



Analyse d'une évaluation entre pairs en algèbre au collège en France

Sylvie COPPÉ

Université de Genève, FPSE, équipe DiMaGe

sylvie.coppe@unige.ch

Résumé : Dans le cadre d'une recherche collaborative associant des enseignants et la chercheuse, autrice de cet article, sur le thème de l'enseignement de l'algèbre au collège en France, nous avons élaboré et expérimenté une séance d'évaluation entre pairs (selon deux modalités) dans laquelle les élèves devaient transformer un programme de calcul en une expression littérale. Une fois leur réponse produite, elles ou ils devaient se prononcer sur la validité de la réponse d'un e autre élève ou de celles de l'ensemble des élèves de la classe. Ensuite, l'enseignant e organisait un débat dans la classe sur les réponses. Dans ce texte, en utilisant les notions de contrat didactique et de milieu, et nos travaux sur les vérifications, nous analysons les réponses et les arguments donnés par les élèves, puis nous montrons la potentialité de ce type d'activités pour la construction des connaissances algébriques.

Mots-clés : algèbre, programmes de calcul, évaluation entre pairs, argumentation

Analyzing algebra students' peer assessment skills at the college (middle school) level in France

Abstract: This article presents the results of a collaborative research project on algebra teaching strategies in middle schools in France. Together with a group of teachers, the author developed and tested a two-step peer assessment exercise in which students were asked to transform a computational program into a literal expression. Once they had produced their answers, students were then asked to assess the validity of either another student's answer or those of the entire class. Next, the teacher organized a class discussion on the answers. We analyze the students' answers and arguments based on our own studies on verification and the notions of didactic contract and didactic milieu, and discuss the potential of this type of activity for building algebra knowledge.

Keywords: algebra, computational program, peer assessment, argumentation

Introduction

Depuis 2002, la chercheuse autrice de cet article a animé un groupe de recherche collaborative (Bednarz, 2013; Desgagné et al., 2001) avec des enseignant·es¹ de collège en France (élèves de 11 à 15 ans) ayant pour but de produire des ressources pour les enseignant·es et formateur·trices de mathématiques sur l'enseignement de l'algèbre (Cherpin et al., 2022; Coppé, 2020). Ce groupe de recherche, nommé SESAMES Algèbre², produit des documents diffusés sur le site PEGAME (<http://pegame.ens-lyon.fr/>) qui comporte deux entrées : « Enseigner » constituée de propositions d'activités pour les classes, classées par niveaux, explicitées et commentées pour les rendre plus accessibles et « Se former », partie dans laquelle nous proposons des documents qui explicitent nos choix pour les activités et qui devraient permettre aux enseignant·s de se les approprier et d'en comprendre les buts et les enjeux.

De 2012 à 2016, tout le groupe a participé au projet européen de recherche ASSIST-ME (*Assess Inquiry in Science, Technology and Mathematics Education*) dont l'objectif était, d'une part, d'analyser l'influence sur les apprentissages des élèves et les pratiques enseignantes de nouveaux dispositifs d'évaluations formatives en lien avec les évaluations sommatives dans le cadre de démarches d'investigations en sciences, mathématiques et technologie et, d'autre part, de concevoir et de diffuser des méthodes d'évaluations formatives en conformité avec la culture et les pratiques de chaque pays. Dans ce cadre, nous avons conçu et expérimenté des séances dans lesquelles nous avons développé des outils d'évaluation formative pour l'enseignement de l'algèbre. Dans cet article, nous décrivons et analysons une de ces séances dans laquelle nous proposons un dispositif d'évaluation entre pairs à propos d'une tâche d'algèbre visant la conversion d'un programme de calcul (registre verbal) en une expression littérale (registre algébrique).

Dans la première partie, nous donnerons quelques éléments sur nos choix en algèbre, puis, dans la deuxième, nous évoquerons quelques éléments sur l'évaluation formative et notamment l'autoévaluation. Enfin, après avoir décrit l'expérimentation, nous analyserons les productions des élèves et leur évolution en nous centrant sur le point de vue de l'argumentation des élèves.

¹ Voici la liste des enseignant·e.s qui participent/ont participé au groupe : Christophe Alves, Olivier Arrouch, Véronique Berger, Serge Betton, Maud Chanudet, Anne Sophie Cherpin, Vincent Duval, Stéphane Garapon, Alexandra Goislard, Sylvie Martin-Dametto, Claire Piolti-Lamorthé, Sophie Roubin, Étienne Spaak.

² Ce groupe a été créé au sein du laboratoire ICAR (CNRS, Université de Lyon, ENS Lyon) et soutenu par l'Institut français de l'éducation (IFE). En 2011, il a intégré le dispositif du réseau des Lieux d'éducation associés (LéA) en association avec le collège Ampère de Lyon.

1. Le contexte de cette étude

Comme nous l'avons mentionné, ce travail collaboratif avait pour but la production de ressources pour les enseignant.es sur l'enseignement de l'algèbre afin de favoriser le développement de connaissances algébriques chez les élèves par le biais notamment de la résolution de problèmes en travaillant sur le sens plutôt que sur les techniques de calcul, mais sans pour autant les négliger. Ainsi, nous avons travaillé dans plusieurs directions. Nous en citons trois qui sont en lien avec cette étude. Tout d'abord, en partant d'un constat partagé dans le groupe selon lequel l'enseignement de l'algèbre restait encore trop formel et ne permettait pas aux élèves de travailler sur le sens, nous avons élaboré des problèmes à mettre en œuvre dans les classes pour introduire l'algèbre. Nous avons suivi en cela ce que Chevallard (1998) appelle les raisons d'être. Pour cela, dans un premier temps, nous nous sommes appuyés sur les travaux de Bednarz et Janvier (1996) qui ont défini quatre perspectives principales d'introduction de l'algèbre :

- l'approche par la généralisation/récurrence,
- l'approche par la résolution de problèmes/mise en équation,
- l'approche par la modélisation,
- l'approche technologique/fonctionnelle.

Puis, plus récemment, nous avons utilisé les travaux de Kieran (2007) qui définit trois aspects de l'activité algébrique avec le modèle GTG pour conceptualiser l'activité algébrique : l'activité générative, l'activité transformationnelle et l'activité globale au niveau méta. Cela nous a donc permis d'élaborer plus particulièrement des problèmes de généralisation et de preuves qui étaient peu proposés dans les classes au début de notre recherche (les problèmes amenant à des équations ont toujours été moins oubliés, mais ils sont souvent proposés après de nombreux exercices techniques). Il est cependant à noter que la situation a commencé à évoluer en France à partir de 2005, à la suite de changements de programmes qui préconisent de plus en plus explicitement les activités de preuve en algèbre (notamment dans le programme de 2016) comme en témoignent, dans l'extrait suivant (figure 1), les deux dernières phrases de chaque colonne.

Utiliser le calcul littéral	
Mettre un problème en équation en vue de sa résolution. Développer et factoriser des expressions algébriques dans des cas très simples. Résoudre des équations ou des inéquations du premier degré. - Notions de variable, d'inconnue.	Comprendre l'intérêt d'une écriture littérale en produisant et employant des formules liées aux grandeurs mesurables (en mathématiques ou dans d'autres disciplines). Tester sur des valeurs numériques une égalité littérale pour appréhender la notion d'équation. Étudier des problèmes qui se ramènent au premier degré (par exemple, en factorisant des équations produits simples à l'aide d'identités remarquables). Montrer des résultats généraux, par exemple que la somme de trois nombres consécutifs est divisible par 3.
Utiliser le calcul littéral pour prouver un résultat général, pour valider ou réfuter une conjecture.	

Figure 1. Extrait du programme français mis en place en 2016, cycle 4 (Ministère de l'Éducation nationale, 2015, p. 73)

Pour élaborer ces problèmes, nous avons largement utilisé les programmes de calcul (Alves et al., 2013). Rappelons qu'un programme de calcul (Drouhard, 1995; Chevallard, 2007; Ruiz Monzon, 2010) se présente sous forme d'un texte qui décrit un calcul à faire sur des nombres. Voici deux exemples de problèmes à partir d'un programme de calcul, le premier peut aboutir à une résolution d'équation et le second à l'établissement d'une preuve.

Je pense à un nombre, j'ajoute 4, puis je multiplie le résultat par 7. J'ajoute le nombre de départ et je trouve 25. Quel était le nombre de départ?

Je pense à un nombre, je le multiplie par 5 puis j'ajoute 3. Je multiplie le résultat par 2 et j'enlève 10 fois le nombre de départ. Fais quelques essais. Que constates-tu? Est-ce toujours vrai? Prouve-le.

Un autre constat portait sur les erreurs récurrentes en calcul littéral. En étudiant les manuels, et grâce à la connaissance du terrain qu'avaient les enseignants, nous avons observé que la propriété de distributivité de la multiplication sur l'addition n'était pas suffisamment utilisée comme un élément de justification des calculs. Quand elle est citée (soit dans les traces écrites soit lors des interactions enseignant·e/élève), il arrive assez souvent que ce soit avec des termes courants comme « transformer » ou « simplifier » (qui ne disent rien des opérations mathématiques à faire) ou bien avec des ostensifs au sens de Bosch et Chevallard (1999) comme des flèches ou des couleurs. Or, comme nous l'avons souligné (Assude et al., 2012), ces termes ou ces ostensifs ne permettent notamment pas des contrôles et vérifications performantes. Par exemple, pour l'erreur suivante assez fréquente, lorsqu'un·e élève développe $3(2x + 5)$ en $6x + 5$, elle ou il a bien transformé le produit en une somme, mais ce n'est pas juste.

