



**HAL**  
open science

## Comparaison de questions posées et votées en ligne dans le cadre d'une classe inversée

Fatima Harrak, François Bouchet, Vanda Luengo

### ► To cite this version:

Fatima Harrak, François Bouchet, Vanda Luengo. Comparaison de questions posées et votées en ligne dans le cadre d'une classe inversée. Environnements Informatiques pour l'Apprentissage Humain, Jun 2019, Paris, France. pp.121-132. hal-02157330

**HAL Id: hal-02157330**

**<https://hal.science/hal-02157330>**

Submitted on 26 Aug 2019

**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.

## Comparaison de questions posées et votées en ligne dans le cadre d'une classe inversée

Fatima Harrak, François Bouchet, Vanda Luengo

Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, LIP6, F-75005 Paris,  
France

[fatima.harrak@lip6.fr](mailto:fatima.harrak@lip6.fr)  
[francois.bouchet@lip6.fr](mailto:francois.bouchet@lip6.fr)  
[vanda.luengo@lip6.fr](mailto:vanda.luengo@lip6.fr)

**Résumé.** Les questions des élèves sont utiles pour leur apprentissage et l'adaptation pédagogique des enseignants. Nous étudions ici les questions posées en ligne par des étudiants de première année de médecine, utilisées par les professeurs pour préparer des sessions de questions-réponses. Comme les étudiants peuvent aussi voter sur les questions posées, on peut s'interroger sur la valeur pédagogique du vote : a-t-il le même impact en termes d'apprentissage, et les étudiants votent-ils sur des questions similaires à celles qu'ils posent ? Pour répondre à ces questions, nous avons développé un schéma de codage de la nature des questions, puis conçu un annotateur automatique que nous avons appliqué à l'ensemble du corpus. La comparaison votants vs. non-votants révèle que les votants réussissent mieux, mais peu de différences apparaissent entre la nature des questions posées et celles votées pour les étudiants qui font les deux. Ce résultat confirme la valeur du vote comme alternative à la formulation de question pour les étudiants sachant déjà formuler leurs propres questions.

**Mots-clés.** Question d'élève, vote d'élève, profil d'apprenant, schéma de codage, classe inversée

**Abstract.** Students' questions are useful for their learning experience as well as to help teachers to adapt their pedagogy. We study here a corpus of questions asked online by 1st year medical students, used by their professors to prepare their questions and answers sessions. As students can also vote on questions already asked, we can wonder about the pedagogical value of such votes: do they help learning as much, and are students voting on questions similar in nature to the ones they ask? To answer, we have developed a coding scheme of questions according to their nature and built an automatic annotator to tag the whole corpus. Comparing voters and non-voters reveals the former perform better than the latter, and there are little differences between questions asked and voted for students who do both. This result confirms the value of voting as an alternative to asking one's own questions for students who know how to ask questions.

**Keywords.** Student's question, student's vote, student's profile, coding scheme, blended learning

## 1 Introduction

Les questions des élèves jouent un rôle important dans le processus d'apprentissage, non seulement pour aider les élèves à mieux apprendre [1], mais aussi pour l'enseignant afin de déterminer ce qui a été compris et d'adapter sa pédagogie en conséquence. Les environnements en ligne et EIAH peuvent éliminer de nombreux obstacles qui empêchent les élèves de poser des questions en classe [2]. Nous nous intéressons ici à une formation hybride dans laquelle les étudiants doivent poser chaque semaine des questions à partir de supports de cours étudiés à distance avant le cours (selon une approche de classe inversée), pour aider les enseignants à préparer leurs séances de questions-réponses en présentiel. Cependant, compte tenu du volume de questions posées, les enseignants n'ont souvent pas assez de temps pour répondre à chaque question et doivent donc sélectionner celles auxquelles ils vont répondre. Pour limiter le nombre de questions et les aider dans ce choix, ils encouragent les étudiants à voter sur les questions déjà posées avant d'en poser de nouvelles. D'un point de vue pédagogique, cela suppose que les étudiants lisent les questions des autres, ce qui peut également avoir un impact positif en les forçant à s'interroger sur leur propre compréhension des points abordés par leurs camarades. Mais on peut aussi penser qu'un vote n'est pas exactement équivalent à une question. En effet, dans le cadre théorique Active-Constructive-Interactive (ACI) proposé par Chi [3], le fait de poser une question est une activité de nature « constructive » (mobilisant des processus cognitifs tels que la recherche de lacunes dans ses connaissances et la restructuration de celles-ci), plus à même d'aider l'apprentissage qu'une activité « active » comme le fait de voter (mettant uniquement en jeu une recherche dans ses connaissances pour savoir si on saurait ou non répondre à cette question). Face à cette contradiction sur la valeur d'un vote, nous avons conduit des analyses afin d'explorer comment les votes sont associés à la réussite des étudiants et aux questions qu'ils posent. Plus précisément, notre objectif était de répondre aux deux questions de recherche suivantes : (QR1) Quel est l'impact du vote sur la performance d'un étudiant, notamment en comparant ceux qui posent ou non également des questions ? (QR2) Y a-t-il des différences dans la nature des questions posées et votées par les étudiants ?

