« Prolongement virtuel » de l'examen sur table par la mise en place d'une évaluation par les pairs dans un but formatif

Bouzidi L'hadi, Département d'Informatique, Faculté des Sciences et sciences de l'ingénieur, Université A/Mira, Béjaïa, Algérie
e-mail : lhadi_bouzidi@yahoo.fr
Fax : 213 34 21 51 05 Tel : 213 63 40 63 32

Jaillet Alain, Directeur ULP MULTIMEDIA, Créateur de la formation UTICEF, Strasbourg, France
e-mail : Alain.jaillet@ulpmm.u-strasbg.fr

Mot(s) clé(s) : évaluation par les pairs, évaluation formative, prolongement virtuel de l'examen sur table.

Résumé : L'examen sur table, en tant qu'évaluation sommative, constitue le plus souvent la seule exigence d'un cursus universitaire. C'est donc un investissement très important pour les étudiants, les enseignants et l'administration d'une institution éducative. Cet investissement permet de noter l'étudiant et se limite à cela. Les copies d'examen et les corrections des enseignants ne constituent pas, en général, une ressource d'apprentissage pour les étudiants. Lors des différentes remises des copies aux étudiants, le plus souvent, seules les notes les intéressent. Cela signifie que dans le cas où l'on soumet les étudiants au même examen, il est possible qu'ils commettent les mêmes erreurs. Nous proposons d'effectuer un « prolongement virtuel » des examens, de sorte de les exploiter au sein d'un campus virtuel pour l'apprentissage, en mettant en place un système d'évaluation formative par les pairs. Notre expérimentation, dont nous livrons les résultats dans cet article, porte sur une promotion de 70 étudiants.

Abstract: The on-table exams, as a sommative evaluation, constitute a very important investment for the students, the teachers and the administration staff of an educational institution. This investment makes it possible to grade the student and is limited to that. The sheets of tests and the corrections of the teachers do not constitute, in general, a resource of training for the students. We noted, during the various handing-over of the grades and the sheets to the students that only the grades interested them. That means that if one subjects the students to the same exams, there are strong chances so that they make the same errors. We propose to carry out a "virtual prolongation" of the exams of kind to exploit them within a virtual campus for the training, by setting up a system of formative evaluation by the peers. We set up experimentation on 70 students. We deliver, in this article the results of this experimentation.

Introduction

L'évaluation des apprentissages est l'une des tâches les plus délicates pour un enseignant. Elle est très importante aux yeux de l'étudiant. L'examen, en classe, induit une évaluation

sommative permettant à l'enseignant de situer le niveau de maîtrise des acquis de ses étudiants à l'issue d'un certain nombre d'enseignements successifs. Cette activité exige un grand investissement de tous les acteurs gravitant autour du processus d'apprentissage (étudiants, enseignants et administration). L'examen final sert, le plus souvent, à récolter des notes pour décider si l'étudiant peut ou non être accepté en regard du règlement des études, soit pour poursuivre dans le niveau supérieur, soit pour se voir décerné le diplôme. L'enseignant peut, par ailleurs, procéder à une évaluation dite formative qui permet de prendre des mesures pour remédier aux lacunes éventuelles de son processus d'enseignement (pour le formateur) et d'apprentissage (pour l'étudiant). Nous pensons que ce dernier mode d'évaluation est essentiel pour un apprentissage efficace. En effet, le fait de permettre à l'étudiant de voir ses propres erreurs et de situer son niveau de maîtrise par rapport aux objectifs du cours, l'aide à prendre les bonnes décisions quant à sa façon d'apprendre. Par ailleurs, il est possible que l'évaluation par les pairs (Topping, 1998) en classe, constitue une technique très efficace pour l'apprentissage. Chacun peut constater que, malgré l'importance de l'investissement qui se fait pour les examens, ces derniers restent dans les archives et constituent plus une tâche administrative qu'une activité pédagogique. Nous proposons de remédier à cette situation en faisant de l'examen un processus au service de la pédagogie, grâce à l'utilisation de l'évaluation par les pairs dans un environnement virtuel d'apprentissage. Ainsi, nous présentons dans ce qui suit, l'état de l'art sur l'évaluation par les pairs en FAD, puis nous expliciterons notre démarche d'implémentation d'un prolongement virtuel des examens effectués en classe impliquant une évaluation par les pairs. Nous présentons, ensuite le processus d'expérimentation que nous avons mis en place et nous livrons, enfin, l'analyse et l'interprétation des données récoltées.

Evaluation par les pairs via Internet

L'utilisation de l'évaluation par les pairs (peer assessment) via Internet est récente. Elle permet à des apprenants d'évaluer les travaux de leurs pairs et éventuellement d'évaluer leur propre travail. A la différence des techniques d'autoévaluation qui, le plus souvent, se limitent aux domaines cognitifs de base (apprendre, comprendre) (Blomm, 1956) (Anderson 2001), l'évaluation par les pairs permet de développer un apprentissage dans des domaines cognitifs élevés (application, analyse, évaluation). Il s'agit d'une approche qui implique les étudiants dans le processus de la révision, de l'évaluation et du feedback sur des travaux en ligne. J. A Gilles Doiron (Gilles, 2003) explique que certains auteurs critiquent l'utilisation des TIC pour l'évaluation par les pairs en argumentant qu'elle est moins rigoureuse que des formes d'évaluations traditionnelles, qu'elle exige trop des étudiants en leur mettant trop de pression, qu'elle n'est pas fiable, et qu'elle n'est pas nécessairement équitable. En réponse à cela, d'autres auteurs comme Bostock (Bostock, 2000) sont convaincus que l'évaluation (formative ou sommative) des travaux d'autres étudiants par les étudiants eux même a beaucoup d'avantages pour l'apprentissage, que ce soit pour l'évaluateur ou l'évalué. Bostock fait remarquer que l'évaluation par les pairs encourage l'autonomie des étudiants et développe des compétences dans des domaines cognitifs élevés. Il reconnaît certaines faiblesses de ce type d'évaluation, mais explique qu'elles peuvent être évitées en mettant en place un système d'anonymat, de multiples évaluations et évaluateurs et une modération par l'enseignant. De plus, Bostock, argumente que l'Internet et les TIC permettent d'assister la gestion d'un grand nombre d'étudiants.

