



Un nouveau processus d'évaluation pour améliorer la qualité des feedbacks dans les tests en ligne

► **Franck SILVESTRE, Philippe VIDAL, Julien BROISIN**
(IRIT, Université Toulouse 3 Paul Sabatier)

■ **RÉSUMÉ** • Nos travaux précédents ont introduit la plateforme Tsaap-Notes dédiée à la génération de questionnaires à choix multiples en recyclant (1) les questions interactives posées pendant le cours par l'enseignant, et (2) les explications saisies par les étudiants pour constituer les feedbacks présentés aux étudiants durant les tests. Dans cet article, nous présentons un nouveau processus baptisé « processus en N phases » et son implantation dans Tsaap-Notes qui visent d'une part à augmenter l'engagement des étudiants dans l'activité de rédaction des explications et, d'autre part à améliorer la qualité des feedbacks restitués au sein des tests générés. Notre approche a été validée par une première expérimentation.

■ **MOTS-CLÉS** • Questions interactives, *feedback*, tests informatisés, évaluation par les pairs, instruction par les pairs, processus en N phases.

■ **ABSTRACT** • *Our previous works introduced Tsaap-Notes, a platform able to generate multiple choice test questions providing feedbacks. It reuses interactive questions asked during the lecture by the teacher, as well as the notes taken by students after the presentation of the results; these are used as the feedbacks integrated into the generated tests. In this paper we introduce a new process called N phases process and its implementation within Tsaap-Notes. This process aims on one hand at increasing substantially the number of contributions of students and, on the other hand, at significantly improving the quality of the feedbacks used in the tests. Our approach has been validated during a first experimentation.*

■ **KEYWORDS** • *Interactive questions, feedback, computer assisted assessment, peer assessment, peer instruction, N phases process.*

1. Introduction

Différents travaux (Miller, 2009 ; Ricketts et Wilks, 2002), montrent que les évaluations informatisées facilitent la mise en œuvre d'évaluations formatives fréquentes pour des cohortes comprenant un grand nombre d'étudiants. Aussi, la qualité du feedback proposé aux étudiants et relatif à leurs travaux ou à leurs résultats d'évaluation est une clé de succès pour l'amélioration de leur apprentissage (Black et Wiliam, 1998 ; Hattie et Jaeger, 1998 ; Higgins *et al.*, 2002). David Nicol et Debra Macfarlane-Dick (2006) préconisent ainsi l'utilisation de tests en ligne pour que le feedback puisse être mis à disposition des apprenants à n'importe quel instant, depuis n'importe quel lieu, aussi souvent que l'apprenant le souhaite.

Cependant, dans les systèmes généralistes fournis par les plates-formes d'enseignement en ligne, telles que Moodle, Chamilo, Claroline, Blackboard ou autres dispositifs compatibles avec la spécification IMS QTI (IMS Global Learning Consortium, 2012), le feedback proposé à l'étudiant est par défaut minimaliste : il comprend, pour chaque question, le score de l'étudiant ainsi que la réponse correcte. La saisie par l'enseignant du feedback complémentaire pour chaque question est consommatrice de temps si bien qu'elle n'est pas souvent réalisée.

Nous avons décrit dans Silvestre *et al.* (2014a) l'approche « *Note as feedback* » mise en œuvre dans notre plate-forme Tsaap-Notes (Silvestre *et al.*, 2014b). Tsaap-Notes permet à l'enseignant de poser des questions interactives pendant un cours en face à face. Les étudiants connectés à la plate-forme peuvent répondre aux questions posées par l'enseignant et prendre des notes sur chacune de ces questions suite à la présentation des résultats obtenus par l'ensemble des participants. Les questions gérées par Tsaap-Notes sont des questions décrites au format Moodle Gift, les rendant nativement compatibles avec la plate-forme Moodle. L'approche « *Note as feedback* » permet à l'enseignant de générer de manière automatique un fichier d'export compatible Moodle recyclant (1) les questions interactives posées par l'enseignant pendant le cours, et (2) les notes prises par les étudiants sur les différentes questions. Les notes prises par les étudiants sur une question donnée constituent le feedback présenté à l'étudiant durant son test de révision. Malgré les résultats encourageants obtenus lors des premières expérimentations, deux limites ont clairement été identifiées dans le dispositif « *Note as feedback* » :

(1) bien que la prise de notes augmente quantitativement lorsque les étudiants savent que leurs contributions seront la seule source de

feedback dans les tests de révision, relativement peu d'étudiants participent à l'activité de rédaction des explications nécessaires à la constitution du feedback ;

(2) toutes les notes saisies par les étudiants sur une question donnée sont récupérées de manière automatique pour constituer le feedback du futur test de révision, elles ne sont ni évaluées ni filtrées. Il n'y a donc pas de « contrôle qualité » du feedback intégré aux tests de révision.

Nous présentons dans cet article un nouveau processus baptisé « processus en N phases », ainsi que son implantation dans Tsaap-Notes. Le processus en N phases s'attaque directement aux deux limites évoquées ci-dessus : les activités mises en œuvre dans le processus maximisent, d'une part, la participation des étudiants à la tâche de rédaction d'une explication et, d'autre part, permettent l'évaluation par les étudiants des différentes explications. Tsaap-Notes intègre alors automatiquement les contributions les mieux notées par les étudiants dans les tests générés. La suite de l'article est organisée de la manière suivante : la section 2 présente les sources scientifiques qui ont inspiré la conception du processus en N phases ; la section 3 détaille le processus en utilisant la notation BPMN (*Business Process Model and Notation*) ; la section 4 présente l'implantation du processus dans Tsaap-Notes ; les résultats issus d'une première expérimentation réalisée auprès de deux groupes d'étudiants inscrits en Master Informatique à l'Université Paul Sabatier (Toulouse III) font l'objet de la section 5. Enfin, nous concluons et présentons les perspectives guidant nos travaux futurs dans la section 6.

2. Pourquoi le processus en N phases ?

La première limite révélée par les expérimentations du dispositif « *Note as feedback* » est la trop faible participation des étudiants à la prise de notes. En effet, les résultats présentés dans Silvestre *et al.* (2014b) indiquent que seuls 25 % des étudiants prennent effectivement des notes, alors qu'en moyenne 75 % des étudiants participent aux questions interactives (Silvestre *et al.*, 2014a). Si cette limite impacte la génération des tests de révision car les feedbacks constitués restent peu nombreux, elle questionne également le déroulement du cours en face à face. En effet, si trois quarts des étudiants sont prêts à répondre à une question à choix multiples, pourquoi ne sont-ils qu'un quart à participer à l'activité de rédaction d'une explication ?