Il en résulte que la propriété de distributivité perd sa prépondérance technologique pour justifier et valider les calculs. Il y a donc un risque que les élèves ne l'utilisent pas et se rabattent sur des techniques portant sur les transformations d'écritures exclusivement basées sur des ostensifs, avec des critères de vérification peu opérationnels portant sur la forme. (Assude et al., 2012, p. 55).

Sur l'utilisation de ces termes courants et des ostensifs, il semble qu'il n'y a pas d'évolution en France puisqu'une étude à grande échelle (étude PRAESCO³) réalisée récemment dans le cadre de la DEPP⁴ (Coppé et al., 2021) montre que 64 % des enseignants interrogés, pour donner une relance sur une erreur de calcul littéral, utilisent des arguments du type « on n'ajoute pas des pommes et des

³ Pratiques d'enseignement spécifiques aux contenus

⁴ Direction de l'évaluation, de la prospective et de la performance

poires » et que les trois quarts d'entre elles ou eux estiment que ce sont des arguments concrets et faciles pour les élèves.

Enfin un troisième axe de notre travail portait sur le développement des contrôles et des vérifications. Dans notre définition des processus de vérifications (Coppé, 1993), nous considérons que ceux-ci font partie des processus de validation à la différence qu'ils peuvent quelquefois ne porter que sur la vraisemblance et pas sur la certitude, mais pour l'élève c'est suffisant à ce moment-là.

Dans une situation de résolution de problème, un élève a identifié un résultat partiel ou final et il se pose la question de la validité de son résultat. Nous appellerons vérification tout argument avancé ou toute action mise en œuvre par l'élève pour limiter l'incertitude sur le résultat, si l'élève en a besoin, à ce moment-là et dans cette situation. Une vérification a pour conséquence, soit d'accroître la vraisemblance et éventuellement d'acquiescer la certitude du résultat, soit d'engendrer un doute plus grand et, éventuellement, de déboucher sur une phase de rectification. (Coppé, 1993, p. 27)

De plus, nous avons montré que les vérifications relevaient souvent de ce que nous avons appelé le travail privé de l'élève (c'est-à-dire ce qui n'est pas donné à voir à l'enseignant), c'est pourquoi on peut avoir l'impression que les élèves ne vérifient pas (Coppé, 1998). Il nous est donc apparu important de donner à voir des démarches et des méthodes de vérifications dans le travail quotidien de la classe, pour que les élèves puissent élargir leur palette. Mais, les vérifications sont fortement liées aux connaissances mathématiques des élèves, comme nous l'avons montré dans une autre expérimentation (Chalancon et al., 2002). Par exemple, pour pouvoir vérifier en testant sur des valeurs numériques, il est nécessaire que l'élève comprenne le statut de variable (c'est-à-dire que dans une expression littérale, x peut prendre n'importe quelle valeur). Le travail fait dans ces séances d'évaluation par les pairs a été l'occasion de développer les vérifications.

2. Mise en place d'évaluations entre pairs

Comme nous l'avons mentionné en introduction, la participation du groupe de recherche au projet ASSIST-ME a été l'occasion pour nous de réfléchir aux processus d'évaluation. Tout au long de notre travail collaboratif, nous avons eu des discussions sur les évaluations sommatives, notamment parce que si on introduit des changements dans les types de tâches proposées aux élèves pendant les séances d'enseignement, on va également modifier ceux proposés en évaluation à visée certificative. Il s'avérerait donc nécessaire de vérifier leur compatibilité avec celles de l'institution soit à l'échelle de l'établissement, soit à celle de l'examen de fin de collège.

Le travail dans le projet ASSIST-ME a été l'occasion de travailler sur l'évaluation formative (à noter que les programmes scolaires français de l'époque de notre expérimentation soulignaient déjà l'importance d'évaluations autres que sommatives) ou plutôt sur le caractère formatif de l'évaluation. En effet, les travaux anglophones et francophones (Allal et Mottier Lopez, 2005; Black et Wiliam, 2009) s'accordent sur le caractère continu de l'évaluation formative, qui porte sur des savoirs et savoir-faire déjà enseignés ou en cours d'enseignement, qui peut se dérouler à tout moment d'une séance d'enseignement. Black et Wiliam (1998) indiquent que l'évaluation devient formative quand elle sert à adapter l'enseignement aux besoins repérés des élèves notamment par la notion de feedback. Comme on peut le noter dans la citation suivante, ils incluent les processus d'autoévaluation ou d'évaluation entre pairs dans l'évaluation formative puisque ces processus permettent bien aux élèves d'avoir un feedback sur leurs apprentissages et aux enseignant.es d'adapter leur enseignement selon les informations recueillies.

Nous utilisons le terme général d'évaluation pour désigner toutes les activités entreprises par les enseignant.es – et par les élèves lors de l'autoévaluation – qui fournissent des informations à utiliser en retour pour modifier les activités d'enseignement et d'apprentissage. Cette évaluation devient une évaluation formative lorsque les données sont effectivement utilisées pour adapter l'enseignement aux besoins des élèves. (Black et Wiliam, 1998, p. 140, traduction libre⁵)

Pour Shavelson et al. (2008), comme indiqué sur la figure 2, il y a un continuum entre des évaluations formelles et planifiées (voire des évaluations « prêtes à l'emploi » produites par des chercheurs : « *Embedded-in-the-curriculum formative assessment* »), celles moins formelles dans l'interaction et celles « *on the fly* » qui peuvent arriver de façon spontanée à l'initiative de l'enseignant.e ou d'un.e élève.

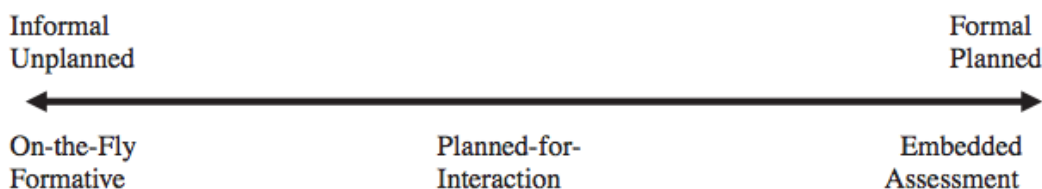


Figure 2. Continuum des évaluations formatives selon Shavelson et al. (2008, p. 300)

⁵ « We use the general term assessment to refer to all those activities undertaken by teachers – and by their students in assessing themselves – that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet student needs » (Black et Wiliam, 1998, p. 140).

En ce qui nous concerne, nous avons choisi de travailler sur l'introduction, dans des séquences portant sur l'algèbre, de dispositifs d'évaluation formative comportant de l'évaluation entre pairs en utilisant des questionnaires élèves de positionnement et/ou des votes. Black et al. (2004) précisent qu'il y a différentes organisations possibles pour l'évaluation entre pairs, le principe étant que les élèves soient placés en position de réfléchir sur la validité de la production (orale ou écrite) d'au moins un·e de leurs camarades. Pour nous, l'organisation régulière de séances incluant des évaluations entre pairs (nous nous situons dans « *embedded assessment* ») peut être un moyen de faire entrer les élèves dans l'autoévaluation en leur donnant des occasions (voire des habitudes) d'argumenter sur la validité de réponses produites (et notamment les leurs).

Cela rejoint les propos de Allal (1999) qui souligne l'importance pour les élèves de développer des compétences leur permettant de se positionner par rapport à leur travail ou par rapport aux réponses des autres dans un but de régulation, ce qui devrait permettre, selon elle, d'accroître l'autonomie de l'élève et une meilleure adaptation au monde.

En s'autoévaluant, l'élève est amené à expliciter ses représentations, à réfléchir sur ses stratégies, à confronter ses démarches avec celles d'autres élèves, à intégrer des critères externes dans ses cadres de référence, à gérer activement les relations entre les différentes tâches à accomplir. (Allal, 1999, p. 43-44)

Ainsi, en demandant aux élèves de fournir un retour sur les réponses d'autres élèves, nous pensons que cela peut favoriser la mise en place de questionnements sur la validité de leurs propres réponses (et/ou démarches), leur donner des outils de vérification et ainsi les entraîner à avoir pour eux-mêmes ce retour sur leur réponse. Bien sûr, c'est un travail qui se construit sur la durée.

Tous ces dispositifs supposent un changement de contrat dans la classe (dans le sens où les élèves ne doivent pas seulement produire des réponses, mais également se questionner sur leur validité) qui, en accordant davantage de responsabilités aux élèves dans les procédures de validation, a des effets sur les responsabilités respectives de l'enseignant·e et des élèves quant à la construction des savoirs.

