



HAL
open science

Évaluation entre pairs en mathématiques : activité d'étudiant•es lors d'évaluations de preuves

Juliette Veuillez–Mainard, Simon Modeste

► To cite this version:

Juliette Veuillez–Mainard, Simon Modeste. Évaluation entre pairs en mathématiques : activité d'étudiant•es lors d'évaluations de preuves. 2025. <hal-04252795v4>

HAL Id: hal-04252795

<https://hal.science/hal-04252795v4>

Preprint submitted on 31 Oct 2025

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.



HAL Authorization

Évaluation entre pairs en mathématiques : activité d'étudiant-es lors d'évaluations de preuves

Juliette Veuillez--Mainard & Simon Modeste

Abstract. This article deals with peer-assessment in the context of higher education teaching in mathematics, and examines the nature of student activity when assessing work, more precisely proofs, produced by peers. After an overview of research on peer assessment, we propose an experiment with students in a specific post-secondary scientific class, and analyze the activity of the students involved. Our analyses are based on *a priori* analyses of the proposed tasks and tools from research on students personal mathematical work. We base on written notes of assessments, observations of pairs in assessment situations, and answers to a questionnaire on the perception of their activity. We have identified some specific student activities during the peer assessment process, in particular related to the analysis and correction of proofs, but also to the awareness of the issues associated with evaluation. This enables us to argue for the potential of peer assessment in higher mathematics education.

Keywords. Didactics of mathematics, Peer assessment, Proof and proving, Students' activity, University education, Real numbers

Résumé. Cet article traite de l'évaluation entre pairs (*peer-assesment*) dans le contexte de l'enseignement supérieur en mathématiques, et étudie la nature de l'activité d'étudiant-es lors de l'évaluation de copies, et notamment des preuves, produites par des pairs. Après un tour d'horizon des recherches sur l'évaluation par les pairs, nous proposons une expérimentation en classe préparatoire MPSI et nous analysons l'activité des étudiant-es impliqué-es. Nos analyses s'appuient sur des analyses a priori des tâches proposées et des outils issus de recherches sur le travail personnel d'élèves en mathématiques. Nous nous basons sur les traces écrites des évaluations, des observations de binômes en situation d'évaluation, et les réponses à un questionnaire sur la perception de leur activité. Nous identifions certaines activités spécifiques des étudiant-es lors du processus d'évaluation de pairs, notamment liées à l'analyse et la correction de preuve, mais aussi liées à la prise de conscience des enjeux associés à l'évaluation. Ceci nous permet d'argumenter du potentiel de l'évaluation entre pairs pour l'enseignement des mathématiques dans le supérieur.

Mots-Clés. Didactique des mathématiques, Évaluation par les pairs, Preuve et démonstration, Activité d'élèves, Enseignement supérieur, Nombres réels

Table des matières

1. Introduction	2
2. Recherches sur l'évaluation et positionnement	3
2.A. Généralités sur l'évaluation	3
2.B. Évaluation par les pairs	3
2.C. Un regard en didactique des mathématiques sur l'évaluation par les pairs	5
2.D. Preuve et évaluation de preuves	5
2.E. Bilan et formulation des questions de recherche	7
3. Expérimentation et méthodologie	8
3.A. Cadrage théorique	8
3.B. Dispositif expérimental	8
3.C. Analyses a priori du problème	11
3.C.a. Analyse a priori des stratégies de résolution	11
3.C.b. Éléments d'analyse <i>a priori</i> des évaluations	12
3.C.c. Retour sur nos choix expérimentaux	14
4. Résultats des analyses de l'activité des étudiant-es	14
4.A. Premières remarques sur la résolution et l'évaluation des exercices	14
4.B. Analyse des entretiens	15
4.C. Quelques résultats sur les évaluations réalisées à la maison	22
4.D. Synthèse des analyses de l'activité des étudiant-es	24
5. Perception des étudiant-es sur leurs activités	24

Ce travail a été financé en partie par l'Agence Nationale de la Recherche (ANR) grâce au projet CaMeLOt ANR-20-CHIA-0001-01.

6. Conclusions	26
6.A. Résultats et perspectives	26
6.B. Quelques conseils et pistes de mise en place	28
Annexe A Grille d'analyse	30
Annexe B Questionnaire	31
Annexe C Éléments de correction	32

1. Introduction

Les apprentissages des élèves et étudiant·es¹ sont évalués dès l'enseignement primaire et tout au long de la scolarité. Ces évaluations peuvent prendre différentes formes en fonction des types de connaissances que l'on souhaite évaluer, de l'âge des élèves, ou encore de leur parcours. Elles sont un outil de mesure des connaissances et des apprentissages des élèves, et peuvent servir au développement et à l'adaptation de l'enseignement. Plusieurs types d'évaluations peuvent être mises en place en fonction des objectifs de l'évaluation, par exemple des devoirs en classe ou à la maison, ou des présentations orales ou affichées. La question de l'implication des élèves dans l'évaluation a été discutée dans la recherche, et diverses études ont montré l'intérêt des processus d'auto-évaluation et d'évaluation par les pairs, notamment leur validité et la diversité des apprentissages qu'ils peuvent favoriser. Nous présentons ces études dans la première partie.

Nous nous intéresserons dans cet article à l'évaluation par les pairs dans l'enseignement supérieur. D'après la définition de [Topping \(2003\)](#), « l'évaluation par les pairs est un dispositif dans lequel les apprenant·es et/ou travailleur·ses sont amené·es à examiner et spécifier la valeur ou la qualité d'un produit ou d'une performance d'autres apprenant·es et/ou travailleur·ses de même statut » ([Topping, 2003](#), p. 65, notre traduction). Dans le cadre d'évaluation par les pairs dans un contexte d'enseignement, les élèves sont donc amené·es à évaluer d'autres élèves du même niveau d'étude (par exemple, des élèves de seconde évaluent d'autres élèves de seconde). Les précédentes recherches sur ce sujet nous invitent à supposer que cette évaluation est un soutien à l'apprentissage. Nous faisons l'hypothèse qu'être mis en situation d'évaluation de pairs change l'activité mathématique des élèves, et nous nous demandons donc quelle activité les élèves mettent en place pendant ce processus, et quels apprentissages cela pourrait alors favoriser. Nous nous plaçons spécifiquement dans un contexte d'activité de preuve, qui tient une place importante à l'entrée dans l'enseignement supérieur ([Gueudet & Vandebrouck, 2022](#); [Selden, 2012](#); [Stylianides et al., 2024](#)).