Pour traiter ces questions de recherche, nous avons défini un schéma de codage adapté aux questions des étudiants et un système d'annotation automatique pour annoter l'ensemble du corpus de questions dont nous disposons. Dans la suite de cet article, nous proposons dans la section 2 un état de l'art sur les différentes taxonomies de questions. Nous présentons dans la section 3 le contexte de l'étude et les données utilisées, avant de décrire dans la section 4 la méthodologie de catégorisation de questions et d'annotation automatique. Enfin, nous présentons dans la section 5 les résultats des analyses effectuées pour répondre aux deux questions de recherche et nous concluons avec quelques perspectives et limites de ce travail en section 6.

## 2 Etat de l'art

Les chercheurs ont étudié le comportement de questionnement des élèves dans divers contextes éducatifs, tels que la classe [4], le tutorat [5] et les environnements

d'apprentissage en ligne [6]. Scardamalia et Bereiter [7] font la distinction entre les questions fondées sur le texte et celles fondées sur le savoir (ces dernières ayant un potentiel éducatif plus fort). Bien que cette distinction soit pertinente, n'ayant pas accès aux transcriptions des vidéos et diapositives avec lesquelles les étudiants ont interagi avant de poser leurs questions, il était difficile d'identifier ceci automatiquement. D'autres chercheurs ont proposé une typologie de questions distinguant celles pouvant faire l'objet d'une investigation scientifique (*e.g.* comparaison, cause à effet, prédiction, exploration) des autres [4]. Bien que notre but ne soit pas d'encourager à poser un certain type de questions, cette distinction pourrait s'appliquer à notre travail, mais est difficile à réaliser sans experts du domaine. Graesser et Person [5] ont pour leur part élaboré une taxonomie de questions posées pendant les séances de tutorat, utilisée pour la génération automatique de questions. Bien que leur taxonomie puisse être pertinente ici, certaines catégories comprenaient des « questions de raisonnement approfondi » de haute qualité, associées à des modèles de raisonnement difficiles à identifier automatiquement. Enfin, des recherches récentes [8] ont utilisé une version réduite de la taxonomie de Bloom [9] pour établir un lien entre rétroaction pratique et rendement de l'apprenant en matière d'évaluation. Cette taxonomie, en raison de son origine, tend toutefois à être plus appropriée aux questions de l'enseignant qu'à celles des élèves.

L'analyse des questions d'apprentissage a été utilisée à des fins très diverses afin d'améliorer l'efficacité de l'enseignement et l'apprentissage des élèves. Ainsi, Graesser et Person [5] ont étudié la relation entre les questions posées et la réussite, constatant une corrélation positive entre la réussite des élèves et la « qualité » de leurs questions (une fois acquise une certaine expérience en tutorat) et une corrélation négative entre la réussite des élèves et le nombre total de questions. Les recherches de Chin [10, 11], ont montré que les questions se situent sur un continuum allant de « superficielles » à « profondes » et que le type de questions que posent les élèves dépend de la façon dont ils abordent leurs tâches d'apprentissage. En se basant sur ces études, on pourrait conclure que poser des questions « superficielles » est généralement peu utile.

Les votes des élèves ont également été étudiés pour analyser le comportement des élèves dans les forums en ligne. Bihani et al. [12] ont utilisé le nombre de votes sur les questions et réponses des étudiants et les réponses de l'enseignant pour révéler les paires de questions/réponses pertinentes pour les futurs cours. Zeng et al. [13] ont également utilisé le nombre de votes pour détecter la confusion au sein de messages de forum de discussion, constatant que les messages contenant de la confusion sont considérés comme importants et reçoivent plus de votes. Jiang et al [14] ont analysé les étudiants considérés comme des “influenceurs” (utilisateurs dont les messages génèrent beaucoup de réponses dans les forums d'un MOOC). Ces influenceurs ont des résultats plus faibles et reçoivent moins de votes que les utilisateurs actifs (ceux qui postent régulièrement sur le forum). De même, Wong et al. [15] ont analysé les votes (positifs et négatifs) sur les messages et les commentaires des utilisateurs actifs. Contrairement à [14], ils ont constaté que les utilisateurs actifs sont aussi des utilisateurs influents qui apportent généralement une contribution positive aux discussions du forum du MOOC. Les votes des élèves ont donc surtout été utilisés pour analyser le comportement des élèves, mais la nature des questions votées n'a apparemment pas encore été explorée.

Dans l'ensemble, les taxonomies proposées jusqu'à présent dépendent essentiellement du contexte, et nous avons décidé de définir un nouveau schéma de codage utilisant une approche fondée sur les données. Dans cet article, nous nous intéressons donc à l'analyse de la nature des questions posées et votées par les étudiants.

