Vers la conception de feedback pour enseignants dans un contexte d'évaluation formative à grande échelle : une approche analytique

Rialy Andriamiseza, Franck Silvestre, Jean-François Parmentier, Julien Broisin

Vers la conception de *feedback* pour enseignants dans un contexte d'évaluation formative à grande échelle : une approche analytique

Rialy Andriamiseza[1], Franck Silvestre[1], Jean-François Parmentier[2], and Julien Broisin[1]

[1] Institut de recherche en Informatique de Toulouse, Université de Toulouse, 118 Route de Narbonne, 31400 Toulouse, France
[2] Toulouse INP-ENSEEIHT, 2 Rue Charles Camichel, 31000 Toulouse

Abstract. Pour résoudre le problème de l'augmentation du nombre d'étudiants dans l'enseignement supérieur, l'évaluation formative assistée par la technologie et ses processus variés ont emergé. Parmi eux, nous nous intéressons aux processus à deux votes et à ses interactions pour identifier des *feedback* à fournir aux enseignants afin de les aider dans leur prise de décision. À partir de la littérature et d'un jeu de données collectées en contexte écologique, nous avons mobilisé différentes techniques d'analyse pour identifier des informations pouvant aider les enseignants à mener des séquences d'évaluation formative dans le contexte des processus à 2 votes : (1) Plus le pourcentage de réponses correctes au premier vote est proche de 50, plus la séquence a des chances d'être bénéfique ; (2) Plus l'évaluation par les pairs est cohérente, plus la séquence a des chances d'être bénéfique ; (3) La cohérence du degré de confiance des apprenants est corrélée avec la cohérence de l'évaluation par les pairs. Les prochains travaux viseront à implémenter ces *feedback* dans notre système.

Keywords: évaluation formative; *learning analytics*; instruction par les pairs; *feedback* pour enseignants

Abstract. To face the increasing number of students in higher education, technology-enhanced formative assessment and its processes emerged. We focus on two-votes-based processes, and its interactions that could be used to identify appropriate feedback to support teachers. Based on literature and using a dataset gathered from an authentic learning contexts, we use learning analytics to identify different information that will help teachers conduct formative assessment sequences in two-votes-based processes: (1) The percentage of correct answers at the first vote should be close to 50; (2) The consistency of peer assessment is correlated with sequence benefits; (3) The consistency of learners' confidence degree is correlated to the consistency of peer assessment. The next step of these works will consist in implementing meaningful guidance tools based on these indicators into our system.

Keywords: formative assessment; learning analytics; peer instruction; teacher feedback

1 Introduction

L'évaluation formative a pour objectif d'améliorer l'enseignement et l'apprentissage en fournissant des *feedback* aux enseignants et apprenants afin de les aider à adapter leurs comportements. Sadler définit le *feedback* comme "une information concernant le succès d'une tâche" [28]. Les *feedback* auxquels nous nous intéressons dans ce papier sont les informations utiles pour les enseignants concernant le succès des tâches accomplies par les apprenants, et ce, afin de faciliter la prise de décision pour la suite de la séquence. Cependant, selon Andersson, l'évaluation formative est souvent utilisée de manière informelle ou approximative [1]. Ellis met l'accent sur la tâche complexe consistant à collecter toutes les interactions d'apprentissage en face à face [14]. Ainsi, fournir des *feedback* de qualité est une tâche complexe, surtout dans les configurations à grand échelle où les interactions d'apprentissage augmentent avec le nombre d'apprenants.

Pour résoudre ce problème et pour faire face à l'augmentation du nombre d'étudiants dans l'enseignement supérieur, l'évaluation formative assistée par la technologie a emergé. Différents systèmes implantent des processus variés permettant aux enseignants de mettre en œuvre des séquences d'évaluation formative dans un contexte d'enseignement de masse. Une famille de processus, à savoir "les processus à deux votes", demande aux apprenants de voter deux fois durant la séquence comprenant 5 phases distinctes : (1) l'enseignant pose une question à la classe ; (2) les apprenants donnent leurs premières réponses ; (3) les apprenants réfléchissent aux réponses de leurs pairs et remettent éventuellement en cause leurs connaissances ; (4) les apprenants donnent leurs secondes réponses à la même question ; (5) l'enseignant et les apprenants échangent sur les résultats restitués. Le Peer Instruction (PI), tel que décrit par Mazur [9], est l'une des premières implantations de ce type de processus d'évaluation formative. Ces travaux sur le Peer Instruction [22] ainsi que des études sur d'autres processus à deux votes [7] ont montré un impact positif sur l'engagement des apprenants et leurs apprentissages. L'objectif des processus à deux votes est d'obtenir une augmentation du nombre de réponses correctes entre le premier et le second vote grâce à la confrontation de points de vue entre apprenants. Lorsque c'est le cas, cela signifie qu'en moyenne, le niveau de compréhension des apprenants a augmenté [32]. Ainsi, dans la suite de ce papier, nous qualifierons de "bénéfique" toute séquence d'évaluation formative à deux votes pour laquelle la proportion de bonnes réponses a augmenté lors du second vote.