Nous présentons dans la première partie de cet article un tour d'horizon succinct de la recherche sur l'évaluation par les pairs, complété par une classification des « processus de vérification en mathématiques chez les élèves » issue de [Coppé \(1993\)](#). Nous présentons également quelques recherches sur l'apprentissage de la preuve au début de l'enseignement supérieur, et sur l'évaluation de preuves. Ceci permettra de formuler nos questions de recherche. Dans la deuxième partie, nous présentons l'expérimentation que nous avons menée et précisons sa méthodologie. Les analyses *a priori* du problème et des évaluations de ce problème sont ensuite développées. Nous exposons dans la troisième partie les analyses de l'expérimentation, avant de présenter les résultats d'un questionnaire dans la quatrième partie. En conclusion, nous revenons sur les principaux résultats de notre étude, les perspectives de recherche, et des idées et conseils pour la mise en place d'évaluations par les pairs.

1. Nous avons choisi d'utiliser le terme générique d'*élève*, à entendre comme incluant l'enseignement supérieur, dans toute la partie de généralités sur l'évaluation entre pairs, et le terme d'*étudiant·es* pour nos expérimentations qui portent sur le début du supérieur.

2. Recherches sur l'évaluation et positionnement

Plusieurs notions sont centrales à notre étude : l'évaluation, l'évaluation par les pairs, et l'évaluation de preuves. Nous présentons dans cette section des éléments généraux sur l'évaluation et sur la preuve, dont certains seront utilisés dans notre recherche.

2.A. Généralités sur l'évaluation

La notion d'évaluation peut être définie de manière générale comme le propose [De Ketele \(1986\)](#) :

L'évaluation est le processus qui consiste à recueillir un ensemble d'informations pertinentes, valides et fiables, puis à examiner le degré d'adéquation entre cet ensemble d'informations et un ensemble de critères choisis adéquatement en vue de fonder la prise de décision. ([De Ketele, 1986](#), p. 266)

Il est courant de distinguer l'*évaluation sommative*, qui a une fonction de certification et de contrôle, de l'*évaluation formative*, qui vise à être un soutien explicite aux apprentissages ([Morrisette, 2014](#)). Le concept d'évaluation formative a été introduit en 1967 par Scriven, et décrit une évaluation mise en place dans le but de fournir des données qui vont permettre des adaptations de l'enseignement ([Allal & Mottier Lopez, 2005](#)). Enfin, lorsque le but premier de l'évaluation est l'apprentissage des élèves plutôt que la mesure de leurs connaissances, on parle d'*évaluation soutien d'apprentissage*. Plus précisément, lors de la troisième conférence internationale sur l'Évaluation soutien d'Apprentissage, cette évaluation a été définie de la manière suivante :

L'Évaluation soutien d'Apprentissage fait partie des pratiques quotidiennes des élèves et enseignants qui, individuellement ou en interaction, recherchent, réfléchissent sur et réagissent à l'information provenant d'échanges, démonstrations et observations afin de favoriser les apprentissages en cours. ([Allal & Laveault, 2009](#), p. 102)

De nombreuses recherches sur l'évaluation formative ont été réalisées, ainsi que plusieurs revues et panorama de ces recherches. On pourra par exemple consulter les synthèses de [Black & Wiliam \(1998\)](#), [Allal & Mottier Lopez \(2005\)](#) et [Morrisette \(2014\)](#), sur lesquelles nous nous appuyons dans la suite, et qui intègrent des travaux allant du primaire au supérieur, toutes disciplines confondues, incluant les mathématiques.

Nous signalons aussi l'ouvrage de l'ADMEE-Europe (Association pour le Développement des Méthodologies d'Évaluation en Éducation) [Detroz et al. \(2017\)](#), ainsi que la conférence de consensus du [Cnesco \(2022\)](#), qui propose une synthèse sur la recherche en évaluation et sur la situation en France. Le site associé propose un certain nombre de documents scientifiques et de synthèses.

2.B. Évaluation par les pairs

L'évaluation par les pairs est une forme d'évaluation dans laquelle les élèves sont amenés à évaluer les productions d'autres élèves de même niveau d'étude. C'est un processus distinct de l'enseignement par les pairs, dans lequel les élèves ne sont pas mis en position d'évaluateur-riche.

Plusieurs éléments peuvent varier dans la mise en place d'une évaluation par les pairs : nature des apprentissages évalués, nature des supports évalués... Nous nous concentrerons ici sur l'évaluation de travaux individuels dans le cadre de l'enseignement. Dans la suite de cette section, nous nous appuyons sur les méta-analyses et revues de littérature de [Falchikov & Goldfinch \(2000\)](#) et [Topping \(2003, 2009\)](#). En cohérence avec la littérature sur le sujet, nous distinguerons l'évaluation par les pairs (peer-assesment) et la notation par les pairs (peer-grading). Deux enjeux principaux se présentent : la *validité* et la *fiabilité* de l'évaluation, et ses apports possibles à l'apprentissage.

Fiabilité et validité. Une partie de la littérature s'interroge sur la capacité des élèves à évaluer et noter leurs pairs. La notion de *validité* réfère à la cohérence de l'évaluation (souvent la notation) des élèves par rapport à celle d'une référence experte (un·e enseignant·e par exemple). La notion de *fiabilité*, en fait assez peu étudiée dans la littérature, réfère à la variabilité des évaluations (souvent des notations) entre élèves-évaluateur·rices, dans des conditions similaires.

Dans le cas de notations, l'étude de la validité s'appuie souvent sur des indicateurs statistiques, notamment la corrélation ou l'effect size. Ce dernier caractérise l'ampleur de la différence de note moyenne entre les élèves et les enseignant·es sur une tâche particulière (Sadler & Good, 2006, p. 6). Les recherches concluent que, sous certaines conditions, les élèves peuvent produire des évaluations valides. Elles identifient des critères influant significativement sur la validité, par exemple la nature du produit évalué ou le degré d'avancement du cours (la significativité de ce second critère est discutée par Falchikov & Goldfinch (2000)). Il semble aussi que la familiarité des élèves avec les critères de notation tend à améliorer la validité de l'évaluation.

