



**HAL**  
open science

# Implanter un processus d'évaluation formative : le défi de l'explicabilité

Rialy Andriamiseza

► **To cite this version:**

Rialy Andriamiseza. Implanter un processus d'évaluation formative : le défi de l'explicabilité. 9èmes rencontres Jeunes Chercheur.e.s en Environnements Informatiques pour l'Apprentissage Humain (RJC EIAH 2022), ATIEF (Association des Technologies de l'Information pour l'Education et la Formation), May 2022, Lille, France. pp.68-74. hal-03894424

**HAL Id: hal-03894424**

**<https://hal.science/hal-03894424>**

Submitted on 12 Dec 2022

**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.

# Implanter un processus d'évaluation formative : le défi de l'explicabilité

Rialy Andriamiseza, 3ème année de thèse<sup>1</sup>[0000–0002–9995–6269]

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

**Résumé.** Ce papier décrit et justifie les choix qui ont été faits concernant l'implantation d'un modèle d'aide à la décision pour enseignant au sein d'un système d'évaluation formative. A partir d'une étude de la littérature, nous identifions des principes fondamentaux à l'explicabilité des systèmes d'enseignement et proposons des artefacts IHM qui mettent en oeuvre ces principes afin de rendre notre système explicable. L'évaluation de ces artefacts sera au coeur de nos prochains travaux.

**Mots-clés:** Explicabilité · tableau de bord · aide à la décision

## 1 Introduction

Lors de précédents travaux, nous avons mené une étude qui analysait les données issues de l'utilisation d'une plateforme d'évaluation formative dans l'enseignement supérieur [2]. Cela nous a permis de concevoir un processus sophistiqué destiné à aider les enseignants à prendre des décisions pour maximiser les bénéfices d'une séquence d'évaluation formative pour les apprenants. La prochaine étape est l'expérimentation de ce nouveau processus pour le mettre à l'épreuve. Pour cela, l'implantation du processus est requise mais nécessite plusieurs réflexions en amont. En effet, la question de l'aide à la décision d'un système par rapport à celle d'un humain est au coeur de la recherche actuelle [13]. Ce domaine est transverse à celui de l'explicabilité qui est un enjeu majeur permettant d'obtenir la confiance de l'utilisateur [8]. Celle-ci est primordiale lorsqu'il y a une contrainte de temps et donc de résultat [7], ce qui est le cas dans les contextes d'éducation. Cependant, une revue de la littérature d'Alamri montre que peu de travaux se sont intéressés à l'explicabilité dans le domaine de l'éducation [1]. Par conséquent, nous souhaitons adresser la problématique suivante : Comment l'implantation d'un processus d'aide à la décision au sein d'un système utilisé par des enseignants doit-il adresser les enjeux d'explicabilité ?

## 2 Contexte

### 2.1 Elaastic

Dans une précédente étude, nous avons étudié le processus à deux votes qui peut s'illustrer par les étapes décrites dans la Figure 1. Le processus proposé



**Fig. 1.** Les cinq étapes du processus à deux votes.

par la plateforme Elaastic <sup>1</sup> est une implantation de ce processus et est celui sur lequel nous avons mené nos analyses. Plus précisément, l'implantation du processus à deux votes proposée par Elaastic se découpe en 3 phases (1) Les apprenants répondent à la question à choix posée par l'enseignant. A cette fin, ils doivent sélectionner le choix qu'ils pensent être le bon, rédiger une justification et fournir leur degré de confiance. (2) Les apprenants voient un ensemble de réponses anonymisées de leurs pairs contenant le choix sélectionné ainsi que la justification associée. Ils doivent évaluer chacune de ces réponses en fournissant une note chiffrée (allant de 1 à 5). Après cela, ils peuvent à nouveau répondre à la question posée par l'enseignant en sélectionnant à nouveau un choix pour éventuellement changer d'avis. (3) Les apprenants et l'enseignant échangent sur les résultats des activités entreprises lors des deux phases précédentes. Pour cela, l'enseignant dispose d'un histogramme résumant les deux votes ainsi que de l'ensemble des justifications écrites rédigées avec leur note moyenne attribuée. Il peut partager ces *feedback* avec les apprenants.