3. Cadre théorique

Pour faire l'analyse de cette expérimentation, nous utilisons tout d'abord les notions de contrat et milieu didactiques (Brousseau, 1986, 1990). Ainsi, cette activité vise à modifier le contrat didactique habituel et fréquent de la classe de mathématiques puisque l'élève doit non seulement répondre à la question, mais aussi se prononcer sur la validité des réponses des autres. Il y a ainsi une redistribution des responsabilités entre l'enseignant·e qui ne donne pas son avis

notamment lors du débat, et les élèves qui doivent valider des réponses et pas seulement les produire. Ce changement de contrat introduit alors un enrichissement volontaire du milieu en y introduisant les réponses des autres, ce qui, selon nous, devrait permettre à chaque élève de se questionner sur des réponses qu'elle ou il n'a pas forcément envisagées et ainsi, mobiliser des connaissances pour produire des arguments de validation. Mais, comme nous le verrons plus loin, à la fois dans l'analyse a priori et a posteriori, ce n'est pas forcément toujours le cas. Par exemple, les élèves pourraient être d'accord sur une réponse fausse et avoir ainsi un feedback qui renforce leur erreur. On retrouve là les distinctions sur le milieu faites par Perrin-Glorian (1999) qui distingue le milieu « potentiel » et le milieu « activé » ou par Schubauer-Leoni et al. (2007) qui évoquent un milieu qui « ne fait pas milieu ». Ces deux citations de Perrin-Glorian (1999) qui, bien qu'elles concernent des situations a-didactiques de résolution de problèmes plus classiques que la nôtre, montrent bien cette distance entre le milieu cognitif et le milieu tel qu'envisagé par l'enseignant e.

Si le rapport personnel de certains élèves n'est pas idoine au rapport institutionnel, ces objets de savoir placés dans le milieu de la situation didactique risquent de ne pas pouvoir apporter les rétroactions attendues aux actions de l'élève. Pour ces élèves, la situation a-didactique ne permettra pas l'apprentissage prévu : la dévolution ne peut se faire convenablement du fait de l'absence dans les connaissances disponibles de l'élève d'éléments supposés dans le milieu. [...] Dans ce cas le milieu avec lequel agit l'élève n'est pas le milieu construit par l'enseignant. (Perrin-Glorian, 1999, p. 294)

L'absence dans le milieu effectif de certains objets que le maître y avait prévu (ou la présence d'objets non prévus) concerne surtout des connaissances auxquelles le sujet a besoin de faire appel de lui-même, ce que j'appellerai la partie cognitive du milieu, ou abrégé, le milieu cognitif. (Perrin-Glorian, 1999, p. 294)

Dans ce cadre, nous faisons l'hypothèse que, dans certains cas, lors de la validation d'une réponse, les connaissances mises en jeu peuvent être différentes de celles qui ont permis de résoudre le problème et, par conséquent, cet enrichissement du milieu par les réponses des autres peut permettre de faire évoluer les connaissances des élèves à certaines conditions que nous allons voir dans les analyses.

Pour analyser les arguments des élèves, nous avons utilisé nos travaux sur les vérifications et notamment la distinction entre les vérifications de type interne (liées au savoir mathématique) et celles externes (liées à des éléments de forme ou au contrat didactique comme « j'ai bien utilisé toutes les données du problème »).

Nous appellerons processus de vérification interne tout processus de vérification mettant en jeu des savoirs ou des savoir-faire typiquement mathématiques, ne

dépendant pas nécessairement de la situation dans laquelle on les utilise. [...] Les autres processus, qui utilisent des connaissances portant sur d'autres savoirs ou savoir-faire moins mathématiques (notamment ceux qui n'utilisent pas seulement la logique du problème mais qui dépendent davantage du contrat) seront appelés des vérifications de type externe. Ces processus ont la propriété d'être généralisables à tous les problèmes ou à des classes de problèmes très étendues [...]. Certains peuvent être très courts et/ou se limiter à des arguments simples. (Coppé, 1993, p. 36)

On peut faire le parallèle entre des vérifications de type externe et ce que Rhéaume et Oliveira (2015) désignent par « des mises entre parenthèses du sens lors de la vérification ».

Ainsi, nous cherchons à répondre aux questions suivantes : dans le cadre de l'enseignement/apprentissage de l'algèbre (au niveau du collège en France), quelles sont les capacités des élèves à se positionner, en justifiant, sur des réponses qu'ils n'ont pas produites? Quels types d'arguments sont avancés pour justifier? Et finalement, de façon plus générale, dans quelle mesure ce type de dispositif d'évaluation entre pairs visant à ajouter une tâche supplémentaire de validation (jeu sur le contrat didactique habituel) et à enrichir le milieu, peut-il favoriser la construction des connaissances en jeu?

4. Description du dispositif d'évaluation entre pairs

Voyons maintenant le dispositif expérimenté dans des classes de collèges en France au niveau 4^e et 3^e (élèves de 13 à 15 ans). Nous présentons la tâche en indiquant les objectifs poursuivis, puis nous faisons une analyse a priori des réponses (correctes et erronées) ainsi que des types d'arguments produits pour de possibles vérifications.

4.1 La tâche proposée

Voici la tâche que nous avons proposée. Elle est le fruit d'une élaboration collective, tant pour son contenu que sa mise en œuvre. Elle est proposée dans les premiers mois de l'année au début du travail sur l'algèbre. Plus précisément, cette tâche est la première d'une séquence de reprise sur l'enseignement de l'algèbre (puisque'en France, cet enseignement commence en classe de 5^e). Dans cette séquence, les élèves vont travailler sur les conversions entre des programmes de calcul et des expressions littérales (dans les deux sens) et sur les équivalences des expressions littérales pour notamment produire des preuves ou ensuite résoudre des équations.

Outre l'utilisation du dispositif d'évaluation entre pairs pour faire entrer les élèves dans l'autoévaluation, comme mentionné plus haut, un second objectif concerne davantage l'enseignant e qui pourra ainsi poser un diagnostic sur la maîtrise de

savoirs ou savoir-faire anciens dont elle ou il sait, par ailleurs, qu'ils sont source de difficulté. Ainsi, nous nous situons bien dans le cas d'une tâche d'évaluation à la fois diagnostique et formative.

Voici un programme de calcul
 Choisir un nombre
 Ajouter 4
 Multiplier la somme par 5
 Soustraire 8 au résultat
 Quelle expression littérale décrit ce programme de calcul?

Figure 3. Le problème proposé aux élèves

Cette tâche, conforme aux programmes français de 2016, est proposée aux niveaux 4^e et 3^e, pour des élèves qui ont déjà rencontré l'algèbre, mais sans que l'on sache de quelles façons, c'est-à-dire avec quelles organisations praxéologiques (au sens de Chevallard, 1998), notamment quels types de tâches ont été travaillés et quels éléments technologiques ont été utilisés. Sachant qu'en France l'algèbre est introduite en classe de 5^e, on pourrait penser que cette activité est trop facile pour des élèves des niveaux que nous avons choisis. Cependant, même si les programmes préconisent ce genre d'activités, il n'est pas sûr que les élèves les aient déjà rencontrées car les pratiques peuvent être très différentes selon les enseignant·es. Là encore nous renvoyons aux résultats de l'enquête PRAESCO (Coppé et al., 2021) qui montre des pratiques très variées concernant l'introduction et le travail sur l'algèbre. Par exemple 40 % des enseignant·es interrogé·es déclarent commencer les chapitres de calcul algébrique par des exercices de technique comme développer, réduire ou factoriser. Si environ la moitié des enseignant·es déclarent passer autant de temps sur les problèmes et sur les techniques, 40 % déclarent passer davantage de temps sur les exercices techniques. Il se peut donc que certains élèves n'aient jamais rencontré ce type de tâches d'autant plus que l'utilisation des programmes de calcul n'est pas encore si fréquente; on peut le constater dans les manuels.

Le choix (en termes de variables didactiques) de la forme du programme de calcul est justifié par les raisons suivantes qui portent sur la mobilisation de la lettre, sur les aspects syntaxiques et sur les justifications des calculs :

- Déterminer si les élèves mobilisent le calcul littéral (ce qui est explicitement demandé) plutôt que le calcul numérique. Ici comme il n'y a pas de finalité autre que celle de produire une expression littérale, on peut penser que les élèves peuvent difficilement utiliser un nombre comme exemple générique comme elles ou ils pourraient le faire dans une activité de preuve.

- Déterminer si les élèves maîtrisent les règles de calcul comme la priorité des opérations et l'usage des parenthèses et également si elles ou ils peuvent écrire une expression littérale plutôt qu'un calcul séquentiel.
- Déterminer si les élèves peuvent justifier l'équivalence de différentes écritures et avec quels arguments (niveau technologique) puisque les élèves peuvent écrire différentes expressions (réduites ou non) et que cela peut engendrer des discussions sur les différentes formes des expressions et sur leurs équivalences.

4.2 Le dispositif expérimental

Nous avons testé cette activité, en utilisant deux protocoles qui diffèrent par l'utilisation de Plickers (<https://plickers.com>) (celles et ceux des enseignants qui n'étaient pas familières ou pas équipés pour l'utilisation de Plickers n'ont pas introduit cet outil), dans neuf classes de la région lyonnaise de 4^e (élèves de 13-14 ans) ou de 3^e (élèves de 14-15 ans) avec cinq professeurs qui font partie du groupe de recherche collaborative, notés P1 à P5. Nous avons choisi à la fois des classes en réseau d'éducation prioritaire (REP)⁶ et des classes non REP. Nous avons filmé la séance dans seulement quatre classes (avec une caméra pointée sur l'enseignant·e, ce qui nous permet d'analyser les débats). Nous avons collecté les feuilles de travail des élèves pour toutes les classes sauf deux (mais nous avons le film avec les réponses des élèves). Nous avons aussi les traces des votes Plickers et certaines copies de tableau blanc. À noter que ce dispositif questionne aussi le travail fait lors des phases de mises en commun (constituées ici d'un débat), dont on sait qu'elles sont difficiles à mettre en œuvre pour les enseignants.

