



HAL
open science

Qualification semi-automatique de contributions d'apprenants pour l'intégration de feedbacks de qualité dans les tests en ligne

Franck Silvestre, Philippe Vidal, Julien Broisin

► **To cite this version:**

Franck Silvestre, Philippe Vidal, Julien Broisin. Qualification semi-automatique de contributions d'apprenants pour l'intégration de feedbacks de qualité dans les tests en ligne. 7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2015), Jun 2015, Agadir, Maroc. pp.210-221. hal-01405950v1

HAL Id: hal-01405950

<https://hal.science/hal-01405950v1>

Submitted on 30 Nov 2016 (v1), last revised 3 Oct 2016 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Qualification semi-automatique de contributions d'apprenants pour l'intégration de *feedbacks* de qualité dans les tests en ligne

Franck Silvestre, Philippe Vidal, Julien Broisin
IRIT, Université Toulouse III, 118 route de Narbonne, F-31062 TOULOUSE CEDEX 9
{franck.silvestre.philippe.vidal.julien.broisin}@irit.fr

Résumé. Nos travaux précédents ont introduit la plateforme Tsaap-Notes dédiée à 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 présentons un nouveau dispositif (NP-Q) qui vise à augmenter considérablement le nombre de contributions des étudiants afin d'améliorer de manière significative la qualité des *feedbacks* restitués au sein des tests générés par Tsaap-Notes. Ce dispositif repose sur le découpage en plusieurs phases du processus de soumission d'une réponse à une question, ainsi que sur un algorithme d'identification et de filtrage des contributions à inclure dans les tests. Notre approche a été validée par une première expérimentation sur un groupe de 51 étudiants inscrits en Master Informatique.

Mots-clés. Questions interactives, qualité du *feedback*, tests informatisés, évaluation par les pairs, conflit sociocognitif.

Abstract. Our previous works have introduced the Tsaap-Notes platform dedicated to the semi-automatic generation of multiple choice test questions providing *feedbacks*. Tsaap-Notes reuses interactive questions asked during the lecture by the teacher; the notes taken by the students after the presentation of the results are used as the *feedbacks* integrated in the generated tests. In this paper we present a new feature of Tsaap-notes called NP-Q which aims at increasing substantially the number of contributions of students in order to significantly improve the quality of the *feedbacks* used in the tests. This feature proposes the split of the submission of an answer in several phases to harvest explanations given by the students. Then Tsaap-Notes uses an algorithm to filter the best contributions to be integrated as *feedbacks* in the tests. Our approach has been validated on a first experimentation on a group of 51 master's students during a computer science course.

Keywords. Interactive questions, *feedback* quality, computer assisted assessment, peer assessment, socio-cognitive conflict.

1 Introduction

Différents travaux [12, 10] 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 [2, 7, 8]. David Nicol et Debra Macfarlane-Dick [11] 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 (tels que Moodle, Chamilo, Claroline, Blackboard ou d'autres dispositifs compatibles avec la spécification IMS QTI [9]), 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 [15] l'approche « Note as feedback » mise en œuvre dans notre plateforme Tsaap-Notes [16]. Tsaap-Notes permet à l'enseignant de poser des questions interactives pendant son cours en face à face. Les étudiants connectés à la plateforme 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 plateforme 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) 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.

(2) 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*.

Nous présentons dans cet article un nouveau dispositif de gestion de questions interactives baptisé « question à soumission de réponse en n-phases » (NP-Q) s'attaquant directement aux deux limites évoquées ci-dessus : les interactions mises en œuvre dans les NP-Q maximisent d'une part la participation des étudiants à l'activité de rédaction d'une explication, et d'autre part permettent l'évaluation par les étudiants des différentes explications. Tsaap-Notes intègre alors automatiquement les

Qualification semi-automatique de contributions d'apprenants pour l'intégration de *feedbacks* de qualité dans les tests en ligne

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 inspirées la conception du dispositif NP-Q ; la section 3 présente la mise en œuvre des NP-Q dans Tsaap-Notes, alors que les résultats issus d'une première expérimentation réalisée auprès d'un groupe de 51 étudiants inscrits en Master Informatique à l'Université Paul Sabatier (Toulouse III) font l'objet de la section 4. Enfin, nous concluons et présentons les perspectives guidant nos travaux futurs dans la section 5.