## 2.2 Processus d'évaluation formative proposé

A partir de ce processus, nous avons proposé une version plus complexe proposant des actions que les enseignants doivent entreprendre lors d'une séquence de questionnement de classe. Comme on peut le voir sur la Figure 2, des indicateurs complexes sont utilisés pour déterminer la meilleure action à entreprendre.

Ces indicateurs sont les suivants :

- $p1$  est la proportion de bonnes réponses à la fin de la phase 1.
- $\rho_{conf}$  est la cohérence du degré de confiance. Cette dernière mesure l'adéquation du degré de confiance exprimé par l'apprenant avec l'exactitude de sa réponse.
- $\rho_{peer}$  est la cohérence de l'évaluation par les pairs. Cette dernière mesure l'adéquation de la note chiffrée qu'une justification a reçue avec l'exactitude de la réponse à laquelle cette justification est attachée.
- $d$  est la taille d'effet de la phase d'évaluation par les pairs sur les réponses des apprenants. Elle se calcule comme suit :  $d = 0.6 \ln \left( \frac{p2}{1-p2} \frac{1-p1}{p1} \right)$  où  $p2$  est la proportion de bonnes réponses à la fin de la phase 2. Lorsque la proportion de bonne réponse a augmenté au second vote ( $p2 > p1$ ), la taille d'effet  $d$  est positive. Dans ce cas, la séquence est dite bénéfique car cela signifie que le niveau de compréhension général des apprenants a augmenté [14].

<sup>1</sup> <https://elaastic.irit.fr/elaastic-questions/>

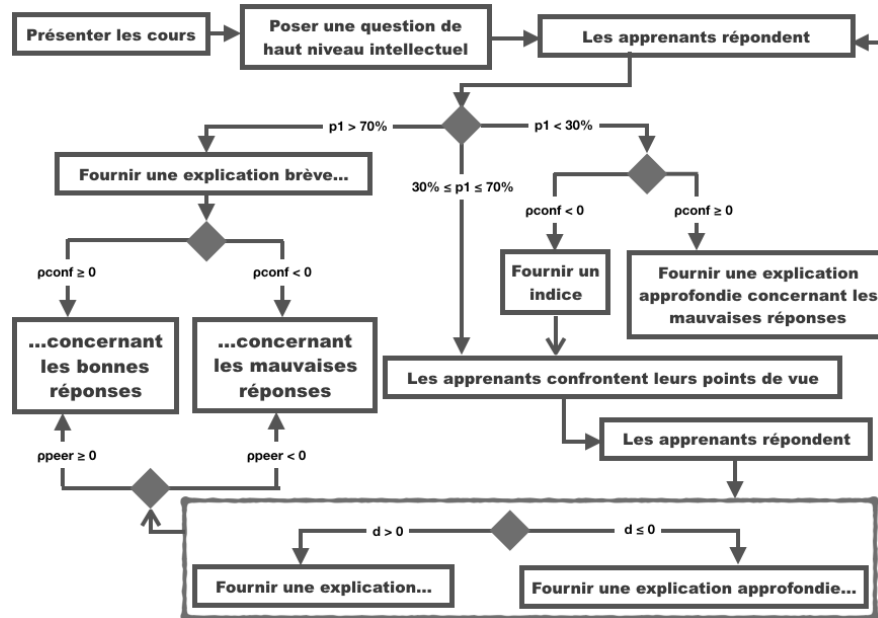


Fig. 2. Un processus d'évaluation formative s'appuyant sur des indicateurs complexes.

Si ce processus s'appuie sur les données pour prendre une décision destinée à maximiser les bénéfices d'une séquence, les enseignants eux ont plutôt tendance à privilégier leur intuition [15]. En effet, dans les contextes sociétaux relevant de l'attitude d'individus, les utilisateurs sont sceptiques concernant la capacité d'un système à prendre la meilleure décision [3]. Ainsi, nous souhaitons exploiter le modèle présenté par ce processus en tant que modèle d'aide à la décision plutôt que de modèle de prise de décision. Si un système souhaite aider l'enseignant à prendre des décisions via des recommandations, il doit être capable de justifier ces recommandations. Cela implique qu'il doit argumenter ses propositions et donc être en mesure d'expliquer le modèle sur lequel il s'appuie. C'est là que l'explicabilité entre en jeu.

