

Évaluer la réutilisabilité d'une question : une utilisation des *learning analytics* dans un contexte d'évaluation formative

Rialy Andriamiseza

Institut de recherche en Informatique de Toulouse, Université Toulouse III,
118 Route de Narbonne, 31400 Toulouse, France
rialy.andriamiseza@irit.fr

Résumé. De nombreux systèmes permettent d'assister l'enseignant dans la mise en oeuvre d'évaluations formatives. Ils permettent la collecte de données et proposent des feedbacks qui visent à faciliter la prise de décision. Ce papier propose d'élargir le périmètre de mobilisation des *learning analytics* dans les systèmes d'évaluation formative. Nous illustrons notre proposition sur une question précise : comment évaluer la réutilisabilité d'une question mobilisée dans le contexte d'une séquence d'évaluation formative ? Nous présentons le cadre théorique, les hypothèses formulées pour répondre à cette question, ainsi que le contexte expérimental qui permettra de les évaluer.

Mots-clefs: évaluation formative, *learning analytics*, Elaastic, instruction par les pairs, objet pédagogique réutilisable, processus en N phases

1 Introduction

Dans un contexte d'évaluation formative supportée par des technologies (*Technology-Enhanced Formative Assessment - TEFA*), les données collectées sont utilisées sur un périmètre qui pourrait être élargi. Les *learning analytics* sont déjà mobilisés afin d'aider l'enseignant dans sa prise de décision et permettent de satisfaire les exigences premières de l'évaluation formative. Par exemple, certains systèmes d'évaluation formative proposent l'agrégation et l'affichage de la distribution des votes concernant une question à choix exclusif sous forme d'histogramme. Nous nous intéressons à un feedback qui contient des données concernant l'impact d'une question sur l'apprentissage et qui permettrait à un enseignant de décider si cette question est réutilisable ou pas. En effet, l'aide à la décision concernant la réutilisabilité d'une question est une possibilité de mobilisation des *learning analytics* peu traitée par les TEFA.

Comment évaluer la réutilisabilité d'une question mobilisée dans le contexte d'une séquence d'évaluation formative ? Nous nous positionnons sur une réutilisabilité évaluée à partir de critères s'intéressant à l'impact de la question sur l'apprentissage. Dans ce contexte, nous faisons la distinction entre 2 niveaux de réutilisabilité : la réutilisation d'une question avec un autre groupe ou une

autre classe d'un même niveau par l'enseignant qui a conçu la question, et la réutilisation d'une question par d'autres enseignants.

Ce papier est structuré comme suit : la première partie introduit l'évaluation formative et présente son articulation actuelle avec les *learning analytics*. La seconde partie décrit le socle théorique sur lequel s'appuient nos travaux, et la troisième partie se concentre sur le contexte expérimental. Enfin, le papier conclut sur les pistes de nos travaux futurs.

2 Travaux connexes

2.1 L'évaluation formative

Si l'évaluation sommative a pour objectif de certifier que l'apprenant maîtrise un sujet donné, une autre forme d'évaluation a pour but d'aider l'enseignant et l'apprenant à prendre des décisions pour améliorer l'apprentissage. Dès 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" [3]. Un feedback représente un élément clé dans l'évaluation formative, et est habituellement défini comme "une information concernant le succès d'une tâche entreprise" [14].

D'après Andersson, l'évaluation formative est souvent pratiquée de manière informelle et approximative [1]. En effet, dans le contexte de cours en face à face et sans assistance, la difficulté de sa mise en oeuvre augmente avec le nombre d'apprenants impliqués.