Scouller (1998) rappelle que l'évaluation influence considérablement l'apprentissage : la manière d'évaluer un étudiant façonne la quantité de choses qu'il va apprendre, les contenus qu'il va apprendre, ainsi que la manière dont il va les apprendre. La plupart des étudiants concentrent leurs efforts dans l'acquisition des connaissances et des aptitudes cognitives sur lesquelles ils seront évalués. Biggs et Tang (2011) indiquent que les pédagogies dites actives insérant l'étudiant dans des activités cohérentes avec les objectifs pédagogiques d'une part, et avec les tâches d'évaluations d'autre part, favorisent une approche d'apprentissage en profondeur quel que soit le profil de l'apprenant. Les deux auteurs mettent en avant le rôle des évaluations formatives apportant à l'étudiant un feedback tout au long de son apprentissage lui permettant de savoir où il en est par rapport aux connaissances ou compétences visées. L'impact positif des systèmes de votes interactifs sur l'engagement des étudiants révélé par différentes études (Caldwell, 2007 ; Gauci et Dantas, 2009 ; Shaffer et Collura, 2009 ; Uhari *et al.*, 2003) peut s'expliquer en grande partie par sa fonction d'évaluation formative : pendant le cours, les étudiants constatent en temps réel, individuellement et par rapport au groupe, leur degré d'acquisition d'une connaissance ou d'une aptitude cognitive.

Pour augmenter le taux de participation à la rédaction d'explications sur une question posée pendant le cours, nous avons émis l'hypothèse que la tâche de rédaction des explications devait être insérée dans une activité fournissant un feedback immédiat à l'étudiant sur sa production. Le feedback proposé à chaque étudiant par le processus en N phases prend plusieurs formes, allant de la confrontation à des réponses alternatives jusqu'à la présentation du résultat de l'évaluation de sa contribution.

Une évaluation rapide d'un grand nombre d'explications ne peut pas être prise en charge par l'enseignant pendant le cours. Pour fournir rapidement à chaque étudiant une évaluation de son travail, nous avons opté pour l'évaluation par les pairs dont les vertus ont été mises en valeur dans différentes études (Boud *et al.* 2014 ; Dochy *et al.*, 1999 ; Topping, 1998).

La deuxième limite révélée par les expérimentations du dispositif « *Note as feedback* » est l'absence de filtrage dans la sélection des contributions des étudiants pour la constitution des feedbacks dans les tests de révision générés par Tsaap-Notes. L'évaluation des explications par les étudiants permet la mise en œuvre d'un algorithme ne sélectionnant que les contributions les mieux notées par les étudiants.

3. Le processus en N phases

Dans cette section, nous décrivons à l'aide de la notation BPMN 2.0 le processus capable d'intégrer, durant les cours en face à face, la tâche de rédaction d'explications. Ce processus assure l'obtention d'un feedback immédiat pour chaque étudiant sur sa production, ainsi que l'évaluation quasi-immédiate de l'ensemble des productions par les pairs.

3.1. La notation BPMN 2.0

Le langage BPMN 2.0 normalisé par l'OMG (OMG, 2011) est un langage visuel de description de modèles de processus métier qui vise deux objectifs principaux :

- obtenir un langage le plus lisible possible pour l'ensemble des parties prenantes (les personnes du métier, les informaticiens, etc.) ;
- fournir un langage standard de visualisation pour des processus métiers définis à l'aide d'un langage exécutable tel que WSBPEL (OASIS, 2007).

L'OMG indique que BPMN 2.0 regroupe l'ensemble des bonnes pratiques ayant émergé de l'utilisation de langages antérieurs (diagrammes d'activités UML, UML EDOC *Business Processes*, IDEF, ebXML BPSS, diagrammes *Activity-Decision Flow*, RosettaNet, etc.). D'autre part, BPMN commence à susciter la curiosité de chercheurs s'intéressant aux environnements informatiques pour l'apprentissage humain (EIAH) : Da Costa et Schneider (2015) invitent à l'expérimentation du langage pour décrire des processus correspondant à des situations d'apprentissage afin de mettre BPMN à l'épreuve dans ce champ de recherche. Les ambitions affichées de l'OMG sur BPMN 2.0 et la démarche proposée par Da Costa et Schneider ont orienté notre choix de BPMN 2.0 pour décrire le processus en N phases.

3.2. Le processus en N phases

La figure 1 fournit un aperçu du processus en N phases au format BPMN dans sa mise en œuvre nominale comprenant quatre phases.

La première phase illustrée par la figure 2 correspond à la pose de la question par l'enseignant, et au recueil des réponses formulées par les étudiants. Le processus est conçu de telle sorte que les réponses puissent contenir des éléments saisis sous forme de texte libre.