Bénéfices pour l'apprentissage. Le processus d'évaluation par les pairs semble favoriser l'apprentissage des élèves. On observe d'une part des progrès du niveau effectif des élèves sur une tâche après la mise en place d'une évaluation par les pairs. Il semble aussi que ce dispositif peut impliquer de nouveaux questionnements chez les élèves, et permet de détecter plus tôt des conceptions erronées. De plus, elle implique en général que les élèves passent plus de temps à pratiquer une tâche (Topping, 2003).

D'autre part, l'évaluation par les pairs permet des gains métacognitifs pour les élèves. La présence d'un retour (feedback) personnalisé permet d'améliorer la capacité d'autoévaluation. En complément d'une évaluation classique, l'évaluation par les pairs peut apporter des feedbacks individualisés plus nombreux et réguliers, ce qui peut être un gain pour l'élève évalué·e. Ce processus permet enfin d'observer des gains sociaux et de communication (Boud & Falchikov, 2007).

Évaluation par les pairs en mathématiques. Les résultats mentionnés jusqu'à présent, même s'ils intègrent des études en mathématiques, ne sont pas spécifiques au contexte de la classe de mathématiques, et encore moins à la nature des connaissances mathématiques qui pourraient être évaluées.

Nous n'entrerons pas dans les détails des travaux existants spécifiques à l'évaluation en mathématiques et en didactique des mathématiques sur l'évaluation, mais nous renvoyons les lectrices et lecteurs intéressé·es vers quelques ressources. Au sujet des mathématiques, mentionnons tout d'abord l'étude ICMI dédiée au sujet (Niss, 1993). Plus récemment, du point de vue de la didactique des mathématiques, on peut signaler le chapitre de Grugeon-Allys (2017) qui développe une réflexion didactique et épistémologie de l'évaluation en mathématiques, et la note d'expertes de Horoks & Pilet (2023) pour la conférence de consensus du Cnesco qui propose une synthèse plutôt tournée vers l'enseignement secondaire et les pratiques enseignantes en France. L'article de Coppé (2018) propose une synthèse et une réflexion sur l'intérêt du point de vue de la didactique des mathématiques pour « apporter un autre éclairage ou [...] affiner des résultats déjà produits » et amener « des études plus fines sur les apprentissages des élèves prenant en compte la nature des tâches qui leur sont proposées » (p. 23), ce à quoi nous souhaitons contribuer dans cet article.

Sur l'évaluation entre pairs en mathématiques, les quelques recherches que nous avons identifiées s'intéressent essentiellement à la fiabilité et la validité des évaluations (par exemple Jones & Alcock, 2014; Baiduri, 2022). Très peu d'études s'intéressent à la nature des activités des élèves pendant le processus d'évaluations par les pairs.

Dans les sections suivantes, nous approfondissons deux perspectives en didactique des mathématiques permettant d'éclairer notre étude. Nous présentons d'abord une recherche sur l'argumentation des élèves au cours d'une évaluation entre pairs en mathématiques. Nous donnons ensuite des éléments de contexte sur le thème spécifique de l'enseignement et l'apprentissage de la preuve et la question de l'évaluation de preuves en didactique des mathématiques. Ceci nous permettra d'explicitier nos questions de recherche.

2.C. Un regard en didactique des mathématiques sur l'évaluation par les pairs : travaux de Coppé et Moulin

Nous nous intéressons ici aux recherches de Coppé & Moulin (2017). Les autrices ont mis en place une séquence d'évaluation par les pairs et de débats argumentatifs dans une classe de sixième. Après avoir cherché un problème, les élèves se positionnent sur la validité des réponses des autres élèves. La classe est ensuite amenée à discuter collectivement des réponses affichées.

Le problème sélectionné porte sur les aires, et fait intervenir des connaissances sur les fractions dans sa résolution et dans les vérifications. Les autrices ont montré que le dispositif permet *a priori* de « provoquer [...] la mobilisation de connaissances autres que celles utilisées pour résoudre le problème » (Coppé & Moulin, 2017, p. 326). Cependant, le problème s'est avéré trop complexe pour certain-es élèves, qui n'ont pas réussi à évaluer des démarches différentes de la leur.

Dans l'analyse des processus de vérification mis en place par les élèves au moment de la correction, les autrices s'appuient sur le travail de thèse de Coppé (1993) définissant une *typologique des processus de vérification*. Deux types sont distingués :

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)

Des exemples de processus de vérification externes sont les vérifications sociales dues au travail de groupe, ou les vérifications faisant référence au texte de l'exercice, comme le type d'unité pour une aire. En effet, ces processus « ne nécessitent pas *a priori* d'entrer dans la stratégie de résolution du problème. » (Coppé & Moulin, 2017, p. 12). D'un autre côté, on retrouve les arguments de vérification internes tels que les vérifications techniques du résultat ou les vérifications par changement de cadre.

La typologie proposée fournira un premier axe d'analyse pour l'expérimentation que nous présentons dans la suite de ce texte.

2.D. Preuve et évaluation de preuves

La preuve en didactique des mathématiques. La didactique des mathématiques s'est intéressée depuis ses débuts à l'enseignement et l'apprentissage de la preuve. Nous présentons d'abord quelques éléments de contexte général sur le thème avant de d'aborder quelques éléments sur la preuve en didactique des mathématiques qui nous semblent éclairer les enjeux de l'évaluation par les pairs.

Il est habituel de distinguer démonstration et preuve (*proof* et *proving* en anglais), le premier désignant le texte produit qui établit la vérité d'un énoncé, et le second l'activité de recherche d'une argumentation en mathématique pour établir cette vérité (ou la rejeter).

Pour une vision générale de l'état actuel des recherches sur preuve et démonstration, nous renvoyons à l'ICMI Study parue en 2012 (Hanna & De Villiers, 2012) sur le thème *Proof and Proving in Mathematics Education*, particulièrement au chapitre de Selden (2012) sur l'enseignement supérieur, ainsi qu'au texte de Stylianides et al. (2024) pour une synthèse très récente sur les enjeux dans l'enseignement supérieur, dans le numéro de ZDM sur le thème *Reviews on important themes in mathematics education*. Plus récemment encore, dans l'introduction du numéro thématique de ZDM *New trends in didactic research in university mathematics education*, Biehler et al. (2024) rappellent que preuve et démonstrations sont des enjeux clés à la transition secondaire-supérieur et un sujet de recherche très actif actuellement, et identifient trois axes importants : la contribution de l'argumentation et la

preuve dans la compréhension des mathématiques, le rôle des nouvelles technologies, et l'évaluation des preuves, notamment par les étudiant·es (nous revenons sur ce troisième point, lié à notre recherche, dans la section suivante).

