

# Contribution d'une évaluation formative pour le développement du savoir dans la classe : le cas de l'évaluation par les pairs

Montpied, Pascale<sup>(1)</sup>, Moulin, Marianne<sup>(1)</sup>, Tiberghien, Andrée<sup>(1)</sup>, Le Hebel, Florence<sup>(1)</sup>, Urgelli, Benoit<sup>(1)</sup>

<sup>(1)</sup> UMR 5191 ICAR, CNRS, ENS Lyon - France

Cette contribution se situe dans le domaine de l'évaluation formative. La position choisie est que l'évaluation formative doit être intégrée dans l'ensemble des activités d'enseignement et donc envisagée lors de la conception de la séquence. L'étude présentée porte sur un type d'évaluation formative en classe de SVT et de physique : l'évaluation par les pairs. Ces évaluations sont intégrées dans des séquences d'enseignement développant des démarches d'investigation. Ces séquences ont été développées sur la base du programme officiel avec un groupe d'enseignants et de chercheurs dans chacune des disciplines et ont été mises en œuvre deux fois en 2014-2015 (8 classes de seconde générale). Les évaluations par les pairs se sont déroulées lors d'activités expérimentales réalisées par des groupes de deux élèves. Les groupes d'élèves échangent leurs productions écrites et évaluent ainsi la production d'un autre groupe à partir de critères qui leur sont fournis. Nous avons recueilli plusieurs types de données : productions écrites de tous les élèves, vidéos de 4 classes en SVT et de 4 classes en physique dans l'année scolaire pendant toutes les séances de la séquence, entretiens des professeurs des classes filmées et de quelques élèves. Dans cette contribution, nous présentons deux résultats principaux. Le premier porte sur la faisabilité de l'évaluation par les pairs d'activités expérimentales mettant en œuvre la compétence d'investigation. Les professeurs ont jugé ce type d'évaluation faisable et pouvant favoriser la compréhension des élèves ; certains ont même été surpris de l'engagement de leurs élèves, même dans des classes au climat assez difficile. Une analyse des données écrites des évaluations formatives montre que les écarts entre les notes données par les élèves et celles données par le professeur ou le chercheur sont faibles. Notre second résultat met également en évidence que même les élèves qui n'ont pas réussi une ou plusieurs composantes de l'activité expérimentale ont pu juger de la production faite par d'autres élèves et de proposer des commentaires valides. Les analyses complémentaires que nous avons réalisées mettent en évidence que la possibilité pour les élèves de proposer des commentaires approfondis repose également sur la manière dont sont formulés les critères et sur la manière dont ils ont été intégrés à l'activité. Cependant, il apparaît que toute activité n'est pas directement adaptée à la mise en place d'une évaluation par les pairs. Une partie de la recherche que nous menons vise à caractériser les activités permettant cette mise en place en vue du développement du savoir dans la classe.

Mots-clés : Évaluation formative ; évaluation par les pairs ; activité d'investigation ; développement du savoir

Actuellement, la plupart des positions sur l'évaluation formative soulignent l'importance de son intégration dans chaque activité de l'enseignement (Allal & Mottier Lopez, 2005). Le projet Assist-Me<sup>1</sup> auquel nous participons a pris non seulement cette position mais vise également à une forte cohérence entre les évaluations formative et sommative faites par le professeur dans sa classe. Cette perspective est ambitieuse. Après la publication de référence de Black & William (2009) sur l'évaluation formative, de nombreux travaux ont été développés avec des définitions variables de cette évaluation. Nous avons choisi de nous appuyer sur le travail de Allal et Mottier Lopez (2005) pour préciser notre approche. Les composantes de l'évaluation formative, le recueil d'information, leur interprétation, la régulation composée du feedback et de la mise en œuvre du feedback peuvent être géré simultanément ou au contraire être décalées dans le temps ; le professeur peut par exemple prendre du temps pour récolter et interpréter les informations sur la compréhension des élèves et ne rétroagir que la séance suivante, voir plusieurs séances après.

