

III.3.4. L'évaluation dans OASYS

Dans OASYS, chaque apprenant répond tout d'abord aux questions qui lui sont posées avant de devenir évaluateur. L'apprenant doit ensuite évaluer trois autres apprenants dont les copies lui sont automatiquement attribuées. Pour l'évaluation de chaque question, l'enseignant a préalablement défini les critères d'évaluation, c'est-à-dire les éléments sur lesquels l'apprenant va devoir se focaliser pour évaluer la réponse. Pour évaluer, l'apprenant dispose d'un questionnaire qui apprécie le respect des critères sélectionnés par l'enseignant. Dans le processus d'évaluation, l'enseignant joue le rôle de modérateur. Il est prévenu si un écart entre les différentes évaluations d'une même copie est important.

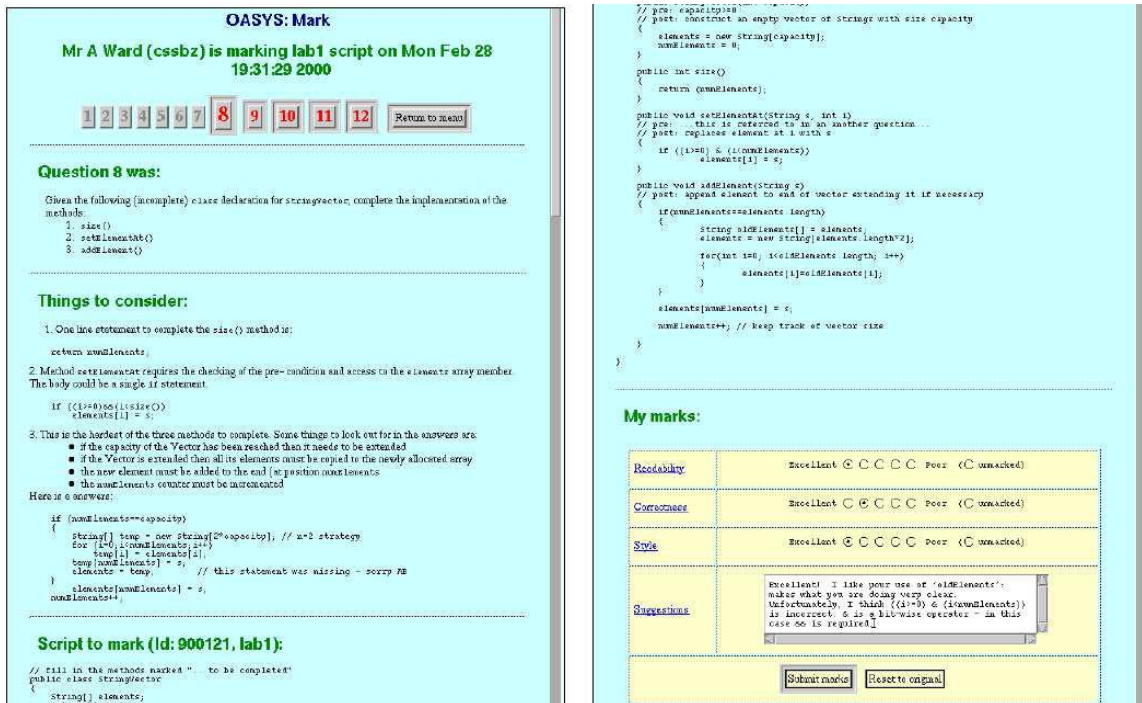


Figure N° 24:OASYS : interface de correction

La version OASYS2 intègre de nouvelles fonctionnalités. Dans la phase d'évaluation, les apprenants peuvent être groupés. Ils évaluent ensemble les réponses d'autres groupes en échangeant leurs points de vue dans un chat. A la fin du processus d'évaluation, chaque apprenant reçoit son résultat.

OASYS est un système évolué d'évaluation par pairs. Il apporte une solution originale à l'évaluation des questions ouvertes. En outre, le fait d'avoir été utilisé par les concepteurs avec leurs élèves a permis l'émergence de la version 2 qui intègre une vision collective de

l'évaluation par pairs. Les résultats issus des expérimentations montrent aussi que l'enseignant est assez rarement sollicité dans son rôle de modérateur.