Pour notre étude, nous retenons des recherches existantes la dialectique entre syntaxe et sémantique dans le travail de preuve et de démonstration, mis en avant par des travaux tels que ceux de [Weber & Alcock \(2004\)](#) ou [Durand-Guerrier et al. \(2012\)](#), qui mettent en avant l'importance à la fois du sens et de la compréhension des objets mathématiques, des règles formelles logiques de démonstrations, et de leur articulation. Ceci permet de prendre en compte aussi les aspects liés à la structure d'une démonstration. Pour les enjeux de preuve et démonstration dans l'enseignement supérieur, nous retenons aussi la distinction entre vérité et validité ([Durand-Guerrier, 2008](#); [Durand-Guerrier et al., 2019](#)), qui permet de distinguer la vérité d'un énoncé mathématique (i.e. vrai ou faux) de la validité de la preuve qui permet de l'attester (i.e. relativement à une théorie). Ces deux points de vue, syntaxe/sémantique et vérité/validité, permettent de prendre en charge une partie des questions de logique sous-jacentes.

Pour intégrer le point de vue de l'étudiant·e, nous proposons de considérer la notion de *valeur épistémique* d'un énoncé ([Duval & Egret, 1993](#)), c'est-à-dire le degré de certitude ou de conviction attaché à une proposition :

Une proposition énoncée apparaît d'emblée évidente, absurde, vraisemblable, arbitraire, nécessaire, ou possible,... Ces valeurs épistémiques doivent être soigneusement distinguées des valeurs de vérité, car elles sont indépendantes. [...] La valeur épistémique d'une proposition énoncée reste souvent implicite. Son explicitation se fait par le recours à des verbes dits "d'attitude propositionnelle", c'est à dire des verbes, ou des expressions, qui introduisent comme complétives la proposition dont ils explicitent la valeur : croire que,... penser que,... imaginer que [...].

Nous étendons cette idée au cas de la validité d'une preuve, et parlerons de valeur épistémique d'un énoncé (évident, absurde, nécessaire...) ou d'une preuve (convaincante, suspecte...).

Évaluation de preuves. Bien que ne concernant pas directement l'évaluation entre pairs, il existe des études sur l'évaluation et la validation de preuves, et en particulier sur les critères d'acceptation de preuves, qui sont pertinentes pour notre sujet ([Selden & Selden, 2003](#); [Sommerhoff & Ufer, 2019](#), par exemple). Comme souligné par [Sommerhoff & Ufer \(2019, p. 717, notre traduction\)](#), « le concept de preuve, ses fonctions, et les normes et valeurs qui s'y rapportent [...] diffèrent d'un contexte et d'une communauté à l'autre ». Par conséquent, l'évaluation de preuves diffère selon la position de l'évaluateur·rice : chercheur·se, enseignant·e, étudiant·e, élève. La pratique de validation des mathématicien·nes expert·es est elle-même un sujet d'étude ([Weber, 2008](#)). Nous retenons pour notre étude certaines catégorisations proposées dans ces articles.

Dans leur recherche, [Selden & Selden \(2003\)](#) étudient les arguments de validation employés par des étudiant·es en licence. Ils montrent que ces étudiant·es se concentrent en priorité sur l'évaluation de détails locaux, plutôt que les problèmes globaux ou structuraux. Les auteur·rices notent également que le *sentiment de compréhension* (*a feeling of understanding*) est un critère important dans l'évaluation de la validité d'une preuve.

[Sommerhoff & Ufer \(2019\)](#) étudient les critères de validation d'élèves, d'étudiant·es, et de mathématicien·nes. Ils catégorisent les critères de validation dans plusieurs familles :

- Les critères portant sur la structure (*structure-oriented criteria*, dimension syntaxique), visant à assurer la validité formelle de la preuve ;
- les critères portant sur le sens (*meaning-oriented criteria*, dimension sémantique), qui servent à assurer la compréhension de l'évaluateur·rice, et peuvent inclure des critères "esthétiques" comme la simplicité d'une preuve ;
- les critères sociaux (*social criteria*), pour lesquels la validation repose sur les décisions et le comportement d'autres personnes. Par exemple, la réputation de l'auteur·rice d'une preuve peut jouer dans la validation de celle-ci.

Les auteurs soulignent que les critères sur la structure ont une fonction de vérification, et que les critères sur le sens ont plutôt une fonction d'explication et de compréhension. Pour l'analyse de leur expérimentation, ils proposent une catégorisation des critères de validation.

Les deux critères les plus que développent le plus les élèves, étudiant·es et mathématicien·nes sont : *logical chain* et *proof structure*. Le premier réfère à la « validité locale de chaque étape d'une preuve », tandis que le second concerne la « structure globale d'une preuve, et sur la question de savoir si elle réduit complètement l'énoncé aux axiomes, aux prérequis et aux énoncés prouvés, et ne contient pas, par exemple, de raisonnement circulaire » (Sommerhoff & Ufer, 2019, p. 720, notre traduction).

Dans notre expérimentation, au cours de l'évaluation de preuves, les étudiant·es devront lire et comprendre des preuves produites par d'autres étudiant·es, déterminer si elles sont valides, et les corriger. Il·elles vont donc être amené·es à utiliser des critères de validation de preuve. Les études présentées dans cette section permettront d'éclairer nos analyses.

2.E. Bilan et formulation des questions de recherche

La littérature étudiée dans cette partie indique que les élèves, sous certaines conditions, fournissent des évaluations de pairs valides. Elle met à jour certains critères qui influent significativement sur la validité des évaluations et sur les apprentissages qui peuvent se produire. L'évaluation par les pairs permet par exemple aux élèves d'améliorer leur performance sur une tâche particulière, ou de développer des capacités métacognitives comme l'autoévaluation. Ce type d'évaluation semble donc orienté vers des fins formatives, pour les élèves évalué·es (rétroactions sur leur travail) mais aussi pour les élèves évaluateur·rices, ce qui fait la spécificité de l'évaluation par les pairs. Notons aussi que plusieurs articles donnent des conseils pratiques quant à la mise en place d'une telle évaluation (nous y reviendrons en fin d'article).