Le tableau suivant rend compte des différents cas.

Tableau 1. Récapitulatif des classes selon l'enseignant·e (Pn, n de 1 à 5 désigne l'enseignant·e), le niveau de classe, le protocole 1 ou 2.

Protocole 1	Protocole 2	Film	Documents élèves
	P1 3 ^e D REP		oui
	P1 3 ^e F REP		oui
	P2 3 ^e REP		oui
	P5 4 ^e		oui
P1 4 ^e A REP		x	oui
P1 4 ^e E REP		x	oui
P3 3 ^e A		x	non
P3 3 ^e B		x	non
P4 3 ^e B REP		x	oui

⁶ Ce sont des établissements qui comportent un nombre important d'élèves de milieux sociaux défavorisés.

Le scénario prévu est le suivant. Tout d'abord, les élèves résolvent le problème individuellement. Ensuite les deux protocoles suivants sont utilisés :

- Protocole 1 : en circulant dans la classe pendant le temps de recherche, l'enseignant·e collecte quatre réponses, puis en utilisant l'application Plickers, elle ou il fait voter les élèves sur la validité de chacune (on ne peut pas voter sur davantage de réponses avec cette application). L'enseignant·e organise alors un débat argumentatif entre élèves, sans intervenir et sans conclure sur la validité des réponses puis propose un nouveau vote. Pendant le débat, l'enseignant·e distribue la parole, reformule les propos des élèves si besoin, mais ne donne pas son avis. Voici la feuille de travail distribuée aux élèves dans laquelle nous avons tenté de recueillir des arguments explicites sur le maintien ou sur le changement de vote.

Exercice Réponse écrite

1^{er} positionnement Plickers

Débat sans validation par l'enseignant

2^{ème} positionnement Plickers

Quelle expression littérale décrit ce programme de calcul ?

Mon 1^{er} vote Plickers : Mon 2^{ème} vote Plickers :

J'ai changé d'avis entre mes deux votes	Je n'ai pas changé d'avis entre mes deux votes
Pourquoi ? Donne un argument qui t'a convaincu :	Pourquoi ? Donne un argument qui t'a convaincu :
.....	
.....	
.....	

Figure 4. Feuille de travail des élèves avec le protocole 1 (vote Plickers)

- Protocole 2 : les élèves échangent leur copie entre voisins·es et se prononcent par écrit sur la réponse de l'autre élève, puis le débat a lieu et ensuite chaque élève exprime son accord ou non avec le positionnement en vrai ou faux.

Mon prénom	Le prénom de mon correcteur :		Mon Prénom
Réponses	Vrai ou Faux	Argumentation :	D'accord ou pas d'accord
Question 1 : Voici un programme de calcul : • Choisir un nombre • Ajouter 4 • Multiplier la somme par 5 • Soustraire 8 au résultat Quelle expression littérale décrit ce programme de calcul ?			
La réponse qui te paraît la plus juste : Une réponse qui pour toi est fausse :			
Question 1			

Figure 5. Feuille de travail des élèves avec le protocole 2
(positionnement sur réponse d'un e autre élève)

En faisant une analyse a priori des réponses et des vérifications en utilisant la distinction vérifications internes/externes, voyons maintenant, comment ces éléments s'articulent pour l'analyse des productions des élèves.

4.3 Analyse a priori des réponses possibles et des vérifications

L'analyse a priori des réponses montre qu'il y a trois réponses correctes possibles selon leur forme plus ou moins réduite :

1. la réponse non réduite $5(x + 4) - 8$ qui traduit pas à pas le programme de calcul en tenant compte des priorités de calcul et donc des parenthèses (nous faisons l'hypothèse qu'elle sera la plus fréquente des réponses justes);
2. la réponse $5x + 20 - 8$ qui correspond au développement de la précédente expression (cette réponse nous semble peu probable car si les élèves développent l'expression, ils réduiront ensuite par effet de contrat didactique);
3. la réponse réduite $5x + 12$.

Nous n'avons aucune exigence sur la forme de cette réponse correcte et nous avons choisi le programme de calcul afin que ces trois réponses apparaissent pour pouvoir aborder le niveau technologique (l'utilisation de la distributivité pour prouver l'équivalence des expressions) en rendant explicites les types d'arguments

utilisés par les élèves. Nous pensions que comme les élèves avaient déjà eu un enseignement de l'algèbre (notamment celles et ceux de 3^e), elles ou ils pouvaient penser à réduire soit par effet de contrat didactique, soit pour avoir une expression « plus simple ».

De nombreuses réponses fausses ou incorrectes sont envisageables :

4. un ou plusieurs calcul(s) numérique(s) comme exemple(s) pour montrer comment fonctionne le programme de calcul, avec éventuellement une réponse numérique;
5. un schéma du programme de calcul en ligne ou colonne avec des flèches pour montrer la succession des calculs (portant sur des nombres ou une variable) sans avoir une seule expression à la fin;
6. des expressions erronées comme $5x + 4 - 8$ ou $x + 4x5 - 8$ qui ne tiennent pas compte des priorités opératoires (simple traduction du programme de calcul) ou des expressions réduites des précédentes et qui sont donc erronées comme $5x + 4$ ou $x + 12$;
7. des expressions comme $5x - 12$ ou bien $20x - 8$ qui proviennent d'une erreur dans la réduction de l'expression correcte $5(x + 4) - 8$;
8. une expression littérale (ou des égalités) avec plusieurs variables pour décrire le programme comme $x + 4 = yx5 = z - 8$.

Ainsi, du fait du choix de la forme du programme de calcul, il y a un nombre important de réponses correctes ou erronées, ce qui est essentiel pour engendrer des discussions sur la validation.

En plus de l'analyse a priori des réponses, nous avons fait une analyse a priori des positionnements en termes de vrai/faux et des arguments associés pour chacune des réponses en utilisant les vérifications internes et externes.

Dans le tableau de l'annexe 1, nous reprenons la liste des réponses possibles avec des arguments a priori qui peuvent servir au positionnement sur la réponse d'un autre élève en distinguant les vérifications, externes ou internes, qui peuvent être utilisées pour valider ou invalider (certains arguments pouvant servir dans les deux cas).

Envisageons quelques cas pour mettre en avant ce qui peut se passer suivant la réponse initiale de l'élève (que nous désignons par A) lorsqu'elle ou il se positionne sur la réponse de l'autre élève (que nous désignons par B) et voir ainsi les effets en termes de milieu.

Tout d'abord considérons que les élèves A et B donnent la même réponse (juste ou erronée). Dans ce cas, l'argument utilisé par A (et par B) pour valider a une grande probabilité d'être « c'est juste car j'ai la même réponse ». Dans ce cas, on constate

que le milieu n'est pas enrichi et que, de fait, cela va même à l'encontre de l'objectif de validation puisque cela risque de renforcer A et B sur la vraisemblance de leur réponse même si elle fausse.

Examinons maintenant les cas 1, 2 ou 3 dans lesquels l'élève A a trouvé une bonne réponse (réduite ou non) et que B a une réponse erronée. On peut alors penser que A pourra utiliser des arguments internes pour invalider la réponse de B et même dans certains cas, qu'il pourra pointer ou qualifier les erreurs, ce qui l'amène à mobiliser des connaissances mathématiques. On peut aussi avoir le cas où A a une réponse non réduite (cas 1 ou 2) et B réduite. Il se peut alors que cela permette à A de faire évoluer sa réponse en mobilisant des connaissances sur la distributivité (lors de la validation) et donc ensuite de revenir sur sa réponse et la faire évoluer. Dans ce cas, l'introduction dans le milieu de la réponse réduite peut avoir un effet positif.

Voyons enfin le cas où A a une réponse incorrecte (cas 4 à 8). Pour A, cette tâche n'est pas problématique puisque A a pu répondre en utilisant des connaissances communes (par exemple en faisant une simple traduction du texte en français par des symboles mathématiques sans prendre en compte les priorités opératoires). Le fait d'être confronté à une réponse de B différente (qui n'est pas une « traduction » aussi directe) et à devoir se prononcer sur sa validité peut l'amener à remettre en cause sa réponse (par exemple parce qu'il ou elle se demande pourquoi il y a des parenthèses ou d'où vient le 12) et donc à mobiliser des connaissances mathématiques. Dans ce cas-là, le milieu joue son rôle pour faire évoluer la réponse de A, voire ses connaissances.

Ces deux derniers exemples illustrent, selon nous, comment, par le biais de cette situation d'évaluation entre pairs, les changements de contrat et milieu peuvent avoir des effets sur les connaissances des élèves et permettre ainsi des évolutions.

Nous analysons maintenant les réponses des élèves dans les différentes classes et suivant les deux protocoles.

5. Analyses des données et discussion

Nous allons analyser les données collectées selon chacun des protocoles (car les types de réponses ne sont pas tout à fait les mêmes). Nous mettrons en avant la diversité de réponses, puis nous analyserons les arguments convoqués par les élèves.