Cela témoigne d'une bonne conception de ce système d'évaluation et d'une réelle autonomie des apprenants.

III.3.5. L'assistance à l'évaluation

S'il est possible d'évaluer relativement finement une production, il n'en est pas de même pour une démarche. Aussi, qu'il s'agisse d'activités individuelles ou collectives, des solutions d'assistance à l'évaluation sont proposées. Elles consistent bien souvent à fournir à l'enseignant des vues sur les activités par le biais d'indicateurs, de mesures (Merceron et Yacef, 2004) qui permettent à l'enseignant d'évaluer le déroulement de l'activité. Dans ce contexte, plus que des outils de mesure (Mazza et Milani, 2005), il existe des systèmes qui proposent de superviser l'activité.

Dans la configuration de la classe, l'enseignant a la possibilité d'observer les réactions de ses élèves et ainsi d'adapter son cours. GISMO (Graphical Interactive Student Monitoring System) (Mazza et Milani, 2005) permet à l'enseignant de maîtriser ce qui se déroule durant l'activité pédagogique, et ce par le biais d'interfaces de visualisation d'indicateurs (Mazza, 1998). Il répond à la question du "retour pour l'enseignant" (Labat, 2002) en lui fournissant des visualisations d'indicateurs. C'est un outil dit d'"awareness".