Dans cette présentation nous nous limiterons à une situation spécifique, l'évaluation par les pairs et plus précisément l'évaluation par un groupe d'élèves de la production écrite d'un autre groupe d'élèves. Précisons que notre travail se situe en seconde d'enseignement général dans deux disciplines enseignées : les SVT et la physique. Notre première question

---

<sup>1</sup> Assess Inquiry in Science, Technology and Mathematics Education ; [www.http://assistme.ku.dk](http://assistme.ku.dk)

derecherche concernela faisabilité d'une telle évaluation. En effet les professeurs qui participent au projet n'avaient jamais réalisé une évaluation de ce type, même si certains mettaient en œuvre la correction d'exercices entre élèves. Dans le cadre des expérimentations menées pour le projet, les élèves devaient évaluer la production écrite d'un autre groupe d'élèves (réalisée pendant une activité expérimentale qu'ils avaient eux même réalisée) avec l'aide d'une grille d'évaluation composée de plusieurs critères déterminés par les chercheurs et les enseignants lors de la construction de la séquence. Notre seconde question de recherche porte sur les liens entre la qualité de la production initiale des évaluateurs, la validité de l'évaluation réalisée et deux composantes de la situation d'enseignement : le type d'activité en jeu (qui conduit à la production évaluée) et les critères fournis aux élèves pour évaluer. Avant d'aborder les résultats nous présentons le contexte de la conception des séquences.

## Intégration de l'évaluation par les pairs lors de la conception d'une séquence d'enseignement

Dans le cadre du projet Assist-me les séquences ont été construites sur la base du programme officiel français, d'une sélection d'une ou plusieurs compétences proposées par le projet européen et d'activités antérieurement utilisées par les professeurs remodelées afin d'avoir une organisation visant particulièrement le développement de ces compétences et permettant les implémentations formatives. La compétence choisie en France est l'investigation aussi bien pour les SVT que pour la physique. Précisons que le travail est mené, dans chacune des disciplines, par un groupe constitué de professeurs (de 6 à 8) et de chercheurs. Les séquences initialement conçues par ces groupes de travail ont été mises en œuvre deux fois l'an dernier (2014-2015), ce qui a permis de faire des modifications grâce aux rétroactions des professeurs et à l'observation des chercheurs dans quelques classes.

Dans le cas de la physique, nous avons choisi les signaux périodiques et les ondes sonores dans la partie « diagnostic médical » du programme. Nous ne présenterons ici que la séquence sur les ondes sonores qui a été conçue autour de la compétence investigation. Les activités portent sur la réflexion, transmission, absorption de matériaux par les ultrasons (US), et par la détermination d'une distance entre l'émetteur des US et le matériau qui les réfléchit. Sur les trois activités conçues, deux donnent lieu à une évaluation par les pairs. Elles visent à mettre en œuvre en particulier les capacités suivantes en lien avec les compétences « officielles » (APP, ANA, REA, COM<sup>2</sup>) : faire des prévisions justifiées par des éléments d'ordre théorique, proposer les étapes d'un protocole (qui permet de savoir si un objet réfléchit les ultrasons), relier des données (montrant la réflexion ou non des US) à des éléments d'ordre théorique, discuter de la validité d'une prévision ou d'un résultat d'une mesure, et bien sûr, des compétences de réalisation comme faire un schéma, un calcul littéral et un calcul numérique. Certaines de ces compétences sont utilisées comme critères donnés aux élèves évaluateurs. Actuellement, pour renforcer l'intégration de l'évaluation dans la séquence, nous essayons d'évaluer les mêmes compétences lors des évaluations sommatives (en DS).

Dans le cas des SVT, la séquence d'enseignement porte sur les combustibles fossiles dans la partie « Enjeux planétaires contemporains » du programme (BO spécial n° 4 du 29 avril 2010). Les deux activités qui nous intéressent ici portent sur la détermination de l'origine des combustibles fossiles et sur les conditions nécessaires à la formation de gisements de pétrole exploitables. Conçues également autour des compétences d'investigation, les élèves avaient à charge de réaliser des observations d'échantillons (extraits de tourbe colorés à la fuchsine, roche encadrant de veines de charbon et présentant des empreintes fossiles de végétaux), d'extraire des informations de documents, de construire un modèle visant à faire des diagnostics sur une situation réelle d'un gisement non exploitable, de faire un rapport en mobilisant des compétences telles que réalisation de schéma, et la production d'un paragraphe

---

<sup>2</sup>ANA= Analyse ; APP : Appropriation ; REA = Réalisation ; COM : Communication

argumenté mettant en lien les observations, les informations extraites des documents pour parvenir à la validation (ou l'invalidation) des hypothèses formulées en classe entière (première activité) ou en binômes (deuxième activité). De la même manière qu'en physique, les critères d'évaluation proposés aux élèves sont basés sur ces compétences.