Ces 5 phases comportent une grande variété d'interactions pouvant générer des traces exploitables. Cependant, étant donné le manque de données relatives aux processus à deux votes [3], peu de travaux ont exploré l'usage de ces interactions pour les transformer en *feedback* pertinents. C'est pourquoi, dans ce papier, nous considérons la question de recherche suivante : Quels *feedback* pourraient être fournis aux enseignants tout au long des différentes phases d'un processus d'évaluation formative à deux votes dispensé à grande échelle, afin d'en améliorer les bénéfices en terme d'apprentissage ? Pour répondre à cette question, nous avons adopté la démarche traditionnelle illustrée par Lebis [20], qui consiste, à

partir des traces de l'EIAH (Environnement Informatique pour l'Apprentissage Humain), à appliquer différentes techniques d'analyse aboutissant à de nouvelles connaissances. Le jeu de données dont nous disposons dans cette étude provient de l'utilisation en contexte écologique, dans l'enseignement supérieur, d'un outil implantant un processus à deux votes.

Le papier est structuré comme suit. La section 2 introduit l'évaluation formative et montre que les *feedback* existants dans les outils d'évaluation formative peuvent être améliorés pour aider les enseignants à prendre des décisions à partir des données collectées au fil de l'activité. La section 3 décrit l'outil d'évaluation formative que nous avons utilisé pour notre étude. La section 4 fournit des détails sur les analyses que nous avons effectuées ainsi que sur les résultats obtenus. Enfin, la section 5 conclut et évoque des pistes pour nos travaux futurs.

2 L'évaluation formative et le *feedback*

2.1 Définitions

Bien que l'évaluation soit souvent utilisée en tant qu'évaluation *de* l'apprentissage, elle peut aussi être utilisé *pour* l'apprentissage [21]. D'un côté, l'évaluation sommative est utilisée pour mesurer le niveau des apprenants à la fin d'une unité d'instruction. De l'autre, l'évaluation formative est cruciale pour permettre aux enseignants d'évaluer la compréhension des apprenants et pour adapter le cours [11]. Hattie évoque l'évaluation formative comme l'une des méthodes à privilégier pour améliorer les résultats des apprenants [17]. En 1998, Black et William définissent l'évaluation formative comme "l'ensemble des activités entreprises par les enseignants, et/ou par leurs élèves, qui fournissent des informations qui seront utilisées comme *feedback* pour modifier les activités d'enseignement et d'apprentissage dans lesquelles ils sont engagés" [4]. Cette définition met en lumière l'importance des données et de leur traitement afin de proposer des *feedback* conçus pour améliorer l'apprentissage et l'enseignement.

Par exemple, lors de cours en face à face, Meltzer et Mannivan ont considéré l'utilisation d'artefacts visuels (tels qu'une feuille de papier ou des cartons) pour permettre aux apprenants de répondre aux questions posées par les enseignants [23]. La présentation des artefacts par les apprenants constituent un *feedback* addressé à l'enseignant lui permettant de prendre connaissance des réponses des apprenants en un coup d'oeil et d'adapter son enseignement. Malheureusement, cette méthode s'applique difficilement à l'enseignement à grande échelle car les enseignants peuvent difficilement visualiser plusieurs dizaines de réponses et les assimiler en temps réel.

2.2 Les *feedback* dans les TEFA

Parmi les solutions qui ont émergé pour appliquer l'évaluation formative à grande échelle avec *feedback* immédiat, les TEFA (*Technology-Enhanced Formative Assessment*) se sont développés [30]. Parmi les systèmes au coeur des TEFA, les CRS (*Classroom Response System*) sont les plus utilisés.