2 Pourquoi le dispositif NP-Q ?

La première limite révélée par les expérimentations du dispositif « Note as feedback » de Tsaap-Notes est la trop faible participation des étudiants à la prise de notes. En effet, dans [16] les résultats présentés indiquent que seuls 25% des étudiants prennent effectivement des notes, alors qu'en moyenne 75% des étudiants participent aux questions interactives [15]. 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 rappelle que l'évaluation influence considérablement l'apprentissage [13] : 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. Dans [1], Biggs et Tang 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 quelque soit le profil de l'apprenant. Les deux auteurs mettent en avant le rôle des évaluations formatives apportant à l'étudiant le *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 [4, 6, 14, 18] 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 dispositif NP-Q prend plusieurs formes, allant de la confrontation à des réponses alternatives pouvant générer un conflit sociocognitif à la présentation du résultat de l'évaluation de sa contribution. Les différentes phases du dispositif NP-Q ont été conçues pour favoriser un apprentissage réflexif.

Une évaluation rapide d'un grand nombre d'explications ne peut pas être prise en charge par l'enseignant pendant le cours. Pour fournir à chaque étudiant une évaluation de son travail rapidement, nous avons opté pour la mise en œuvre d'un dispositif d'évaluation par les pairs dont les vertus ont été mises en valeur dans différentes études [3, 5, 17].

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 notes par les étudiants permet la mise en œuvre d'un algorithme ne sélectionnant que les contributions les mieux notées par les étudiants.

3 Conception et mise en œuvre des NP-Q

Dans cette section, nous décrivons le dispositif conçu et mis en œuvre dans le cadre de nos recherches pour intégrer la tâche de rédaction des explications dans une activité permettant (1) l'obtention d'un *feedback* immédiat pour chaque étudiant sur sa production, et (2) l'évaluation rapide de l'ensemble des productions par les pairs.

3.1 Principes généraux

L'activité proposée aux étudiants pour les inciter à participer à la rédaction d'une explication relative à une question à choix multiples consiste à découper la soumission de la réponse en plusieurs phases.

The image shows a screenshot of a questionnaire form. At the top, it says '@FranckSilvestre Master/Informatique/OL/2013_2015/DCL/1' and '13 janvier 2015 15:32'. The title is 'La licence GPL et Javascript'. The question is: '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 ?'. There are two radio buttons: 'Oui' and 'Non'. Below them is a text area with the prompt 'Give an explanation to your choice'. At the bottom, there is a confidence scale: 'What is the degree of confidence in your response (1: not confident to 5:very confident) ?' with a dropdown menu showing '1'. A blue 'Submit' button is on the left, and 'Reply', 'Favorite', and 'Learn' icons are on the right.

Fig. 1. Formulaire correspondant à la première phase de soumission d'une réponse à une NP-Q

La première phase consiste à proposer à l'étudiant un formulaire permettant (1) de répondre à la question à choix multiples, (2) de saisir l'explication justifiant son choix, et (3) d'indiquer son degré de confiance en sa réponse. Fig. 1 présente le formulaire tel qu'il a été implémenté dans Tsaap-Notes. Dès cette première étape l'étudiant est mis en situation d'apprentissage réflexif : en indiquant le degré de

Qualification semi-automatique de contributions d'apprenants pour l'intégration de *feedbacks* de qualité dans les tests en ligne

confiance en sa réponse, l'étudiant est amené à réfléchir sur ses propres connaissances et/ou compétences. Cet indice est un des éléments clés de l'algorithme sous-tendant la deuxième phase.