De nombreuses recherches se concentrent donc sur la validité et la fiabilité de l'évaluation entre pairs, à différents niveaux et dans différentes disciplines. Nous nous focalisons sur l'évaluation entre pairs en mathématiques, plus spécifiquement sur l'activité des étudiant·es au cours de ce processus, un sujet qui semble peu abordé dans la littérature.

Nous faisons l'hypothèse que l'activité des étudiant·es est différente lors d'une évaluation de pairs par rapport à la réalisation de la tâche évaluée elle-même, en particulier dans le travail de preuve. Nous nous concentrons dans cet article sur l'évaluation par les pairs, en mathématiques, dans les premières années de l'enseignement supérieur. Ceci amène la problématique suivante : Quelle est l'activité d'étudiant·es en situation d'évaluer des pairs ? Et en quoi diffère-t-elle de celle dans des situations ordinaires d'élaboration de preuves ? L'étude de cette question amène les sous-questions suivantes :

- Quels éléments les étudiant·es prennent-ils-elles en compte dans leurs évaluations ?
- Quelles corrections et annotations les étudiant·es apportent-ils-elles à une copie qu'il·elles évaluent ?
- À quelles ressources font appel les étudiant·es en situation d'évaluer des pairs ?
- De quelle manière la position de correcteur·rice contribue-t-elle à l'apprentissage des étudiant·es ?

Pour apporter des éléments de réponse à ces questions, nous avons expérimenté un processus d'évaluation entre pairs en première année d'enseignement supérieur scientifique (plus précisément en CPGE MPSI²), et avons recueillis différentes données : des résolutions d'exercices et les évaluations de ces exercices par les étudiant·es, des vidéos d'entretiens d'explicitation de binômes d'étudiant·es en situation d'évaluation de pairs, ainsi que les réponses des étudiant·es à un questionnaire interrogeant leur perception de l'activité d'évaluation de pairs. Nous détaillons la méthodologie de cette étude dans la section suivante.

2. Classe Préparatoire aux Grandes Écoles, filière Mathématiques-Physique-Sciences de l'Ingénieur

3. Expérimentation et méthodologie

Cette section est consacrée à la présentation de l'expérimentation que nous avons menée. Nous commençons par expliciter notre positionnement théorique.

3.A. Cadrage théorique

Notre projet est d'étudier l'activité d'étudiant·es mis·es en situation d'évaluation de pairs. Notre hypothèse, en cohérence avec les recherches existantes, est que mettre les étudiant·es dans une telle situation a des effets sur la nature de leurs activités, et notamment engendre des activités non-usuelles (sur les plans mathématique et cognitif, mais aussi méta-mathématique et méta-cognitif) que nous souhaitons identifier. Pour cela, nous nous positionnons implicitement, sans nécessairement en reprendre les outils, dans le paradigme de la théorie de l'activité, telle qu'elle peut-être utilisée en didactique des mathématiques (Robert & Vandebrouck, 2017). Ainsi, si nous adoptons principalement un modèle d'élève sujet épistémique, nous ne nous interdisons pas d'intégrer d'autres dimensions du sujet (psychologiques, sociales...) dans notre analyse, bien que nous ne cherchons pas à les catégoriser.

Nous nous appuyerons sur une méthodologie de type ingénierie didactique (Artigue, 1988), qui permet un contrôle sur la conception et l'expérimentation de la tâche proposée, offre dans un mode de validation interne basé sur la comparaison analyse *a priori* - analyse *a posteriori*. Ceci nous permettra aussi d'étudier les activités possibles et activités réelles des étudiant·es face à la tâche proposée. Nous intégrerons à ces analyses la typologie de Coppé (1993), déjà utilisée par Coppé & Moulin (2017) pour étudier des processus d'évaluation entre pairs, que nous adapterons à notre situation. Cette analyse *a priori* est également éclairée par certaines catégories de critères d'acceptation, introduites par Sommerhoff & Ufer (2019) et Selden & Selden (2003).

3.B. Dispositif expérimental

Comme annoncé, notre expérimentation a été réalisée auprès d'un public en début de l'enseignement supérieur, en classe préparatoire MPSI, en apprentissage d'une pratique avancée des mathématiques. L'expérimentation s'est déroulée dans une classe de 42 étudiant·es. Nous sommes conscients qu'un contrat didactique d'évaluation bien spécifique existe en CPGE (évaluations écrites régulières, attention forte à la rédaction, préparation aux "normes d'évaluation" des concours, rôle formatif de l'évaluation...), et nous en discuterons en conclusion. Cependant ce contexte peut aussi être pertinent pour notre étude, puisqu'il s'agit d'identifier l'activité que peuvent développer des étudiant·es dans un contexte où il·elles ont construit un rapport à l'évaluation propice à entrer dans un processus d'évaluation entre pairs.

Pour concevoir l'expérimentation et sélectionner les exercices, nous nous sommes basés sur les programmes de MPSI (définis à l'échelle nationale), puis le dispositif a été soumis à l'enseignant pour approbation. Nous avons proposé des exercices aux étudiant·es, puis les avons mis·es en situation d'évaluation entre pairs. L'expérimentation en elle-même s'est déroulée pendant trois séances d'une heure de cours, complétées d'évaluations réalisées à la maison pour la majorité des étudiant·es. Nous nous sommes appuyés sur les conseils de la littérature examinée précédemment, notamment concernant l'implication des étudiant·es dans une discussion sur les critères d'évaluation, au cours de la deuxième séance du dispositif. Les séances se sont déroulées de la manière suivante :

- Pendant la première séance, l'expérimentation a été présentée aux étudiant·es et il·elles ont cherché et rédigé les exercices proposés. Nous avons ensuite récupéré et numérisé les copies.
- Au cours de la deuxième séance une discussion collective a été menée, d'une part sur la résolution des exercices, et d'autre part sur l'évaluation. Pour le premier point, nous avons fait expliciter aux étudiant·es les stratégies qu'il·elles avaient menées pour résoudre les deux exercices. Pour le second point, nous avons demandé aux étudiant·es de quoi était constituée, selon eux, l'évaluation d'une copie. Les points suivants ont été discutés :

- la présentation de la copie ;
 - le "passif" de l'étudiant·e dont on corrige la copie (dans l'année, ou dans la copie)³ ;
 - le raisonnement dans la copie : principe de "mort subite" utilisé par l'enseignant dans ses corrections, raisonnement correct, incorrect, incomplet ;
 - l'évaluation d'un calcul (résultat, validité des étapes...).
- Durant la semaine suivante, les étudiant·es ont tous·tes reçu deux copies à évaluer en travail à la maison, et certain·es étudiant·es volontaires ont réalisé leurs évaluations en binôme sous observation. Toutes les copies corrigées ont été récupérées et nous avons complété l'évaluation pour que les corrections rendues soient valides et complètes.
- La troisième et dernière séance a été consacrée à une correction complète des exercices, et à des commentaires sur les erreurs récurrentes rencontrées dans les copies.