Les systèmes technologiques interactifs supportant les processus d'évaluation formative (référéncés dans la suite du papier par l'expression Systèmes d'Évaluation Formative ou SEF) apportent un début de solution. Dans un premier temps, les Clickers [2] font leur apparition. Ils proposent une agrégation simple et générique des réponses des apprenants (par exemple, sous forme d'histogramme ou de diagramme circulaire représentant la répartition des votes sur une question à choix exclusif). Par la suite, le développement d'applications web telles que Plickers [12] et Poll Everywhere [6] a permis d'augmenter les interactions avec des systèmes qui s'affranchissent des limites liées aux dispositifs matériels propriétaires. Parmi ces systèmes, Elaastic [13] et myDalite [13] (évolutions respectives de Tsaap-Notes et Dalite) s'inspirent de l'instruction par les pairs [8,15] pour proposer une version étendue de l'évaluation formative avec une confrontation de point de vue sur la base d'une argumentation écrite.

Les systèmes d'évaluation formative nécessitent la collecte des traces des apprenants pour la restitution de feedbacks plus riches. Ils mobilisent donc ce que la communauté scientifique désigne sous l'expression *learning analytics*.

2.2 Les *learning analytics*

Les *learning analytics* sont encore un champ de recherche relativement récent [17]. En 2011, date de la toute première conférence LAK (Learning Analytics

and Knowledge), une définition est proposée : “Les *learning analytics* sont la mesure, la collecte, l’analyse et la communication de données sur les apprenants et leurs contextes, dans le but de comprendre et d’optimiser l’apprentissage et les environnements dans lesquels il se produit” [18].

Les *learning analytics* sont déjà exploités par les SEF pour répondre aux exigences fondamentales de l’évaluation formative telles que la présentation des résultats des apprenants. L’objectif visé est la fourniture d’un feedback immédiat à l’aide d’une visualisation adaptée. L’amélioration continue de l’apprentissage visée par l’évaluation formative inscrit naturellement les *learning analytics* dans un cycle vertueux tel que celui proposé par Clow [7]. Sur la Figure 1, l’élément

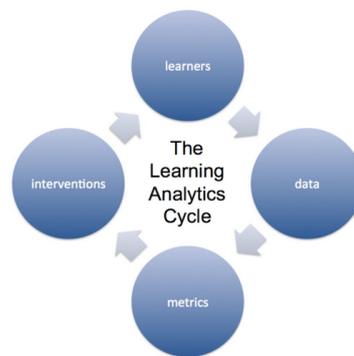


Fig. 1. Le cycle des *learning analytics* [7]

”learners” représente les apprenants qui vont générer des informations. “data” correspond aux traces collectées par le système, “metrics” représente les indicateurs calculés à partir des traces, et “interventions” comprend toutes les actions qui ont effet sur l’apprenant. Ces interventions peuvent être la restitution du feedback aux apprenants (sous forme de tableaux de bord par exemple) ou une intervention directe de l’enseignant auprès de l’apprenant. À plus grande échelle encore, une intervention peut être la modification d’une technologie et dans notre cas d’un système d’évaluation formative.

La plateforme MyDalite propose déjà des feedbacks qui vont au-delà des exigences fondamentales précédemment citées. Ces feedbacks présentent par exemple le nombre d’apprenants qui sont passés d’une réponse correcte à une réponse incorrecte, et le nombre d’apprenants étant passés d’une réponse incorrecte à une réponse correcte. Ces indicateurs représentent un bon moyen pour l’enseignant de mesurer l’impact de la confrontation de points de vue [5].

À partir de ces premiers travaux, nous souhaitons étendre l’utilisation des *learning analytics* en nous intéressant à la qualité des questions posées aux apprenants. En effet, d’après Fisher [9], elles sont un élément clé de la réussite d’une séquence d’évaluation formative. Plus précisément, nous proposons d’étudier

comment les *learning analytics* peuvent aider l'enseignant à évaluer la qualité d'une question en vue de la réutiliser ou non pour une séquence d'évaluation formative future. Pour identifier des indicateurs qui serviront notre objectif, nous allons nous appuyer sur un cadre théorique après avoir introduit le processus d'évaluation formative dans lequel nous inscrivons nos travaux.