5.1 Protocole 1

Les réponses par classe de l'ensemble des élèves avant et après le débat se trouvent en annexe 2. Un premier résultat montre la variété des réponses, on retrouve ainsi

toutes celles prévues dans l'analyse a priori. Il nous paraît important de souligner que dans 3 classes sur 5, avant le débat, on voit encore apparaître des réponses qui n'utilisent pas les expressions littérales, ce qui montre qu'une part (faible) des élèves (7 sur 105) ne sont pas encore entrés dans le domaine algébrique. De plus, on constate que de façon très minoritaire, certaines réponses sont apparues, mais lors du premier vote, les élèves qui les avaient produites n'ont pas voté pour elles. On peut penser que la seule exposition des réponses a pu amener certains élèves à changer d'avis.

Toutefois, selon les classes, la variété est plus ou moins grande. Ainsi, dans la classe P4 3^e A, toutes les réponses étaient correctes, données sous forme d'expressions littérales plus ou moins réduites. Lors du second vote, seuls deux élèves votent pour l'expression correcte mais non réduite. Ici l'enrichissement du milieu par les réponses réduites a permis des discussions sur la transformation des expressions et sur la propriété de distributivité comme élément de justification (ce qui était un objectif pour être un appui pour la suite de la séquence). On peut donc dire que le milieu potentiel et le milieu activé étaient assez proches pour la plupart des élèves. De point de vue de l'évaluation formative, cela donne à l'enseignant un renseignement important sur les connaissances des élèves à ce moment-là qui va lui permettre de poursuivre sa séquence de classe. Finalement on peut sûrement penser que dans cette classe, cette tâche était trop facile pour les élèves, mais c'est le seul cas.

Dans les deux autres classes de 3^e de P3, on peut noter que l'expression réduite n'est pas apparue, ce qui nous porte à nous interroger. Peut-être est-ce dû à un effet de contrat didactique local pour cette activité, à savoir que les élèves ont produit une réponse et n'ont pas continué. Ou bien elles ou ils ne considèrent pas le calcul littéral comme un outil et donc s'arrêtent à la première expression obtenue si on ne leur dit rien de plus. Pourtant, il est sûr que ces élèves de 3^e ont déjà pratiqué le calcul littéral et ont été entraînés aux types de tâches de développement et réduction.

Dans la classe de P1 4^e E, aucune bonne réponse n'est apparue à la fin du travail individuel. Comme c'est une classe de 4^e, on peut penser que cette moins bonne réussite est un effet du niveau de classe. Certainement pour cette même raison, dans l'autre classe de 4^e, il n'y a pas eu d'expression réduite.

Dans cette classe de P1 4^e E, lors du débat qui a été très riche, les élèves ont, d'une part, collectivement décidé de supprimer la réponse basée sur un exemple de calcul $4 + 5 = 9 \times 5 = 45 - 8 = 37$ en avançant l'argument que la question demandait « une expression littérale et qu'ici il n'y avait pas de lettres » et d'autre part, ils ont corrigé collectivement l'écriture de ce calcul en faisant référence à des

connaissances sur les priorités de calcul et sur le signe $=$. Cette discussion sur l'organisation des calculs a amené la question suivante d'un élève : « est-ce que dans les calculs avec des lettres, les priorités existent aussi ? » L'enseignante ayant répondu par l'affirmative (les enseignant.es avaient pour consigne de ne pas se prononcer sur la validité des réponses, mais ils ou elles pouvaient répondre à d'autres questions), il a alors proposé la nouvelle réponse $(n + 4) \times 5 - 8$ pour laquelle les élèves ont ensuite voté de façon assez massive. Nous soulignons donc le fait que, lors de ce débat sur la validation, cet élève a pu mettre en discussion un élément de savoir qui était resté complètement implicite pour lui, à savoir que les règles de calcul qu'il connaissait dans le cadre numérique étaient vraies dans le cadre algébrique. Nous rapprochons ce résultat de ce que Pilet (2012) nomme des besoins d'apprentissages implicites. Dans cette même séance, nous avons également noté, à la suite de questions d'élèves : « si n est le nombre choisi, il ne peut pas être égal à $n + 4$, il reste le nombre de départ » ou encore « $5x$ c'est 5 fois x ou 5 plus x ? ». Pour conclure sur ce débat dont nous avons déjà rendu compte (Coppé et Roubin, 2019), nous avons dégagé deux types d'objets de savoir qui ont été discutés :

- des éléments portant sur des savoirs enseignés anciens comme la multiplication des relatifs, les priorités opératoires et les parenthèses. Ainsi, cette modalité d'évaluation formative a bien permis à l'enseignant d'avoir des informations sur l'état des connaissances de ses élèves dans cette phase de rappel.
- des objets paramathématiques (Chevallard et Joshua, 1991) comme le signe égal (« il doit y avoir la même chose de chaque côté du signe $=$ ») ou la lettre (« on n'est pas obligé de mettre x , on peut mettre une autre lettre » ou « x ça représente tous les nombres »).

Par conséquent, dans ces classes de 4^e, le milieu activé était en quelque sorte plus éloigné du milieu potentiel puisque les réponses attendues sont moins apparues, voire pas apparues comme dans la classe de P1 4^e E. Ainsi, dans cette classe, il aurait pu arriver que le milieu ne fasse pas milieu parce que les élèves n'auraient eu que le choix entre des réponses fausses, ce qui, en plus, les aurait conforté.es dans cette voie. Finalement, c'est l'intervention de l'élève sur la priorité des opérations en calcul littéral qui a enrichi le milieu et a permis d'avancer. Dans une autre expérimentation (Coppé et Moulin, 2017) en classe de 6^e sur les fractions, dans un cas semblable, nous avons décidé d'introduire la bonne réponse en évoquant un élève fictif et nous avons montré que cela n'avait pas eu d'effet puisqu'aucun élève n'avait pu défendre cette réponse et qu'elle n'avait donc pas été choisie lors du vote. Ceci est un point particulièrement délicat lors de ces séances.

Enfin, le dernier résultat que nous pouvons mettre en avant est qu'il y a une évolution entre les deux votes dans le sens des réponses correctes même s'il reste quelques erreurs. Sur la feuille de réponse, nous demandions aux élèves de dire si elles ou ils avaient changé d'avis et pourquoi. Nous n'avons ces feuilles de réponses que pour trois classes. Voici les arguments utilisés par les élèves pour justifier leur changement ou non de vote, que nous avons classés selon trois types :

- T1 : Des arguments ne mettant pas en jeu des éléments de savoir (par exemple, « je n'ai pas changé d'avis, car c'était ma première réponse »);
- T2 : Des arguments sur la conformité de la réponse qui correspondent aux vérifications externes (il y a des lettres, calcul en ligne, « je n'avais pas remarqué que c'était une écriture littérale »);
- T3 : Des arguments mathématiques mettant en jeu des propriétés qui correspondent aux vérifications internes (par exemple, « il faut mettre une parenthèse pour qu'on fasse le calcul en priorité »).

Tableau 2. Les arguments des élèves classés selon les trois types (le nombre entre parenthèses indique le nombre d'occurrences de l'argument)

		J'ai changé d'avis	Je n'ai pas changé d'avis
Classe de P1 4 ^e A	T1	Il y a une faute (2)	Simple et juste, je ne voulais pas changer d'avis (3)
	T2		Conformité de la réponse (3)
	T3	Utilisation des parenthèses (3)	Priorité des calculs, utilisation des parenthèses (7)
Classe de P1 4 ^e E	T1	Plus facile, pas le même résultat (3)	Pas vrai, j'ai gardé ma réponse (2)
	T2	Conformité expression littérale (4)	
	T3	Priorité des calculs, utilisation des parenthèses (9)	
Classe de P4 3 ^e A	T1		Simple, rapide... (10)
	T2		
	T3	Transformations d'écritures, distributivité (2)	Transformations d'écritures, distributivité (4)

Bien sûr ce tableau est à interpréter en faisant des liens avec les réponses obtenues dans chaque classe. Ainsi, pour les deux classes de P1 où il y avait une variété de réponses erronées, on trouve davantage d'arguments, dont une bonne moitié du troisième type, ce qui est à mettre en lien avec la richesse du débat. Mais certains utilisent tout de même encore ceux des deux premiers types, notamment des

arguments portant sur la forme de la réponse dans la classe de P4 3^e A (dans laquelle les réponses étaient justes). Dans ce cas, nous ne pouvons pas exclure que ce grand nombre de réponses de type 1 peut être le signe de la trop grande facilité de cette tâche.

Si l'on rassemble les résultats par types d'arguments (tableau 3), on constate que ce sont les arguments de type 1 et 3 qui sont majoritaires. Nous retenons que les arguments de type 3 apparaissent assez fortement, ce qui est un bon signe de la mobilisation de connaissances mathématiques et notamment de la propriété de distributivité. On peut aussi penser que les élèves mobilisent cet argument parce que les enseignant.es qui participent au groupe de recherche ont l'habitude de l'utiliser dans leurs classes.

Tableau 3. Récapitulatif des arguments pour le protocole 1

	J'ai changé d'avis	Je n'ai pas changé d'avis	total
Argument de type 1 T1	5	15	20
Argument de type 2 T2	4	3	7
Argument de type 3 T3	5	11	16

Pour conclure sur ce premier type d'expérimentation, nous voulons souligner l'investissement des élèves dans les débats à la fois par leur attitude active et concentrée mais aussi cognitivement par les arguments avancés.