### Mise en œuvre de l'évaluation par les pairs et sa faisabilité

Au cours de l'année 2014-2015, les séquences conçues en SVT et en physique ont chacune été mises en œuvre deux fois dans quatre classes de seconde d'un LGT. Les publics étaient hétérogènes (lycée centre ville, périphérique, lycée de ville moyenne). Ces classes ont été observées et filmées pendant les activités donnant lieu à une évaluation par les pairs. Pour ces activités nous disposions de deux caméras (l'une orientée vers le professeur et l'autre vers deux élèves, avec un champ assez large); de plus nous disposions de 8 micros complémentaires nous permettant d'enregistrer le travail de 9 groupes au maximum.

Pour préparer la modification possible du contrat didactique dans ce type d'évaluation, nous avons conçu un texte intitulé « code de bonne conduite » qui a été distribué et/ou commenté par le professeur (Extrait figure 1).

<p><b>COMMENT FAIRE DES COMMENTAIRES ?</b>          Pour atteindre les buts proposés, il faut respecter certaines règles ; les voici.</p> <ul style="list-style-type: none"> <li>• Etre d'accord sur les règles</li> <li>• Fournir des commentaires constructifs</li> <li>• Prendre conscience des besoins des autres</li> </ul> <p><b>REGLE DE BASE</b>  <b>Pour chaque point négatif que vous commentez, vous pouvez proposer un point positif et éviter toute remarque discriminatoire ou désobligeante</b> (par exemple <i>c'est nul, t'as rien compris</i>)</p> <ul style="list-style-type: none"> <li>• Donnez <b>toujours une raison de votre avis</b> pour que le commentaire soit constructif</li> <li>• Evitez des commentaires qui n'apportent aucune aide : <i>bien, d'accord, c'est bon, OK ou faux, à revoir</i> <ul style="list-style-type: none"> <li>• Respectez le travail des autres et proposez un commentaire adapté.</li> </ul> </li> </ul>
---

Figure n°1 : Extrait du code de bonne conduite donnée aux élèves

En effet, l'activité d'évaluation contribue au contrat didactique mais elle en dépend également (Venturini & Tiberghien, 2015). Ainsi, l'évaluation par les pairs donne aux élèves la responsabilité de juger d'éléments de savoir mais également d'éléments de compétences argumentatives et communicationnelles proposés par d'autres, cette responsabilité est souvent assurée par le professeur ; dans certaines classes il peut donc y avoir une modification du contrat qui peut être ponctuelle pour ces activités, ou plus large.

Aussi bien en SVT qu'en physique, après les premières implémentations, les professeurs ont jugé ce type d'évaluation faisable et pouvant favoriser la compréhension des élèves ; certains ont même été surpris de l'engagement de leurs élèves même dans des classes au climat assez difficile. L'application des élèves afin de permettre une meilleure compréhension par leurs pairs, le sens de la responsabilité lors de l'évaluation d'autrui avec un retour sur les documents et consignes non utilisés pour leur propre travail sont des comportements que les professeurs ont pu observer même chez des élèves en « décrochage ». Les données analysées pendant la réalisation en classe et les entretiens menés en fin d'année avec les élèves ont confirmé ces premiers retours. Cependant des difficultés liées à l'organisation de la classe et au savoir sont apparues. Par exemple, une difficulté apparaît quand les élèves évaluateurs considèrent comme correcte une proposition non pertinente qui est déterminante pour la réalisation de l'activité. Cette difficulté, qui nous a amenée à procéder à des ajustements au cours de l'année montre que toute activité n'est pas directement adaptée à l'évaluation par les pairs.