3 Cadre théorique

3.1 Le processus en N phases

En 2009, Black et William [4] énoncent 5 activités et 5 stratégies caractérisant la mise en oeuvre d'une séquence d'évaluation formative. Par exemple, une des activités évoquées est le questionnement de la classe, et une des stratégies est d'amener les apprenants à devenir des ressources d'instruction pour les autres apprenants. Dans le cadre de sa thèse [19], Silvestre s'est appuyé sur ces travaux pour proposer un processus d'évaluation formative qui, en comparaison à d'autres processus s'appuyant sur les technologies tels que les SEPT (Systèmes d'Echange et de Partage de Texte) ou encore les SVI (Systèmes de Vote Interactif), supporte l'ensemble de ces activités et stratégies. Il s'agit du processus en N phases qui, comme son nom l'indique, se décompose en plusieurs phases.

La première phase est celle où les apprenants répondent à la question en cochant la/les bonnes réponses et en rédigeant une justification de celle(s)-ci. Si la question est une question ouverte, seule une justification écrite est demandée. La seconde phase est celle où les apprenants se voient attribuer une ou plusieurs réponses (accompagnées de leur justification) fournies par d'autres apprenants. Ils/elles doivent évaluer ces réponses sur une échelle de 1 à 5. Cette confrontation de points de vue permet aux apprenants de porter un second regard sur leur propre réponse et d'éventuellement changer d'avis. Enfin, la troisième phase est la phase de restitution au sein de laquelle l'enseignant échange avec les apprenants sur les réponses et les évaluations réalisées entre pairs. Le processus a été implanté initialement dans la plateforme Tsaap-Notes¹ puis dans la plateforme Elaastic² activement maintenue à ce jour.

Nous inscrivons nos travaux de recherche dans le contexte d'utilisation du processus en N phases car l'utilisation de ce dernier est une occasion de l'évaluer et de l'analyser. De plus, le support exhaustif des activités et stratégies caractérisant l'évaluation formative par le processus en N Phases induit une grande diversité d'interactions qui rend possible la collecte de données variées : les votes, les degrés de confiance, les changements d'avis, les arguments des apprenants, les évaluations par les pairs. Elles sont autant de dimensions exploitables pour répondre à notre question de recherche.

¹ <http://notes.tsaap.eu/tsaap-notes/>

² <https://elaastic.irit.fr/elaastic-questions/>

3.2 Réutilisabilité d'une question dans le contexte de l'évaluation formative

Pour assister l'enseignant dans l'évaluation de la réutilisabilité d'une question, le SEF utilisé pouvoir qualifier cette question. En ce qui concerne l'évaluation formative, nous assimilons une question posée dans un processus en N phases à un objet pédagogique réutilisable (ou RLO pour *Reusable Learning Object*).

En 2011, les travaux de Sanz Rodriguez [16] proposaient des indicateurs de réutilisabilité pour les *learning objects*. Les métriques proposées par ce modèle permettent de spécifier les conditions d'utilisation de la ressource mais elles ne donnent pas d'indication sur l'impact de la ressource en terme d'apprentissage. Sanz Rodriguez définit la réutilisabilité comme "le degré auquel un *learning object* peut fonctionner (...) pour différents utilisateurs pour différents environnements digitaux et dans différents contextes éducationnels". Notre objectif est d'évaluer la réutilisabilité d'une question sur la base de l'effet de la question sur l'apprentissage quand elle se déroule dans une séquence d'évaluation formative.