La forme la plus simple d'évaluation formative peut être mise en oeuvre avec des CRS tels que Clickers [6] ou avec des platformes disponibles sur le web telles que Poll Everywhere [8]. Elle permet aux enseignants de poser une question à choix et aux apprenants de sélectionner une réponse. Des *feedback* représentant la répartition des votes sont ensuite présentés aux enseignants. Ils ont pour but de faciliter l'échange entre les enseignants et apprenants à la fin de la séquence. Kahoot [18], Socrative [2] et Plickers [19] supportent le même processus mais proposent en plus d'une vue d'ensemble du vote des apprenants, un *feedback* représentant une vue d'ensemble de la séquence ainsi que les réponses par apprenant sur différentes séquences.

Tirant partie d'une des cinq stratégies caractéristiques de l'évaluation formative identifiées dans la théorie de Black et William [5], ComPAIR [27] utilise les apprenants comme ressources d'apprentissage au sein de son processus d'évaluation formative. Plus précisément, il permet aux apprenants de fournir une réponse textuelle à une question ouverte posée par les enseignants. Par la suite, une phase de revue par les pairs s'engage. Les apprenants doivent fournir un *feedback* textuel sur les réponses de deux de leurs pairs ainsi qu'une justification concernant la raison pour laquelle une des réponses leur paraît plus pertinente qu'une autre. Pendant et après cette phase, les enseignants accèdent à un *feedback* présentant, pour chaque apprenant, la réponse choisie, le *feedback* textuel qu'il a fourni et les comparaisons soumises pour chaque paire de réponses présentée.

Elaastic [13] et myDalite [7] supportent tous deux un processus à deux votes illustré par la Figure 1. Il consiste à demander aux apprenants de voter pour une première réponse et de fournir une explication textuelle pour justifier leur choix. Par la suite, les apprenants peuvent voter une seconde fois. Concernant cette seconde tentative, les deux plateformes se distinguent. myDalite permet aux apprenants de choisir une justification écrite par un pair comme second vote. Ils peuvent également ne pas changer d'avis et garder leurs justifications du premier vote. Les enseignants disposent d'un *feedback* qui détaille le nombre d'apprenants qui sont passés d'une réponse incorrecte à correcte, correcte à incorrecte, incorrecte à incorrecte et correcte à correcte. De son côté, Elaastic implique les apprenants dans une phase d'évaluation avant qu'ils puissent soumettre leur second vote. Il leur est demandé d'évaluer les justifications écrites de leurs pairs. À tout moment, les enseignants peuvent consulter les premiers et seconds votes des apprenants et accéder à l'ensemble de leurs justifications écrites ainsi qu'à la note moyenne attribuée par les pairs à chacune d'entre elles. Plus de détails sont donnés dans la section 3.1.

Les *feedback* présentés dans cette section ont pour objectif de restituer aux enseignants une synthèse du déroulement de la séquence lorsque cette dernière est terminée. Ils ne sont pas conçus pour aider les enseignants à orchestrer des séquences d'évaluation formative en leur fournissant, par exemple, tout au long des différentes phases du processus à deux votes, des informations sur le comportement des apprenants. Pour combler cette lacune, nous souhaitons étudier comment la variété de données générées par les différentes interactions

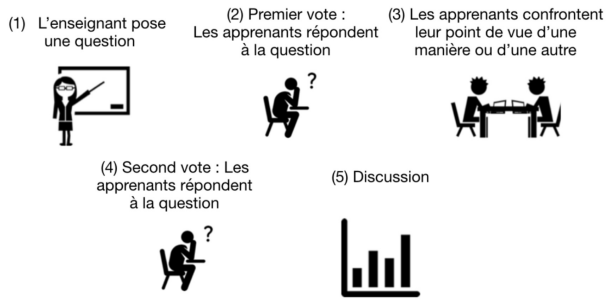


Fig. 1: Les 5 phases du processus à deux votes.

des processus à deux votes constitue une opportunité de construire des *feedback* utiles à la prise de décision.

3 Conception du jeu de données

Nous présentons ici la plateforme d'évaluation formative utilisée ainsi que le jeu de données collectées par son utilisation dans l'enseignement supérieur.

3.1 Elaastic, une technologie au service de l'évaluation formative

Elaastic est une plateforme web [26] utilisée depuis 2015 dans différents contextes d'enseignement dans le supérieur (face à face, distance ou hybride) pour des disciplines telles que la physique, l'informatique et le management de projet.