	Compétences (critère)	A : maitrisée	B : partiellement maitrisée	C : non maitrisée	Je ne comprends pas
Q2	Faire des prévisions et les justifier				

Raisons
<b>Exemples de raisons (avec l'orthographe de la copie de l'élève) :</b> <ul style="list-style-type: none"> <li>• Les justifications ne sont pas très cohérentes ("faire un aller retour car ils sont isolants").</li> <li>• Les prévisions sont faites et argumentées mais ne semble pas avoir de rapport avec le sujet principal</li> <li>• La réponse n'est pas assez développée car on n'a pas assez de détail sur le choix proposé</li> <li>• Leurs prévisions sont précises et justifiées.</li> </ul>

Figure n°2a : Extrait de grille d'évaluation pour un critère en physique

Critère : Le schéma de la modélisation permet de comprendre ce qu'il se passe sur le site de la crèche : les liens entre la réalité et les éléments du modèle sont bien présents, mouvement du pétrole bien représenté (par une flèche par exemple).		Justification :
		
		
	Pas de réponse	
<b>Exemples de justifications (avec l'orthographe de la copie de l'élève) :</b> <ul style="list-style-type: none"> <li>• Ils n'ont pas précisé que le pétrole s'échappe</li> <li>• La modélisation est bien faite</li> <li>• Si nous ne sommes pas au courant de ce qu'il se passe sur le site on ne peut pas comprendre le schéma</li> </ul>		

Figure n°2b : Extrait de grille d'évaluation pour un critère en SVT

### Liens entre qualité, validité et pertinence de l'évaluation par les élèves, et la situation d'enseignement.

Une analyse des données écrites des évaluations formatives montre que les élèves sont capables d'attribuer une note appropriée (A, B, C, D en physique ou smiley en SVT) à leur camarade et ceci même lorsqu'ils n'ont pas réussi l'activité expérimentale (Moulin & Al., 2015). En effet (Figure 3), les écarts entre notes données par les élèves et celles données par le professeur/le chercheur sont faibles. Les écarts les plus élevés (égaux à 2 ou 3) ne correspondent pas forcément à des évaluateurs ayant eu des difficultés durant la réalisation de l'activité expérimentale. Cela montre, à notre sens, la capacité des élèves à saisir les attentes de l'enseignant en termes de compétences et de réussites des activités disciplinaires.

	SVT – A1 (116 élèves ; 2 critères)	SVT – A2 (92 élèves ; 3 critères)	Physique – A2 (118 élèves ; 4 critères)	Physique – A3 (118 élèves ; 5 critères)	Total
Écart de 0	53	76	89	105	323
Écart de 1	36	43	105	73	257
Écart de 2	3	8	29	29	69
Écart de 3	4	4	2	9	37

Figure n°3 : Ecart entre les notes données par les élèves évaluateurs et celles données par les chercheurs/professeurs dans deux activités en SVT et en physique.

Les analyses en cours (Figure 4) semblent montrer que lorsque les élèves proposent des commentaires pour justifier leur note, ceux-ci sont en grande partie valides (Figure 4a). La difficulté pour les élèves semble être de pouvoir proposer un commentaire approfondi et apportant une aide. Cependant, si la validité des commentaires augmente avec le niveau de réussite à l'activité (variable liée à la diversité des apprenants), l'aptitude à proposer un commentaire approfondi semble être également dépendant de la formulation des critères (variable liée à l'expérimentation). L'hypothèse du lien entre un commentaire approfondi et la formulation des critères résulte des différences de résultats en physique et en SVT (Figure 4b), la construction des questions et la relation question / critère n'ayant pas été construite de la même manière dans les deux disciplines.