Pour guider les étudiant·es dans l'évaluation, en amont de la séance 2, nous avons réalisé une feuille d'"éléments de corrections", disponible en annexe C. Elle a été conçue comme un support de réflexion pour l'évaluation et présente différentes démarches pour résoudre les exercices. Nous n'avons pas à ce moment là fourni de correction complète, pour ne pas normaliser les évaluations des étudiant·es, et éviter que l'évaluation se résume à une vérification phrase par phrase. Pendant l'évaluation, tous·tes les étudiant·es ont également eu accès à leurs notes de cours. À l'issue de la séance 3, nous avons transmis aux étudiant·es un document présentant différentes résolutions possibles des exercices proposés, ainsi qu'une compilation des erreurs les plus fréquemment rencontrées.

Nous détaillons dans la suite les données recueillies, les méthodes d'analyse, les exercices proposés, ainsi que les analyses *a priori*.

Données recueillies. À l'issue de la première séance, nous avons numérisé toutes les copies produites par les étudiant·es. Sur le plan du dispositif, cela permet de faire évaluer une même copie par plusieurs étudiant·es, et de faire évaluer plusieurs copies à chacun·e. Sur le plan de la recherche, cela nous permet d'accéder au travail initial produit par chaque étudiant·e.

À l'issue de la deuxième séance, nous avons donné aléatoirement à chaque étudiant·e deux copies à évaluer, préalablement anonymisées. La consigne était d'évaluer les copies et de les annoter à destination de leur auteur·rice. Ces copies corrigées ont été numérisées avant d'être rendues à leur auteur·rice. Ces données sont complétées par des formes d'entretiens d'explicitations, pour les étudiant·es volontaires. Une dizaine d'étudiant·es ont été rassemblé·es en binômes, que nous avons observés en train d'évaluer des copies, avec une consigne d'explicitation à voix haute ce qu'il·elles font. Avec l'aide de l'enseignant nous avons formé des binômes d'étudiant·es évaluateur·rices de niveaux proches, afin qu'un échange entre les deux étudiant·es ait lieu, plutôt qu'une correction par celui ou celle de meilleur niveau. Ce dispositif a un intérêt sur le plan de la recherche : les échanges à voix haute des étudiant·es et la consigne d'explicitation ce qu'il·elles font nous permettent d'accéder, au moins partiellement, à l'activité des étudiant·es pendant qu'il·elles évaluent leurs pairs. Ces séances de correction en binôme ont été filmées.

Précisons qu'après une rapide revue des copies, nous avons effectué une sélection pour celles que nous avons fournies aux étudiant·es volontaires pour les entretiens. Nous avons fait en sorte que certaines de ces copies présentent des stratégies de résolution classiques et originales, et qu'elles fassent apparaître différents types d'erreurs que nous avons identifiés dans les copies.

Au final, notre corpus est composé de :

- 44 copies originales,
- 64 copies évaluées à la maison par 30 élèves,
- 6 entretiens d'explicitations au cours desquels ont été évaluées entre 2 et 4 copies.

3. Le passif de l'étudiant·e est un critère social (*social criteria*), qui n'a pas été rencontré en pratique dans les évaluations.

Enfin, ce corpus a été complété d'un questionnaire (donné en annexe B) rempli par 38 étudiant·es, permettant de recueillir leurs impressions sur le processus d'évaluation entre pairs qu'il·elles ont vécu. Sur le plan de la recherche, ces données nous permettent d'accéder à la perception de la tâche par les étudiant·es et à des discours sur leur activité, possiblement sur les plans méta-cognitifs et méta-mathématiques.

Méthodologie d'analyse. Après avoir sélectionné les tâches données aux étudiant·es (section 3.B.) nous avons réalisé une analyse *a priori* des stratégies de résolution des exercices, et une analyse *a priori* des activités possibles lors de leur évaluation (section 3.C.). Ceci nous a permis de construire une grille d'analyse des évaluations, complétée par la typologie des arguments de vérification de Coppé & Moulin (2017). Une fois les différentes données recueillies, la grille nous a permis d'identifier l'activité des étudiant·es dans les évaluations en binômes. Ces premiers résultats nous fournissent des indications quant à l'activité des étudiant·es en évaluation à la maison, activité à laquelle on ne peut accéder directement.

À la fin de l'expérimentation, nous avons fait passer un questionnaire aux étudiant·es, afin de recueillir leurs impressions sur l'expérimentation et la perception qu'il·elles ont de leur activité d'évaluation (section 5.). Nous analyserons ces réponses au regard des résultats des analyses qui les précèdent.

Tâches proposées. Pour concevoir les exercices, nous avons d'abord sélectionné des notions familières des étudiant·es, déjà pratiquées en cours. Les chapitres "Réels", sur le thème des bornes supérieures et inférieures, et "Ensembles" ont été choisis. Nous avons ensuite cherché des exercices qui auraient pu être posés en interrogation écrite ou orale au moment de l'expérimentation.

Au moment de l'expérimentation, l'enseignant avait déjà traité le programme sur ces deux chapitres. Dans le chapitre sur les nombres réels et les suites numériques, les notions de bornes supérieure et inférieure, leur existence, le théorème de la limite monotone, et la caractérisation séquentielle des bornes sup et inf sont au programme. Le programme sur les ensembles inclut en particulier les notions d'appartenance, d'inclusion, d'opérations sur un ensemble, et de produit cartésien. L'enseignant nous a confirmé que les exercices que nous avons sélectionnés relèvent de types de tâche déjà abordés en travaux dirigés. De plus, ces types d'exercices sont classiques du programme de première année de l'enseignement supérieur.