5.2 Protocole 2

Cette modalité a été testée dans quatre classes de trois enseignant.es (voir l'ensemble des résultats dans l'annexe 3). Pour faire cette analyse, nous allons considérer les réponses de chaque élève, les évaluations en vrai ou faux de l'autre élève et les arguments produits pour valider ou non ces réponses.

Comme pour dans la modalité 1, les réponses que nous avons prévues dans l'analyse a priori sont apparues avec des différences selon les classes, qui s'expliquent encore une fois par ce que nous désignons par le parcours algébrique antérieur des élèves. Ainsi, dans la classe de P1 3^e F, on note 18 réponses erronées sur 23 alors que dans l'autre classe de cette même enseignante P1 3^e D, il y a 10 réponses correctes sur 21. Enfin, il est à souligner qu'un nombre important d'élèves (20 élèves sur 73) de chaque classe restent au niveau du calcul numérique.

Nous n'avons les réponses finales retenues par les élèves que dans deux classes. À la fin de l'activité, le nombre de réponses déclarées correctes augmente tout de même, ce qui est encourageant, mais il reste quelques réponses erronées.

En ce qui concerne le positionnement sur la réponse d'un autre élève, nous observons que tous les cas se produisent, à savoir des réponses correctes évaluées

comme fausses ou le contraire. Même si une grande majorité d'élèves se déclarent d'accord avec la position de l'élève évaluateur·trice, il en reste certains·es qui ne veulent pas changer d'avis, notamment dans le cas des réponses fausses déclarées fausses, ce qui peut signifier que ces élèves ont confiance en leur réponse et donc en leurs connaissances et que les arguments d'un pair ne sont pas suffisants pour les convaincre. Cette résistance peut poser un vrai problème d'enseignement.

Voyons maintenant les arguments qui sont employés. Pour cela nous reprenons les trois types utilisés dans l'analyse précédente. Bien sûr certains arguments sont différents puisque l'objectif de l'argumentation n'est pas le même mais on retrouve des similitudes. Par exemple, les arguments de conformité de la réponse peuvent suffire dans un premier temps quand la réponse n'est pas une expression littérale.

On trouve également dans chaque classe les arguments de type 1 qui consistent à comparer la réponse de l'autre élève avec la sienne. De la même façon, si l'élève a produit la même réponse, cet argument est suffisant. Dans le cas où la réponse est évaluée comme fausse parce qu'elle est différente de sa propre réponse, les élèves ne mobilisent pas vraiment de connaissances et le milieu n'est donc pas suffisant. Dans la classe de P1 3^e F, il est à noter qu'il y a encore quatre élèves qui ont la conception erronée bien connue du signe égal comme un résultat et qui la mobilisent ici pour invalider la réponse de leur camarade (« c'est faux parce qu'il n'y a pas de résultat »).

Enfin, dans le cas de réponses erronées évaluées comme fausses, un nombre important d'élèves évaluateur·trices proposent également la rectification en expliquant soit ce qui est faux soit en donnant la réponse qu'elles ou ils estiment correcte.

Tableau 4. Les arguments des élèves classés selon les trois types

		Réponse évaluée comme JUSTE	Réponse évaluée comme FAUSSE
Classe de P1 3e D	T1	Même chose (4) C'est bien fait (4)	Donne leur réponse (3)
	T2	Conformité : il faut une écriture littérale (1)	Non-conformité : il faut une écriture littérale (5)
	T3		Repérage de l'erreur (2) Énoncé d'un élément de connaissance (utilisation des parenthèses, etc.) (1)
Classe de P1 3e F	T1	Même chose (1) C'est bien fait (5)	Donne leur réponse (2)
	T2		Non-conformité : il faut une écriture littérale (3) Pas de résultat (4)
	T3	Énoncé d'un élément de connaissance (1)	Repérage de l'erreur (4)
Classe de P2 3e	T1	Même chose (1) C'est bien fait (4)	Donne leur réponse (3) Tu n'as pas compris (2)
	T2	Conformité : il faut une écriture littérale (1)	Non-conformité : il faut une écriture littérale (3)
	T3	Énoncé d'un élément de connaissance « tu as bien utilisé la distributivité » (1)	
Classe de P5 4e	T1		Pas même résultat (2) Donne leur réponse (1)
	T2	Conformité : il faut une écriture littérale (1)	Non-conformité : il faut une écriture littérale (2)
	T3	Énoncé d'un élément de connaissance « tu as bien utilisé la distributivité » (1)	Repérage de l'erreur (2)

En faisant la somme de ces arguments sur toutes les classes (tableau 5), on peut remarquer que les plus nombreux sont ceux de type 1. Ainsi, quand il s'agit de réponses considérées comme vraies, les élèves peuvent estimer qu'ils n'ont pas à chercher trop d'arguments, les plus basiques pouvant suffire. Comme nous l'avons expliqué ci-dessus, les arguments de type 2 sur la non-conformité peuvent suffire lorsque la réponse est numérique. Enfin on trouve tout de même des arguments mobilisant des connaissances mathématiques comme des vérifications explicites étape par étape avec soit des indications d'erreurs (par exemple : « $4 + x$ ne font pas $4x$ ») soit une vérification plus globale des points importants (par exemple : « tu as bien mis les parenthèses, tu as bien utilisé la distributivité et les calculs sont justes »). Mais bien sûr, ce sont les élèves qui ont trouvé la réponse qui utilisent ce genre d'arguments.

Tableau 5. Récapitulatif des arguments protocole 2

	Réponse évaluée comme JUSTE	Réponse évaluée comme FAUSSE	Total
Argument de type 1	19	13	32
Argument de type 2	3	17	20
Argument de type 3	3	9	12

En conclusion de cette seconde modalité, nous voyons que les élèves sont capables de mobiliser différents types d'arguments allant de certains assez peu appuyés sur des connaissances mathématiques à d'autres qui pourraient aider les élèves à rectifier leurs réponses initiales. Mais ce travail doit être complété soit par un débat (c'est la modalité que nous avons choisie, mais dont nous n'avons pas de trace) soit par des discussions entre les deux élèves pour se mettre d'accord sur une solution.

Conclusion

Nous avons décrit et analysé un dispositif de classe orienté sur une évaluation entre pairs sur le thème de l'algèbre au collège en France. Nous retenons plusieurs points des analyses faites. Le premier porte sur les compétences en algèbre des élèves de ces classes. Cette tâche, proposée comme une tâche de rappel servant à introduire le chapitre d'algèbre et que nous avons analysée comme devant être assez peu problématique à ces niveaux de classe, reste encore non maîtrisée par une proportion importante d'élèves à la fois en classe de 4^e et de 3^e (davantage tout de même en 4^e) alors que cela fait deux à trois ans que ces derniers travaillent l'algèbre. En annexe 4, le récapitulatif des réponses sur toutes les classes montre qu'environ 50 % des élèves échouent, dont 17 % environ qui ne mobilisent pas du tout le calcul littéral, restant dans le cadre numérique. Les autres échouent parce qu'ils ne considèrent pas les priorités opératoires, ce qui montre que les aspects syntaxiques sont encore peu maîtrisés, les élèves restant sur des procédures de calcul en étapes. Toutes ces remarques rejoignent encore une fois nos questionnements sur le travail mathématique proposé aux élèves en ce qui concerne la mobilisation d'une lettre comme variable ou comme nombre généralisé (si ce n'est pas indiqué explicitement) et sur la mobilisation des propriétés, notamment la distributivité de la multiplication sur l'addition comme élément de justification.

Le deuxième point concerne les questions de recherche. Nous nous demandons si les élèves étaient capables de se prononcer sur la validité de réponses produites par les autres en donnant des arguments et nous avons vu que c'était le cas même si des arguments de différents types, dont certains peu mathématiques, cohabitent encore. Dans toutes les classes, même celles réputées comme étant difficiles, nous

avons noté une grande implication des élèves pour ce travail tant au niveau de la résolution de la tâche en elle-même que dans les débats. Par ces raisons, nous pensons que ce dispositif d'évaluation par les pairs est tout à fait adapté pour des élèves de collège. De façon globale, sur l'ensemble des classes, la tâche proposée semble être plus adaptée pour les élèves de 4^e que de 3^e.

Nous avons également noté, avec les deux protocoles, une évolution positive des réponses avant et après le débat. Ainsi, l'hypothèse faite sur les effets de l'enrichissement du milieu par les réponses des autres élèves nous semble en grande partie validée, moyennant cependant certaines conditions, notamment celle de l'existence d'au moins une réponse juste produite par un élève (pour qu'elle puisse être donnée à voir et discutée) et celle portant sur les appariements de réponses (ne pas mettre ensemble deux élèves qui ont la même réponse fausse). On peut penser que c'est par la répétition de ce type d'activités que les élèves pourront développer des aptitudes à l'autoévaluation.