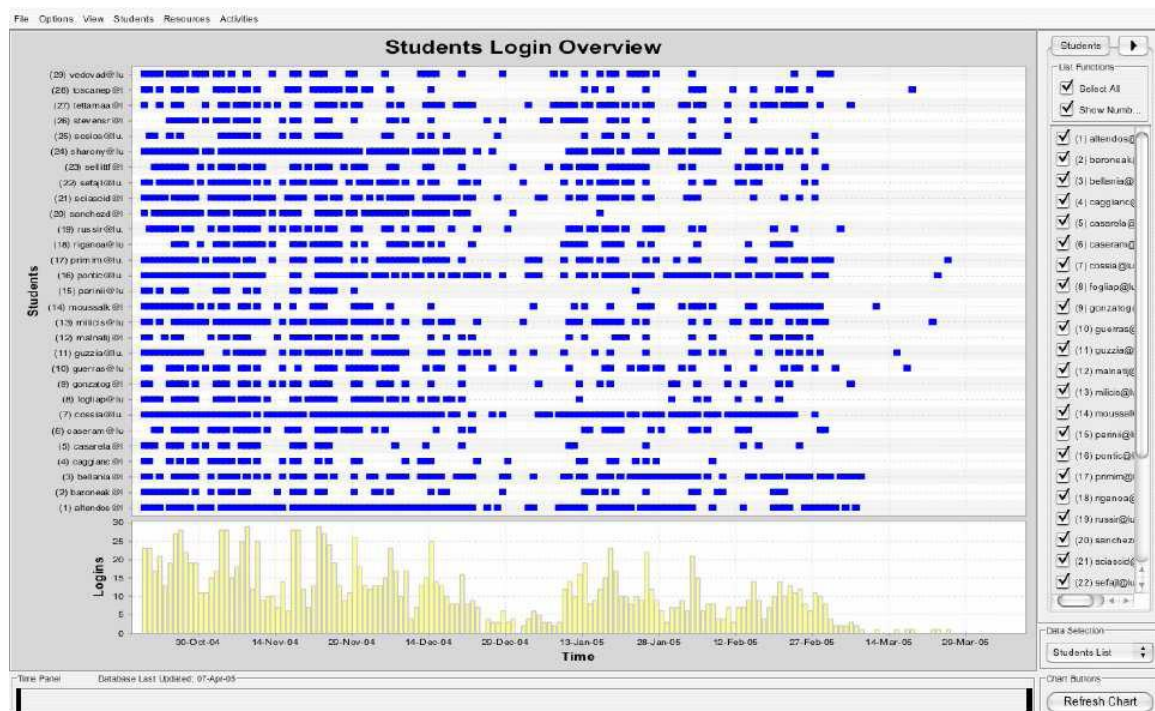


Figure N° 25: GISMO: visualisation des connexions des apprenants au système

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

Le système de gestion d'apprentissage utilisé par GISMO est Moodle. Moodle est un Espace Numérique de Travail (ENT) particulier. C'est un Système de Gestion de l'Apprentissage (SGA). Cette sous-catégorie très spécialisée d'ENT permet de créer et de jouer des questionnaires et des cours. Moodle dispose d'outils de communication et d'espaces partagés. GISMO permet de visualiser les connexions à Moodle, les accès aux ressources contenues dans Moodle, les contributions dans le forum (création d'un sujet de discussion, postage d'un commentaire, etc.) et l'état des exercices proposés aux apprenants (en cours, terminé, etc.).

III.3.6. L'évaluation dans GISMO

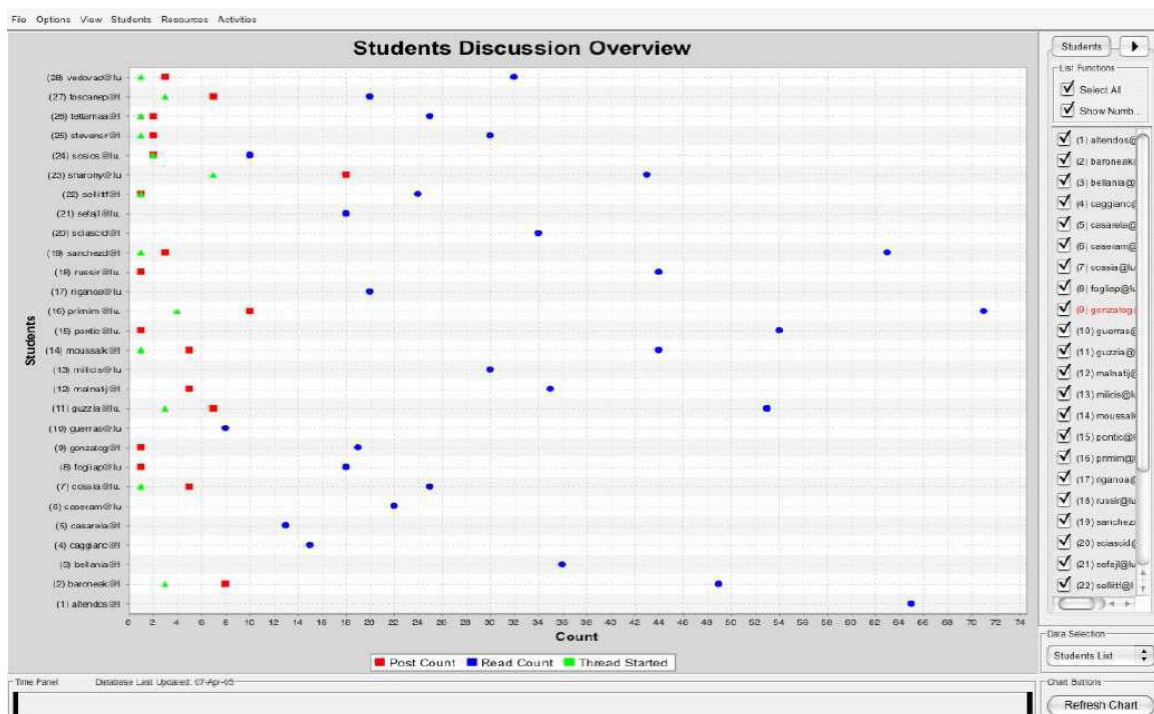


Figure N° 26: GISMO : visualisation des interventions des apprenants dans le forum

GISMO fournit à l'enseignant des indicateurs qui l'aident dans sa tâche d'évaluation et de régulation du travail des apprenants dans Moodle. Mis à part les tests qui sont corrigés de manière automatique dans Moodle, l'évaluation formative des apprenants est à la charge de l'enseignant. A partir des observations fournies par GISMO, l'enseignant peut interagir et guider les apprenants. GISMO est un outil de visualisation de l'activité à travers des traces prédéfinies par les concepteurs du logiciel. Il n'est pas possible pour l'enseignant de visualiser d'autres traces. C'est pourquoi GISMO offre une vision orientée de l'activité même si les concepteurs ont défini les observables avec des enseignants. Cependant, GISMO atténue

l'effet "boîte noire" des EIAH et est un premier pas certain pour aider l'enseignant à superviser une activité.