Il existe peu d'outils ou systèmes d'évaluations par les pairs en ligne. En premier exemple, nous citons le programme CPR (Criterion Peer Review™) développé par Orville L. Chapman

et Michael A. Fiore à UCLA (USA). Ce programme a été introduit en 1999 et incorpore un ensemble d'outils logiciels permettant de gérer le processus de correction, d'analyse des travaux des étudiants et de préparation des rapports pour les étudiants et l'enseignant (Chapman, 2001). Ce programme engage l'étudiant dans la correction de courts essais. Après la soumission électronique d'un essai, chaque étudiant, lors d'une phase dite de calibrage, assigne des notes à 3 travaux : un premier est un exemple (modèle) écrit par un expert, les deux autres sont des documents comportant des omissions et des erreurs de conception. Après cette phase, le programme CPR évalue la performance de chaque étudiant, et si cette performance n'est pas satisfaisante, l'étudiant reçoit des instructions plus précises. Ce dernier doit recommencer cette phase de calibrage jusqu'à obtenir un résultat jugé satisfaisant par le programme CPR. Cette phase sert à mettre en situation les étudiants et à leur conférer des compétences d'évaluateur. En seconde phase, les étudiants seront invités à évaluer et noter 3 essais anonymes en plus de leur travail. En dernière phase, le programme génère un rapport montrant les commentaires des évaluateurs et les notes. Un second exemple de système d'évaluation par les pairs en ligne est OASYS développé par A. Bhalerao et A. Ward (Bhalerao, 2001) à l'université Warwick au Royaume-Uni. La particularité de ce système est, qu'en plus d'une notation automatique des questions à choix multiples, il traite les questions à réponses ouvertes en contrôlant leur affectation aux étudiants, en vue d'une évaluation par les pairs. Il s'agit donc d'un système hybride tirant, à la fois, des avantages des outils d'autoévaluation et de la technique d'évaluation par les pairs. « Peer Grader » (Gehringer, 2001) est un autre exemple de système d'évaluation par les pairs en ligne. Il a été développé à l'Université de l'Etat de la Caroline du Nord. La particularité de ce système est la possibilité qu'il offre à l'étudiant pour évaluer les corrections des pairs de sorte à assurer, globalement, une meilleure fiabilité des évaluations. « NetPeas » (Lin, 2001) est un système d'évaluation par les pairs se distinguant par le fait que le processus d'évaluation se fait en deux phases différentes distinguant les feedback du processus de notation. Ainsi, les étudiants sont invités dans une première phase, à évaluer des travaux de leurs pairs en donnant des feedbacks. Ces derniers sont ensuite redistribués aux étudiants qui vont procéder à la notation.

Un dernier exemple d'outil d'évaluation par les pairs est le module « Atelier » de la plate forme Moodle[1]. Cet outil permet de mener une évaluation par les pairs en plusieurs phases. La première phase est la mise en place du travail : L'enseignant décide des paramètres de l'atelier (échéances, choix des méthodes d'évaluation, définition des critères d'évaluation, barèmes, etc.). L'enseignant peut présenter ensuite des échantillons de travaux qui vont servir d'exemples et de contre-exemples. Les étudiants devront les évaluer afin de bien comprendre les critères d'évaluation avant de créer et de remettre leurs propres travaux. Cependant, avant que les étudiants ne puissent remettre leurs travaux, l'enseignant devrait évaluer lui-même ces exemples. Il aura ainsi à sa disposition des spécimens d'évaluations pour comparaison avec les évaluations de ces exemples faits par les étudiants. La seconde phase est l'ouverture de la remise des travaux. Les étudiants peuvent déposer leur travail sous forme de texte HTML auquel ils peuvent joindre des fichiers divers. Les travaux déposés seront distribués pour évaluation dans la phase suivante qui est l'ouverture des évaluations, où les étudiants peuvent évaluer les travaux de leurs pairs, mais ne peuvent plus remettre leur travail ou d'autres versions de leur travail. Vient ensuite la phase du calcul des notes finales : Le travail des étudiants est alors terminé, le système calcule les notes des évaluations en se basant sur des paramètres fixés au préalable par l'enseignant. Il faut noter ici que deux types de notes sont calculées : les notes sur les travaux déposés (calcul de la moyens des notes données par les étudiants par exemple), et les notes sur le qualité des évaluations. Là l'étudiant se voit affecté

[1] http://www.moodle.org

une note sur la pertinence de son évaluation. Les notes finales des étudiants comportent trois composantes : la note accordée par l'enseignant (au travail déposé), la moyenne des notes accordées par les pairs et une note représentant la qualité des évaluations faites par l'étudiant. Cette dernière note peut inclure une « note moyenne » donnée par l'enseignant en conséquence des commentaires de l'étudiant dans ses évaluations. L'enseignant fixe un coefficient à chaque composante de la note avant que le calcul soit effectué. La dernière phase est évidemment l'affichage des notes pour les étudiants.

Tous les environnements d'évaluation par les pairs présentés englobent non seulement un processus d'évaluation, mais surtout un environnement de travail collaboratif impliquant tout les ingrédients de la collaboration où le feedback, l'autocritique et la pensé critique prennent une dimension importante (Trahasch, 2004).

Processus d'évaluation en ligne par les pairs des examens effectués en classe

A la différence des systèmes existants, nous proposons d'étendre l'évaluation par les pairs en ligne aux travaux réalisés en classe (hors ligne), notamment pour les examens surveillés. Ainsi, le processus que nous proposons s'articule autour des phases suivantes :

1. organiser un examen en classe
2. récupérer puis numériser les copies d'examen
3. traiter les copies numérisées pour assurer l'anonymat des étudiants
4. déposer les copies numérisées dans le campus virtuel pour chaque étudiant
5. préparer l'activité sur le campus virtuel en indiquant le barème, les critères d'évaluation et divers paramètres pour le calcul des notes et la distribution des copies aux étudiants.
6. ouvrir les évaluations (pour l'enseignant et pour les pairs)
7. remettre les notes et les feedback aux étudiants et ouvrir les discussions.

