInstructionalAuthoring and Distribution Networks for Europe).

Le projet européen ARIADNE regroupe une quinzaine d'universités européennes et a pour objectifs de favoriser et mutualiser la production de documents pédagogiques hypermédias et d'expérimenter des scénarios pédagogiques avec des étudiants en situation d'apprentissage autonome. Le CAFIM peut se définir à la fois comme une unité de réflexion sur les usages pédagogiques de l'ordinateur et comme un centre de développement des outils et des ressources nécessaires à la mise en place d'enseignements médiatisés.

Un exercice d'auto-évaluation créé avec GenEval offre aux étudiants un cadre pour résoudre un exercice, avec des accès hypermédias à des informations pertinentes pour la compréhension des concepts étudiés, avec des orientations et des aides méthodologiques progressives. GenEval propose à l'apprenant une suite de questions auxquelles il doit répondre. Il dispose dans cet exercice d'indices qu'il peut ou non utiliser pour répondre aux

questions. Il répond à toutes les questions puis évalue sa maîtrise des compétences mises en jeux dans le questionnaire.

### III.3.2. L'évaluation dans GenEval

Dans GenEval, deux types d'évaluation sont menés conjointement :

- Une évaluation individuelle sommative. Lorsque l'apprenant répond à une question, le système GenEval donne une correction ;
- Une auto-évaluation. L'apprenant évalue sa maîtrise des compétences. Cette évaluation favorise l'auto-régulation de ses apprentissages en s'insérant dans le cadre d'une évaluation formative (Schunk, 1990).



Figure N° 22: GenEval : phase d'auto-évaluation de l'apprenant

Contrairement aux autres pratiques d'évaluation, l'auto-évaluation est peu représentée. Pourtant, comme le soulignent certains psychologues (Schunk, 1990), cette évaluation s'intègre idéalement dans une logique formative de l'évaluation, en responsabilisant l'apprenant dans sa construction du savoir. C'est donc une pratique d'évaluation légitime et intéressante dans cette optique.

### III.3.3. L'évaluation par pairs

Une des premières pratiques d'évaluation des activités collaboratives en EIAH est l'auto-évaluation par pairs (Maor, 1998). Cette évaluation peut être :

- Individuelle. Un apprenant évalue un autre apprenant (Bhalerao et Ward, 2001).
- Collective. Un groupe d'apprenants évalue un apprenant ou un groupe d'apprenants (Ward, Sitthiworachart et Joy, 2004), (juwah, 2003).

Dans le cas où l'évaluation est menée par un groupe d'apprenants et porte sur une production réalisée par un autre groupe, une des difficultés peut provenir de la difficulté pour

## Chapitre II : Environnements Informatiques pour l'Apprentissage Humain (EIAH) et Evaluation des apprenants

les membres du groupe qui évaluent à se mettre d'accord sur les résultats de l'évaluation [Juw03].

Le développement d'OASYS (Bhalerao et Ward, 2001) (On-line Assessment System) a débuté en 2000 et s'est terminé en 2002 avec une nouvelle version OASYS2 (Ward, Sitthiworachart et Joy, 2004). Cet outil a été conçu par des chercheurs de l'Université de Warwick à Coventry en Angleterre. OASYS est un outil WEB d'évaluation par pairs. Il permet de créer et diffuser auprès d'apprenants des questionnaires. Les questionnaires sont composés de questions à choix multiples et de questions ouvertes.

The image shows two side-by-side screenshots of the OASYS web interface. The left screenshot displays 'Question 7' with a text-based question about binary tree traversal and a 'My answer:' section containing a 'Potential answer' table with four binary tree diagrams. The right screenshot displays 'Question 2' with a diagram of a binary tree and a 'My answers:' section with a 'Potential answer' table containing seven options.

**Question 7:**  
Assuming an inorder traversal, which of the following is the correct diagram for a binary tree that represents the expression:  $a + b * c$

**My answer:**

A	B	Potential answer
<input checked="" type="radio"/>	<input checked="" type="radio"/>	Unanswered
<input type="radio"/>	<input type="radio"/>	leaf and child of P
<input type="radio"/>	<input type="radio"/>	root and parent of Q
<input type="radio"/>	<input type="radio"/>	root
<input type="radio"/>	<input type="radio"/>	internal node and parent of S
<input type="radio"/>	<input type="radio"/>	leaf and child of Q
<input type="radio"/>	<input type="radio"/>	child of S

**Question 2:**  
The nodes labelled on this binary tree are:

Node 1 points to node Q, Node 2 points to node S.

A: Node 1 is a  
B: Node 2 is a

**My answers:**

A	B	Potential answer
<input checked="" type="radio"/>	<input checked="" type="radio"/>	Unanswered
<input type="radio"/>	<input type="radio"/>	leaf and child of P
<input type="radio"/>	<input type="radio"/>	root and parent of Q
<input type="radio"/>	<input type="radio"/>	root
<input type="radio"/>	<input type="radio"/>	internal node and parent of S
<input type="radio"/>	<input type="radio"/>	leaf and child of Q
<input type="radio"/>	<input type="radio"/>	child of S

Figure N° 23: OASYS : exemple de question fermée

Les concepteurs ont choisi l'évaluation par pairs pour évaluer les réponses aux questions ouvertes dont l'évaluation est difficilement automatisable. OASYS a été expérimenté en 2002 dans le cadre de cours en programmation. Dans cette expérimentation, la possibilité d'utiliser des questions ouvertes était indispensable pour permettre aux apprenants d'écrire des programmes informatiques.