### 3 Contexte

Nous avons considéré l'ensemble des questions posées par des étudiants de 1ère année de médecine et pharmacie de l'université Joseph-Fourier de Grenoble en 2012-2013. 1608 étudiants étaient inscrits cette année-là, bien que tous n'aient pas posé de questions. La Faculté de médecine dispose d'un système de formation hybride pour ses étudiants de 1ère année (PACES). Chaque semestre se termine par un concours (en janvier et mai) éliminatoire (seule une partie des étudiants qui ont échoué au concours est autorisée à repasser l'année une seule fois, les autres devant se réorienter). Chaque cours est composé de deux à six séquences de 4 semaines (*cf.* Figure 1).

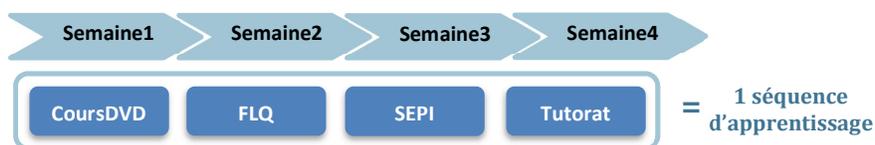


Fig. 1. Les quatre activités d'une séquence d'apprentissage sur quatre semaines

Dans chaque séquence, la première semaine consiste à étudier le cours sur DVD-ROM ou sur le site Medatice (diapositives + vidéos du professeur). La deuxième semaine est consacrée à la Formulation en ligne des questions (FLQ) pour les enseignants : ces questions concernent exclusivement les cours multimédias étudiés la semaine précédente. Les élèves peuvent voir les questions posées par les autres élèves de leur groupe et voter pour celles auxquelles ils veulent aussi une réponse, mais ils ne peuvent pas y répondre. En fin de semaine, les questions sont envoyées aux enseignants intervenant la troisième semaine, qui les utilisent pour structurer leurs sessions d'enseignement interactives en classe. Au cours de ces sessions, l'enseignant répond à certaines questions posées en ligne par les étudiants. La quatrième semaine est consacrée à des séances de tutorat afin de tester les connaissances acquises lors de la séquence de formation par le biais d'un auto-test utilisant des questions à choix multiples (QCM), qui sont ensuite corrigées par un professeur auxiliaire. Il y a deux séances de tutorat de 2 heures/semaine. Chaque étudiant peut vérifier individuellement ses notes et son classement par rapport à l'ensemble de la promotion, et il est nécessaire s'être connecté à la plateforme de questions pour pouvoir consulter ses notes. Ces résultats de tutorat permettent aux élèves d'évaluer la qualité de leur apprentissage.

Pour chacun des 13 cours, nous avons donc 2 à 6 ensembles de questions (un par séquence) posées par 429 étudiants (6457 questions au total) et votées par 672 étudiants (10951 votes) pendant la deuxième semaine de chaque séquence. La répartition des questions est inégale (*cf.* Table 1), avec plus de questions au 1er semestre, notamment car certains étudiants sont obligés d'arrêter à la fin de celui-ci. On note que seul un élève

sur quatre a posé au moins une question, ce qui peut s'expliquer en partie par l'encouragement à voter au lieu de poser des questions (pour forcer à lire les questions des autres mais aussi pour réduire le nombre de questions reçues par courriel).

**Table 1.** Distribution des questions posées par cours

BCH	BPH	HBD	BCE	ANT	PHS	SSH	ICM	MAT	Spec.
19%	17%	15%	11%	10%	9%	8%	6%	3%	1%

BCH = Biochimie, BPH = Biophysique, HBD = Histoire et biologie du développement, BCE = Biologie cellulaire, ANT = Anatomie, PHS = Physiologie, SSH = Santé, société, humanité, ICM = Initiation à la connaissance du médicament, MAT = Mathématique, Spécialité = Pharmacie, Odontologie, Maïeutique

## 4 Catégorisation et annotation de questions

### 4.1 Schéma de codage

Afin d'identifier la nature des questions posées par les étudiants, nous avons décidé de développer notre propre schéma de codage. Nous avons travaillé sur un échantillon de 800 questions (12% du corpus) issues de deux cours (BCH et HBD), considérés par l'équipe pédagogique comme étant parmi les plus difficiles et ayant suscité le plus de questions (*cf.* Table 1). Cet échantillon a été divisé en 4 sous-échantillons de 200 questions pour appliquer 4 étapes successives de catégorisation.

(1) L'**étape de découverte** consistait à regrouper empiriquement des phrases ayant des similitudes pour en extraire des concepts significatifs. Bien que les enseignants demandent aux élèves de poser des questions simples (c.-à-d. d'éviter des questions comme « Pourriez-vous expliquer à nouveau X ? De plus, Y n'était pas clair »), 40% des questions pouvaient être divisées en plusieurs questions indépendantes. Une fois les phrases segmentées en questions dites simples, nous avons regroupé celles dont la structure (ex : « qu'est-ce que X ? » et « qu'est-ce que Y ? ») et la sémantique (ex : « qu'est-ce que X ? » et « pourriez-vous définir X ? ») semblent similaires. Des groupes de questions ont ensuite reçu des "étiquettes" (ex : « définition d'un concept »), puis nous avons identifié les exclusions mutuelles entre étiquettes (ex : une question simple ne peut pas être à la fois une vérification et une demande de ré-explication). Cela nous a conduit à définir le concept de « dimensions », ensembles d'étiquettes de types de questions similaires mais mutuellement exclusives. Chacune de ces étiquettes individuelles ("vérification", "ré-explication"... ) sont des valeurs pouvant être associées à une dimension. Chaque question simple peut alors être associée à une annotation dans ce schéma de codage en choisissant une valeur par dimension. Une annotation de question peut donc être vue comme un vecteur de 4 valeurs (ex : "Pourriez-vous réexpliquer la différence entre un composé ionisable et un partiellement ionisable ?" est représenté par le vecteur [Ree,0,Lie,0], *cf.* Table 2).