Durant la phase 1, les enseignants posent une question. Si la question est fermée, elle peut être à choix exclusif ou à choix multiples. Durant la phase 2, les apprenants répondent à la question et fournissent une explication écrite pour justifier leur(s) choix. Ils doivent également fournir leur degré de confiance sur une échelle de Likert à 4 items (voir Figure 2). Durant la phase 3, les apprenants s'engagent dans une activité d'évaluation par les pairs [31]. Plus précisément, il leur est demandé d'exprimer leur degré d'accord avec chaque justification. Ce degré d'accord est fourni sur une échelle de Likert à 5 items (voir Figure 3). Afin de réduire les biais sociaux, les justifications présentées aux apprenants sont anonymes. Les enseignants peuvent configurer le nombre de justifications (5 maximum) présentées à chaque apprenant. Ensuite, la phase 4 commence et les apprenants ont l'opportunité de voter une seconde fois pour la(les) réponse(s) qu'ils pensent être la(les) bonne(s). Enfin, les enseignants démarrent la phase 5. Les votes des apprenants, l'ensemble des justifications écrites ainsi que leurs notes moyennes attribuées par les pairs sont communiquées aux étudiants pour une discussion.

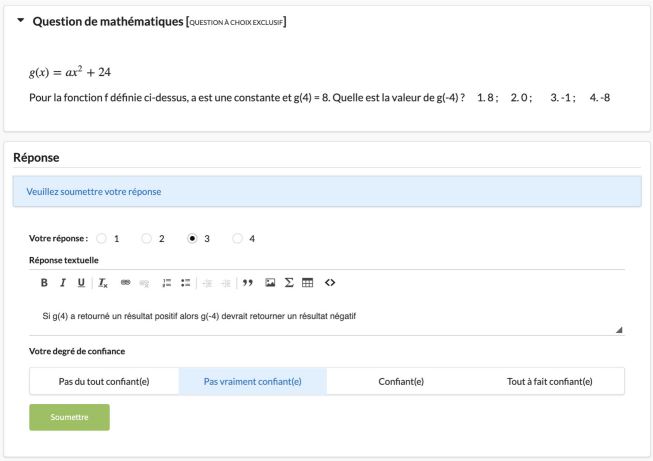


Fig. 2: Elaastic : formulaire de soumission du premier vote.

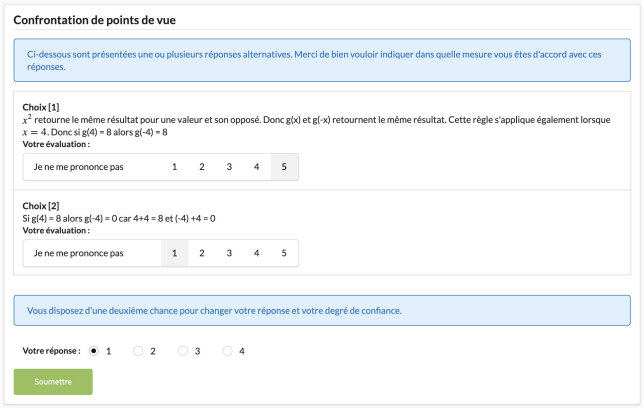


Fig. 3: Elaastic : formulaire de soumission du second vote.

3.2 Description du jeu de données

Jusqu'à présent, la plateforme a été utilisée lors de 623 séquences menées par 53 enseignants au sein desquelles 1769 apprenants ont fourni 8757 réponses ainsi que 9256 évaluations de leurs pairs.

Une séquence est caractérisée par un contexte d'apprentissage (face-à-face, à distance ou hybride), les réponses soumises avant et après la confrontation de point de vue entre pairs, ainsi que le nombre de participants. Pour chaque réponse, les données suivantes ont été collectées : l'identifiant de l'apprenant,

la justification textuelle, le score de la réponse, ainsi que le choix sélectionné dans le cas d'une question à choix. Concernant les réponses correspondant à un premier vote, des informations additionnelles sont collectées telles que la note moyenne attribuée par les pairs et le degré de confiance de l'apprenant qui a fourni la réponse. Les questions sont composées de leur contenu textuel, du type (ouverte, fermée, à choix exclusif, à choix multiples, etc.) et, dans le cas des questions à choix, du nombre de différents choix proposés aux apprenants. Enfin, pour chaque évaluation par les pairs, la justification écrite, la note attribuée, l'identifiant de l'évaluateur et l'identifiant de l'évalué sont fournis.