### 3 Etat de l'art sur l'explicabilité

Dans une synthèse des définitions de l'explicabilité [11], il est indiqué qu'un système est explicable (i) s'il prend en compte le destinataire de l'explication et (ii) s'il répond à la question "Pourquoi ?".

Au niveau de la prise en compte du destinataire, les explications doivent être adaptées à ce dernier. En effet, lorsqu'une explication donne trop de détails, n'est pas interactive, demande trop de connaissances techniques ou est insuffisante, cela peut supprimer la curiosité et surtout introduire des conceptions erronées chez l'utilisateur [7]. Une explication doit satisfaire la curiosité, celle-ci

est déclenchée lorsque l'utilisateur manque de connaissance, ne comprend pas le système ou observe un comportement contraire à ses attentes [9]. Etant donné que cette curiosité est spécifique à chaque individu, un système idéal devrait présenter des explications plus ou moins détaillées selon sa curiosité. Dans notre cas, la curiosité de l'utilisateur (c'est-à-dire l'enseignant) n'est pas mesurable. Nous décidons donc de laisser le choix aux enseignants du niveau de détail des explications dont ils veulent bénéficier.

Concernant la réponse à la question "Pourquoi ?", plusieurs études [4, 6] proposent de montrer à l'utilisateur une situation similaire dont l'issue est connue (modèle du voisin le plus proche). Dans notre situation, cela reviendrait à montrer à l'enseignant une instance antérieure du processus pour servir de points de comparaison. Cependant, les probabilités et généralisations statistiques ne sont pas une explication satisfaisante pour eux à moins que cette dernière ne soit accompagnée d'une explication qualitative [10]. Par conséquent, nous décidons de mettre en avant les arguments qualitatifs pour nos explications et de les accompagner d'un argument quantitatif ponctué par une représentation graphique des indicateurs déterminants.

A partir de ces décisions, nous sommes en mesure de proposer des artefacts IHM dans notre système d'évaluation formative.

## 4 Implantation du processus dans Elaastic

Dans cette section, nous décrivons le travail que nous avons effectué lors de l'implantation du processus et la mise en oeuvre de son explicabilité. Les explications sont présentées aux enseignants à la fin des phases 1 et 2. A titre d'exemple, si à la fin de la phase 1,  $p1 < 30\%$  et  $\rho_{conf} < 0$ , une recommandation apparaît (voir Figure 3). Elle est accompagnée d'une brève explication.

Nous vous recommandons de sauter la phase 2 qui ne sera pas bénéfique pour les apprenants et de passer directement à la phase 3. [En savoir plus...](#)

**Fig. 3.** Recommandation : explication brève.

L'enseignant peut obtenir une explication plus détaillée si sa curiosité n'est pas satisfaite. En cliquant sur "En savoir plus", une fenêtre contextuelle telle que sur la Figure 4 apparaît. Elle propose une explication textuelle contenant des arguments qualitatifs (identifiés lors de notre précédente étude) précédés des valeurs des indicateurs pertinents. Ces indicateurs calculés à partir de variables latentes dont les modèles sont difficiles à expliquer n'ont pas de signification intuitive [5]. Pour contourner ce problème, nous avons donc exploité leurs propriétés mathématiques. Dans cet exemple, nous souhaitons indiquer à l'enseignant que le degré de confiance des apprenants est incohérent ( $\rho_{conf} < 0$ ). Cela implique plus concrètement que les apprenants ayant fourni une réponse correcte sont en moyenne moins confiants que ceux ayant fourni une réponse incorrecte.