(2) L'**étape de consolidation** consistait à annoter le deuxième sous-échantillon pour valider les dimensions et les valeurs précédemment identifiées. Cela a conduit à divers ajustements des dimensions pour s'assurer qu'elles étaient bien indépendantes les unes

des autres (par exemple l'ajout de la valeur "correction" dans Dim2, non identifiée précédemment). Parallèlement, les dimensions identifiées ont été revues et validées par un professeur expert enseignant dans le cadre de PACES, qui a estimé que les catégories étaient potentiellement pertinentes pour analyser les questions des étudiants.

(3) Dans l'étape de validation, nous avons effectué une double annotation pour valider l'ensemble de nos catégories sur le troisième sous-échantillon de 200 phrases. Ensuite, deux annotateurs humains ont utilisé comme référence unique le schéma de codage créé à la fin de l'étape précédente pour annoter chaque segment (238 au total). A l'issue de l'étape précédente, trois dimensions avaient été identifiées : Dim1 (relative au type de question), Dim2 (relative à la modalité d'explication), Dim4 (facultative, annotée uniquement si la question est une vérification, relative à la nature de ce qui est vérifié). La dimension appelée "Dim3" plus loin n'existait pas encore à cette étape. Les annotateurs humains ont fait deux annotations distinctes et indépendantes sur chaque dimension, et leur accord a été évalué à l'aide du Kappa de Cohen ( $K1=0,72$ ,  $K2=0,62$  où  $K1$  et  $K2$  correspondent respectivement au Kappa de Dim1 et Dim2). Pour Dim4, en raison de son caractère facultatif, les deux annotateurs n'ont pas nécessairement annoté les mêmes questions (annotateur 1 : 82 questions ; annotateur 2 : 68 questions) : sur les 68 en commun, le kappa valait 0,66. Puis ils se sont rencontrés pour discuter et résoudre les désaccords, ce qui a conduit à un affinement final des catégories (par exemple, séparation des catégories Dim1 et Dim4, ajout de la catégorie Dim3). Finalement, tout l'échantillon (600 phrases) a été ré-annoté sur les 4 dimensions pour tenir compte des changements et fournir une référence à laquelle comparer l'annotation automatique. Cette version finale du schéma de codage est présentée dans la Table 2.

Table 2. Schéma de codage créé à partir de l'annotation manuelle

Dim1	Type de question	Description
Ree	Ré-expliquer / redéfinir	Demander de revenir sur un concept déjà expliqué
App	Approfondir un concept	Approfondir une connaissance, clarifier une ambiguïté ou demander plus de détails pour mieux comprendre
Ver	Validation / vérification	Vérifier ou valider une hypothèse
Dim2	Modalité d'explication / Sujet de question	Description
Exe	Exemple	Exemple d'application (cours/exercice)
Sch	Schéma	Schéma d'application ou explication sur ce dernier
Cor	Correction	Correction d'un exercice en cours/examen
Dim3	Type d'explication	Description
Def	Définir	Définir un concept ou un terme
Man	Manière (comment ?)	La manière comment procéder
Rai	Raison (pourquoi ?)	Demander la raison
Rol	Rôles (utilité ?)	Demander l'utilité / fonction
Lie	Lien entre concepts	Vérifier le lien entre deux concepts, le définir
Dim4	Type de vérification (facultatif)	Description
Err	Erreur / contradiction	Détecter une erreur/contradiction en cours ou dans l'explication de l'enseignant
Con	Connaissances du cours	Vérifier une connaissance
Exa	Examen	Vérifier une connaissance attendue à l'examen

(4) Finalement, dans l'**étape d'évaluation**, le dernier sous-échantillon de 200 segments a été annoté manuellement par les deux annotateurs experts (avec un kappa accru de 0,83 sur Dim1, 0,76 sur Dim2 et 0,47 sur Dim3). Ce sous-échantillon, non utilisé pour l'entraînement de l'annotateur automatique, a été utilisé pour son test.

## 4.2 Annotation automatique

Afin d'annoter l'ensemble de questions posées par les étudiants (et à terme pour l'utiliser en ligne pour analyser les questions collectées), un outil semi-automatique à base de règles et de mots clés pondérés manuellement a été utilisé antérieurement pour segmenter et annoter les questions automatiquement. Bien qu'efficace sur les questions qu'il annoté (kappa élevé), une proportion importante de questions ne sont pas annotées par cet outil [10], ce qui nous amène à utiliser ici une annotation entièrement automatisée basée sur des techniques d'apprentissage automatique sur le corpus des questions (*cf.* Figure 2 pour un résumé des différentes étapes décrites ci-dessous).