La deuxième phase de soumission 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. Fig. 2 illustre l'interface correspondant à cette phase au sein de Tsaap-Notes. Cette phase propose un premier *feedback* à l'étudiant à travers la proposition d'une réponse et d'une explication alternatives. L'étudiant est alors mis en situation de conflit sociocognitif : il doit confronter sa réponse à la réponse différente proposée par l'un de ses pairs. L'étudiant peut alors conforter son choix et son explication, ou modifier sa réponse. Notons qu'aucune indication sur l'identité de l'auteur de la réponse alternative n'est présentée à l'étudiant afin d'écarter tout phénomène parasite d'influence relationnelle.

The screenshot shows a forum post from user @FranckSilvestre. The post contains an alternative response to a question, followed by a quiz question about the GPL license and JavaScript. The quiz question asks if users can demand the source code of a version L2 of a software library. The user has selected 'Non' (No) and provided an explanation: '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.' The confidence level is set to 3 out of 5. The post also includes a 'Submit' button and social media icons for Reply, Favorite, and Learn.

Fig. 2. Formulaire correspondant à la deuxième phase de soumission d'une réponse à une NP-Q

La troisième et dernière phase illustrée par Fig.3 présente les résultats obtenus par l'ensemble des étudiants ayant participé au questionnaire (la barre verte indique le taux de bonnes réponses, la barre rouge indique le taux de mauvaises réponses) ; chaque étudiant prend connaissance du score qu'il a obtenu à la question en cours ; il peut se situer par rapport à l'ensemble des étudiants en consultant les taux de bonnes et mauvaises réponses. 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, et chaque explication correspondant à une réponse correcte est évaluée au moins trois fois.

Une fois cette phase terminée, 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 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.



Fig. 3. Interface correspondant à la troisième phase de soumission d'une réponse à une NP-Q

3.2 Focus sur l'algorithme d'association de réponses pour conflit sociocognitif

La phase 2 du dispositif NP-Q, dont l'objectif est de générer un conflit sociocognitif, nécessite la mise en œuvre d'un algorithme capable d'associer de manière pertinente deux réponses obtenues à l'issue de la phase 1. Alors le principe sous-tendant notre algorithme consiste à associer une réponse a_i à une réponse a_j lorsque les deux

Qualification semi-automatique de contributions d'apprenants pour l'intégration de *feedbacks* de qualité dans les tests en ligne

réponses ont fait l'objet de choix distincts et que l'explication associée à a_i est de taille raisonnable ; 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 triplet

$$a = (s; d; e) \text{ où}$$

- 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.

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

$$\forall (a_i = (s_i; d_i; e_i), a_j = (s_j; d_j; e_j)) \in A^2, a_i \geq a_j \Leftrightarrow d_i \geq d_j$$

est une relation d'ordre sur A .

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

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

où MaxScore est le score maximum qu'un étudiant peut obtenir à une réponse 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 un conflit sociocognitif est l'ensemble

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

où TailleMinExp est la taille minimum requise d'une explication pour qu'une réponse soit considérée comme associable lorsqu'il s'agit de générer un conflit sociocognitif.

Définition 5. L'ensemble des réponses correctes à q pouvant être proposées pour générer un conflit sociocognitif est l'ensemble

$$C_c = C \cap A_c$$

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

Définition 7. Soit $A_c = \{a_1, a_2, a_3, \dots, a_n\}$ où $\forall (i, j) \in \{1, \dots, n\} \quad i < j \Rightarrow a_i \geq a_j$ l'ensemble ordonné des réponses correctes à q .

Soit $\overline{C_c} = \{c_1, c_2, c_3, \dots, c_p\}$ où $\forall (i, j) \in \{1, \dots, p\} \quad i < j \Rightarrow c_i \geq c_j$ l'ensemble ordonné des réponses incorrectes à q pouvant être proposées pour constituer un conflit.

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 $\overline{A_c} = \{a_1, a_2, a_3, \dots, a_n\}$ où $\forall (i, j) \in \{1, \dots, n\} \quad i < j \Rightarrow a_i \geq a_j$ l'ensemble ordonné des réponses incorrectes à q .

Soit $C_c = \{c_1, c_2, c_3, \dots, c_p\}$ où $\forall (i, j) \in \{1, \dots, p\} \quad i < j \Rightarrow c_i \geq c_j$ l'ensemble ordonné des réponses correctes à q pouvant être proposées pour constituer un conflit.