III.3.7. L'évaluation de la participation

Bien que commune à toute activité collective, l'évaluation de la participation se retrouve quasi essentiellement dans le cadre d'activités utilisant des outils de communication. Deux cas de figures sont rencontrés :

– L'évaluation porte sur la qualité des discussions grâce à des techniques de "textmining" qui analysent le rapport entre mots clés (ceux attendus par l'enseignant) et nombre d'échanges. C'est le cas par exemple du logiciel Ibee (S.Fujitania, T.Mochiduki, Y.Isshiki et al, 2003), qui assiste l'enseignant pour gérer l'activité des apprenants dans un forum. En évaluant la qualité du discours des participants, le logiciel aide à l'évaluation de la maîtrise de compétences attendues par l'enseignant.

– L'évaluation est menée au sein d'un outil de communication, par l'analyse des actes de langage des apprenants. C'est le cas dans SPLACH (George, 2004), où à chaque communication, le participant doit au préalable pré-sélectionner un acte de langage (Question, Réponse, etc) parmi la liste proposée. Le système est alors capable d'évaluer le profil de l'apprenant, c'est-à-dire s'il est actif, en retrait etc.

Dans ce cas la qualité du discours n'est pas évaluée. On ne cherche pas à vérifier des compétences mais l'engagement (Durand et Vignollet, 2003) des participants dans l'activité. SPLACH (Support d'une pédagogie de Projet pour l'Apprentissage Collectif) est un environnement informatique support d'une pédagogie par projet permettant l'apprentissage collectif à distance (George, 2001). Il fait partie des Environnements Interactifs d'Apprentissage à Distance (EIAD). SPLACH alterne des phases de travail synchrones et asynchrones. Deux prototypes ont vu le jour : le premier a permis à des collégiens d'établissements différents de réaliser la construction de robots. Les collégiens étaient en concurrence pour fournir le meilleur robot possible. Le second permettait à des étudiants de la Télé université du Québec (Téluq) de développer un programme informatique en commun.