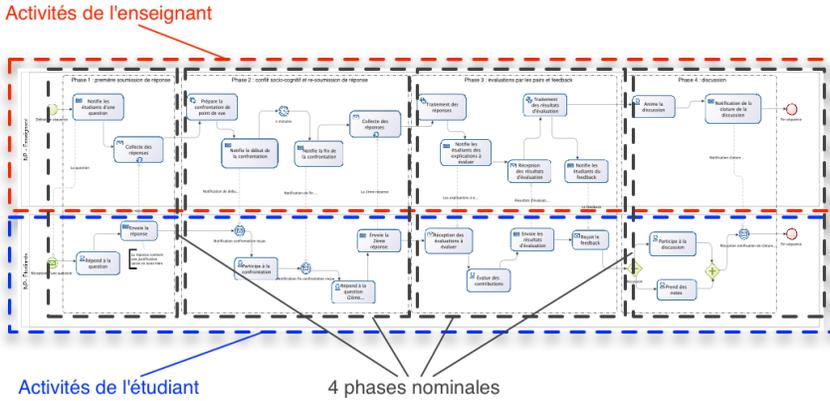


Figure 1 • Aperçu global du processus en N phases

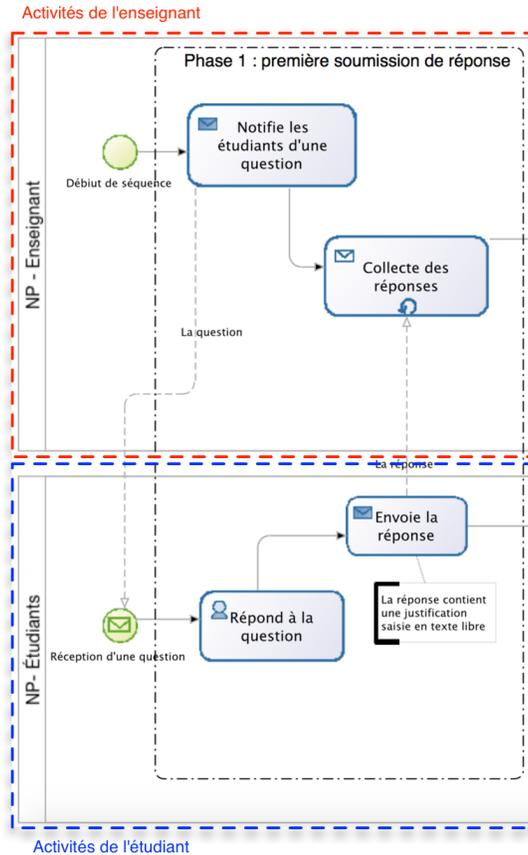


Figure 2 • Phase 1 du processus en N phases

La deuxième phase détaillée sur la figure 3 a été conçue pour intégrer explicitement la confrontation de points de vue au cœur du processus : à chaque étudiant est proposée une réponse différente de la sienne, et fournie par un autre étudiant. Le symbole  sur l'activité de préparation de la confrontation indique que la tâche est automatisée. Cette phase propose un premier feedback textuel à l'étudiant à travers la proposition d'une réponse alternative. L'étudiant confronte sa réponse à la réponse différente proposée par l'un de ses pairs. Il peut alors conforter ou modifier sa réponse.

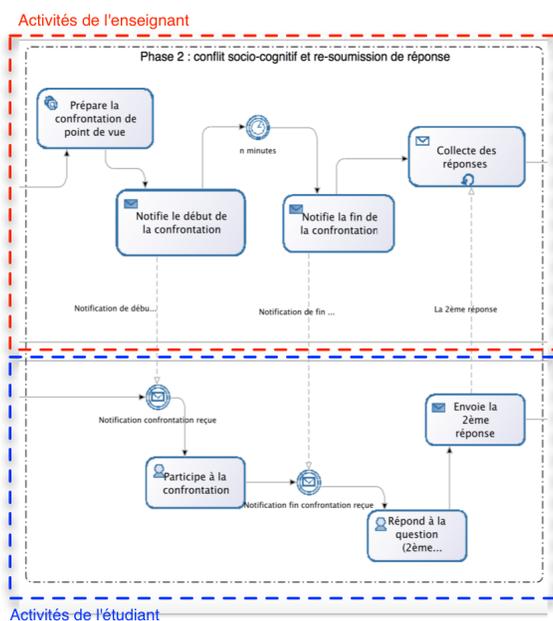


Figure 3 • Phase 2 du processus en N phases

Cette phase vise le même objectif que la phase de discussion proposée par Mazur dans son protocole d'« Instruction par les pairs » : augmenter le nombre de bonnes réponses grâce à la confrontation des points de vue entre pairs. En effet, le protocole de Mazur consiste, pour une même question, à sonder les étudiants 2 fois : une première fois après une phase de réflexion individuelle, puis une deuxième fois après que les étudiants ont été invités à convaincre leurs voisins du bien-fondé de la réponse qu'ils ont fournie à l'issue du premier sondage. Les résultats du deuxième sondage révèlent un plus grand nombre de bonnes réponses (Crouch et Mazur, 2001).

La troisième phase exposée sur la figure 4 est dédiée à l'évaluation par les pairs des réponses. L'activité d'évaluation des réponses textuelles est déléguée aux étudiants. Chaque étudiant reçoit un lot de réponses à évaluer ; les résultats des évaluations réalisées par les étudiants sont traités afin de fournir un feedback immédiat. Cette approche règle la question de l'évaluation en temps réel de réponses textuelles en parallélisant ce traitement auprès des différents étudiants.

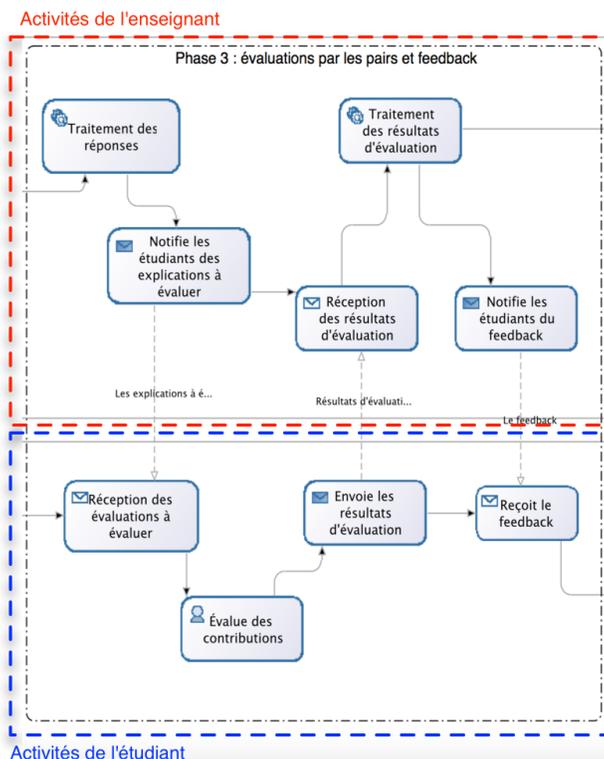


Figure 4 • Phase 3 du processus en N phases

La quatrième phase (figure 5) est la phase de discussion animée par l'enseignant suite à la présentation du feedback aux étudiants. Ces derniers participent à la discussion en prenant éventuellement des notes.

Une variante de la mise en œuvre du processus est la répétition ou la suppression de certaines phases. Par exemple, la seconde phase peut être répétée pour confronter chaque étudiant à différentes réponses alternatives, ou être supprimée dans le cas où les étudiants ont répondu à l'unani-

mité correctement à la question posée. Dans ce cas, le processus met en œuvre N phases où N dépend du nombre d'exécutions de chaque phase.

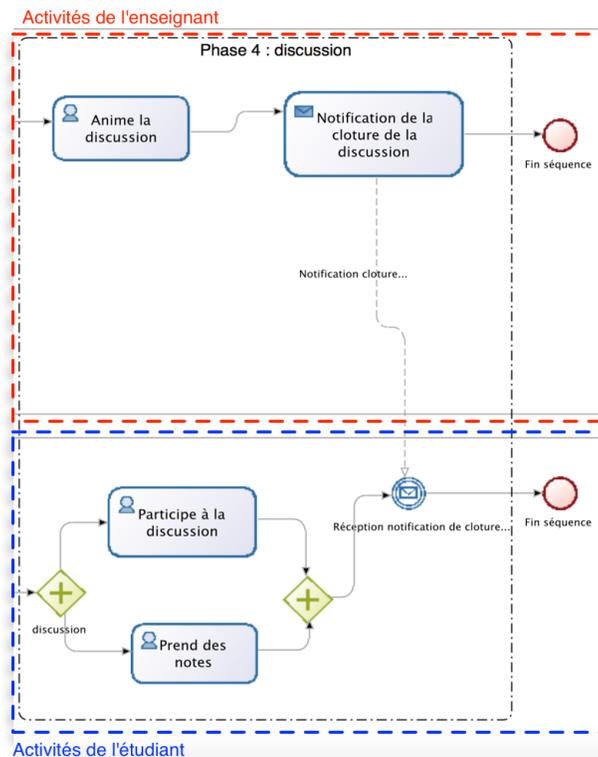


Figure 5 • Phase 4 du processus en N phases

4. Implantation du processus en N phases

Cette section présente le processus en N phases sous l'angle de son implantation dans Tsaap-Notes.