4 Analyse de données

Afin de réduire l'influence de facteurs externes et des valeurs aberrantes, nous avons filtré notre jeu de données. Dans un premier temps, nous n'avons gardé que les questions à choix afin de pouvoir distinguer de manière automatique une réponse fausse d'une réponse juste. Dans notre analyse, nous considérons une réponse comme correcte si le score obtenu correspond au score maximum. De plus, nous n'avons gardé que les séquences en face-à-face. Puis nous avons écarté les séquences comportant moins de 10 apprenants car nous souhaitons nous concentrer sur les contextes correspondant à un plus grand nombre d'apprenants. Finalement, nous considérons deux variables $p1$ et $p2$ qui sont respectivement la proportion d'apprenants ayant répondu correctement au premier et second vote. Les séquences où $p1 = 0$ ont été enlevées car la confrontation de point de vue ne peut pas avoir lieu (il n'y a pas de justification associée à une bonne réponse pour convaincre ceux qui se sont trompés). Les séquences où $p2 = 1$ ou $p1 = 1$ ont également été enlevées. En effet, elles représentent des questions considérées comme trop faciles pour mesurer un effet. Après avoir nettoyé nos données, nous obtenons 104 séquences menées par 21 enseignants au sein desquelles 616 apprenants ont fourni 1981 réponses ainsi que 4072 évaluations de leurs pairs.

Pour mener nos analyses, nous avons utilisé l'estimation de la taille d'effet de Cohen d proposée par Parmentier [25] qui s'appuie sur $p1$ et $p2$ et est calculée comme suit : $d = 0.6ln\left(\frac{p2}{1-p2}\frac{1-p1}{p1}\right)$. Il est important de mentionner qu'une séquence est bénéfique lorsque $d > 0$ car cela implique que $p1 < p2$. La taille d'effet [29] permet de mesurer l'impact de l'évaluation par les pairs sur les votes et, ainsi, de quantifier les bénéfices apportés par une séquence. De plus, notre volume de données est suffisamment grand (> 40) pour utiliser des tests paramétriques sans satisfaire la contrainte de normalité [16]. À partir de ces éléments, nous avons pu exprimer et vérifier nos hypothèses.

4.1 Hypothèse 1 : Les bénéfices d'une séquence sont liés à la distance de $p1$ à 50%

Watkins et Mazur [22] ont constaté que leur application du Peer Instruction était la plus bénéfique pour les apprenants lorsqu'entre 30% et 70% d'entre eux ont répondu correctement à la question lors du premier vote. Selon eux, un trop

petit pourcentage de réponse correcte indique des connaissances insuffisantes sur la question pour entamer des discussions productives. À l'inverse, un trop haut pourcentage de bonnes réponses signifie que la question est trop facile pour provoquer une confrontation. À partir de ces travaux, nous émettons l'hypothèse que les bénéfices apportés par une séquence sont liés à la distance de $p1$ à 50%.

Les résultats de la Figure 4 indiquent que la taille d'effet diminue lorsque la distance de $p1$ à 50% augmente. La corrélation de Pearson de $|p1 - 0.5|$ avec d

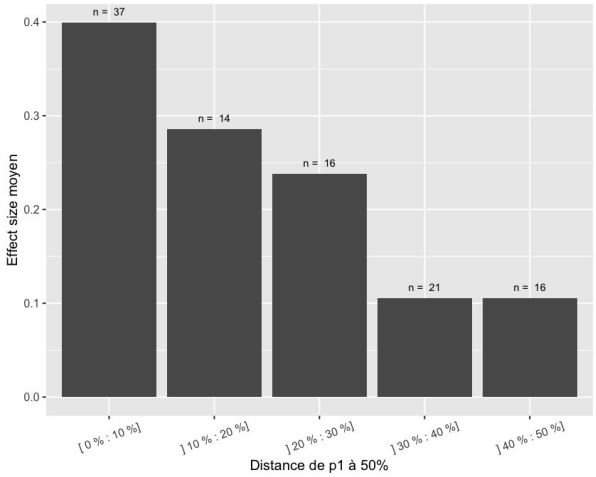


Fig. 4: Taille d'effet moyenne d selon la distance de p1 à 50%.

est égale à -0.31 avec une p-value égale à 0.001 et un intervalle de confiance à 95% égal à [-0.48:-0.13]. Ainsi, plus la proportion de bonne réponse s'éloigne de 50%, plus les bénéfices de la séquence ont des chances de diminuer.