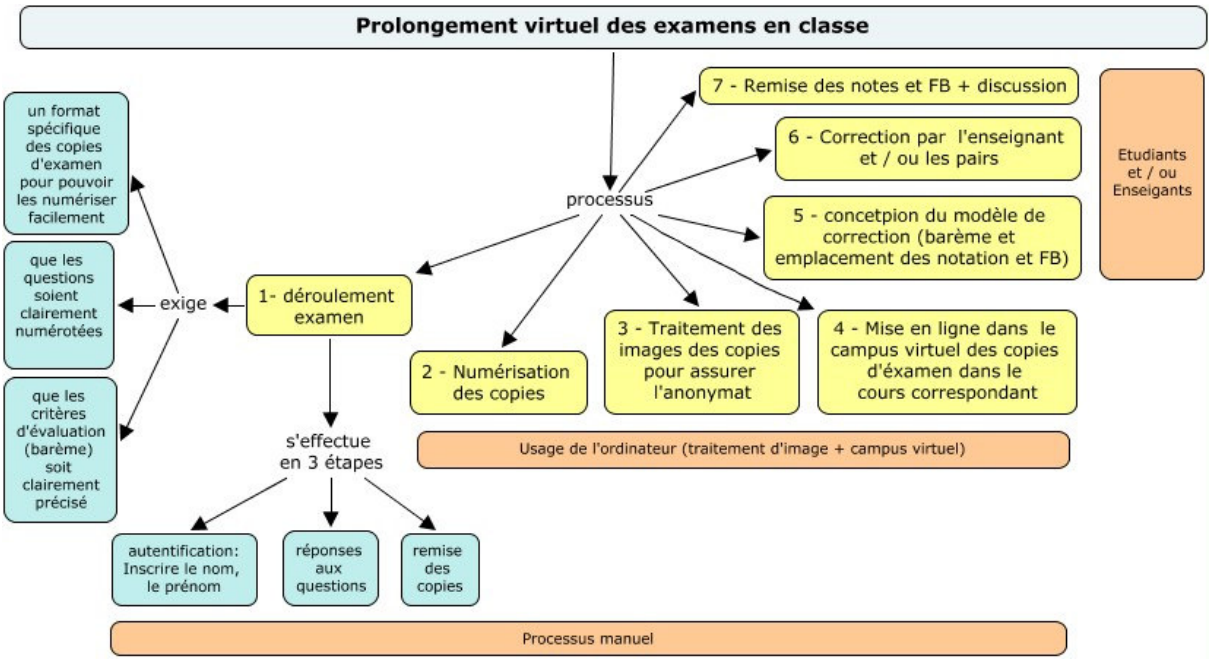

Figure 1 – Processus d'évaluation en ligne par les pairs des examens effectués en classe

Expérimentation

L'expérimentation a été réalisée sur 68 étudiants de 2ème année ingénieur en informatique. L'examen choisi concerne l'architecture des ordinateurs et a duré 2 heures. Pour cette expérimentation, nous avons choisi un format papier permettant de numériser facilement les copies d'examen et facilitant l'anonymat des étudiants. Par la suite, nous avons crée un atelier (d'évaluation par les pairs) dans le campus virtuel de l'université de Bejaia[2] où sont déjà inscrits nos étudiants. Nous avons numérisé les copies avec un scanner manuel (format A4). Nous avons, ensuite, traité chacune des images pour cacher les noms des étudiants et insérer à la place des numéros assurant ainsi l'anonymat. Sachant que la plate forme Moodle ne permet pas de récupérer les copies numérisées d'une façon automatique, l'enseignant a procédé à leurs dépôts comme si c'était les étudiants qui l'avaient fait. Un paramétrage de l'atelier a fait en sorte que chaque étudiant puisse évaluer 4 copies de ses pairs en plus de la sienne. Les étudiants ne pouvaient pas identifier les auteurs des copies qu'ils devaient évaluer. L'atelier a été paramétré de sorte à ce que la qualité des évaluations des étudiants soit notée. La note associée à cette qualité est calculée automatiquement (par l'intermédiaire de la plate forme Moodle) à la fin du processus d'évaluation par les pairs. Ainsi, nous nous sommes basé sur trois hypothèses : La première stipule que la note donnée par l'enseignant est plus fiable que celles données par les étudiants (dans notre cas, nous avons fixé le poids de la note de l'enseignant à 5 fois celui d'un étudiant). La seconde hypothèse stipule que la qualité d'une évaluation est bonne lorsqu'elle engendre une note qui s'approche de la moyenne des notes attribuée par les pairs et l'enseignant (la note attribuée par l'enseignant étant pondérée d'un coefficient égal à 5) et, inversement, cette qualité est mauvaise lorsqu'elle engendre une note qui s'éloigne de cette moyenne. La troisième hypothèse stipule que la marge acceptable entre la note souhaitée (moyenne des notes attribuées à la copie par les pairs et l'enseignant) et la note attribuée à la copie par un étudiant donné est dépendante du domaine des sciences dans lequel on se place. Si on se place dans le domaine des sciences exactes, cette marge doit être très petite ou nulle. Dans le cas des sciences humaines, elle peut être plus large. Comme nous avons effectué un examen en sciences exactes (informatique), nous avons, donc fixé le paramétrage de notation de l'évaluation à « très stricte »[3] c'est-à-dire que l'évaluation effectuée par un étudiant sur une copie ne se verra affecté une bonne note pour sa qualité que si elle est très proche ou égale à l'évaluation souhaitée (celle qui est égale ou très proche de celle de l'enseignant).