4.1. Du processus en N phases à Tsaap-Notes

La première phase consiste à proposer à l'étudiant un formulaire lui permettant de (1) de répondre à une question à choix multiples, (2) de saisir l'explication justifiant son choix, et (3) d'indiquer son degré de confiance en sa réponse. La figure 6 présente le formulaire tel qu'il a été implanté dans Tsaap-Notes. Notons que le degré de confiance indiqué par chaque étudiant est un des éléments clés de l'algorithme sous-tendant l'implantation de la deuxième phase.

@FranckSilvestre Master/Informatique/DL/2013_2015/DCLL/1 13 janvier 2015 15:32

La licence GPL et Javascript

Soit L une librairie javascript distribuée sous licence GPL . Soit E une entreprise qui modifie la librairie L pour en faire une version L2. L'entreprise E propose à des utilisateurs son application serveur qui utilise la version L2 de la librairie. Est ce que les utilisateurs peuvent exiger de récupérer le code source de la version L2 du logiciel ?

Oui
 Non

Give an explanation to your choice

What is the degree of confidence in your response (1: not confident to 5:very confident) ? 1

Submit

Reply Favorite Learn

Figure 6 • Formulaire correspondant à la première phase

@FranckSilvestre Master/Informatique/DL/2013_2015/DCLL/1 13 janvier 2015 15:32

Here is an alternative response to yours.
Please examine this response and then feel free to change your response, explanation or confidence degree if necessary.

Oui
Le javascript étant exécuté côté client, l'utilisateur a une copie, la contamination du GPL force donc la possibilité d'accès au code à l'utilisateur.

Réponse alternative

La licence GPL et Javascript

Soit L une librairie javascript distribuée sous licence GPL . Soit E une entreprise qui modifie la librairie L pour en faire une version L2. L'entreprise E propose à des utilisateurs son application serveur qui utilise la version L2 de la librairie. Est ce que les utilisateurs peuvent exiger de récupérer le code source de la version L2 du logiciel ?

Oui
 Non

Give an explanation to your choice

En tant qu'utilisateurs de l'application serveur, on se retrouve dans le même cas que précédemment, on n'a pas de copie du logiciel. On ne peut donc toujours pas exiger récupérer le code.

What is the degree of confidence in your response (1: not confident to 5:very confident) ? 3

Submit

Second formulaire de soumission : une chance de modifier la première réponse

Reply Favorite Learn

Figure 7 • Formulaire correspondant à la deuxième phase

La deuxième phase (*i.e.* la confrontation de points de vue et la resoumission d'une réponse) consiste à présenter à l'étudiant le même formulaire, mais cette fois pré-rempli avec les éléments de sa précédente réponse, et accompagné d'une réponse différente de la sienne qui a été fournie par un autre étudiant participant à la même session interactive.

La figure 7 illustre l'interface correspondant à cette phase au sein de Tsaap-Notes. Notons qu'aucune indication sur l'identité de l'auteur de la réponse alternative n'est présentée à l'étudiant afin d'écartier tout phénomène parasite d'influence relationnelle.

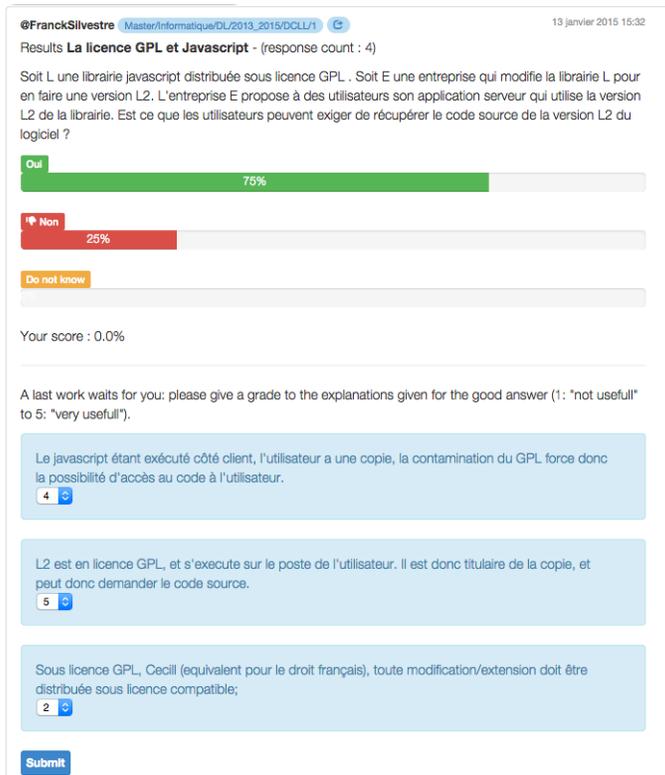


Figure 8 • Formulaire correspondant à la troisième phase

La troisième phase (*i.e.* évaluation par les pairs et feedback) illustrée par la figure 8 présente les résultats obtenus par l'ensemble des étudiants ayant participé au questionnaire ; chaque étudiant prend donc connaissance du score qu'il a obtenu à la question en cours. Dans cette même interface, chaque étudiant est invité à évaluer, à l'aide d'une échelle de Lickert graduée de 1 à 5, trois explications distinctes correspondant chacune à la bonne réponse. Durant cette phase, chaque étudiant est en situation d'évaluateur des productions provenant de ses pairs. L'algori-

thme utilisé pour la répartition des explications à évaluer a été conçu pour maximiser le nombre d'évaluateurs différents par explication.

Pour la quatrième phase (*i.e.* discussion), Tsaap-Notes calcule la note moyenne de chaque explication évaluée et présente les résultats du processus d'évaluation sous la forme d'une liste ordonnée par ordre décroissant des explications évaluées. La discussion est animée par l'enseignant. Les étudiants peuvent prendre des notes à l'aide de Tsaap-Notes au cours de la discussion.