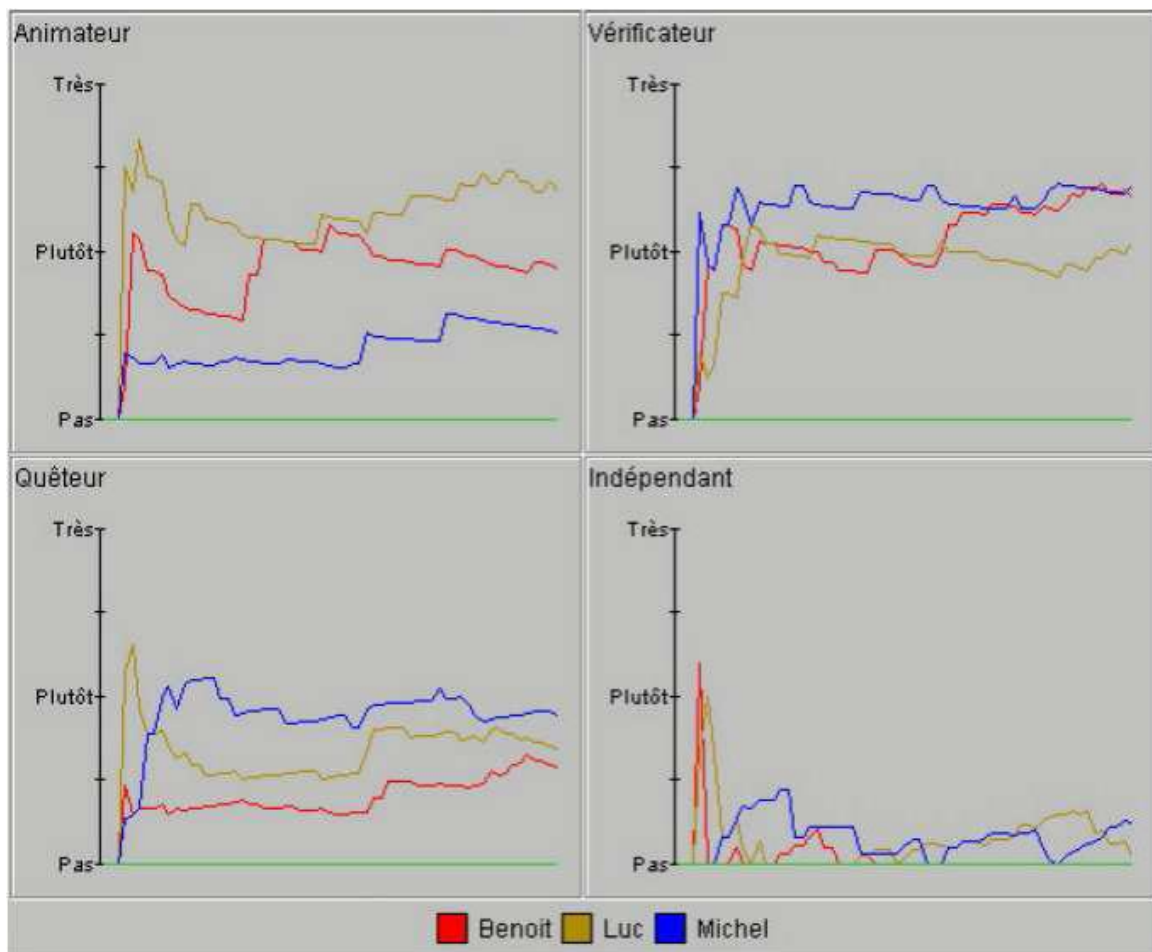
SPLACH dispose de différents outils de communication : forum, webmail, espace de réunion. L'espace de réunion est un chat amélioré, fonctionnant par tour de parole et requérant des actes de langage (répondre, demander,..) pour obtenir une conversation structurée. Les apprenants rédigent différents documents et questionnaires qu'ils sauvegardent dans un espace "Documents", qui est partagé en "documents de l'équipe", "documents personnels" et "documents des équipiers". Suivant les prototypes, différents outils ont été

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

ajoutés. Dans le premier prototype, on trouve un outil de description de robot, un outil de programmation et une interface de pilotage de robot. Dans le second, un outil permettant de programmer en Pascal dans l'environnement SPLACH a été développé.

III.3.8. L'évaluation dans SPLACH

L'évaluation des apprenants dans SPLACH est formative. L'enseignant supervise l'activité et agit sur son déroulement. Il a ainsi la possibilité de vérifier la participation au projet et d'évaluer les productions des apprenants. Le point fort de SPLACH est son outil d'évaluation des profils des apprenants basé sur les conversations dans un chat structuré. Chaque apprenant intervient dans le chat en utilisant un acte de langage. A partir de la théorie sur les actes de langage de Pléty (1996), SPLACH (George, 2004) permet de profiler l'apprenant comme animateur, vérificateur, quêteur ou indépendant de l'activité de groupe.



L'intérêt de SPLACH réside dans l'assistance à l'évaluation, qu'il propose en permettant à l'enseignant de superviser le déroulement du projet mais aussi en permettant de connaître le rôle social joué par chaque apprenant.

III.3.8.1. Le portfolio

Les systèmes de portfolio sont très répandus en Amérique du Nord. Un groupe de travail de la commission scolaire des Premières-Seigneuries au Québec définit le portfolio comme étant un outil d'évaluation des apprentissages qui permet de recueillir et de conserver des échantillons des réalisations de l'élève. Il s'inscrit dans une démarche d'évaluation formative continue et est réalisé en collaboration avec l'élève.

Nous nous intéressons ici au portfolio numérique de l'apprenant, système logiciel permettant le stockage et l'annotation de ressources numériques issues des réalisations de l'apprenant. Charge est à l'apprenant de choisir les réalisations qu'il ajoute à son portfolio, ce qui l'aide à s'auto-évaluer, observer ses progrès, s'impliquer dans une réelle démarche de réflexion métacognitive.