Voici les échéances qui ont été suivies pour mener à bien notre expérience :

- Début de la remise des travaux: vendredi 7 avril 2006
- Fin de la remise des travaux: dimanche 14 avril 2006
- Début des évaluations: vendredi 14 avril 2006
- Fin des évaluations: jeudi 18 mai 2006

La figure-2 montre comment le processus d'évaluation a été réalisé. Cette figure présente une fenêtre comportant deux cadres : un cadre destiné à l'évaluation et un autre permettant l'observation de la copie d'examen à évaluer. Le premier cadre comporte une zone d'identification de la copie d'examen et de l'évaluateur et une zone énumérant les différents éléments d'évaluation (dans notre cas, nous avons utilisé 17 éléments chacun associé à une question de l'examen). Pour chaque élément, on permet à l'évaluateur d'indiquer une note et un feedback. La note est indiquée par un clic sur une zone de boutons à option allant de

[2] http://elearning.univ-bejaia.dz

[3] L'atelier d'évaluation par les pairs de Moodle (Workshop) permet d'indiquer 5 types de comparaisons des évaluations allant de très stricte à très indulgente.

« correcte » jusqu'à « incorrecte ». La valeur de la note associée à un élément d'évaluation est calculée à base d'un barème et d'un coefficient fixé à l'avance lors de la création de l'atelier d'évaluation par les pairs. La note finale de l'examen est calculée automatiquement, par l'addition des notes affectées aux 17 éléments d'évaluation multipliées par leurs coefficients respectifs. Le second cadre permet à l'évaluateur d'observer la copie d'examen.

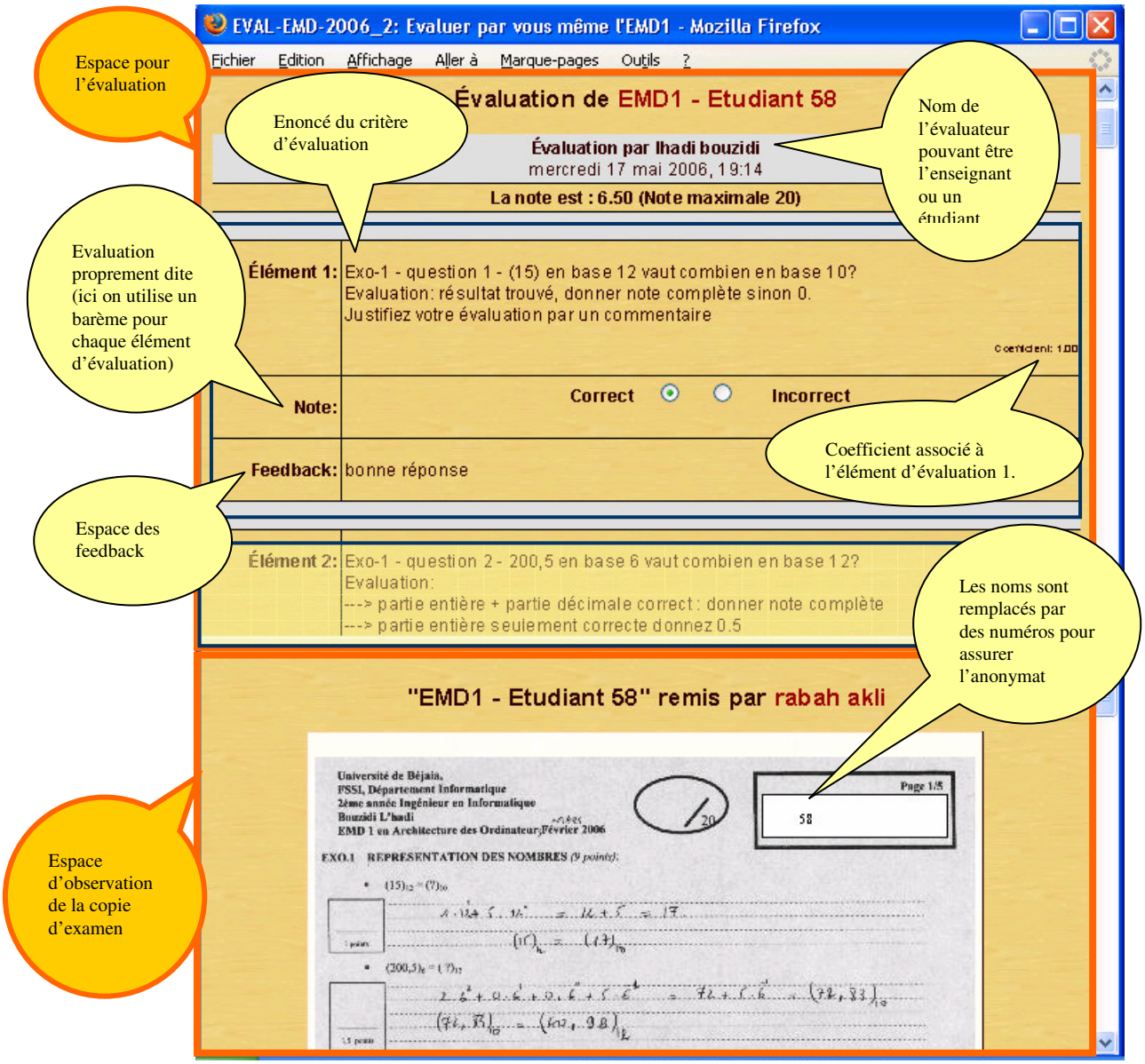

Figure 2 – Exemple d'évaluation par les pairs

Récolte des données

Les données quantitatives : 68 étudiants ont participé à l'expérience. 49 parmi ces 68 étudiants ont évalué leur propre travail. Les données quantitatives ont été récoltées depuis le campus virtuel à la suite de la délivrance des notes finales. Nous avons récupéré, pour chaque copie : Une note d'autoévaluation, une à 4 notes attribuées par les pairs et une note attribuée par l'enseignant.