La génération automatique du fichier d'export Moodle contenant les questions et feedbacks du test de révision a été modifiée afin que les feedbacks ne soient constitués, pour une question donnée, que des explications ayant une évaluation moyenne supérieure à 2,5.

4.2. L'algorithme d'association des réponses pour la confrontation de points de vue

La phase 2 du processus en N phases, dont l'objectif est de provoquer une confrontation de points de vue, nécessite la mise en œuvre d'un algorithme capable d'associer de manière pertinente deux réponses obtenues à l'issue de la phase 1. Le principe sous-tendant notre algorithme consiste à associer une réponse a_i à une réponse a_j lorsque les deux réponses ont fait l'objet de choix distincts et que l'explication associée à a_i est de taille minimale ; si plusieurs réponses peuvent être associées à a_j , l'algorithme privilégie la diversité dans l'association des réponses afin d'exposer des associations différentes aux apprenants, tout en favorisant les réponses ayant le plus fort degré de confiance.

Afin de présenter notre algorithme, nous proposons dans ce qui suit une formalisation du problème et des éléments de la solution retenue.

Définition 1. Une réponse a à une question à choix multiples est un quadruplet $a = (s ; d ; e ; r)$ où

- s est le score calculé en fonction de la réponse de l'étudiant ;
- d est le degré de confiance fourni par l'étudiant sur sa réponse ;
- e est le nombre de caractères composant l'explication fournie par l'étudiant.
- r est le rang de la réponse dans la liste des réponses collectées.

Dans la suite du document, q désigne une question à choix multiples.

Propriété 1. Soit A l'ensemble des réponses obtenues à q . La relation \geq sur A définie par

$$\forall (a_i = (s_i; d_i; e_i; r_i), a_k = (s_k; d_k; e_k; r_k)) \in A^2,$$

$$a_i \geq a_k \Leftrightarrow d_i > d_k \text{ ou } d_i = d_k \text{ et } r_i \geq r_k$$

est une relation d'ordre sur A permettant d'ordonner les réponses obtenues suivant leur degré de confiance.

Définition 2. L'ensemble des réponses correctes à q , noté A_c , est défini comme suit

$$A_c = \{a_i = (s_i; d_i; e_i; r_i) \in A | s_i = \text{MaxScore}\}$$

où MaxScore est le score maximum qu'un étudiant peut obtenir à une question q .

Définition 3. L'ensemble des réponses incorrectes à q , noté $\overline{A_c}$, est le complémentaire de A_c dans A .

Définition 4. L'ensemble des réponses à q pouvant être proposées pour constituer une confrontation de points de vue est l'ensemble

$$C = \{a_i = (s_i; d_i; e_i; r_i) \in A | e_i > \text{TailleMinExp}\}$$

où TailleMinExp est la taille minimum requise d'une explication pour qu'une réponse soit considérée comme associable lorsqu'il s'agit de déclencher une confrontation de points de vue.

Définition 5. L'ensemble des réponses correctes à q pouvant être proposées pour déclencher une confrontation de points de vue est l'ensemble $C_c = C \cap A_c$

Définition 6. L'ensemble des réponses incorrectes à q pouvant être proposées pour constituer une confrontation, noté $\overline{C_c}$, est le complémentaire de C_c dans C .

Définition 7.

Soit $A_c = \{a'_1, a'_2, \dots, a'_n\}$ où $\forall (i; j) \in \{1, \dots, n\}^2, i < j \Rightarrow a'_i > a'_j$

l'ensemble réordonné des réponses correctes à q .

Soit $\overline{C_c} = \{c'_1, c'_2, \dots, c'_p\}$ où $\forall (i; j) \in \{1, \dots, p\}^2, i < j \Rightarrow c'_i > c'_j$

l'ensemble réordonné des réponses incorrectes à q pouvant être proposées pour constituer une confrontation.

La fonction $f: A_c \longrightarrow \overline{C_c}$ est la fonction associant à tout élément $a'_i \in A_c$ l'élément $f(a'_i) = c'_{i \bmod p}$ où $i \bmod p$ est le reste de la division euclidienne de i par p .

Définition 8.

Soit $\overline{A}_c = \{a''_1, a''_2, \dots, a''_m\}$ où $\forall (i; j) \in \{1, \dots, m\}^2, i < j \Rightarrow a''_i > a''_j$

l'ensemble réordonné des réponses incorrectes à q , ordonnées suivant le degré de confiance.

Soit $C_c = \{c''_1, c''_2, \dots, c''_l\}$ où $\forall (i; j) \in \{1, \dots, l\}^2, i < j \Rightarrow c''_i > c''_j$

l'ensemble réordonné des réponses correctes à q pouvant être proposées pour constituer une confrontation, ordonnées suivant le degré de confiance.

La fonction $g: \overline{A}_c \rightarrow C_c$ est la fonction associant à tout élément $a''_i \in \overline{A}_c$ l'élément $g(a''_i) = c''_{i \bmod l}$.

À partir de ces définitions, la figure 9 illustre les différentes étapes de l'algorithme développé dans Tsap-Notes sur un exemple simplifié, où le 4e élément du quadruplet (rang) correspond à l'indice i de la réponse a_i .

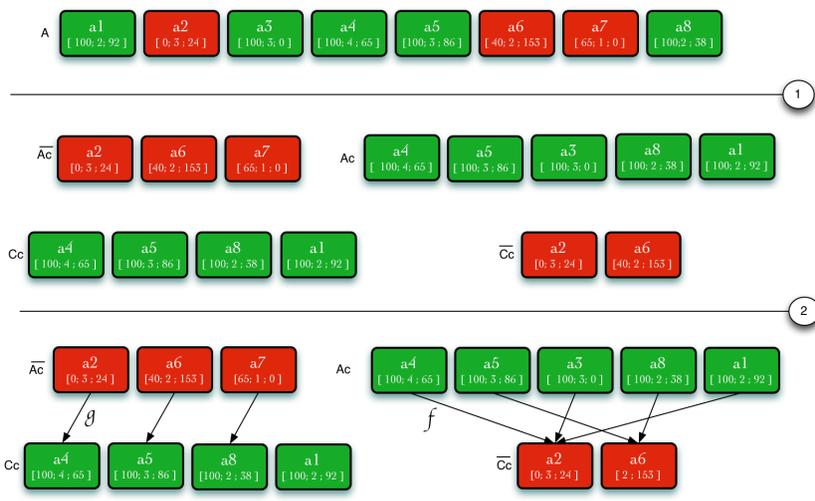


Figure 9 • Les étapes de l'algorithme d'association de réponses pour la constitution d'une confrontation de points de vue