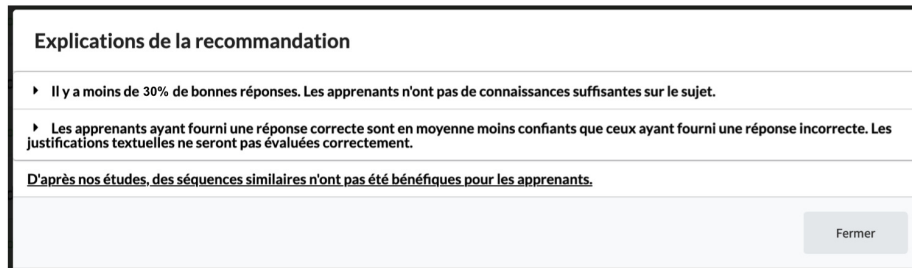


Fig. 4. Recommandation : explication textuelle.

Si ces explications ne sont toujours pas suffisantes pour les enseignants, ils peuvent cliquer sur une affirmation pour visualiser un diagramme illustrant la valeur de l'indicateur concerné. La Figure 5 présente le diagramme qui s'affiche lorsque l'enseignant clique sur l'affirmation concernant la proportion de bonnes réponses au premier vote ( $p1 < 30\%$ ). Bien que la plateforme Elastic possède

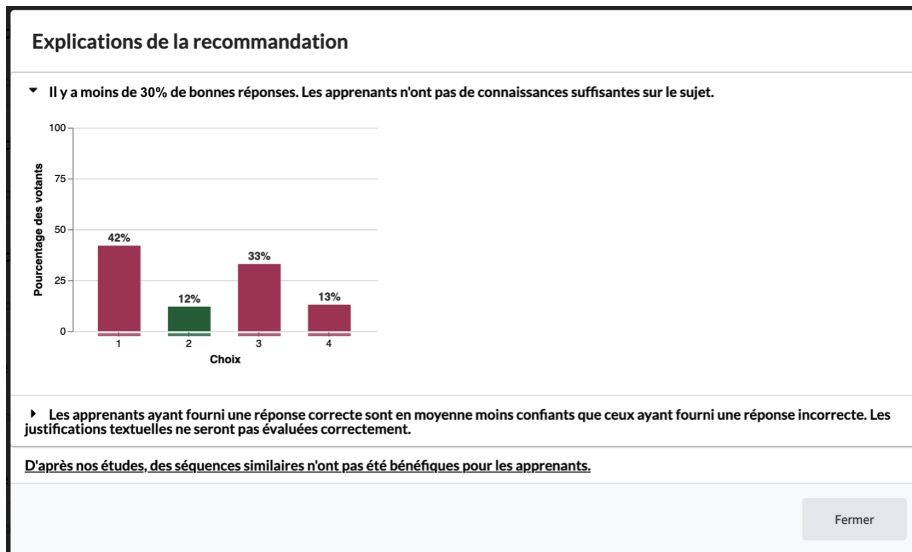


Fig. 5. Recommandation : explication textuelle avec diagramme.

déjà des diagrammes visant à restituer les résultats des votes (et donc les indicateurs  $p1$  et  $d$ ), elle ne possède aucun diagramme visant à restituer la cohérence du degré de confiance ni la cohérence de l'évaluation par les pairs (et donc les indicateurs  $\rho_{conf}$  et  $\rho_{peer}$ ). En combinant l'étude de Robbins sur les avantages et inconvénients des différentes manières de représenter des données récupérées sur une échelle de Likert [12] avec les résultats que nous avons obtenus lors d'un ate-

lier centré utilisateur impliquant 3 enseignants, nous avons décidé de représenter  $\rho_{conf}$  tels que dans la Figure 6. Cette figure restitue la répartition des différents degrés de confiance exprimés par les apprenants et regroupés en fonction de l'exactitude de leurs réponses. Concernant  $\rho_{peer}$  un diagramme similaire sera

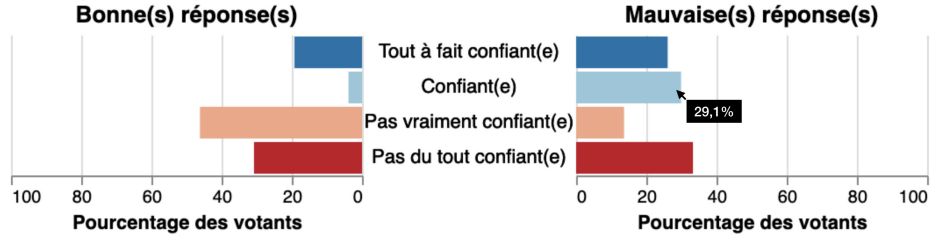


Fig. 6. Diagramme pour illustrer l'indicateur  $\rho_{conf}$ .

utilisé.