La fonction $g: \overline{A_c} \longrightarrow C_c$ est la fonction associant à tout élément $a_i \in \overline{A_c}$ l'élément $g(a_i) = c_{i \bmod p}$.

À partir de ces définitions, Fig. 4 illustre les différentes étapes de l'algorithme développé dans Tsaaq-Notes.

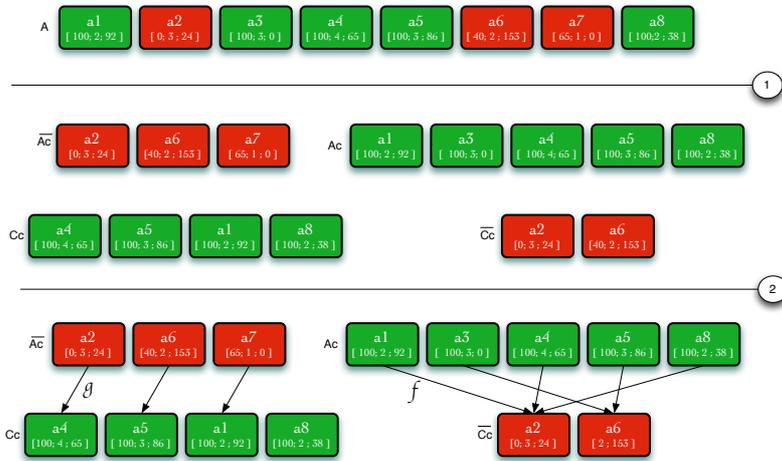


Fig. 4. Les étapes de l'algorithme d'association de réponses pour la constitution d'un conflit sociocognitif.

La première étape prend en entrée l'ensemble A et crée les 4 ensembles $A_c, \overline{A_c}, C_c$ et $\overline{C_c}$. Ces ensembles sont implémentés sous la forme de listes dans

Qualification semi-automatique de contributions d'apprenants pour l'intégration de *feedbacks* de qualité dans les tests en ligne

lesquelles les éléments sont ordonnés selon la relation d'ordre introduite par la Propriété 1. Puisque dans l'implémentation actuelle de Tsaap-Notes le score maximum qui peut être obtenu à une question est toujours 100, 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 générer un conflit sociocognitif a été fixé à 10. La deuxième étape de l'algorithme récupère dans un tableau associatif le résultat de l'application de la fonctions f (respectivement g) sur tous les éléments de l'ensemble A_c (respectivement $\overline{A_c}$). L'algorithme a pu être testé en situation auprès d'un groupe de 51 étudiants. La section suivante présente le détail de l'expérimentation.

4. Résultats issus de la première expérimentation

Le dispositif NP-Q fourni par Tsaap-Notes a été expérimenté sur un groupe de 51 étudiants inscrits en 1^{ère} année de Master Informatique dans le cadre de l'unité d'enseignement intitulée « Développement collaboratif et Logiciels libres ». L'expérimentation a été menée sur deux séances de 2 heures où 3 questions à soumissions en N phases ont été posées aux étudiants. Le tableau 1 montre les statistiques obtenues au cours des deux premières phases de soumission.

Tableau 1. Synthèse des résultats obtenus sur deux séances.

	Question 1		Question 2		Question 3	
	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
Nombre de réponses soumises	26	24	37	34	42	42
Nombre d'explications fournies	22	22	32	31	34	35
Nombre de réponses correctes	6	9	29	27	24	29
Nombre d'explications correctes	6	9	25	25	22	26

Le premier constat notable à la lecture du tableau est que la participation des étudiants n'a cessé d'augmenter au fil des questions. Le taux de soumission de réponses est passé de 49% à la première question à 82% à la troisième question. La baisse du nombre de réponses concernant les deux premières questions entre la première et la seconde phase s'explique sans doute par le fait que certains étudiants ne souhaitant pas modifier leur réponse entre les deux phases ne pensaient pas avoir à soumettre de nouveau leur réponse. Ce point ayant été clarifié pour la troisième question, le nombre de réponses soumises est identique à l'issue des phases 1 et 2 de la dernière question. Le taux de participation à la tâche de rédaction d'explications a suivi la même tendance positive, évoluant de 43% à 69%.