Les réponses correctes et incorrectes sont respectivement représentées par les rectangles verts et rouges. La première étape prend en entrée l'ensemble A de toutes les réponses et crée les 4 ensembles A_c , \overline{A}_c , C_c et \overline{C}_c . L'ensemble A_c contient la liste des réponses correctes ordonnées suivant leur degré de confiance ; l'ensemble \overline{A}_c contient les réponses qui pourront être affectées aux réponses correctes pour la confrontation de points de

vue. De manière symétrique, l'ensemble $\overline{A_c}$ contient les réponses incorrectes; l'ensemble C_c contient les réponses qui pourront être affectées aux réponses incorrectes. Dans l'implémentation actuelle de Tsap-Notes le score maximum qui peut être obtenu à une question est toujours 100. En conséquence, le paramètre *MaxScore* a été fixé à cette même valeur. Le paramètre *TailleMinExp* correspondant à la taille minimum requise d'une explication pour qu'une réponse soit considérée comme associable pour déclencher une confrontation de points de vue a été fixé à 10.

La deuxième étape de l'algorithme récupère dans un tableau associatif le résultat de l'application de la fonction f (respectivement g) sur tous les éléments de l'ensemble A_c (respectivement $\overline{A_c}$). Ce tableau associatif stocke donc, pour chaque réponse, la réponse qui lui est associée dans la phase de confrontation de points de vue.

L'algorithme a pu être testé en situation auprès de deux groupes d'étudiants. La section suivante présente le détail de l'expérimentation.

5. Résultats issus de la première expérimentation

L'expérimentation avait pour objectif la validation des hypothèses suivantes :

- le fait de proposer une évaluation en temps réel des réponses textuelles engage fortement les étudiants dans les tâches de rédaction de texte (rédaction des explications);
- la mise en situation de confrontation de points de vue a un impact positif sur la quantité de bonnes réponses fournies à la deuxième soumission ;
- l'évaluation par les pairs suscite un fort engagement des étudiants et est valide vis-à-vis de l'usage qui en est fait (i.e., la sélection du feedback dans les tests de révision);
- les étudiants perçoivent Tsap-Notes comme ayant une bonne utilisabilité.

L'implantation du processus en N phases fourni par Tsap-Notes a été expérimentée avec deux groupes d'étudiants dans le cadre de deux cours de Master 1 mention Informatique. L'expérimentation s'est déroulée sur 3 sessions de 2 heures pendant lesquelles 5 questions mobilisant le processus en N phases, ont été posées aux étudiants.

5.1. Résultats quantitatifs

Le tableau 1 montre un taux moyen de participation des étudiants proche de 88 % sur l'activité de soumission d'une réponse à une question ; ce taux de participation élevé est cohérent avec les autres études relatives aux systèmes de votes interactifs.

Tableau 1 • Synthèse des résultats obtenus sur deux séances

Code cours	Code question	Nombre de présents	Nombre de réponses soumises	Taux de participations
DCLL	DCLL-1	51	37	72,5%
DCLL	DCLL-2	51	41	80,4%
MA	MA-1	34	34	100%
MA	MA-2	34	34	100%
MA	MA-3	34	32	94,1%
Moyenne	-	40,8	35,6	87,8%

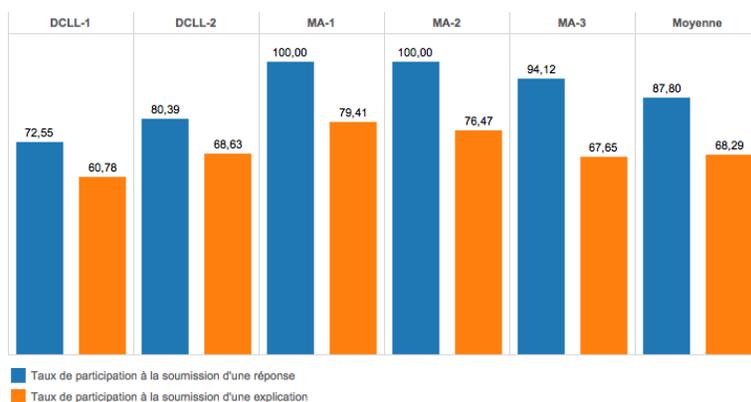


Figure 10 • Les taux de participation sur les soumissions de réponses et d'explications

Un indicateur remarquable est le taux de participation des étudiants dans la soumission d'une explication : la figure 10 montre un taux moyen de participation de 68,29 % pour l'activité de soumission d'une explication pour justifier leur réponse. Même si le taux de participation est en dessous de celui observé sur la soumission d'une réponse (87,8 %), il représente une très importante augmentation du taux de participation à une activité exigeant la saisie de texte libre (au mieux 25 % constaté dans (Silvestre *et al.*, 2014a)).

La figure 11 révèle, d'autre part, qu'entre la première phase et la deuxième, en moyenne 30 % des étudiants ayant soumis une mauvaise réponse en première intention ont modifié leur réponse et leur explication. Ceux-ci ont donc pleinement bénéficié de la confrontation de points de vue, qui les a amenés à modifier leur croyance sur ce qui était vrai ou faux. Cette augmentation du nombre de réponses correctes est similaire à celle constatée dans les mises en œuvre du protocole d'« Instruction par les pairs ». En effet, dans leur étude de 2001 (Crouch et Mazur, 2001), les auteurs recensent une augmentation moyenne du nombre de bonnes réponses de 32 % entre le premier et le deuxième sondage.

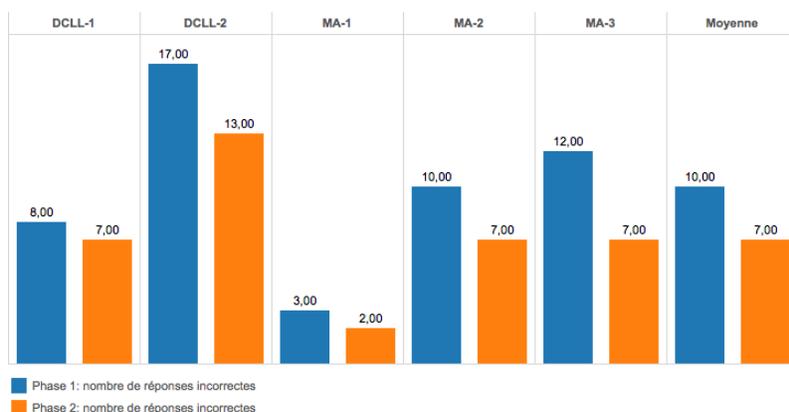


Figure 11 • Nombre de mauvaises réponses en phase 1 et en phase 2

Enfin, la figure 12 indique que 75,6 % des étudiants ont participé à la tâche d'évaluation de trois explications fournies par les pairs. Cette tâche est proposée en phase 3 du processus et n'est pas obligatoire. Le nombre moyen d'évaluateurs par explication est de 4,17 pour une moyenne de 40,8 étudiants présents et de 22,8 explications à évaluer. L'écart-type moyen des notes distribuées par les étudiants (comprises entre 1 et 5) sur les explications est de 0.98. Si cet écart-type moyen est non négligeable puisqu'il signifie que les notes d'un étudiant à l'autre varient de 20 % en moyenne, il nous semble raisonnable pour l'usage que nous en faisons, à savoir la sélection des explications les mieux notées pour générer le feedback dans les tests de révision.