Enfin, il était intéressant que les stratégies de résolution de ces exercices soient variées, et amènent une diversité d'erreurs potentielles dans les raisonnements. De cette manière, les étudiant·es peuvent être confronté·es à des raisonnements classiques et d'autres plus originaux, mais aussi à une diversité de raisonnements erronés.

Ceci a mené au choix des deux exercices suivants :

Exercice 1. On considère l'ensemble $S = \left\{ \frac{n - \frac{1}{n}}{n + \frac{1}{n}}, n \in \mathbb{N}^* \right\}$.

S admet-il une borne supérieure ? une borne inférieure ? un minimum ? un maximum ? Si oui, les déterminer.

Exercice 2. Soient E et F deux ensembles, A_1, A_2 des parties de E , et B_1, B_2 des parties de F . On note $A \times B = \{(a, b), a \in A \text{ et } b \in B\}$.

1. Montrer que $(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$.
2. A-t-on toujours $(A_1 \times B_1) \cup (A_2 \times B_2) = (A_1 \cup A_2) \times (B_1 \cup B_2)$?

Dans l'exercice 2, la première question demande une preuve très formelle et requiert une bonne maîtrise de la syntaxe mathématique. La deuxième peut amener à de nombreux contre-exemples. Cependant, nous avons remarqué *a posteriori* que les étudiant·es avaient très peu réussi cet exercice et qu'il avait amené à de nombreuses erreurs de résolution et de correction. L'enseignant nous a confirmé

que ce type d'exercice n'était pas encore maîtrisé par la plupart des étudiant-es⁴. Nous avons donc choisi de ne pas présenter en détail les analyses du second exercice et de nous concentrer sur l'exercice 1.

Dans la partie suivante, nous présentons les analyses *a priori* de l'expérimentation.

3.C. Analyses a priori du problème

Dans toute la suite, nous nous concentrons sur le premier des deux exercices :

Exercice 1. On considère l'ensemble $S = \left\{ \frac{n - \frac{1}{n}}{n + \frac{1}{n}}, n \in \mathbb{N}^* \right\}$.

S admet-il une borne supérieure ? une borne inférieure ? un minimum ? un maximum ? Si oui, les déterminer.

Nous analysons d'abord les stratégies liées à la tâche de résolution de l'exercice, puis nous analysons la tâche d'évaluation associée.

3.C.a. Analyse a priori des stratégies de résolution

Nous présentons ici les principales stratégies de résolution du premier exercice. Il peut être résolu en adoptant un point de vue "suites numériques" ou un point de vue "ensembles" sur les objets en jeu. Bien entendu, des méthodes hybrides entre ces deux points de vues sont envisageables.

Notons pour tout $n \in \mathbb{N}^*$, $s_n = \frac{n - \frac{1}{n}}{n + \frac{1}{n}}$.

En adoptant le point de vue des suites numériques, on montre que la suite (s_n) est croissante. Ceci est possible par exemple :

- en étudiant la différence $s_{n+1} - s_n$;
- en étudiant le quotient s_{n+1}/s_n ;
- en étudiant la croissance de la fonction $f : x \in \mathbb{R} \mapsto \frac{x - \frac{1}{x}}{x + \frac{1}{x}}$.

L'ensemble des termes de la suite est donc minoré par le premier terme s_1 , ce qui prouve l'existence de la borne inférieure et du minimum de l'ensemble S , et permet de trouver sa valeur : $\inf(S) = \min(S) = s_1 = 0$. Pour montrer l'existence de la borne supérieure de S , on peut commencer par montrer que (s_n) converge vers une limite finie :

$$\text{Pour tout } n \in \mathbb{N}^*, \text{ on a : } s_n = \frac{n - \frac{1}{n}}{n + \frac{1}{n}} = \frac{1 - \frac{1}{n^2}}{1 + \frac{1}{n^2}} \text{ donc } s_n \xrightarrow{n \rightarrow +\infty} 1.$$

Le théorème de la limite monotone permet de conclure que l'ensemble S a une borne supérieure, égale à 1. Il reste alors à montrer que cette borne supérieure n'est pas un maximum. Supposons par l'absurde qu'il existe un entier $n \in \mathbb{N}^*$ tel que $\frac{n - \frac{1}{n}}{n + \frac{1}{n}} = 1$. Cette égalité est équivalente à $n - \frac{1}{n} = n + \frac{1}{n}$, elle-même équivalente à l'égalité $-1 = 1$, ce qui est une contradiction. Aucun élément de l'ensemble S n'est donc égal à 1, la borne supérieure n'est pas un maximum.

4. Ceci confirme la nécessité que l'évaluation entre pairs porte sur des notions suffisamment installées chez les étudiant-es. Nous y reviendrons en conclusion.

Retrouvons ce résultat avec un point de vue "ensembles". On peut encadrer directement les éléments de l'ensemble. En effet, pour tout $n \in \mathbb{N}^*$, on a :

$$n - \frac{1}{n} \geq 0 \text{ et } n + \frac{1}{n} > 0 \text{ donc } \frac{n - \frac{1}{n}}{n + \frac{1}{n}} \geq 0$$

et de plus

$$n - \frac{1}{n} < n + \frac{1}{n} \text{ donc } \frac{n - \frac{1}{n}}{n + \frac{1}{n}} < 1.$$

Ceci montre que pour tout $n \in \mathbb{N}^*$, $0 \leq s_n < 1$. L'ensemble S admet donc une borne inférieure et une borne supérieure, qui vérifient $\inf(S) \geq 0$ et $\sup(S) \leq 1$. Or $s_1 = 0$, donc $\inf(S) = \min(S) = 0$. Pour montrer $\sup(S) = 1$, on peut utiliser la caractérisation séquentielle de la borne supérieure, et montrer comme ci-dessus que $s_n \xrightarrow{n \rightarrow +\infty} 1$. On peut également faire appel à la caractérisation "avec des

epsilons" de la borne supérieure. Soit donc $\varepsilon > 0$, montrons qu'il existe $n_0 \in \mathbb{N}^*$ tel que $1 - \frac{n_0 - \frac{1}{n_0}}{n_0 + \frac{1}{n_0}} < \varepsilon$.