N° de l'étudiant	Notes : autoévaluation	Notes: évaluation par les pairs				Notes: évaluation par l'enseignant
1	16,00	12,50	12,00			**16,00**
2	14,00	12,00				**12,50**
4		09,50	09,00	07,50	07,00	**07,50**
5		05,50	05,50	04,00	03,00	**04,50**
6	10,00	10,50	07,00	07,00	07,00	**07,50**
7	08,50	07,50	04,50	02,00		**06,50**
8	12,00	13,00				**12,50**
9		09,00	09,00	08,50	07,50	**09,00**
10	14,00	10,50	09,50			**08**
11	07					
12						
13						

Tableau 1 : Exemple de données quantitatives récoltées.

Données qualitatives : Il s'agit de données permettant d'avoir une appréciation globale de l'expérience à travers les avis des étudiants. Ces données ont été récoltées à l'issue d'un questionnaire publié dans le campus virtuel et dont l'objectif est d'estimer l'apport pédagogique de notre expérimentation. Ce questionnaire englobe 8 questions que voici :

Q1. Pensez-vous que le fait d'évaluer par vous même quelques copies de vos camarades vous permet d'apprendre plus?
Q2. Pensez-vous que cette méthode développe l'esprit d'autocritique chez vous?
Q3. Pensez vous que cette méthode permet une meilleur interaction entre les étudiants?
Q4. Pensez vous que cette méthode est lourde et vous fait perdre votre temps?
Q5. Souhaitez vous que cette méthode soit utilisée ultérieurement pour d'autres EMD du même module?
Q6. Souhaitez-vous que cette méthode soit utilisée ultérieurement pour d'autres modules?
Q7. Quelles sont les difficultés que vous avez rencontré (difficultés de compréhension, difficultés techniques, autres).
Q8. Avez-vous un commentaire général sur l'activité "corriger votre EMD1", si oui donnez-le (nous souhaitons avoir vos suggestions si vous en avez).

Les réponses aux questions précédentes ont été récoltées dans des forums ouverts dans le campus virtuel. Le tableau 2 donne une idée sur leur tendance.

Questions	Nombre d'étudiants ayant répondu	Nombre de réponses affirmatives	Nombre de réponses négatives
Q1	33	32	1
Q2	28	28	0
Q3	31	21	10
Q4	25	22	3
Q5	25	23	2
Q6	22	20	2
Q7	38	-	-
Q8	22	-	-

Tableau 2 – Réponses au questionnaire

Analyse

Analyse des données quantitatives : On suppose que la note de l'enseignant est fiable et va constituer un repère à l'analyse qui va suivre. On pose :

Na(i) : note attribuée par l'étudiant *i* à sa propre copie

Ne(i) : note attribuée par l'enseignant à la copie *i*.

Np(i) : vecteur comportant les notes attribuées par les pairs à la copie *i*.

$\sum(Np(i))$ *:* somme des notes attribuées par les pairs pour la copie *i.*

MNp(i) : moyenne des notes attribuée par les pairs (1 à 4 étudiants) à la copie *i*. On appellera cette moyenne « note des pairs ».

Nn(i) : nombre de notes qui ont été attribuées à la copie *i*.

MNep(i) : moyenne des notes attribuée par les pairs (1 à 4 étudiants) et l'enseignant à la copie *i*. La note de l'enseignant étant pondérée d'un coefficient égal à 5.

$$MNep(i) = (\sum(Np(i)) + Ne(i)*5)/(Nn(i)+4)$$

$Max_{np}(i)$: la plus grande note attribuée par les pairs à la copie *i*.

$Min_{np}(i)$: la plus petite note attribuée par les pairs à la copie *i*.

Eea(i) : écart entre la note attribuée par l'enseignant à la copie *i* et la note attribuée par l'auteur de la copie *i* à sa propre copie (autoévaluation).

$$\boxed{Eea(i) = Ne(i)- Na(i)}$$

Eep(i) : écart entre la note attribuée par l'enseignant à la copie *i* et la moyenne des notes attribuées par les paires à la même copie *i*.

$$\boxed{Eep(i) = Ne(i)-MNp(i)}$$

Ep(i): Ecart des notes attribuées par les pairs à la copie *i*.

$$Ep(i) = Max_{np}(i) - Min_{np}(i)$$

$F(i) = Eep(i) - Ep(i).$

La fonction *F* permet de comparer, pour une copie *i* donnée, la distribution des notes attribuées par les pairs et la distance de la note des pairs par rapport à celle attribuée par l'enseignant.

A partir du tableau 1, nous avons généré trois graphiques (figures 3, 4 et 5) permettant de comparer les trois types d'évaluation qui ont eu lieu durant l'expérimentation : l'autoévaluation, l'évaluation par les pairs et l'évaluation par l'enseignant.

A – Données comparatives entre l'autoévaluation et l'évaluation par l'enseignant : Le graphique de la figure 3 montre les notes attribuées par l'enseignant (*Ne(i)*) et celle attribuées par les étudiants à leur propre travail (*Na(i)*). En bas du graphique une courbe montre les écarts entre les deux types d'évaluation (*Nea(i)*).

B – Données comparatives entre l'évaluation par les pairs et l'évaluation par l'enseignant : Le graphique de la figure 4 montre le moyennes des notes attribuées par les pairs (*MNp(i)*) et celles attribuées par l'enseignant (*Ne(i)*). En bas du graphique, on présente les écarts entre ces deux types d'évaluations (*Eep(i)*).

Le graphique de la figure-5 montre les écarts entre les évaluations effectuées par les pairs (*Ep(i)*) comparés aux écarts entre les évaluations par l'enseignant et celle effectuées par les pairs (*Eep(i)*).