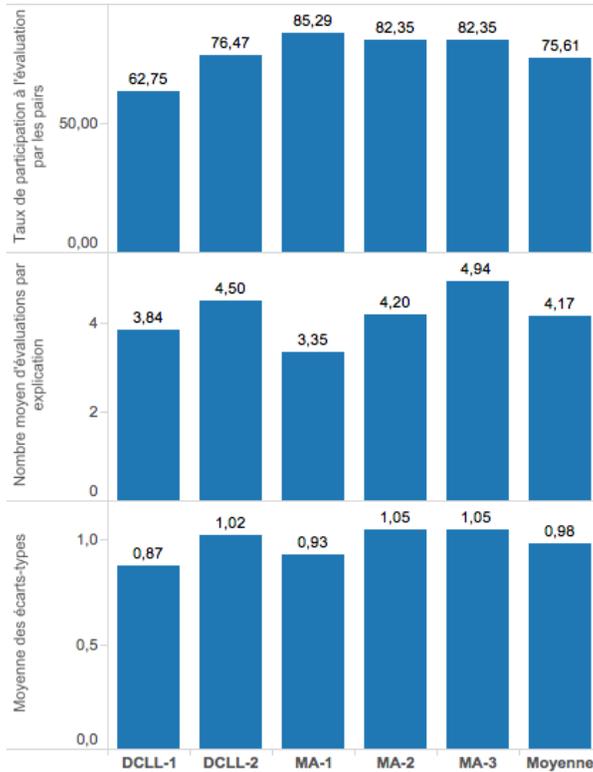


Figure 12 • Taux de participation à la tâche d'évaluation par les pair

5.1. Résultats qualitatifs

Tous les étudiants ayant participé à l'expérimentation ont été contactés pour répondre au questionnaire *System Usability Scale* (SUS) (Brooke, 1996). Le SUS s'est imposé comme un standard de l'industrie pour évaluer de manière simple et efficace l'utilisabilité d'un système. Le SUS est composé de dix items dont l'évaluation attendue se répartit sur une échelle de Lickert à cinq niveaux: le premier niveau correspond à l'évaluation « Pas du tout d'accord » et le cinquième niveau correspond à l'évaluation « Tout à fait d'accord ». Le tableau 2 présente les dix items du SUS dans leur version originale en anglais et dans leur version traduite en français (version proposée aux étudiants de notre expérimentation).

Tableau 2 • Les items du System Usability Scale

	Version originale	Version française
1	<i>I think that I would like to use this system frequently.</i>	Je pense que je voudrais utiliser ce système fréquemment.
2	<i>I found the system unnecessarily complex.</i>	J'ai trouvé le système inutilement complexe.
3	<i>I thought the system was easy to use.</i>	J'ai pensé que le système était facile à utiliser.
4	<i>I think that I would need the support of a technical person to be able to use this system.</i>	Je pense que j'aurais besoin du support d'une personne technique pour être capable d'utiliser le système.
5	<i>I found the various functions in this system were well integrated.</i>	J'ai trouvé que les différentes fonctions du système étaient bien intégrées.
6	<i>I thought there was too much inconsistency in this system.</i>	J'ai pensé qu'il y avait trop de d'incohérence dans ce système.
7	<i>I would imagine that most people would learn to use this system very quickly.</i>	J'imagine que la plupart des gens apprendraient à utiliser le système très rapidement.
8	<i>I found the system very cumbersome to use.</i>	J'ai trouvé le système très lourd à utiliser.
9	<i>I felt very confident using the system.</i>	Je me suis senti à l'aise en utilisant le système.
10	<i>I needed to learn a lot of things before I could get going with this system.</i>	J'ai eu besoin d'apprendre beaucoup de choses avant de pouvoir utiliser le système.

Pour calculer le score SUS, il faut commencer par additionner le score de chaque item. Pour les items 1, 3, 5, 7 et 9, le score est le niveau de l'échelle de Lickert moins 1. Pour les items 2, 4, 6, 8 et 10, le score est 5 moins le niveau de l'échelle de Lickert. Chaque score d'item varie de 0 à 4, la somme des scores des 10 items donne un résultat sur 40. Il faut donc multiplier le résultat obtenu par 2,5 pour obtenir le score SUS sur 100. Les recherches sur l'utilisation du SUS (Brooke, 2013) ont conduit à considérer un système comme étant en dessous de la moyenne quand il obtient un score en dessous de 68. Le SUS permet de mesurer l'utilisabilité perçue d'un système avec un nombre restreint de participants au sondage : entre 8 et 12 (Tullis et Stetson, 2004). Nous avons recueilli 43 réponses au questionnaire ce qui est donc tout à fait recevable pour la mesure de l'utilisabilité de Tsaap-Notes. Le score SUS obtenu pour Tsaap-Notes est de 77,8. Il peut donc être considéré comme un système perçu comme ayant une bonne utilisabilité par les étudiants (Bangor *et al.*, 2009). Nous n'avons pas pu proposer le questionnaire aux enseignants car actuellement, en dehors du concepteur de la plateforme, un seul enseignant utilise Tsaap-Notes.

5.2. Synthèse des résultats

Les résultats issus de l'expérimentation montrent que les étudiants s'engagent largement dans les tâches de rédaction d'explications. La mise en situation de confrontation de points de vue a un impact positif sur la quantité de réponses correctes fournies après la deuxième soumission. L'évaluation par les pairs engage la grande majorité des étudiants, permettant ainsi la sélection des meilleures explications en vue de leur recyclage dans les tests de révision. Enfin, les réponses apportées par les étudiants au questionnaire SUS montrent que le système Tsaap-Notes est perçu comme ayant une bonne utilisabilité.