En 2017, les travaux de Koh et Linge sur les RLO [11] proposent un modèle évaluant la réutilisabilité d'un objet pédagogique en s'appuyant sur la mesure du niveau d'engagement des étudiants dans l'activité qui intègre l'objet. La mesure de cet engagement s'effectue sur 5 dimensions proposées par Howland et al. [10] pour caractériser l'activité d'apprentissage : active, constructive, authentique, intentionnelle et collaborative. Dans notre contexte d'activités d'évaluation formative, ce modèle s'intéresse à l'effet de la question sur l'apprentissage. Plus précisément, il mesure l'engagement cognitif induit par la ressource et se concentre sur la performance des apprenants (feedback, compétences/connaissances). Nous pouvons appuyer nos travaux sur ce modèle. Ainsi, nous émettons l'hypothèse qu'une question est réutilisable quand l'activité d'apprentissage qui l'intègre (la mise en œuvre du processus en N phases pour cette question dans notre cas) est évaluée positivement sur les 5 dimensions de Koh et Linge. Le Tableau 1 propose une interprétation des 5 dimensions dans ce contexte. Nous avons identifié, pour chacune des dimensions, des indicateurs susceptibles de participer à la qualification de l'exécution du processus en N phases pour une question donnée.

Maintenant que des indicateurs ont été identifiés, nous devons d'une part vérifier qu'ils permettent de mesurer la réutilisabilité d'une question et, si c'est le cas, identifier les valeurs significatives des indicateurs dans chaque dimension. Le paragraphe suivant présente le cadre expérimental permettant l'étude de nos indicateurs en contexte écologique.

4 Cadre expérimental

Le projet B4MATIVE! né d'une collaboration entre l'Académie de Nancy-Metz et l'IRIT (Institut de Recherche et d'Informatique de Toulouse) vise à offrir l'opportunité aux enseignants de prendre des décisions de réingénierie pédagogique. Le projet bénéficie d'un terrain d'expérimentation, le "lycée 4.0" comprenant 50 établissements pilotes, dont l'objectif est d'équiper l'ensemble des lycées

Table 1. Les indicateurs de la réutilisabilité d'une question dans le contexte du processus en N phases

Dimension	Description	Indicateurs
Active	L'apprenant participe à une activité : sélectionne un item ou évalue ses pairs.	Le taux de participation est élevé à la phase 1 (part de participants inscrits au devoir qui a soumis une réponse) et à la phase 2 (part de participants ayant soumis une réponse à la phase 1 qui a soumis une réponse à la phase 2)
Constructive	L'apprenant utilise ses capacités cognitives	L'argumentation textuelle est de taille cohérente avec le temps passé entre la lecture et la réponse. Les apprenants expriment leur degré de confiance
Intentionnelle	L'apprenant a des opportunités de remédiation après feedback	La tendance de changement d'avis est significative. Le taux de réussite augmente après la phase 2
Authentique	L'apprenant utilise des compétences liées aux connaissances que le cours lui a apportées	Les apprenants ont un bon ressenti concernant l'impact de la séquence sur leur apprentissage
Collaborative	La séquence favorise les confrontations de point de vue et les discussions	La tendance de changement d'avis est positive. Les taux de participation sont élevés. L'écart-type moyen des évaluations de l'ensemble des réponses est significatif

du Grand Est avec des outils numériques capables de répondre aux besoins d'enseignement d'aujourd'hui. Pour accomplir cela, les enseignants ont exprimé le besoin d'un outil numérique favorisant une évaluation formative individualisée. La plateforme Elaastic, développée à l'IRIT depuis 2015, est déployée dans le cadre du projet B4MATIVE!. Elle permet de poser une question et de l'exécuter dans le contexte du processus en N phases. Elaastic convient à dans n'importe quel niveau d'enseignement et quelque soit la matière. Un travail de développement sera réalisé pour adapter la plateforme à ces expérimentations car dans sa version actuelle, Elaastic ne collecte pas tous les indicateurs mentionnés.