Le tableau 1 révèle d'autre part qu'entre la première phase et la deuxième, plusieurs étudiants ont modifié leur réponse et leur explication. Ceux-ci ont donc pleinement bénéficié du conflit sociocognitif les ayant amené à modifier leur croyance sur ce qui était vrai ou faux. Ce constat est particulièrement fort sur la

troisième question où le conflit sociocognitif a permis à 5 étudiants de modifier leur réponse. L'un de ces 5 étudiants avait soumis la réponse « je ne sais pas » lors de la première phase, et a proposé une réponse correcte à l'issue du conflit.

Le tableau 2 présente les résultats relatifs à la dernière phase de soumission. Cette phase correspond à l'évaluation par les étudiants des explications correspondant à la bonne réponse. Le tableau révèle que le taux de participation à l'évaluation a aussi augmenté au fil des questions, passant de 39% pour la première question à 76% pour la troisième question, chaque étudiant ayant évalué 3 explications. Le nombre moyen d'évaluateurs par explication est de 4,55. L'écart type moyen des notes distribuées par les étudiants (comprises entre 1 et 5) sur les explications est de 0.97. 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 révisions.

Tableau 2. Synthèse des résultats obtenus à la troisième phase.

	Question 1	Question 2	Question 3
Nombre d'explications évaluées	9	27	26
Nombre d'étudiants ayant soumis une évaluation	20	32	39
Nombre d'explications évaluées par étudiant	3	3	3
Nombre moyen d'évaluateurs par explication	6,67	3,84	4,5
Écart type moyen des notes fournies par les étudiants	1,07	0,87	1,02

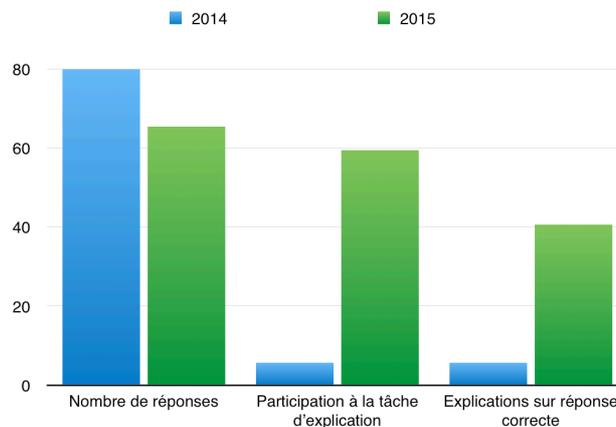


Fig. 5. Étude comparée du taux de participation à la tâche de rédaction d'explications.

Enfin, afin de mesurer la progression en termes de participation à la tâche de rédaction d'une explication, nous avons comparé le taux de participation à cette tâche sur les questions posées en mode NP-Q (en 2015) par rapport aux questions posées en mode « traditionnel » en 2014 ; nous rappelons que dans le mode traditionnel, les étudiants répondent à une question à choix multiples et sont invités à fournir des explications uniquement après la présentation des résultats. Notons également qu'en 2014, l'expérimentation a été menée sur un groupe de 35 étudiants.

Qualification semi-automatique de contributions d'apprenants pour l'intégration de *feedbacks* de qualité dans les tests en ligne

Fig. 5 synthétise les résultats de cette comparaison et montre que le taux de participation à la tâche de rédaction d'explications a massivement augmenté : seulement 6% des étudiants en moyenne ont participé à cette tâche en 2014, contre près de 60% en 2015. Au final, le nombre de contributions pouvant être retenues pour le *feedback* dans les tests de révision est nettement supérieur sur l'expérience de 2015 : les contributions de 40% des étudiants sont sélectionnables en 2015, contre seulement 6% de contributions en moyenne dans les expériences de 2014.