6. Conclusion et perspectives

La plate-forme Tsaap-Notes permet la génération semi-automatique de questionnaires à choix multiples pourvus de feedbacks en recyclant (1) les questions interactives posées pendant le cours par l'enseignant, et (2) les notes prises par les étudiants pour constituer les feedbacks présentés aux étudiants durant les tests. Dans cet article, nous avons présenté le processus en N phases que nous avons conçu et décrit son implantation dans Tsaap-Notes. L'objectif de ce processus consiste d'une part à augmenter le nombre de contributions des étudiants et, d'autre part, à améliorer de manière significative la qualité des feedbacks restitués au sein des tests générés. Le processus présenté découpe la soumission d'une réponse en plusieurs phases permettant la collecte, la confrontation et l'évaluation des explications textuelles fournies par les étudiants. Les explications les mieux évaluées par les pairs accompagnant les réponses correctes sont celles qui sont intégrées comme feedback lors de la génération des tests par Tsaap-Notes, améliorant ainsi la qualité globale des évaluations formatives. Les résultats d'une première expérimentation menée auprès de deux groupes d'étudiants inscrits en Master Informatique sont très encourageants : ils montrent une augmentation accrue de la participation des étudiants à la tâche de rédaction d'explications, permettant ainsi à l'algorithme de filtrage des contributions pour la constitution de *feedback* de s'appliquer sur un volume de données significatif.

Au cœur de l'implantation du processus en N phases, l'algorithme détaillé dans la section 3 permet d'associer, à une réponse donnée, une autre réponse dans l'objectif de déclencher une éventuelle confrontation de points de vue. Aujourd'hui, cet algorithme ne prend en considération que très peu d'informations relatives aux apprenants : seuls le score

obtenu à la question courante, la taille de l'explication et son degré de confiance sur la réponse fournie sont utilisés pour déterminer une réponse alternative. Les résultats obtenus sur les différentes questions, l'aptitude à convaincre (i.e. à faire changer d'avis un autre étudiant suite à la confrontation de points de vue) sont autant de données qui pourraient être exploitées pour améliorer l'algorithme.

D'autre part, ce dispositif est aujourd'hui mis en œuvre dans le contexte de cours dispensés en face à face. Il nous paraît intéressant d'adapter et de tester le dispositif dans le cadre d'un Cours Ouvert Massif en Ligne (MOOC). En effet, les MOOC intègrent déjà des dispositifs tels que les forums de discussion permettant aux apprenants de s'accompagner mutuellement. Cependant, ce type de dispositif ne garantit pas la participation de tous les apprenants et donc ne garantit pas que tous les apprenants bénéficient d'un accompagnement par les pairs. La mise en situation de confrontation de points de vue telle qu'elle est proposée dans le processus en N phases tire parti des productions de l'ensemble des apprenants, pour l'ensemble des apprenants. Nous envisageons donc, dans la suite de nos travaux, d'adapter si nécessaire et d'expérimenter notre dispositif dans le cadre de MOOC, avant d'améliorer l'algorithme en exploitant la masse importante de données qui peut être recueillie dans un tel contexte d'apprentissage.

BIBLIOGRAPHIE

BANGOR, A., KORTUM, P. et MILLER, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123.

BIGGS, J. et TANG, C. (2011). *Teaching for quality learning at university: What the student does (4^e éd.)*. McGraw-Hill Education (UK).

BLACK, P. et WILLIAM, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74, doi:10.1080/0969595980050102

BOUD, D., COHEN, R. et SAMPSON, J. (2014). *Peer learning in higher education: Learning from and with each other*. Routledge.

BROOKE, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189-194

BROOKE, J. (2013). SUS : a retrospective. *Journal of Usability Studies*, 8(2), 29-40.

CALDWELL, J. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education*, 6, 9-20. Repéré à <http://www.lifescied.org/content/6/1/9.short>

CROUCH, C. H. et MAZUR, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics*, 69(9), 970-977.

DA COSTA, J. et SCHNEIDER, D. K. (2015). Modélisation et implémentation de dispositifs pédagogiques avec BPMN 2.0. Dans Actes de la 7ème conférence sur les *Environnements Informatiques pour l'Apprentissage Humain (EIAH 2015)* (p. 282-287). Repéré à http://atief.fr/sitesConf/eiah2015/uploads/Actes_EIAH2015.pdf

DOCHY, F., SEGERS, M. et SLUIJSMANS, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, 24(3), 331-350.

GAUCI, S. et DANTAS, A. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, 33(1), 60-71. Repéré à <http://advan.physiology.org/content/33/1/60?rss=1&ssource=mfr>

HATTIE, J. et JAEGER, R. (1998). Assessment and classroom learning: A deductive approach. *Assessment in Education: principles, policy & practice*, 5(1), 111-122.

HIGGINS, R., HARTLEY, P. et SKELTON, A. (2002). The conscientious consumer: reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1). doi:10.1080/0307507012009936

IMS Global Learning Consortium. (2012). *IMS Question & Test Interoperability Specification*. Repéré à <https://www.imsglobal.org/question/index.html>

MILLER, T. (2009). Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, 34(2), 181-192.

NICOL, D. et MACFARLANE-DICK, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218.

OASIS W. (2007). *Web Services Business Process Execution Language Version 2.0*. Repéré à <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>

OMG, B. (2011). *Business Process Model And Notation (BPMN) Version 2.0*. Repéré à <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>

RICKETTS, C. et WILKS, S. (2002). Improving student performance through computer-based assessment: Insights from recent research. *Assessment & Evaluation in Higher Education*, 27(5), 475-479. doi:10.1080/026029302200000934

SCOLLER, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*. 35(4), 453-472. Repéré à <http://link.springer.com/article/10.1023/A:1003196224280>

SHAFFER, D. et COLLURA, M. (2009). Evaluating the effectiveness of a personal response system in the classroom. *Teaching of Psychology*, 36(4), 273-277.

SILVESTRE, F., VIDAL, P. et BROISIN, J. (2014a). Génération semi-automatique de tests d'auto-évaluation pourvus de feedback résultant de la prise de notes collaborative. Dans *Actes du Colloque Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2014)* (p. 144-155). Repéré à <https://hal.inria.fr/hal-01187103/file/actes-TICE.pdf>

SILVESTRE, F., VIDAL, P. et BROISIN, J. (2014b). Tsaap-Notes--An Open Micro-blogging Tool for Collaborative Notetaking during Face-to-Face Lectures. Dans *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT 2014)* (p. 39-43). Repéré à http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6901393

TOPPING, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, 68(3), 249-276.

TULLIS, T. S. et STETSON, J. N. (2004). A comparison of questionnaires for assessing website usability. Dans *Proceedings of the Usability Professional Association Conference*, 1-12.

UHARI, M., RENKO, M. et SOINI, H. (2003). Experiences of using an interactive audience response system in lectures. *BMC Medical Education*, 3(1), 12.