Selon l'usage qui en est fait, on distingue trois types de portfolio de l'apprenant selon (Eyssautier-Bavay, 2004), (Ministère de l'Éducation du Québec, 2002) :

- le portfolio d'apprentissage, qui centralise les travaux de l'apprenant et ses commentaires sur ses propres travaux ;
- Le portfolio de présentation, qui répertorie les meilleurs travaux de l'apprenant. C'est l'équivalent du "book" de l'artiste ;
- Le portfolio d'évaluation, qui permet d'évaluer la progression de l'apprenant dans ses apprentissages.

Il contient des productions, des résultats d'évaluation, les observations de l'apprenant et parfois ses auto-évaluations.

Ces types de portfolio ne sont pas exclusifs les uns par rapport aux autres. En général, un portfolio numérique peut engendrer ces trois stéréotypes. C'est l'enseignant qui choisit l'usage du portfolio. Le développement du portfolio APOM a débuté en 1998 par un groupe de travail de la Coopération régionale de développement pédagogique (CDR) de Montérégie (Québec). Les concepteurs ont voulu développer un portfolio numérique en ligne, simple, convivial et ne nécessitant pas de compétences techniques.

En 2003, APOM est devenu un service du système de gestion de l'apprentissage COLLABA et en 2005 d'EDU-GROUPE. Dans APOM, l'apprenant gère des fiches de

¹COLLABA, <http://www.collaba.ca/fr/>

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

travail (voir figure ci-dessous) dans lesquelles il indique une description du travail réalisé, les compétences mises en jeu, ses remarques et observations ainsi que la date et la discipline. Il est possible d'associer aux fiches un document numérique (pdf, text). L'enseignant peut consulter les fiches et les modifier. L'apprenant peut décider de partager ses fiches avec d'autres apprenants.



Figure N° 28: APOM: interface élève

III.3.8.2 L'évaluation dans APOM

L'évaluation dans un portfolio peut être de nature différente selon les désirs de l'enseignant (Ministère de l'Éducation du Québec, 2002). Dans le portfolio d'apprentissage et de présentation, une large place est laissée à l'apprenant qui autoévalue ses productions en les choisissant et en les commentant. En revanche, dans le cas du portfolio d'évaluation, l'évaluation principale est celle de l'enseignant qui évalue les compétences de l'apprenant grâce aux documents qu'il a ajoutés. Un portfolio d'évaluation peut être utilisé, par exemple, pour l'obtention d'un diplôme.

Les systèmes de portfolio très répandus en Amérique du Nord, offrent un contexte d'apprentissage riche. Avec ces outils, les situations d'apprentissage envisageables sont diverses. Du point de vue de l'évaluation, le portfolio permet de mener des évaluations formatrices dans lesquelles l'apprenant prend conscience de ses progrès. Cela lui permet de mieux s'impliquer dans ses apprentissages.

Mais ces types sont rarement présents simultanément dans un même EIAH. Or, un enseignant peut avoir besoin, sur une même activité pédagogique, de prévoir et de mettre en place des évaluations de différents types. S'il veut, par exemple, constituer des groupes de lecture en les fondant sur la capacité de l'élève à lire un texte, il aura besoin d'introduire une méthode d'évaluation de type diagnostique. Si sur la lecture proposée de ce même texte, il s'aperçoit que les élèves ont du mal à comprendre une notion, il devra introduire une méthode d'évaluation de type formatif. Si enfin il veut s'assurer que le sujet dont traite le texte a été compris et qu'il n'a pas l'intention d'y revenir, il devra prévoir une évaluation de type sommative.

Les évaluations ne sont pas utiles uniquement pour mesurer les progrès réalisés par les élèves. Elles peuvent aussi servir à faciliter ces progrès à condition de pouvoir être utilisées par les enseignants comme un outil pédagogique à part entière, ce qui implique qu'ils puissent les organiser à leur guise.