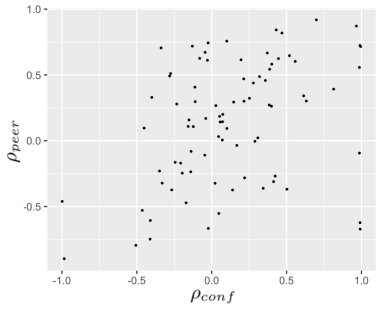
En synthèse, la distance de $p1$ à 50% est un indicateur utile pour prédire si une séquence a des chances d'être bénéfique. Un *feedback* restituant $p1$ pourrait être fourni aux enseignants après le premier vote pour les aider à prendre une décision. Plus précisément, si la distance de p1 à 50% est supérieure à 30%, la Figure 4 montre que la taille d'effet moyenne est significativement plus basse. Dans ce cas, l'enseignant peut sauter la phase de confrontation entre les pairs.

4.2 Hypothèse 2 : La cohérence de l'évaluation par les pairs est liée aux bénéfices d'une séquence

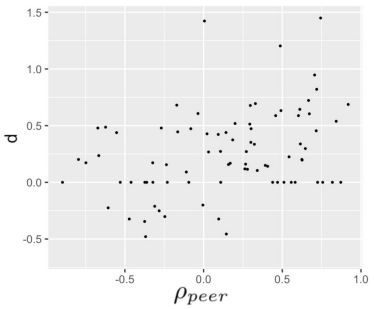
Double et al. montrent que l'utilisation de l'évaluation par les pairs comme pratique formative permet d'améliorer les performances académiques [12]. Ils affirment que réfléchir sur les réponses des pairs favorise l'augmentation du

pourcentage de réponses correctes. Etant donné qu'il est attendu des apprenants ayant fourni une bonne réponse qu'ils convainquent les autres apprenants, nous émettons l'hypothèse que la cohérence de la phase d'évaluation par les pairs est liée aux bénéfices d'une séquence.

Afin de mesurer la cohérence de l'évaluation par les pairs lors d'une séquence, nous définissons ρ_{peer} qui est la corrélation d'un degré d'accord d'un pair à une justification avec l'exactitude de la réponse correspondante. Ces deux variables étant des variables latentes [15], la corrélation polychorique est l'outil statistique approprié [24]. Par exemple, si les justifications des apprenants qui se sont trompés sont bien notées et que celles des apprenants ayant répondu correctement sont mal notées, alors ρ_{peer} sera plutôt proche de -1. À l'inverse, si lors d'une séquence les justifications liées à des réponses correctes sont bien notées et que les justifications liées à des réponses incorrectes sont mal notées, alors ρ_{peer} sera proche de 1. La Figure 5a montre la taille d'effet d en fonction de ρ_{peer}.



(a) Taille d'effet d selon la cohérence de l'évaluation par les pairs ρ_{peer}.



(b) Cohérence de l'évaluation par les pairs ρ_{peer} selon la cohérence du degré de confiance ρ_{conf}.

Fig. 5: Taille d'effet d selon ρ_{peer}, et ρ_{peer} selon ρ_{conf}.

La corrélation de Pearson de ρ_{peer} avec d est égale à 0.34 avec une p-value inférieure à 0.002 et un intervalle de confiance à 95% égal à [0.14:0.52]. Ainsi, plus l'évaluation par les pairs est cohérente, plus les bénéfices de la séquence ont des chances d'être élevés. De plus, ρ_{peer} est indépendante de la distance de $p1$ à 50% (p-value = 0.25), ce qui signifie que les Sections 4.1 et 4.2 ont permis d'identifier deux prédicteurs distincts des bénéfices d'une séquence.

À partir de la valeur de ρ_{peer}, nous proposons la fourniture d'un *feedback* aidant les enseignant à sélectionner les réponses à commenter lors de la phase 5. Par exemple, si $\rho_{peer} < 0$, alors les réponses incorrectes sont plus populaires que les réponses correctes. Ces réponses incorrectes doivent être adressées par les enseignants car elles sont potentiellement des conceptions erronées.

4.3 Hypothèse 3 : la cohérence du degré de confiance est liée avec la cohérence de l'évaluation par les pairs

Curtis a utilisé le degré de confiance des apprenants concernant leur réponse comme un moyen d'identifier les apprenants mal informés [10] qu'il définit comme étant des apprenants très confiants mais ayant fourni une mauvaise réponse. Etant donné qu'il est attendu des apprenants ayant répondu incorrectement qu'ils soit convaincus par les autres, nous pensons que des apprenants mal informés ne peuvent pas correctement noter les justifications de leurs pairs. Plus généralement, nous émettons l'hypothèse que la cohérence du degré de confiance des apprenants est liée à la cohérence de l'évaluation par les pairs.