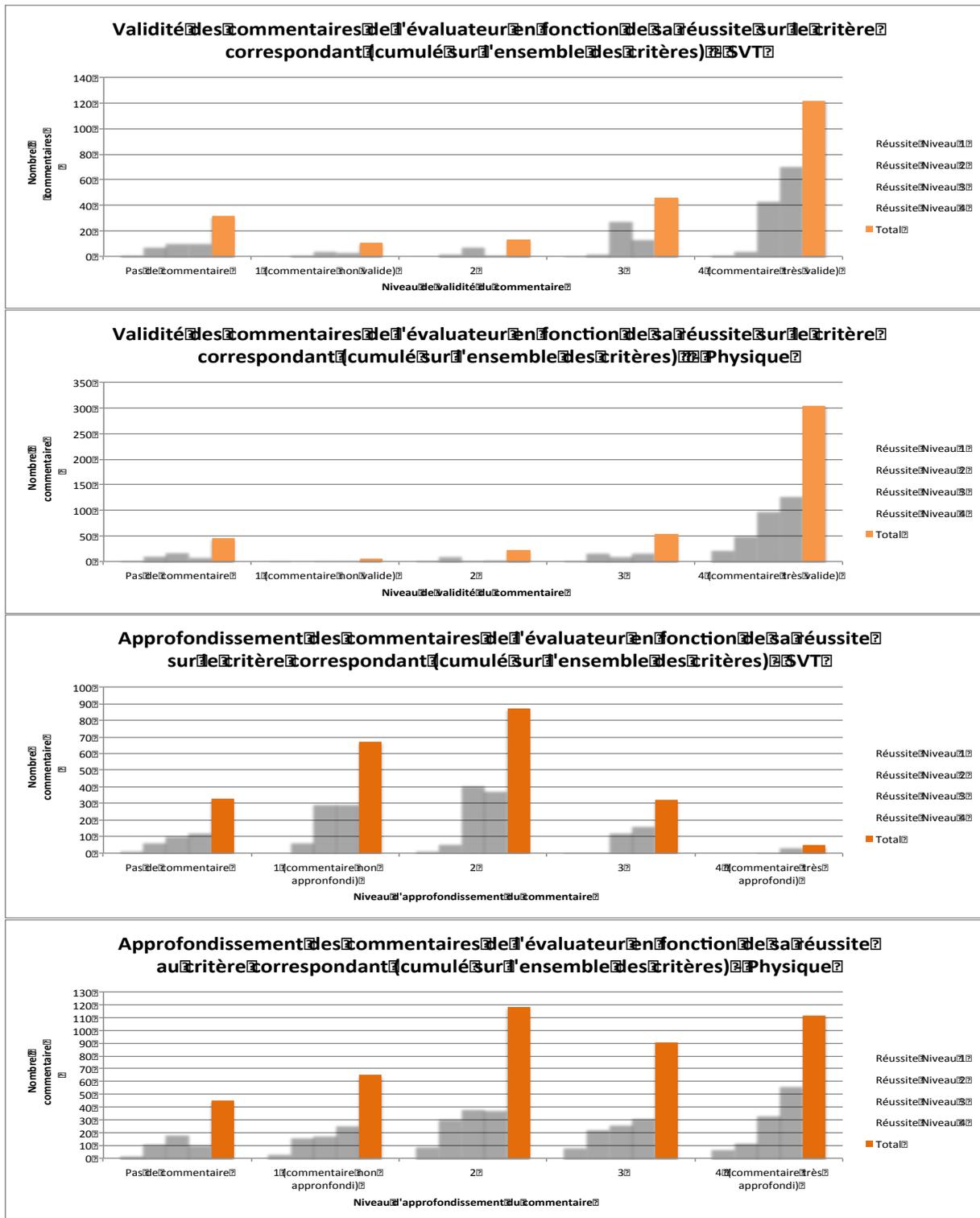


Figure n°4 : Éléments d'analyse en physique et SVT

En conclusion, nos résultats suggèrent que l'évaluation par les pairs est faisable et peut favoriser l'apprentissage. Un travail actuel que nous menons vise à caractériser le type d'activité adapté à l'évaluation par les pairs. Il nous semble qu'un des critères principaux est que la production des élèves soit principalement fondée sur un raisonnement (même si toutes les étapes du raisonnement ne sont pas évalué) et peu sur des compétences factuelles (ou alors des compétences disponibles par le biais des documents) ou le cas échéant que les critères d'évaluation explicitent d'une manière ou d'une autre les connaissances mise en jeu dans l'activité. Pour favoriser la mise en œuvre d'une évaluation par les pairs permettant le

développement du savoir dans la classe, il semble donc nécessaire d'aider les élèves, notamment grâce à un travail en amont sur le choix des critères et de leur formulation, puis avec les élèves grâce à l'explicitation des attentes si on veut les faire entrer dans l'évaluation de certaines connaissances.

## Références bibliographiques

- Allal, L. & Mottier-Lopez, L. (2005). L'évaluation formative : Revue de publications en langue française. In L'évaluation formative pour un meilleur apprentissage dans les classes secondaires (pp. 265-290). Paris: OCDE
- Black, P., & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal*, 16(2), 249-261.
- Venturini, P., & Tiberghien, A. (2015). Analysis at different timescales of the didactic contract in a physics classroom. In 11th biannual Conference of the European Science Education Research Association (ESERA). Helsinki.
- Moulin, M., Tiberghien, A., Le Hebel, F. (2015), Peer-assessment: Relation between un/success in the task and un/ability to assess classmates' written work. In 11th biannual Conference of the European Science Education Research Association (ESERA). Helsinki.