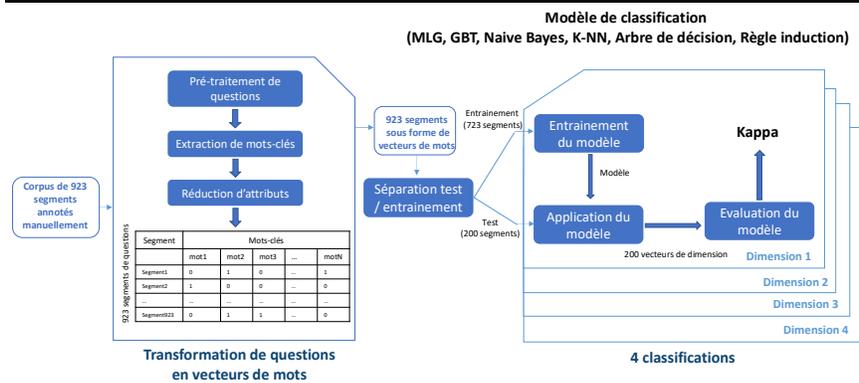
La première étape a consisté à transformer les 923 segments annotés manuellement en vecteurs de mots. Tout d'abord, nous avons utilisé la version française de WordNet [17], base de données lexicale reliant des concepts sémantiques entre eux dans une ontologie selon une variété de relations sémantiques (tels que synonymes et hyperonymes). L'objectif est de ramener différentes expressions synonymes à une même expression dans les questions (ex : pour la valeur « Rai » dans dim 3, les mots synonymes « cause », « raison » et « motif » sont remplacés dans le texte par « pourquoi »), pour diminuer la diversité lexicale et renforcer certaines expressions pour le traitement qui suit. Nous avons effectué par la suite un ensemble de pré-traitements classiques sur le corpus de 923 segments : tokenisation, stemmatisation, suppression de ponctuation et de stopwords (mots creux non porteurs de sens), etc. Puis, nous avons extrait tous les unigrammes et bigrammes, avec une approche de type sac de mots, et compté leurs occurrences dans l'échantillon de 600 questions (723 segments) de l'étape de validation. Chaque segment est représenté par un vecteur de mots (nombre d'occurrences de chaque unigramme/bigramme (n-grammes) possible sur chaque segment). Le nombre de n-grammes étant très important par rapport au nombre de segments, nous avons réduit celui-ci pour conserver les mots-clés les plus importants et les plus significatifs en utilisant une technique de sélection d'attributs (suppression des n-grammes les moins fréquents et corrélés).

La deuxième étape a consisté à entraîner un classifieur pour automatiquement annoter chaque valeur de dimension (par ex. « réexpliquer »). Nous avons essayé 6 techniques de classification différentes tels qu'implémentées dans RapidMiner (Modèle linéaire généralisé, Gradient Boosted Trees, Arbre de décision, K-NN, Règle d'induction et Naïve Bayes) sur chaque dimension séparément (les dimensions étant conçues comme indépendantes). Chaque classifieur est entraîné en prenant en entrée un ensemble de vecteurs de mots représentant les 723 segments de l'ensemble d'entraînement, et l'étiquette à prédire est la valeur associée manuellement au segment dans cette dimension. Le modèle est ensuite évalué sur un échantillon indépendant de 200 segments sans étiquettes, afin d'assurer une bonne estimation de la performance

sur des données non vues. Enfin, nous avons calculé les valeurs Kappa entre les valeurs prédites par le classifieur et les valeurs correspondantes trouvées par l'annotation manuelle. Les meilleurs résultats ont été obtenus par l'algorithme Gradient Boosted Trees avec un Kappa moyen sur chaque dimension de 0,70 (cf. Table 3) – une valeur suffisamment élevée pour appliquer l'annotation automatique au corpus complet.

**Table 3.** Kappas obtenus des différentes techniques de classification utilisées

Dimension	Modèle linéaire généralisé (GLM)	Gradient Boosted Trees (GBT)	Naive Bayes	K-NN (k=2)	Arbre décision (C4.5)	Règle induction
Dim1	0.68	<b>0.70</b>	0.29	0.57	0.36	0.70
Dim2	0.17	0.77	0.10	0.43	<b>0.79</b>	0.37
Dim3	<b>0.69</b>	0.63	0.37	0.61	0.63	0.58
Dim4	0.62	<b>0.66</b>	0.38	0.60	0.13	0.66



**Fig. 2.** Processus d'annotation à base d'apprentissage automatique

## 5 Différences de nature entre questions posées et votées

A l'issue de la section précédente, nous disposons donc d'un corpus complet de 6457 questions annoté automatiquement. Nous nous sommes concentré sur quatre cours ayant généré le plus de questions (cf. Table 1) et considérés par les enseignants comme les plus difficiles : BCH, HBD, BCE et ANT. Les trois premiers ont lieu au premier semestre et ANT au second semestre. Nous n'avons pas fusionné les questions des différents cours, car des études antérieures [16, 18] sur ces données ont montré des différences significatives entre les cours. De plus, considérer les cours séparément permet de vérifier si des tendances similaires apparaissent d'un cours à l'autre.