"

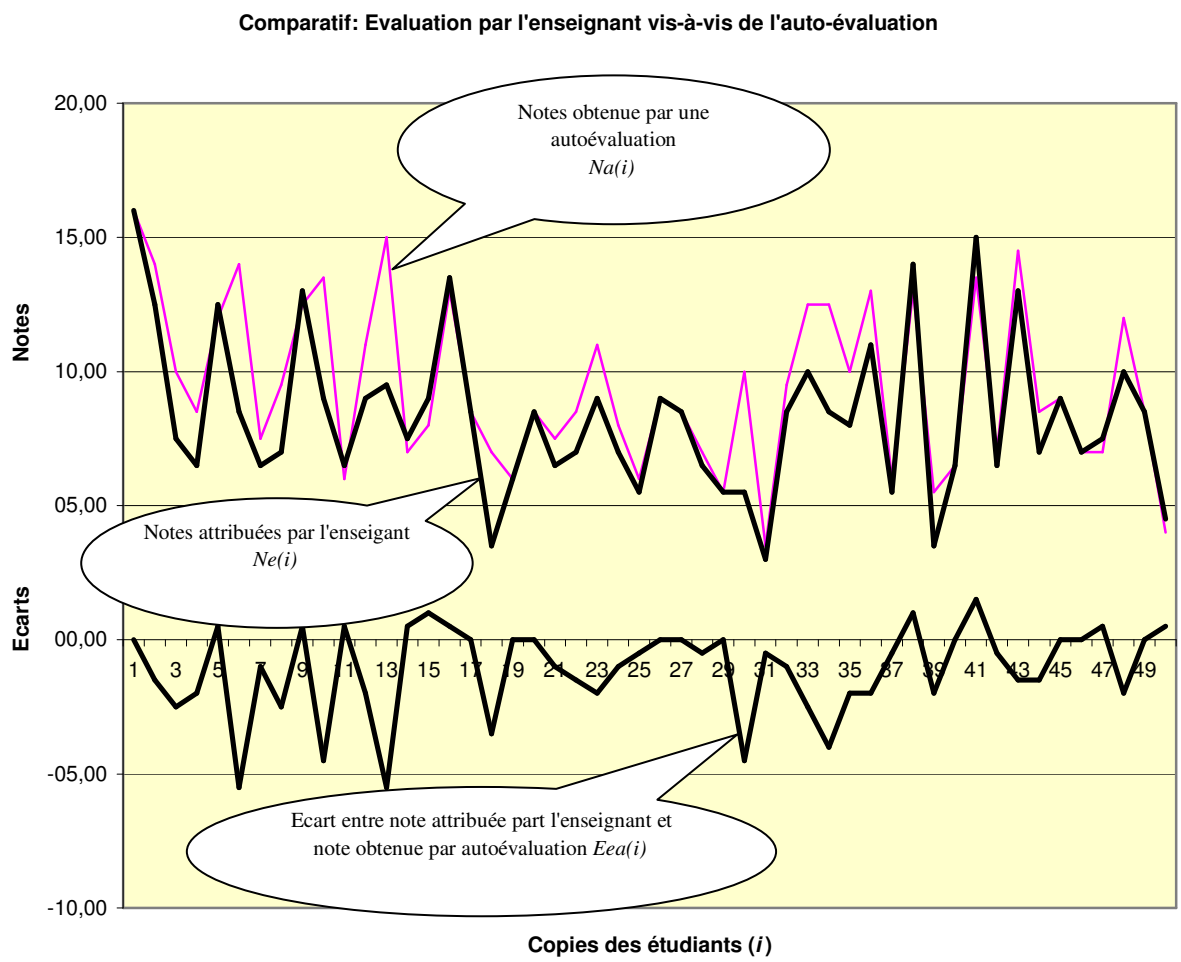

Figure 3 – comparatif entre l'évaluation par l'enseignant et l'autoévaluation

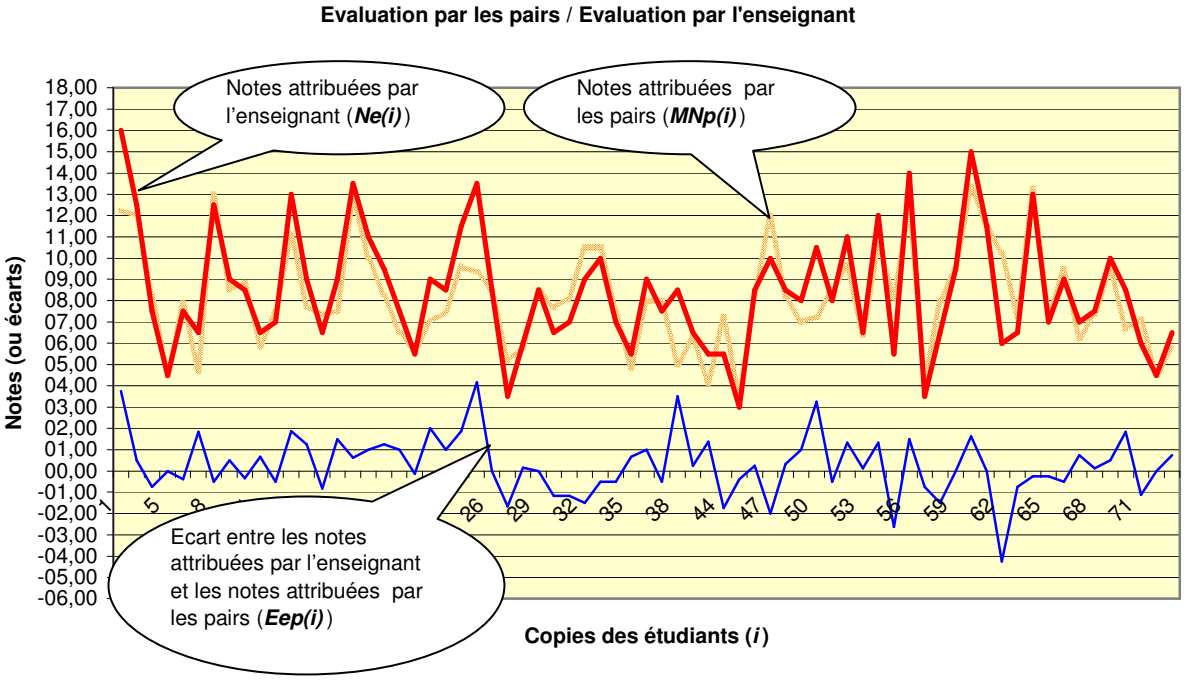

Figure 4 – comparatif entre l'évaluation par l'enseignant et l'évaluation par les pairs