5. 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 décrit le dispositif NP-Q, une évolution de la plate-forme qui vise à augmenter considérablement le nombre de contributions des étudiants afin d'améliorer de manière significative la qualité des *feedbacks* restitués au sein des tests générés. Le dispositif NP-Q repose sur le découpage en plusieurs phases du processus de soumission d'une réponse, et place l'étudiant en situation d'apprentissage réflexif, de conflits sociocognitifs et d'évaluation. Les explications les mieux évaluées 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 d'un groupe de 51 étudiants inscrits en Master Informatique sont très encourageants : ils démontrent 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 du dispositif NP-Q, l'algorithme détaillé dans la section 3 permet d'associer à une réponse donnée une autre réponse dans l'objectif de déclencher un éventuel conflit sociocognitif. 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 lors du conflit sociocognitif) sont autant de traces 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 génération d'un conflit sociocognitif tel qu'elle est proposée dans le dispositif NP-Q tire parti des productions de l'ensemble des apprenants, pour l'ensemble des apprenants. Nous envisageons

donc, dans la suite de nos travaux, d'expérimenter l'apport du dispositif dans le cadre de MOOC avant d'améliorer l'algorithme en exploitant les traces massives exploitables dans un tel contexte d'apprentissage.

Références

1. Biggs, J. and Tang, C.: Teaching for quality learning at university. (2011).
2. Black, P. and Wiliam, D.: Assessment and Classroom Learning, *Assess. Educ. Princ. Policy Pract.*, vol. 5, no. 1, pp. 7–74, Mar. 1998.
3. Boud, D., Cohen, R., and Sampson, J.: Peer learning in higher education: Learning from and with each other, (2014).
4. Caldwell, J.: Clickers in the large classroom: Current research and best-practice tips, *CBE-Life Sci. Educ.*, (2007).
5. Dochy, F., Segers, M., and Sluijsmans, D.: The use of self-, peer and co-assessment in higher education: A review, *Stud. High. Educ.*, (1999).
6. Gauci, S. and Dantas, A.: Promoting student-centered active learning in lectures with a personal response system, *Adv. ...*, (2009).
7. Hattie, J. and Jaeger, R.: Assessment and classroom learning: A deductive approach, *Assess. Educ.*, (1998).
8. Higgins, R., Hartley, P., and Skelton, A.: The conscientious consumer: reconsidering the role of assessment feedback in student learning, *Stud. High. Educ.*, vol. 27, no. 1, (2002).
9. IMS Global Learning Consortium: IMS Question & Test Interoperability Specification. (2012).
10. Miller, T.: Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning, *Assess. Eval. High. Educ.*, (2009).
11. Nicol, D. and Macfarlane-Dick, D.: Formative assessment and self-regulated learning : A model and seven principles of good feedback practice ., *Stud. High. Educ.*, (2006) 1–19.
12. Ricketts, C. and Wilks, S.: Improving student performance through computer-based assessment: Insights from recent research, *Assess. Eval. High. Educ.*, vol. 27, no. 5, (2002).
13. Scouller, K.: The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay, *High. Educ.*, (1998).
14. Shaffer, D. and Collura, M.: Evaluating the effectiveness of a personal response system in the classroom, *Teach. Psychol.*, (2009).
15. Silvestre, F., Vidal, P., and Broisin, J.: Génération semi-automatique de tests d'auto-évaluation pourvus de feedback résultant de la prise de notes collaborative, conférence des Technologies de l'Information et de la Communication pour l'Enseignement, (2014).
16. Silvestre, F., Vidal, P., and Broisin, J.: Tsaap-Notes--An Open Micro-blogging Tool for Collaborative Notetaking during Face-to-Face Lectures, *International Conference on Advanced Learning Technologies (ICALT)*, (2014).
17. Topping, K.: Peer assessment between students in colleges and universities, *Rev. Educ. Res.*, (1998).
18. Uhari, M., Renko, M., and Soini, H.: Experiences of using an interactive audience response system in lectures, *BMC Med. Educ.*, (2003).