Pour que les enseignants aient confiance dans les évaluations qui leurs sont proposées, il faut qu'ils comprennent comment elles sont bâties, sur quels critères elles s'appuient, par quels moyens les résultats qu'elles proposent sont obtenus, quels sont leurs référents, leurs référés et les relations qui existent entre les deux, pour employer la terminologie utilisée en Sciences de l'Éducation.

Les élèves, eux aussi, ont le droit de connaître les modalités d'une évaluation. Par exemple, si la grammaire et la forme générale d'un devoir écrit revêtent de l'importance pour les objectifs que l'enseignant s'est assigné, les élèves doivent être informés de ces critères lorsque le travail à faire leur est donné. Si leur degré d'autonomie est un élément qui compte dans l'évaluation de leurs travaux, il importe qu'ils le sachent. Si l'enseignant a l'intention d'observer ses élèves au cours de l'activité, il vaut mieux le leur indiquer.

D'une manière générale, le "scénario" prévu pour l'évaluation doit être compréhensible et communicable.

2.5 Limites des évaluations dans les EIAHs

La plupart des produits EIAHs souffrent toujours de quelques déficits. Voici d'une manière succincte quelques limites :

2.5.1 Limites par rapport à l'outil d'évaluation

- ***Unicité du modèle d'évaluation***

Chaque EIAH utilise au moins un modèle d'évaluation qui est généralement en adéquation avec son utilité pédagogique. Ainsi, il existe des EIAH qui utilisent la co-évaluation dans les activités de groupe, de l'auto-évaluation pour stimuler l'implication des apprenants, de l'évaluation sommative pour certifier des compétences, etc. (Campanale, 2001). Mais dans tous les cas de figure, l'enseignant est contraint par l'EIAH dans le choix du modèle d'évaluation. Les EIAH ne permettent pas de changer de modèle d'évaluation. On peut dès lors se demander si le modèle d'évaluation peut être au moins paramétré.

- ***Fixité de la méthode d'évaluation***

Le modèle d'évaluation utilise une méthode. Mais, là encore cette méthode est statique. Une critique récurrente faite aux exercices de mathématiques est bien souvent de mal-corriger. De circulaires en réformes, ce qui est juste à une époque ne l'est plus toujours à une autre. Il est alors possible qu'un enseignant considère la méthode de correction fautive sans pouvoir la modifier. Toujours dans la méthode d'évaluation, le barème est aussi problématique. L'enseignant n'est pas libre de choisir les caractéristiques qu'il souhaite évaluer et encore moins de leur associer des coefficients.

- ***Résultats d'évaluation peu compréhensibles***

En outre la méthode d'évaluation de l'EIAH, utilise bien souvent un algorithme de calcul de score. Dans TDmaths le score final est pondéré par la difficulté des questionnaires ainsi que par le nombre de tentatives. Dans Pépite (Delozanne et Grugeon 2004), l'enseignant dispose d'un bilan très complet des compétences en algèbre de l'apprenant sans pour autant connaître la manière dont ce bilan est obtenu. Lors de la publication des résultats, l'EIAH n'exprime pas cet algorithme. Il est alors difficile de donner du sens à un score lorsque sa méthode de calcul n'est pas explicite.

- ***Interopérabilité des EIAH***

Enfin, l'EIAH comme entité logicielle autonome et très spécialisée se transforme de plus en plus en une simple fonctionnalité d'un Espace Numérique de Travail (ENT) ou d'un Système de Gestion de l'Apprentissage (SGA). Dans ces dispositifs, de nouveaux acteurs apparaissent tels que les parents mais aussi les services administratifs. Les résultats

d'évaluation produits par l'EIAH doivent donc être utilisables par d'autres services. Par exemple, les scores obtenus en mathématiques sont nécessaires à l'enseignant mais aussi au service de scolarité. Du point de vue de l'évaluation, ces services doivent être interopérables. Cette interopérabilité passe par un format de résultat commun qui n'existe pas aujourd'hui.

2.5.2 Limites par rapport à la connaissance évaluée

En pédagogie, on distingue habituellement savoirs déclaratifs (savoir) et savoirs procéduraux (savoir-faire), bien que cette distinction soit parfois controversée (Bentolila, 1995). En termes de connaissance, on parle de connaissances déclaratives et de connaissances procédurales. Les connaissances déclaratives sont beaucoup plus faciles à apprendre que les connaissances procédurales, car apprendre à un apprenant à réfléchir, à analyser et résoudre des problèmes, est plus délicat qu'à lui apprendre à appliquer une règle bien définie.