Pour répondre aux questions de recherche, nous avons distingué 4 sous-populations sur chacun des cours considérés en fonction de l'activité des étudiants en distinguant : Q pour les étudiants « ayant posé au moins une question » (NQ sinon), et V pour les étudiants « ayant voté sur au moins une question » (NV sinon). En croisant les deux, cela donne donc les 4 sous-populations suivantes : QV, QNV, NQV et NQNV.

### 5.1 Impact du vote sur la performance

**Méthode.** Pour examiner la QR1, pour évaluer la performance, nous avons considéré pour chaque étudiant et sur chaque cours : (1) la note moyenne obtenue sur les QCM du cours (NotMoy, sur 20), qui peut donner une mesure à court terme de l’impact des questions posées, et la note finale obtenue au concours à cette matière (NotFin, sur 20), qui peut donner une mesure de l’effet à plus long terme.

Nous avons ensuite fait des comparaisons 2 à 2 entre QV et QNV (pour mesurer l’impact du vote chez les étudiants posant des questions) et entre NQV et NQNV (pour l’impact du vote chez les étudiants ne posant pas de questions), sur chacun des 4 cours, pour les deux variables. Pour ces 2 variables, nous avons utilisé des tests Mann-Whitney U au lieu de t-tests (les distributions ne suivant pas une loi normale). Nous avons effectué 2 fois 7 tests (2 comparaisons de population avec 4 cours et 2 variables à chaque fois, sauf NotFin, manquante pour BCE) et les seuils de significativité ont été corrigés par la méthode de Holm–Šidák pour éviter les erreurs de type I.

**Résultats.** En ce qui concerne les QV par rapport à QNV, 1 seul résultat significatif a été obtenu sur les cours du premier semestre (BCH, HBD, BCE) où parmi les étudiants ayant posé des questions, ceux n’ayant pas voté ont eu une meilleure note finale que les étudiants ayant voté (pour HBD). En revanche, pour ANT, en dépit d’effectifs plus réduits, les étudiants qui ont posé des questions et votés ont mieux réussi que ceux n’ayant fait que poser des questions, aussi bien aux QCM du cours (*cf.* Table 4) qu’au concours final (*cf.* Table 5).

En ce qui concerne les NQV par rapport à NQNV, 3 résultats significatifs ont également été obtenus (sur 7 tests) : pour BCH et HBD, les étudiants n’ayant pas posé de questions et n’ayant pas non plus voté ont eu une meilleure note sur les QCM que ceux ayant uniquement voté. Ce résultat se retrouve également au niveau du concours final pour HBD. Aucune différence n’a été observée au cours du deuxième semestre (ANT). En résumé, lorsqu’une différence a été observée, les élèves qui ont voté (sans poser de question) ont obtenu des résultats inférieurs, tant dans le cours que dans l’ensemble.

**Table 4.** Caractérisation des votants en termes de Note Moyenne (NotM) pour les étudiants qui posent des questions (QV vs. QNV) et ceux qui n’en posent pas (NQV vs. NQNV) et statistiques descriptives (quartiles et médiane) des 4 sous-populations sur chacun des 4 cours

	QV vs. NQV vs.		QV				QNV				NQV				NQNV			
	QNV	NQNV	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3	N	Q1	Md	Q3
BCH	.049	.001*	181	6.50	8.50	11.50	59	7.64	9.67	12.50	217	5.60	8.00	10.17	980	5.50	8.00	10.87
HBD	.176	.000*	154	7.69	10.12	13.75	53	8.25	11.67	14.00	252	6.00	8.67	11.27	956	6.00	9.25	12.50
BCE	.128	.039	83	7.47	10.40	12.55	47	6.40	9.00	11.70	117	6.00	8.00	10.80	1133	5.00	7.67	10.75
ANT	.001*	.005	42	11.20	13.80	15.35	46	6.43	10.20	13.35	23	8.57	10.00	12.70	968	5.00	8.00	11.80

\* significatif avec  $p < .05$  après correction de Holm–Šidák (.000 signifie  $p < .001$ )

**Table 5.** Idem Table 4 pour la Note Finale (NotF)

	QV vs.	NQV vs.	QV				QNV				NQV				NQNV			
	QNV	NQNV	<i>p</i>	<i>N</i>	Q1	Md	Q3	<i>N</i>	Q1	Md	Q3	<i>N</i>	Q1	Md	Q3	<i>N</i>	Q1	Md
BCH	.080	.015	168	5.25	8.25	11.50	56	6.31	10.38	13.31	197	4.25	6.50	9.75	880	3.75	6.75	10.50
HBD	.002*	.000*	144	6.50	9.25	12.12	44	8.19	11.50	13.06	229	4.50	7.25	10.50	886	4.00	7.75	10.94
BCE	-	-	83	N/A	N/A	N/A	49	N/A	N/A	N/A	118	N/A	N/A	N/A	1368	N/A	N/A	N/A
ANT	.000*	.706	42	13	15.38	16.19	45	7.50	11.50	14.25	22	8.13	11.88	14.69	1116	3.5	7.00	12.00