Pour les même raisons que dans la section 4.2, la cohérence du degré de confiance peut-être calculée en utilisant ρ_{conf} qui est la correlation polychorique du degré de confiance des apprenants avec l'exactitude de leurs réponses au premier vote. La figure 5b est un nuage de point de la cohérence de l'évaluation par les pairs ρ_{peer} en fonction de la cohérence du degré de confiance ρ_{conf}.

La corrélation de Pearson de ρ_{conf} avec ρ_{peer} est égale à 0.38 avec une p-value inférieure à $4e-4$ et un intervalle de confiance à 95% égal à [0.18:0.55]. Cela suggère que la cohérence du degré de confiance des apprenants est liée à la cohérence de l'évaluation par les pairs. Ainsi, plus le degré de confiance des apprenants est cohérent, plus l'évaluation par les pairs a des chances de l'être aussi.

À partir de la valeur de ρ_{conf}, il est envisageable de concevoir un *feedback* pour les enseignants leur indiquant si leurs apprenants sont mal informés. Si $\rho_{conf} < 0$ alors les apprenants ayant répondu incorrectement sont plus confiants que ceux ayant répondu correctement. Dans ce cas, dès la fin du premier vote, un *feedback* pourrait avertir les enseignants que la phase d'évaluation par les pairs est susceptible d'être de mauvaise qualité et ainsi les aider à décider s'ils veulent immédiatement traiter la conception erronée ou raccourcir la séquence.

5 Conclusion

Ce papier s'intéresse à l'évaluation formative, plus précisément, aux bénéfices et défis que constitue la fourniture de *feedback* aux enseignants dans des contextes d'enseignement massif. Nous avons exploré la contribution de la technologie pour adresser les défis de l'évaluation formative et nous avons constaté que les *feedback* conçus pour assister les enseignants pour une prise de décision peuvent être améliorés. Nous avons introduit les processus à deux votes comme une famille de processus au sein de laquelle un vote a lieu avant et après une confrontation de points de vue. Nous nous sommes intéressés à l'identification de nouveaux *feedback* à proposer aux enseignants lors des différentes phases du processus à deux votes afin d'en améliorer les bénéfices. À cette fin, nous avons émis 3 hypothèses en nous appuyant sur une revue de la littérature et avons exploré les données issues de Elaastic pour les vérifier.

Nos résultats suggèrent que les bénéfices d'une séquence dépendent de la cohérence de la phase d'évaluation par les pairs ainsi que de la proportion de

réponses correctes lors du premier vote. De plus, la cohérence de l'évaluation par les pairs dépend de la cohérence du degré de confiance. Ainsi, nous pouvons utiliser ces indicateurs pour accompagner les enseignants dans leur prise de décision au fil de la séquence. Après le premier vote, nous pouvons avertir les enseignants qu'il y a une proportion trop basse ou trop haute de réponses correctes pour que la séquence soit bénéfique. Nous pouvons également utiliser la cohérence du degré de confiance afin d'avertir les enseignants que la phase d'évaluation par les pairs peut ne pas provoquer l'effet attendu. Puis, après l'évaluation par les pairs, nous pouvons dire à l'enseignant si cette dernière était cohérente ou non afin de détecter et traiter les conceptions erronées.

Dans cette étude, nous nous sommes concentrés sur l'identification de *feedback* utiles aux enseignants. Les prochains travaux s'intéresseront à l'implantation de ces *feedback* au sein d'Elaastic. Nous mesurerons ainsi leurs l'impact sur la prise de décisions des enseignants pour mener les séquences. De plus, l'impact de ces décisions sur les bénéfices des séquences sera également étudié.