En cliquant sur la dernière phrase soulignée en bas de la Figure 4, un nouvel onglet s'ouvre sur l'étude complète publiée.

## 5 Conclusion

Nous avons présenté notre processus d'évaluation formative et montré que son implantation nécessitait un travail préalable. Nous nous sommes donc interrogés sur la manière dont l'implantation d'un processus au sein d'un système utilisé par des enseignants doit adresser les enjeux d'explicabilité. Pour répondre à cette problématique, nous avons identifié des principes d'explicabilité applicables aux systèmes d'enseignement. A partir de ces systèmes, nous avons conçu des artefacts IHM que nous avons mis en oeuvre dans le cadre de notre implantation. Notre proposition n'ayant pas encore fait l'objet d'études en contexte réel, les travaux futurs viseront à mesurer l'efficacité de cette implantation selon plusieurs métriques telles que la confiance des utilisateurs, l'expérience utilisateur ou les bénéfices que nous observerons chez les apprenants.

Si les résultats sont concluants, l'utilisation étendue de ces artefacts IHM dans tout autre système proposant des recommandations pour aider les enseignants à prendre des décisions pourrait faire l'objet d'études. Dans le cas contraire, nous pourrions nous intéresser aux raisons qui poussent les enseignants à suivre ou non les recommandations.

## Références

1. Alamri, R., Alharbi, B.: Explainable student performance prediction models: a systematic review. IEEE Access (2021)

2. Andriamisa, R., Silvestre, F., Parmentier, J.F., Broisin, J.: Recommendations for orchestration of formative assessment sequences: a data-driven approach. In: European Conference on Technology Enhanced Learning. pp. 245–259. Springer (2021)
3. Araujo, T., Helberger, N., Kruikemeier, S., De Vreese, C.H.: In ai we trust? perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY* **35**(3), 611–623 (2020)
4. Chen, G.H., Shah, D., et al.: Explaining the success of nearest neighbor methods in prediction. Now Publishers (2018)
5. Cheng, W., Shen, Y., Huang, L., Zhu, Y.: Incorporating interpretability into latent factor models via fast influence analysis. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 885–893 (2019)
6. Herlocker, J.L.: Understanding and improving automated collaborative filtering systems. University of Minnesota (2000)
7. Hoffman, R.R., Mueller, S.T., Klein, G., Litman, J.: Metrics for explainable ai: Challenges and prospects. arXiv preprint arXiv:1812.04608 (2018)
8. Li, L., Zhang, Y., Chen, L.: Generate neural template explanations for recommendation. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. p. 755–764. ACM (Oct 2020)
9. Maheswaran, D., Chaiken, S.: Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of personality and social psychology* **61**(1), 13 (1991)
10. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* **267**, 1–38 (2019)
11. Ribera, M., Lapedriza, A.: Can we do better explanations? a proposal of user-centered explainable ai. Joint Proceedings of the ACM IUI 2019 Workshops p. 8 (2019)
12. Robbins, N.B., Heiberger, R.M., et al.: Plotting likert and other rating scales. In: Proceedings of the 2011 joint statistical meeting. vol. 1 (2011)
13. Selwyn, N.: Should robots replace teachers?: AI and the future of education. John Wiley & Sons (2019)
14. Smith, M.K., Wood, W.B., Adams, W.K., Wieman, C., Knight, J.K., Guild, N., Su, T.T.: Why peer discussion improves student performance on in-class concept questions. *Science* **323**(5910), 122–124 (Jan 2009). <https://doi.org/10.1126/science.1165919>
15. Vanlommel, K., Van Gasse, R., Vanhoof, J., Van Petegem, P.: Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research* **83**, 75–83 (2017)