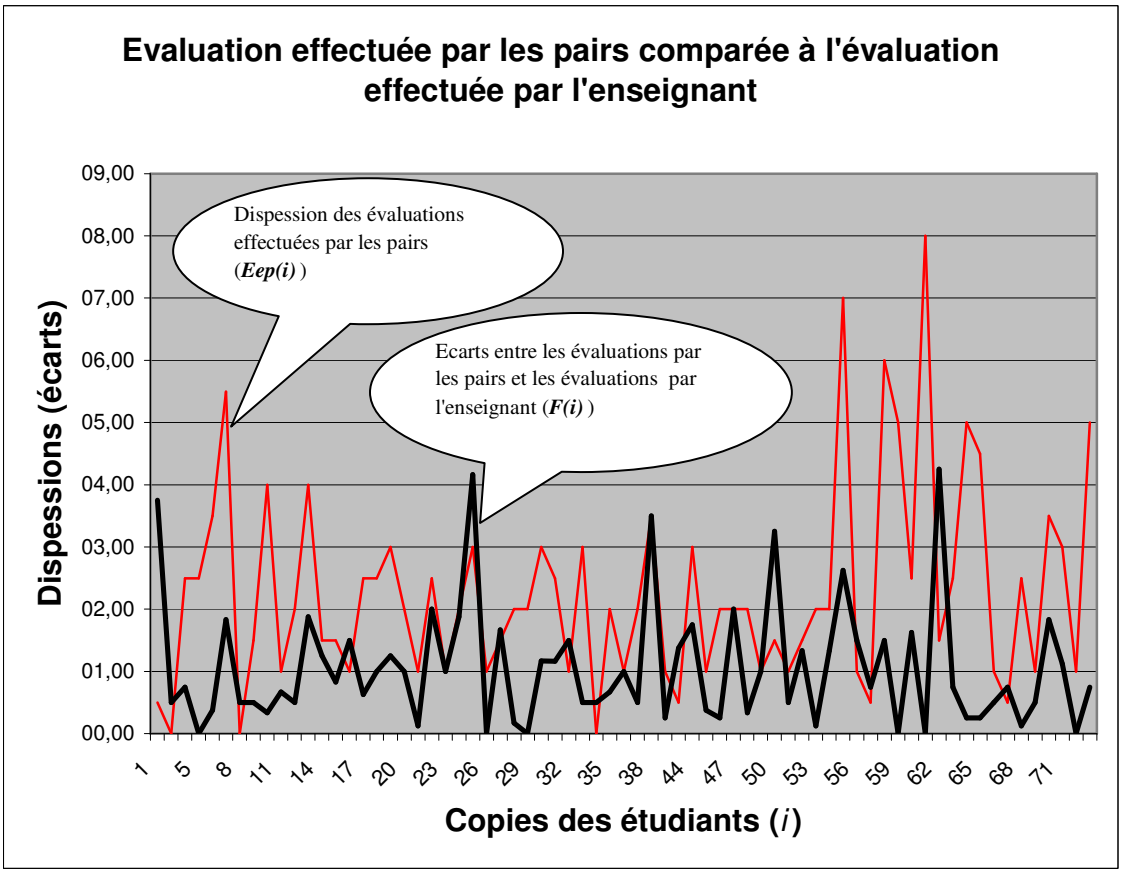

Figure 5 – Comparatif entre les dispersions des évaluations effectuées par les pairs et les écarts entre les évaluations par les pairs et celles effectuées par l'enseignant

Analyse des données qualitatives

Le tableau 2 présente l'analyse du questionnaire.

Question	Analyse
Q1	Les étudiants expliquent en majorité que l'expérience leur permet de corriger leurs erreurs, de mieux comprendre la solution de l'examen et de ne pas refaire leurs erreurs dans le futur. Ils ajoutent qu'en corrigeant les erreurs de leurs camarades, ils apprennent plus.
Q2	Tous les étudiants confirment que l'expérience qu'ils ont suivi permet de développer l'esprit d'autocritique, mais plusieurs d'entre eux expliquent que c'est difficile d'accepter les mauvaises notes. Certains ont expliqué que ça leur permet de ne pas trop critiquer l'enseignant.
Q3	2/3 des étudiants qui ont répondu à cette question pensent que cette expérience permet une meilleure interaction entre eux alors que 1/3 pensent que non. Ces derniers expliquent cela par le fait que les copies soient anonymes.
Q4	La majorité des étudiants trouvent que la méthode expérimentée n'est pas lourde et ne leur fait pas perdre leur temps. Ils se plaignent cependant, des problèmes d'accès au campus virtuel. Ils trouvent aussi que l'expérience leur prend beaucoup de temps
Q5	La majorité des étudiants pensent que ça vaut la peine de renouveler l'expérience, car son apport pédagogique est très important.
Q6	La majorité des étudiants pensent qu'il serait intéressant de généraliser cette expérience sur d'autres modules, à condition de garantir une bonne connexion au campus virtuel.
Q7	Tous les étudiants se sont plaints des problèmes de connexion au campus virtuel. Un tiers a trouvé des difficultés pour lire sur écran les copies numérisées. Un étudiant sur cinq a rencontré des difficultés dans l'application du barème qu'il juge trop sévère. Quelques étudiants expliquent qu'ils ont eu des difficultés sur la manière de rendre des feedbacks lors de leurs corrections.
Q8	La majorité des étudiants pensent que l'évaluation par les pairs qu'ils ont expérimenté est profitable pour eux dans la mesure d'un apprentissage plus riche appuyé par le fait qu'ils sont confrontés à leurs propres erreurs. Ils soulignent, cependant, qu'il faut garantir de bonnes conditions d'accès au campus virtuel. Certains, pensent que ça leur fait du bien d'être momentanément à la place du professeur.