L'objectif des expérimentations sera d'analyser des données d'utilisation pour vérifier la corrélation entre les valeurs des indicateurs mentionnés dans la section 3 et la réutilisation réelle des questions par les enseignants d'une classe à une autre. Pour mesurer cette réutilisation réelle, la plateforme Elaastic doit pouvoir fournir deux indicateurs. En effet, comme mentionné précédemment (voir Section 1), nous considérons deux niveaux de réutilisabilité. Ainsi, pour une question, la réutilisabilité par le même enseignant avec un autre groupe sera mesurée par un indicateur dont la valeur sera le nombre d'utilisation tandis que la réutilisabilité de la question par un enseignant autre que le concepteur sera mesurée par un indicateur dont la valeur sera le nombre de partage. Chaque question doit pouvoir

être réutilisée et partagée. C'est-à-dire qu'un enseignant qui utilise une question doit avoir l'occasion de la réutiliser avec un autre groupe ou une autre classe de même niveau ou de la partager pour permettre à d'autres enseignants de l'utiliser.

Si cette corrélation est établie, nous nous concentrerons sur l'identification de valeurs "seuils" des indicateurs. Nous nous donnons 6 mois de collecte de données afin que ces dernières soient significatives, ainsi qu'un minimum de 30 enseignants.

5 Conclusion

Nous avons introduit les concepts d'évaluation formative et de *learning analytics* et nous avons montré que leur association concerne un périmètre que nous souhaitons élargir. A cette fin, la principale contribution de ce papier est la définition d'indicateurs pour la réutilisabilité d'une question dans le cadre du processus en N phases. Enfin, nous avons décrit le cadre expérimental permettant de valider ces indicateurs en situation réelle.

Nous avons proposé, dans ce papier, un seul nouvel axe de mobilisation des *learning analytics* dans le contexte de l'évaluation formative. Pourtant, le périmètre de mobilisation des *learning analytics* pourrait s'étendre à d'autres problématiques posées par l'utilisation des SEF. Par exemple, il serait intéressant qu'un apprenant puisse disposer d'un tableau de bord lui permettant de faciliter ses séances de révisions.

Références

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. Beatty, I.D., Gerace, W.J.: Technology-enhanced formative assessment: A research-based pedagogy for teaching science with classroom response technology. *Journal of Science Education and Technology* **18**(2), 146–162 (2009)
3. Black, P., Wiliam, D.: Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice* **5**(1), 7–74 (Mar 1998)
4. 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)
5. 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 (2019)
6. Clark, S.: *Enhancing active learning: Assessment of poll everywhere in the classroom* (2017)
7. Clow, D.: *The learning analytics cycle: closing the loop effectively* (2012)
8. Crouch, C.H., Mazur, E.: Peer instruction: Ten years of experience and results. *American journal of physics* **69**(9), 970–977 (2001)

9. Fisher, D., Frey, N.: *Checking for understanding: Formative assessment techniques for your classroom*. ASCD (2014)
10. Howland, J.L., Jonassen, D.H., Marra, R.M.: *Meaningful Learning with Technology: Pearson New International Edition*. Pearson Higher Ed (2013)
11. Koh, J.H.L.: Designing and integrating reusable learning objects for meaningful learning: Cases from a graduate programme. *Australasian Journal of Educational Technology* **33**(5) (2017)
12. Krause, J.M., O’Neil, K.: *Assessment Tool for K–12 and PETE Professionals* p. 7 (2017)
13. 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 (2019)
14. Sadler, D.R.: Formative assessment and the design of instructional systems. *Instructional Science* **18**(2), 119–144 (Jun 1989)
15. Sadler, P.M., Good, E.: The impact of self-and peer-grading on student learning. *Educational assessment* **11**(1), 1–31 (2006)
16. Sanz-Rodriguez, J., Doderio, J.M., Sanchez-Alonso, S.: Metrics-based evaluation of learning object reusability. *Software Quality Journal* **19**(1), 121–140 (2011)
17. Siemens, G.: Learning analytics: The emergence of a discipline. *American Behavioral Scientist* **57**(10), 1380–1400 (2013)
18. Siemens, G., Long, P.: Penetrating the fog: Analytics in learning and education. *EDUCAUSE review* **46**(5), 30 (2011)
19. Silvestre, F.: *Conception et mise en oeuvre d’un système d’évaluation formative pour les cours en face à face dans l’enseignement supérieur*. Ph.D. thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier (2015)