Évaluer des savoirs faire est une tâche délicate. Le savoir-faire est une mise en œuvre d'un savoir et d'une habileté pratique maîtrisée dans une réalisation spécifique. Évaluer le savoir-faire d'un apprenant, c'est évaluer sa créativité... imaginons un peu la difficulté. Si on rajoute à cette difficulté, les limites de l'évaluation dans un EIAH (déjà citées), on se rend compte de la complexité à évaluer des connaissances procédurales dans un EIAH.

IV. Conclusion

L'évaluation automatisée recouvre un large éventail de techniques. Elle se base sur des outils assez rudimentaires comme les questionnaires à choix multiples (QCM) ou les textes à trous aussi bien que sur des techniques plus avancées comme les programmes de test de codes ou d'évaluation automatisée de copies. On peut distinguer les exercices d'application du cours des tests de mémorisation de contenu. Ces derniers ont pour fonction de s'assurer qu'une information a été mémorisée. Dans les MOOC, on les retrouve fréquemment au sein même des vidéos de cours, les séquences de cours magistraux de quelques minutes alternant avec des séquences de test.

Les exercices d'application ont une fonction tout autre. Leur difficulté n'est pas nécessairement liée à la simplicité apparente du type de test. Un simple QCM à quatre solutions peut nécessiter des heures de calculs. Parmi les exercices d'application, les applications numériques sont particulièrement utilisées dans les cours de sciences fondamentales (physique, statistiques, etc.). La réponse donnée par le participant n'est considérée comme correcte que si elle se trouve dans une gamme de valeurs fixée par l'équipe

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

pédagogique. L'inconvénient de cette approche est qu'elle permet de considérer comme correct un résultat juste même si le raisonnement suivi pour l'obtenir est faux, et un résultat comme faux même si le raisonnement suivi est le bon. Ce sont les limites de l'évaluation automatisée.

L'évaluation par des programmes de test est une forme d'évaluation automatisée utilisée dans les cours impliquant la rédaction de programmes informatiques. Les programmes de test permettent d'analyser de manière automatique le contenu d'un code, d'en détecter les erreurs et de faire des retours éventuels à son concepteur. Ce type de programme est bien antérieur à l'avènement des MOOC ; les plates-formes comme edX ou Coursera n'ont fait qu'intégrer ce concept. Les enseignants sont donc libres de rédiger leurs propres programmes de test ou d'utiliser des programmes existants, selon leurs besoins.

Pour terminer sur la question de l'évaluation automatisée, nous aimerions souligner que ces techniques connaissent un essor rapide au sein des plateformes, qui inclue désormais de petits laboratoires virtuels. Les exercices basés sur l'apprentissage par essai et erreur vont sans doute se multiplier dans les années à venir.

Nous nous sommes placés jusqu'à présent dans le cadre de l'évaluation de productions individuelles, les productions collectives peuvent également être notées via l'évaluation par les pairs. Et ce pour l'évaluation des projets eux-mêmes, qui rencontre les mêmes problématiques que l'évaluation des productions individuelles, aussi bien que pour l'évaluation des différents membres de l'équipe au sein du projet.

Nous n'avons eu, dans ce qui précède, qu'un bref aperçu des différentes méthodes d'évaluation automatique, sans être exhaustifs, et sans nous préoccuper outre mesure des contextes dans lesquels ces techniques sont appliquées. Celles-ci peuvent être combinées entre elles, scénarisées d'une infinité de façons ; une thèse ne suffirait pas à faire le tour de la diversité de leurs applications. Imaginez par exemple que l'on puisse, en fonction des réponses que vous avez données à tel ou tel exercice, sélectionner en temps réel un test qui soit parfaitement adapté à votre profil, vos connaissances et vos objectifs d'apprentissage. Tests adaptatifs, évaluation en deux étapes, il y a sans doute beaucoup à écrire encore sur la question ...