Tableau 2 : Analyse du questionnaire

Les réponses au questionnaire nous révèlent les points suivants :

1. un bon apport pédagogique (autocritique, autonomie, plus grande responsabilité)
2. des problèmes de connexion au campus virtuel
3. une interaction non satisfaisante
4. des difficultés pour rendre les feedback
5. des difficultés pour appliquer les barèmes
6. des problèmes de lisibilité de certaines copies d'examen

Interprétation

Il ressort du graphique de la figure 3 que les étudiants ont tendance à attribuer, à leur propre travail, des notes plus élevées que celles données par l'enseignant. Le plus grand écart entre la

note de l'enseignant et celle de l'étudiant est de 5,5 points. La moyenne des écarts est de -0,98. Sachant que les copies évaluées relèvent d'un domaine des sciences exactes, nous pensons que l'écart entre les évaluations ne devrait pas dépasser 1 point et la moyenne des écarts doit être proche de 0.

Le graphique de la figure 4 révèle, à la différence de l'autoévaluation, que l'évaluation par les pairs a tendance à se rapprocher de celle effectuée par l'enseignant. Le plus grand écart entre la note donnée par l'enseignant et celle donnée par les pairs est de 4,17 points. La moyenne des écarts est de +0,27. Cette moyenne, étant positive, indique que les pairs ont tendance (légèrement) à attribuer des notes inférieures à celles attribuées par l'enseignant. Sachant que cette valeur est assez petite, on peut considérer que l'évaluation effectuée par les pairs est fiable.

Le graphique de la figure 5 révèle que les écarts des évaluations effectuées par les pairs (*Eep(i)*) sont légèrement en corrélation avec les écarts qu'il y a entre l'évaluation de l'enseignant et celles des pairs (fonction *F* comparée à la fonction *Eep*). Cela veut dire qu'en général, plus les notes attribuées par les pairs se rapprochent les unes des autres, plus la moyenne de ces notes s'approche de la note de l'enseignant, donc plus fiables. Pour avoir une évaluation globale fiable, il serait intéressant d'éliminer automatiquement, pour une copie *i* donnée, les notes des pairs qui s'écartent trop de la moyenne *MNep(i).*

Les réponses au questionnaire permettent d'affirmer que l'expérience a été bénéfique pour les étudiants, mais suggère une remédiation :

1. garantir les bonnes conditions d'accès au campus virtuel
2. bien concevoir les barèmes.
3. mieux préparer les étudiants sur la façon de corriger les examens, notamment sur la manière de rendre des feedbacks.

Conclusion

Les résultats obtenus démontrent la faisabilité de l'évaluation par les pairs sur des examens réalisés en classe. Le processus de numérisation et de dépôt des copies dans le campus virtuel constitue une contrainte pour l'enseignant et suggère, de ce fait, de le rendre automatique. Le résultat stipulant que l'autoévaluation engendre des notes surévaluées permet de conclure qu'il ne faut pas s'en servir comme évaluation sommative, mais que ça constitue un très bon exercice pour l'évaluation formative. Le manque d'interactions entre les étudiants, expliqué par l'anonymat des copies, suggère de dévoiler, après le processus d'évaluation par les pairs, l'identité des auteurs des copies en prévoyant une phase d'interaction. L'évaluation formative impliquant des interactions fréquentes entre l'enseignant et ses étudiants et l'élaboration de feedbacks, de consignes et d'orientations, alourdie la charge de travail de l'enseignant. La méthode proposée dans cet article, peu alléger considérablement cette charge en confiant une partie de cette évaluation aux étudiants. Les résultats obtenus ne peuvent pas être généralisés car l'expérimentation a été réalisée sur un seul domaine scientifique (sciences exactes) et que l'enseignant impliqué est lui-même à l'origine de l'idée de cette expérience (ils pourrait avoir tendance à favoriser la réussite plus que l'échec). Il faut donc que l'expérience soit renouvelée dans des conditions plus larges (l'étendre à d'autres domaines scientifiques comme les sciences humaines) et pour un effectif étudiant et enseignant plus important.

Références

Anderson, W, A. (2001). A taxonomy for learning, teaching, and assessing; a revision of Bloom's taxonomy of educational objectives. Longman.

Bhalerao, A & Ward, A. (2001). Toward Electronically Assisted Peer Assessment: A Case Study'. ALT-J, Vol. 9, No. 1, pp. 26-37.

Bloom, B, S. (ed.) (1956). Taxonomy of Educational objectives: The Classification of Educationnal Goals. Handbook I: Cognitive Domain. New York: Longmans Green.

Bostock, S. (2000). Student Peer Assessment, Learning Technology. Keele University, UK, (Electronic Citation), http://www.keele.ac.uk/depts/aa/landt/lt/docs/bostock_peer_assessment.htm (dernière visite 19 mai 2006).

Chapman, O. L. (2001), 'Calibrated Peer Review™, The White Paper. A Description of CPR', (Electronic Citation), http://cpr.molsci.ucla.edu (dernière visite le 19 mai 2006).

Gehringer, E., F. (2001). Electronic peer review and peer grading in computer science course », Proceeding of the thirty second SIGCSE technical symposium on Computer Science Education, Vol. 33, No 1, 2001, pp. 139-143.

Gilles Doiron, J. A. (2003). The value of Online Student Peer Review, Evaluation and Feedback in Higher Education, Centre for Development of Teaching and Learning, National University of Singapore (NUS), Vol.6 No.9, September 2003.

Lin, S. S., Liu, E. Z. and Yuan, S. (2001) "Web BasedPeer Assessment: Attitude and Achievement", IEEE Transactions on Education, Vol 9, No 4.

Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. Review of Educational Research, 68(3), 249-276.

Trahasch, S. (2004). From Peer Assessment Towards Collaborative Learning. 34th ASEE/IEEE Frontiers in Education Conference, Savannah, GA, October 20 – 23, 2004.