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

## 5.2 Comparaison de nature de questions posées et votées

**Méthode.** Pour comparer la nature des questions que posent les étudiants à la nature des questions sur lesquelles ils votent, nous avons dû nous concentrer sur la population des étudiants qui font les deux (QV). Pour ces étudiants, sur chacun des 4 cours, nous avons considéré toutes les questions sur lesquelles ils ont voté pour calculer la proportion de chaque type de question votée dans chaque dimension. Par exemple, si dans BCH, un élève a voté sur une question de ré-explication et une autre de vérification (étiquetées [Ree,0,Sch,0] et [Ver,0,0,Con]), sur la dimension 1, il aurait voté à 50 % sur des questions de valeur "Ree" (ré-explication) et à 50% de ses questions de valeur "Ver" (validation). Ces proportions sont codées entre 0 et 1, de sorte que pour chaque étudiant, sur chaque cours, on obtient un *vecteur de vote* composé de 14 (3+3+5+5+3) valeurs comprises entre 0 et 1, et en suivant la même approche pour les questions posées, on peut également obtenir un *vecteur de question* posée de 14 valeurs.

Une fois le pré-traitement effectué, la comparaison des questions votées aux questions posées consistait à comparer pour chaque cours, pour chaque valeur d'une dimension (ex : la valeur "Ree" de la dimension 1), la distribution de la proportion des questions posées par les étudiants et celle de questions votées dans cette dimension. En d'autres termes, comparer deux distributions (non distribuées normalement) entre 0 et 1 pour la même population d'élèves, ce qui a été fait en effectuant 14 tests de Wilcoxon, en utilisant comme en 5.1 la méthode Holm-Šidák pour ajuster la valeur  $p$  critique.

**Résultat.** Un seul test sur 56 a révélé un résultat statistiquement significatif (cf. Table 6), répondant ainsi négativement à la QR2.

**Table 6.** Différences entre voter et poser une question selon la nature des questions

	Ree	App	Ver	Exe	Sch	Cor	Def	Man	Rai	Rol	Lie	Err	Con	Exa
BCH	.176	.427	.11	.31	.411	.017	.92	.000*	.236	.352	.295	.514	.259	.078
HBD	.382	.851	.717	.755	.809	.225	.093	.728	.007	.285	.003	.043	.941	.706
BCE	.476	.067	.015	.515	.723	.929	.89	.652	.797	.681	.118	.51	.686	.033
ANT	.826	.087	.551	.204	.795	.18	.076	.485	.212	.198	.039	.611	.691	.701

\* significatif avec  $p < .05$  après correction de Holm-Šidák (.000 signifie  $p < .001$ )

## 6 Discussion et conclusion

La comparaison des questions posées et votées par les étudiants révèle donc que, dans notre cas, **le vote ne semble pas être un bon substitut au fait de poser des questions**. En effet, quand un effet est mesuré, il est plutôt négatif pour les étudiants ayant voté, surtout lorsqu'ils n'ont pas posé de questions par ailleurs. Cette tendance semble cependant s'inverser plus tard dans l'année, où le fait de voter en complément de poser des questions entraîne de meilleurs résultats, tant aux QCM de cours qu'à l'examen final. Dans notre contexte, il semble donc que des activités « actives » (au sens de Chi – ici, voter) complémentaires à des activités « constructives » (ici, poser des questions) soient plus efficaces que des activités « constructives » seules, mais qu'une activité « active » seule soit plus négative en termes d'apprentissage que de la passivité. La différence entre le premier et le second semestre pourrait être liée au fait que les étudiants les plus en difficulté ont été forcés à arrêter à la fin du premier semestre, ainsi qu'à la baisse générale d'activité sur la plateforme : les étudiants qui continuent à y participer sont donc probablement plus motivés que la moyenne.

Le second résultat est que **pour les étudiants qui votent et posent des questions**, c'est-à-dire les plus impliqués et ceux qui réussissent le mieux (notamment au second semestre), il n'y a globalement **pas de différence de nature entre les questions sur lesquelles ils votent et celles qu'ils posent eux-mêmes**. Ce résultat est intéressant car il suggère une interprétation possible du résultat précédent, à savoir que les votes de ces étudiants correspondent à des questions qu'ils se posent vraiment. Il est possible que des étudiants se contentant de voter ne font pas l'effort de formuler leurs propres questions, et que s'ils le faisaient, elles seraient d'une nature différente.

Ce travail est exploratoire et présente donc plusieurs limites : même si tous les étudiants se connectent à la plateforme de questions, nous n'avons pas accès à des logs permettant de savoir s'ils ont vraiment lu les autres questions posées. Les étudiants qui se connectent en premier n'ont également pas de questions sur lesquelles ils peuvent voter, sauf s'ils se reconnectent par la suite pour voir les nouvelles questions. Une expérience davantage contrôlée dans laquelle les étudiants doivent poser des questions et/ou voter sur des questions précédemment posées après avoir vu une vidéo, pour pouvoir passer à la suite, permettrait de vérifier les interprétations précédentes. Néanmoins le travail préalable réalisé ici rend désormais possible ce type d'expérience.