Du point de vue des pratiques enseignantes, introduire ce dispositif d'évaluation formative nous semble, d'une part, pertinent pour les apprentissages (ici en algèbre, mais on pourrait l'adapter à d'autres domaines), car il permet des interactions de qualité entre les élèves amenant à des discussions argumentatives tout à fait intéressantes et, d'autre part, assez faciles à mettre en place dans les classes parce que la tâche proposée peut être mise en œuvre sur un temps assez court (moins d'une séance). Ce sont des critères importants pour que des activités de ce type soient diffusées plus largement dans les pratiques. Cependant, la gestion de la mise en commun et du débat nécessite à la fois un grand investissement de l'enseignant·e pour être à l'écoute des arguments des élèves, mais aussi pour être suffisamment en retrait afin de ne pas donner son avis. Cette mise en œuvre peut donc se révéler délicate, comme c'est souvent le cas pour les mises en commun. L'introduction de ces nouveaux dispositifs devrait être un objet de formation comme, de façon plus générale, la gestion des interactions.

Références

Allal, L. (1999). Impliquer l'apprenant dans le processus d'évaluation : Promesses et pièges de l'autoévaluation. Dans C. Depover et B. Noël (dir.), *L'évaluation des compétences et des processus cognitifs, modèles, pratiques et contextes* (p. 35 56). De Boeck.

Allal, L. et Mottier Lopez, L. (2005). Formative assessment of learning : A review of publications in French. Dans OCDE (dir.), *Formative Assessment – Improving Learning in Secondary Classrooms*. (p. 241-264). OCDE.

Alves, C., Coppé, S., Duval, V., Goislard, A., Kuhman, H., Martin Dametto, S., Piolti Lamorthe, C. et Roubin, S. (2013). Utilisation des programmes de calcul pour introduire l'algèbre au collège. *Repères IREM*, 92, 9-30.

Assude, T., Coppé, S. et Pressiat, A. (2012). Tendances de l'enseignement de l'algèbre élémentaire au collège. Atomisation et réduction. *Recherches en didactique des mathématiques, Hors-série*, 41-62.

Bednarz, N. (2013). Regarder ensemble autrement : Ancrage et développement des recherches collaboratives en éducation au Québec. Dans N. Bednarz (dir.), *Recherche collaborative et pratique enseignante. Regarder ensemble autrement* (p. 13-30). L'Harmattan.

Bednarz, N. et Janvier, B. (1996). Emergence and development of algebra as a problem-solving tool: Continuities and discontinuities with arithmetic. Dans N. Bednarz, C. Kieran et L. Lee (dir.), *Approaches to algebra: Perspectives for research and teaching* (p. 115-136). Kluwer Academic Publishers.
https://doi.org/10.1007/978-94-009-1732-3_8

Black, P., Harrison, C., Lee, C., Marshall, B. et Wiliam, D. (2004). Working inside the Black Box: Assessment for Learning in the Classroom. *Phi Delta Kappan* 86(1), 8-21.

Black, P. et Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.

Black, P. et Wiliam, D. (2009). Developing the theory of formative assessment. *Educational assessment, evaluation and accountability*, 21(5), 5-31.
<https://doi.org/10.1007/s11092-008-9068-5>

Bosch, M. et Chevallard, Y. (1999). La sensibilité de l'activité mathématique aux ostensifs. Objet d'étude et problématique. *Recherches en didactique des mathématiques*, 19(1), 77-124.

Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en didactique des mathématiques*, 7(2), 33-116.

Brousseau, G. (1990). Le contrat didactique et le concept de milieu : Dévolution. *Recherches en didactique des mathématiques*, 9(3), 309-336.

Chalancon, F., Coppé, S. et Pascal, N. (2002). Les vérifications dans les équations, inéquations et en calcul littéral. *Petit x*, 59, 23-41.

Cherpin, A. S., Coppé, S., Goislard, A., Piolti Lamorthe, C. et Roubin, S. (2022). Production de ressources pour l'enseignement de l'algèbre au collège. Le LéA réseau d'écoles et collèges Ampère. Dans R. Monod-Ansaldi, B. Gruson et C. Loisy (dir.), *Le réseau des lieux d'éducation associés à l'Institut français de l'éducation : Un instrument pour la recherche* (p. 183-201). Presses universitaires de Rennes.

Chevallard, Y. (1998). Analyse des pratiques enseignantes et didactique des mathématiques : L'approche anthropologique. Dans R. Noirfalise (dir.), *Analyse des pratiques enseignantes et didactique des mathématiques. Actes de l'université d'été de La Rochelle* (p.91-119). IREM de Clermont-Ferrand.

Chevallard, Y. (2007). *Séminaire de didactique des mathématiques PLC2, année universitaire 2006/2007*. IUFM d'Aix-Marseille. http://yves.chevallard.free.fr/spip/spip/IMG/pdf/Seminaire_2006-2007.pdf

Chevallard, Y. et Joshua, M. A. (1991). *La transposition didactique. Du savoir savant au savoir enseigné*. Éditions La pensée sauvage.

Coppé, S. (1993). *Processus de vérification en mathématiques chez les élèves de première scientifique en situation de devoir surveillé* [Thèse de doctorat inédite]. Université Claude Bernard Lyon 1.

Coppé, S. (1998). Composantes privées et publiques du travail de l'élève en situation de devoir surveillé en mathématiques. *Educational studies in mathematics*, 35(2), 129-151.

Coppé, S. (2020). Conception collaborative de ressources pour l'enseignement de l'algèbre élémentaire : Une entrée par les programmes de calculs. Dans H. Squalli, I. Oliveira, A. Bronner, et M. Larguier (dir.), *Le développement de la pensée algébrique à l'école primaire et au début du secondaire. Recherche et perspectives curriculaires* (p. 21 -43). Livres en ligne du CRIRES. https://lel.crires.ulaval.ca/sites/lel/files/le_developpement_de_la_pensee_algebrique_a_lecole_primaire_et_au_debut_du_secondaire.pdf

Coppé, S., Grugeon- Allys, B., Horoks, J., Pilet, J., Solnon, A., Raffaëlli, C. et Charpentier, A. (2021). *Premiers résultats de l'enquête sur les pratiques d'enseignement des mathématiques, Praesco en classe de troisième en 2019* [note d'information n° 21.00]. Ministère de l'Éducation nationale, de la Jeunesse et des Sports.

Coppé, S. et Moulin, M. (2017). Évaluation entre pairs et débat argumenté dans le cadre d'un problème complexe en mathématiques. *Canadian Journal of Sciences, Mathematics and Technology Education*, 17, 308-327. <https://doi.org/10.1080/14926156.2017.1378832>

Coppé, S. et Roubin, S. (2019). Intégrer des évaluations entre pairs dans les séances de mathématiques : un exemple en algèbre au collège. Dans M. Abboud (dir.), *Mathématiques en scène des ponts entre les disciplines. Actes du colloque Espace mathématique francophone* (p. 953-962).

Desgagné, S., Bednarz, N., Lebuis, P., Poirier, L. et Couture, C. (2001). L'approche collaborative de recherche en éducation : un rapport nouveau à établir entre recherche et formation. *Revue des sciences de l'éducation*, 27(1), 33-64.

Drouhard, J. P. (1995). Algèbre, calcul symbolique et didactique. Dans R. Noirfalise et M.-J. Perrin-Glorian (dir.), *Actes de la 8^e école d'été de didactique des mathématiques* (p. 325-344). IREM de Clermont-Ferrand.

Kieran, C. (2007) Learning and teaching algebra at the middle school through college levels: building meaning for symbols and their manipulation. Dans F. K. Lester (dir.), *Second handbook of research on mathematics teaching and learning* (p. 707-762). Information Age.

Perrin Glorian, M.-J. (1999). Problèmes d'articulation de cadres théoriques ; l'exemple du concept de milieu. *Recherches en didactique des mathématiques*, 19(3), 279-321.

Pilet, J. (2012). *Parcours d'enseignement différencié en algèbre élémentaire* [Thèse de doctorat, Université Paris Diderot-Paris 7]. TEL. <https://theses.hal.science/tel-00784039v2>

Rhéaume, S. et Oliveira, I. (2015). « Je me suis vérifié... j'ai relu la question pis j'ai reregardé ma réponse » : illustrations de mises entre parenthèses du sens lors de la vérification. Dans A. Adihou, L. Bacon, D. Benoit et C. Lajoie (dir.), *Regards sur le travail de l'enseignant de mathématiques. Actes du colloque du Groupe de didactique des mathématiques du Québec* (p. 167-180). Université de Sherbrooke.

Ruiz Munzón, N. (2010). *La introducción del álgebra elemental y su desarrollo hacia la modelización funcional* [Thèse de doctorat inédite]. Université autonome de Barcelone.

Schubauer-Leoni, M. L., Leutenegger, F., Ligozat, F. et Fluckiger, A. (2007). Un modèle de l'action conjointe professeur-élèves : Les phénomènes didactiques qu'il peut/doit traiter. Dans G. Sensevy et A. Mercier (dir.), *Agir ensemble. L'action didactique conjointe du professeur et des élèves* (p. 51-91). Presses universitaires de Rennes.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz Primo, M. A., Tomita, M. K. et Yin, Y. (2008). On the role and impact of formative assessment on science inquiry teaching and learning. *Applied Measurement in Education*, 21(4), 295-314.