References

1. Andersson, C., Palm, T.: The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. Learning and Instruction **49**, 92–102 (Jun 2017)
2. Awedh, M., Mueen, A., Zafar, B., Manzoor, U.: Using socrative and smartphones for the support of collaborative learning. arXiv preprint arXiv:1501.01276 (2015)
3. Bhatanagar, S., Zouaq, A., Desmarais, M.C., Charles, E.: A dataset of learnersourced explanations from an online peer instruction environment. International Educational Data Mining Society **13**, 350–355 (2020)
4. Black, P., Wiliam, D.: Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice **5**(1), 7–74 (Mar 1998)
5. Black, P., Wiliam, D.: Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education) **21**(1), 5 (2009)
6. Bruff, D.: Classroom response systems ("clickers") (2015)
7. Charles, E.S., Lasry, N., Bhatnagar, S., Adams, R., Lenton, K., Brouillette, Y., Dugdale, M., Whittaker, C., Jackson, P.: Harnessing peer instruction in-and out-of class with mydalite. In: Education and Training in Optics and Photonics. p. 11143_89. Optical Society of America, SPIE, Quebec City, Quebec, Canada (2019)
8. Clark, S.: Enhancing active learning: Assessment of poll everywhere in the classroom. Tech. rep., University of Manitoba (2017)
9. Crouch, C.H., Mazur, E.: Peer instruction: Ten years of experience and results. American journal of physics **69**(9), 970–977 (2001)
10. Curtis, D.A., Lind, S.L., Boscardin, C.K., Dellinges, M.: Does student confidence on multiple-choice question assessments provide useful information? Medical education **47**(6), 578–584 (2013)
11. Davis, M.: Technology fed growth in formative assessment. Education Week p. 11 (2015)
12. Double, K.S., McGrane, J.A., Hopfenbeck, T.N.: The impact of peer assessment on academic performance: A meta-analysis of control group studies (2020)

13. Elaastic: `https://elaastic.irit.fr`, dernière consultation: 2021-04-23.
14. Ellis, C.: Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. British Journal of Educational Technology **44**(4), 662–664 (2013)
15. Everett, B.: An introduction to latent variable models. Springer Science & Business Media (2013)
16. Ghasemi, A., Zahediasl, S.: Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism **10**(2), 486 (2012)
17. Hattie, J.: Visible learning for teachers: Maximizing impact on learning. Routledge, 711 Third Avenue, New York, NY 10017 (2012)
18. Ismail, M.A.A., Mohammad, J.A.M.: Kahoot: A promising tool for formative assessment in medical education. Education in Medicine Journal **9**(2) (2017)
19. Krause, J.M., O'Neil, K.: Assessment Tool for K–12 and PETE Professionals. Strategies, A Journal for Physical and Sport Educators **30**, 7 (2017)
20. Lebis, A.: Capitaliser les processus d'analyse de traces d'apprentissage: modélisation ontologique & assistance à la réutilisation. Ph.D. thesis, Sorbonne Université (2019)
21. Martinez, M.E., Lipson, J.I.: Assessment for learning. Educational Leadership **46**(7), 73–75 (1989)
22. Mazur, E., Watkins, J.: Just-in-time teaching and peer instruction. In: Just-in-time Teaching: Across the Disciplines, Across the Academy, pp. 39–62. Stylus Publishing, LLC, 22883 Quicksilver Drive, Sterling, Virginia 20166-2102 (2010)
23. Meltzer, D.E., Manivannan, K.: Transforming the lecture-hall environment: The fully interactive physics lecture. American Journal of Physics **70**(6), 639–654 (2002)
24. Olsson, U.: Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika **44**(4), 443–460 (1979)
25. Parmentier, J.F.: How to quantify the efficiency of a pedagogical intervention with a single question. Physical Review Physics Education Research **14**(2), 020116 (2018)
26. Parmentier, J.F., Silvestre, F.: La (dé-)synchronisation des transitions dans un processus d'évaluation formative exécuté à distance : impact sur l'engagement des étudiants. In: 9ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2019). pp. 97–108. ATIEF, Sorbonne Universite, LIP6Paris, France (2019)
27. Potter, T., Englund, L., Charbonneau, J., MacLean, M.T., Newell, J., Roll, I., et al.: Compair: A new online tool using adaptive comparative judgement to support learning with peer feedback. Teaching & Learning Inquiry **5**(2), 89–113 (2017)
28. Sadler, D.R.: Formative assessment and the design of instructional systems. Instructional Science **18**(2), 119–144 (Jun 1989)
29. Sawilowsky, S.S.: New effect size rules of thumb. Journal of Modern Applied Statistical Methods **8**(2), 26 (2009)
30. Spector, J.M., Ifenthaler, D., Sampson, D., Yang, J.L., Mukama, E., Warusavitarana, A., Dona, K.L., Eichhorn, K., Fluck, A., Huang, R., et al.: Technology enhanced formative assessment for 21st century learning. International Forum of Educational Technology and Society **19**(3), 58–71 (2016)
31. Topping, K.J.: Peer assessment. Theory into practice **48**(1), 20–27 (2009)
32. Tullis, J.G., Goldstone, R.L.: Why does peer instruction benefit student learning? Cognitive Research: Principles and Implications **5**(1), 15 (Dec 2020)