Dans notre contexte, ces résultats suggèrent qu'encourager les étudiants à formuler leurs questions, plutôt que de se contenter de voter sur les questions des autres, serait une stratégie positive pour l'apprentissage. En effet, il est possible que pour certains étudiants, voter donne le sentiment de faire ce qui est attendu d'eux, sans pour autant développer les stratégies métacognitives mises en jeu lorsqu'on se pose ses propres questions (identifier les concepts clés, tester sa compréhension, résumer ce qui a été appris...). Cela pourrait être fait en encourageant les étudiants à poser une question avant de pouvoir consulter celles des autres. Du point de vue des enseignants, cela signifie qu'il est d'autant plus critique de leur proposer une visualisation plus efficace que le « mur de questions » actuel, ce qui est une des perspectives de ce travail.

En résumé, voter est une bonne stratégie pour les étudiants sachant déjà formuler leurs propres questions. En revanche, pour ceux en difficulté, cela peut retarder la prise de conscience de leurs lacunes et leur capacité à les combler activement.

**Remerciements :** Nous remercions Pierre Gillois de nous avoir fourni les données.

## References

1. Sullins, J., McNamara, D., Acuff, S., Neely, D., Hildebrand, E., Stewart, G., Hu, X.: Are you asking the right questions: The use of animated agents to teach learners to become better question askers. In *International Florida Artificial Intelligence Research Society Conf., FLAIRS* (2015).
2. Otero, J., Graesser, A.C.: PREG: Elements of a model of question asking. *Cognition and instruction*. 19, 143–175 (2001).
3. Chi, M.T.H.: Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*. 1, 73–105 (2009).
4. Chin, C., Kayalvizhi, G.: Posing Problems for Open Investigations: What questions do pupils ask? *Research in Science & Technological Education*. 20, 269–287 (2002).
5. Graesser, A.C., Person, N.K.: Question asking during tutoring. *American educational research journal*. 31, 104–137 (1994).
6. Li, H., Duan, Y., Clewley, D.N., Morgan, B., Graesser, A.C., Shaffer, D.W., Saucerman, J.: Question asking during collaborative problem solving in an online game environment. In: *International Conference on Intelligent Tutoring Systems*. pp. 617–618. Springer (2014).
7. Scardamalia, M., Bereiter, C.: Text-Based and Knowledge Based Questioning by Children. *Cognition and Instruction*. 9, 177–199 (1992). [https://doi.org/10.1207/s1532690xci0903\\_1](https://doi.org/10.1207/s1532690xci0903_1).
8. Supraja, S., Hartman, K., Tatinati, S., Khong, A.W.: Toward the Automatic Labeling of Course Questions for Ensuring their Alignment with Learning Outcomes. In: *Proc. of The 10th International Conf. on Educational Data Mining*. pp. 56–63. Xiangen Hu, Tiffany Barnes, Arnon Hershkovitz, Luc Paquette, Wuhan, China (2017).
9. Bloom, B.S., Engelhart, M.B.: Furst, E.J. Hill, WH, & Krathwohl, DR Taxonomy of educational objectives. The classification of educational goals. *Handbook I: Cognitive domain*. New York: Longmans Green. (1956).
10. Chin, C., Brown, D.E.: Student-generated questions: A meaningful aspect of learning in science. *International Journal of Science Education*. 24, 521–549 (2002).
11. Chin, C., Osborne, J.: Students' questions: a potential resource for teaching and learning science. *Studies in science education*. 44, 1–39 (2008).
12. Bihani, A., Ullman, J.D., Paepcke, A.: FAQtor: Automatic FAQ generation using online forums. In: *International Conference on Educational Data Mining*. pp. 529–532. Kristy Elizabeth Boyer, Michael Yudelson (2018).
13. Zeng, Z., Chaturvedi, S., Bhat, S.: Learner Affect Through the Looking Glass: Characterization and Detection of Confusion in Online Courses. In: *International Conference on Educational Data Mining*. pp. 272–277 (2017).
14. Jiang, Z., Zhang, Y., Liu, C., Li, X.: Influence Analysis by Heterogeneous Network in MOOC Forums: What Can We Discover?. In: *Int. Conf. on Educ. Data Mining* (2015).
15. Wong, J.-S., Pursel, B., Divinsky, A., Jansen, B.J.: An analysis of MOOC discussion forum interactions from the most active users. In: *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. pp. 452–457. Springer (2015).
16. Harrak, F., Bouchet, F., Luengo, V., Gillois, P.: Profiling Students from Their Questions in a Blended Learning Environment. In: *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. pp. 102–110. ACM, New York, NY, USA (2018).
17. Sagot, B., Fišer, D.: Building a free French wordnet from multilingual resources. In: *OntoLex* (2008).
18. Harrak, F., Bouchet, F., Luengo, V.: From Students' Questions to Students' Profiles in a Blended Learning Environment. 1. 6, 54–84 (2019).