Annexe 1

Les réponses de l'élève A et les évaluations de ces réponses de l'élève B en fonction des types et de la nature des vérifications

Réponse de l'élève A	Évaluation de l'élève B	Évaluation de l'élève B
	Accord Type et nature de vérifications	Désaccord Type et nature de vérifications
1 Correcte proche du PC $5(x + 4) - 8$	Externe Même réponse ⁷ Il y a une lettre	Externe Pas même réponse ⁸
	Interne Retrouver l'expression à partir du programme Il y a bien les parenthèses	Interne Pas terminé (car pas réduite)
2 Correcte développée $5x + 20 - 8$	Externe Même réponse Il y a une lettre	Externe Pas même réponse Ne correspond pas au PC à cause du 20
	Interne Retrouver l'expression à partir du programme et développer Refaire seulement le développement	Interne Pas terminé (car pas réduite)
3 Correcte réduite $5x + 12$	Externe Même réponse Il y a une lettre	Externe Pas la même réponse (car réduite)
	Interne Retrouver l'expression à partir du programme et la réduire Refaire seulement la réduction	Interne Montrer erreur de calcul dans le développement
4 Un calcul/plusieurs calculs comme exemple(s)	Externe Même technique	Externe Pas le même exemple
	Interne Refaire le/les calcul(s)	Interne C'est une formule qui est demandée

⁷ « Même réponse/technique » signifie que l'élève qui évalue considère que la réponse de l'autre élève est juste car il a trouvé la même.

⁸ « Pas même réponse » signifie que l'élève qui évalue considère que la réponse de l'autre élève est fausse car il n'a pas trouvé la même.

5	Un schéma de programme en ligne ou colonne	Externe	Externe
		Même technique	Pas même réponse
		Interne	Interne
6	$5x + 4 - 8$ ou $x + 4 \times 5 - 8$ ou $5x + / 4$ ou $x + 12$ Réponse erronée car pas de prise en compte des parenthèses	Refaire le/les calcul (s)	Pas forme demandée
		Vérifier pas à pas les calculs	Refaire le/les calculs
		Externe	Externe
7	$5x - 12$ ou $20x - 8$ Réponse erronée car erreur dans la réduction à partir d'une expression littérale correcte	Même réponse	Pas même réponse
		Il y a des lettres	Interne
		Interne	Retrouver l'expression à partir du programme
8	Des formules avec plusieurs variables ou la variable ne désigne pas toujours le même nombre	Retrouver l'expression à partir du programme en se trompant	Pointer erreur de parenthèses
		Externe	Pointer erreur de calcul
		Externe	Externe
		Même réponse	Pas même réponse
		Il y a des lettres	Interne
		Interne	Refaire et comparer
		Refaire en se trompant	repérer erreurs de calcul
		Externe	Externe
		Externe	Externe
		Même réponse	Pas même réponse
		Il y a des lettres	Il ne peut y avoir qu'une lettre
		Interne	Interne
		Refaire en se trompant	Indiquer que la variable n'est pas correctement utilisée
		Externe	Externe
		Externe	Externe

Annexe 2

Les réponses et les votes dans les classes utilisant le protocole 1 (total 105 élèves)

P1 4^e A : 18 élèves		
Réponses	Votes avant le débat	Votes après le débat
$x + 4 \times 5 - 8$	7	2
Des calculs sur un exemple numérique	0	0
$(x + 4) \times 5 - 8$	11	16

P1 4^e E : 19 élèves		
Réponses	Votes avant le débat	Votes après le débat
$22 + 4 = 26$ $26 \times 5 = 130$ $130 - 8 = 122$	5	0
$n + 4 = n$	2	0
$n \times 5 - 8 = n$		
$n + 4 \times 5 - 8$	10	2
$4 + 5 = 9 \times 5 = 45 - 8 = 37$ remplacée ensuite par $(n + 4) \times 5 - 8$	2	16

P3 3^e A : 26 élèves		
Réponses	Votes avant le débat	Votes après le débat
$x + 4 \times 5 - 8$	0	1
$x + 4 = x \times 5 - 8$	4	0
$7 + 4 = 11$ $11 \times 5 = 55$ $55 - 8 = 47$	Enlevée car pas de lettres	
$(x + 4) \times 5 - 8$	19	25
$(x + 4) \times 5 - 8x$	3	0

P3 3^e B : 25 élèves		
Réponses	Votes avant le débat	Votes après le débat
$x \rightarrow \cdot 4x \rightarrow \cdot 20x \rightarrow \cdot 12x$	Enlevée car pas conforme	
$(15 + 4) \times 5 - 8$	1	1
$(x + 4) \times 5 - 8$	18	20
$X + 4 \times 5 - 8$	2	1
$X + 4 = x \times 5 = x - 8$	4	3

P4 3^e A : 17 élèves		
Réponses	Votes avant le débat	Votes après le débat
$(x + 4) \times 5 - 8$	6	2
$12 + 5x$	10	14
$20 + 5x - 8$	0	0
$((x + 4) \times 5 - 8)$	1	0

Annexe 3

Les réponses et évaluations en vrai/faux dans les classes utilisant le protocole 2 (total 73 élèves)

A : accord avec élève qui évalue

D : désaccord avec élève qui évalue

R : la réponse (évaluée comme fausse) est rectifiée par l'évaluateur.trice

Classe de P5 4 ^e (9 élèves)				
Types de réponses	Avant évaluation entre pairs	Évaluée comme JUSTE	Évaluée comme FAUSSE	Après évaluation entre pairs
Correcte proche PC $(5x + 4) - 8$	2	1	1	AA
Correcte développée $5x + 20 - 8$				
Correcte réduite $5x + 12$	1	1		A
Un calcul/plusieurs calculs comme exemple(s)	3		3R	DAA
Un schéma de programme en ligne ou colonne				
Réponse erronée $5x + 4 - 8$ ou $5x + / - 4$ ou $x + 20 - 8$ ou $x + 12$	2	1	1R	AA
Réponse erronée $5x - 12$	1		1R	A
Des formules avec plusieurs variables				

Réponses déclarées justes après le débat

$5(x + 4) - 8$: 8 élèves

$(x + 4 \times 5) - 8$: 1 élève

Classe de P1 3 ^e D (21 élèves)				
Types de réponses	Avant évaluation entre pairs	Évaluée comme JUSTE	Évaluée comme FAUSSE	Après évaluation entre pairs
Correcte proche PC $5(x + 4) - 8$	6	5	1	AAAAAA
Correcte développée $5x + 20 - 8$				
Correcte réduite $5x + 12$	4	4		AAAA
Un calcul/plusieurs calculs comme exemple(s)	6		6R	AAADAA
Un schéma de programme en ligne ou colonne	1		1R	A

Analyse d'une évaluation entre pairs en algèbre au collège en France

Réponse erronée $5x + 4 - 8$ ou $5x + / - 4$ ou $x + 20 - 8$ ou $x + 12$	3	1	2R	ADA
Réponse erronée $5x - 12$ ou $20x - 8$	1	1		A
Des formules avec plusieurs variables				

Réponses déclarées justes après le débat

$5(x + 4) - 8$: 11 élèves

$5x + 12$: 4 élèves

$x + 4 \times 5 - 8$: 1 élève

$5x - 8 + 12$: 1 élève

Un calcul numérique : 1 élève

Classe de P1 3e F (23 élèves)				
Types de réponses	Avant évaluation entre pairs	Évaluée comme JUSTE	Évaluée comme FAUSSE	Après évaluation entre pairs
Correcte proche PC $5(x + 4) - 8$	1	1		A
Correcte développée $5x + 20 - 8$				
Correcte réduite $5x + 12$	4	4		AAAA
Un calcul/plusieurs calculs	4		1 3R	AADA
Un schéma de programme en ligne ou colonne				
Réponse erronée $5x + 4 - 8$ ou $5x + / - 4$ ou $x + 20 - 8$ ou $x + 12$	14	4	10R	DAAADDD DDAAAAA
Réponse erronée $5x - 12$				
Des formules avec plusieurs variables				

Classe de P2 3 ^e (20 élèves)				
Types de réponses	Avant évaluation entre pairs	Évaluée comme JUSTE	Évaluée comme FAUSSE	Après évaluation entre pairs
Correcte proche PC $5(x + 4) - 8$	4	3	1R	AAAD
Correcte développée $5x + 20 - 8$				
Correcte réduite $5x + 12$	2	1	1	AD
Un calcul/plusieurs calculs comme exemple(s)	7 1 non évaluée	3	3R	AAADDD
Un schéma de programme en ligne ou colonne	1		1	A
Réponse erronée $5x + 4 - 8$ ou $5x + /-4$ ou $x + 20 - 8$ ou $x + 12$	6	4	1 1R	DAA-AA
Réponse erronée $5x - 12$				
Des formules avec plusieurs variables				

Annexe 4

La répartition des réponses pour l'ensemble des élèves (105 élèves pour le protocole 1 et 73 pour le protocole 2)

Types de réponses	Total 105 + 73 = 178	Pourcentage
Correcte proche PC $5(x + 4) - 8$	55 + 13 = 68	38,2 %
Correcte réduite $5x + 12$	10 + 11 = 21	11,8 %
Un calcul/plusieurs calculs comme exemple(s)	8 + 20 = 28	15,7 %
Un schéma de programme en ligne ou colonne	0 + 2 = 2	1,1 %
Réponse erronée $5x + 4 - 8$ ou $5x + /-4$ ou $x + 20 - 8$ ou $x + 12$	26 + 25 = 51	28,7 %
Réponse erronée $5x - 12$ ou $20x - 8$	0 + 2 = 2	1,1 %
Des formules avec plusieurs variables	6	3,3 %