Or si $n \in \mathbb{N}^*$, on a les équivalences :

$$1 - \frac{n - \frac{1}{n}}{n + \frac{1}{n}} < \varepsilon \iff \frac{2}{n^2 + 1} < \varepsilon \iff \frac{2}{\varepsilon} - 1 < n^2.$$

Si $\frac{2}{\varepsilon} - 1 \leq 0$, $n_0 = 1$ convient. Sinon, on pose n_0 la partie entière supérieure de $\sqrt{\frac{2}{\varepsilon} - 1}$, qui vérifie bien l'inégalité demandée. On a donc montré que la borne supérieure de S est égale à 1. De plus, l'inégalité $s_n < 1$ vraie pour tout $n \in \mathbb{N}^*$ permet de conclure que l'ensemble S n'a pas de maximum.

Notons qu'on peut réécrire les éléments de l'ensemble grâce à l'égalité $s_n = 1 - \frac{2}{n^2 + 1}$. Ceci permet de montrer rapidement que la suite (s_n) est croissante, qu'elle tend vers 1, ou encore que 1 n'est pas un élément de l'ensemble.

Retour sur nos contraintes. Cet exercice vérifie donc les contraintes que nous nous sommes données : les visions "suites" ou "ensembles" permettent de varier les types de raisonnements rencontrés dans les copies. Au sein de la vision "suites", différentes démonstrations sont possibles pour montrer une même propriété, comme la croissance de la suite. Nous nous attendons donc à rencontrer divers raisonnements et permettant des activités différentes lors des évaluations. L'énoncé de cet exercice pourrait être considéré comme favorable à une activité multipreuve, ou *multiple proof tasks* (Dreyfus et al., 2012).

3.C.b. Éléments d'analyse *a priori* des évaluations

Une analyse *a priori* de l'activité des étudiant·es et des arguments d'évaluation est présentée dans cette partie, et a servi de grille d'analyse des évaluations. Cette analyse s'articule en plusieurs niveaux. Cinq grands types d'activités sont définis, chacun étant décliné en deux sous-niveaux, qui précisent et détaillent les niveaux précédents. Une description des arguments d'évaluation adaptée de Coppé & Moulin (2017) est ensuite utilisée.

Classification générale des évaluations. Nous avons distingué trois catégories d'évaluations : celles portant sur le fond du raisonnement mathématique, celles sur la forme, et celles ne portant pas sur le raisonnement mathématique. Au sein de ces catégories, on retrouve les types d'activités suivantes :

Catégorie 1 Activités relevant de l'évaluation d'une preuve proposée dans la copie ;

Catégorie 2 Activités relevant de l'évaluation d'une preuve envisagée par l'étudiant·e corrigé·e, selon l'évaluateur·rice ;

Catégorie 3 Activités relevant de l'évaluation du résultat d'une preuve proposée dans la copie ;

Catégorie 4 Activités relevant de l'évaluation de la rédaction ou de la syntaxe d'une preuve ;

Catégorie 5 Activités ne portant pas sur l'évaluation du raisonnement mathématique.

Notons qu'une distinction est faite entre l'évaluation d'une preuve telle qu'elle est proposée dans la copie, et l'évaluation d'une interprétation de cette preuve, faite par l'évaluateur-riche.

Détails du premier niveau de classification. Chaque catégorie a été déclinée en plusieurs familles d'activités. Par exemple, lorsque les étudiant·es évaluent une preuve proposée dans la copie - catégorie 1 -, il·elles sont amené·es à :

- Z Lire la preuve ;
- A Essayer de comprendre la preuve ;
- B Évaluer la valeur épistémique de la preuve ;
- C Évaluer la validité de la preuve ;
- D Corriger la preuve.

Ces activités peuvent apparaître simultanément ou séparément. Par exemple, certains étudiant·es lisent toute la copie, puis évaluent la validité des raisonnements tout en les corrigeant. D'autres pourront lire et évaluer la validité dans un même temps, puis corriger chaque exercice.

Notons que ces catégories renvoient à des fonctions différentes de l'évaluation : la compréhension du raisonnement, la validation ou l'invalidation du raisonnement, puis la correction. Ceci fait écho à la catégorisation des critères de validation de [Sommerhoff & Ufer \(2019\)](#) : les critères portant sur la structure, qui permettent la validation, et les critères portant sur le sens, qui permettent la compréhension.

Détails du deuxième niveau de classification. Au sein de chaque famille, nous avons enfin identifié plusieurs activités que les étudiant·es peuvent mettre en place lors de l'évaluation. Détaillons l'ensemble *Évaluer la validité de la preuve*. Les étudiant·es vont être amené·es à valider ou invalider la preuve complète (*proof structure*) ou seulement une partie de celle-ci (*logical chain*, [Sommerhoff & Ufer, 2019](#)) par exemple un calcul, une implication ... Les étudiant·es s'interrogent aussi sur la démarche, "l'idée de la preuve", entamée dans la copie. Nous incluons dans cette famille d'activités les questionnements sur la validité de la preuve, y compris lorsqu'il n'aboutissent pas.

Décrivons à présent la famille *Essayer de comprendre la preuve*. Les étudiant·es peuvent chercher à identifier les différentes étapes de la preuve qu'il·elles évaluent. Dans notre exercice, un·e étudiant·e pourra d'abord montrer que la borne supérieur existe en majorant les éléments de l'ensemble, puis montrer que la suite est croissante pour conclure sur la valeur de la borne supérieure. Pendant ce processus, les évaluateur·rices sont amené·es à identifier les concepts mathématiques mobilisés, implicitement ou explicitement, dans la copie. Lors des entretiens, les étudiant·es peuvent expliquer le raisonnement mené dans la copie (ou une partie de celui-ci) à leur camarade. Enfin, certain·es étudiant·es peuvent anticiper la suite du raisonnement en même temps qu'il·elles lisent et détaillent la preuve.

Une classification des arguments de vérification. Pour enrichir l'analyse de l'activité des étudiant·es, nous avons introduit un deuxième niveau de description, basé sur la classification de ([Coppé & Moulin, 2017](#)). Dans les entretiens en binômes, nous identifions dans le discours des étudiant·es des arguments à visée de validation ou d'invalidation de preuves, de correction ou encore d'explication d'une preuve. Pour formuler ou étayer ces arguments, les étudiant·es peuvent être amené·es à faire appel à des références comme le cours, la feuille de correction, ou leur propre expérience.

Nous appelons argument ou référence *internes* ceux qui font intervenir intervenir un savoir ou de savoir faire mathématique. Parmi les sous-catégories identifiées par Coppé, pour l'analyse *a priori* de notre problème, nous avons retenu les vérifications suivantes :

1. les vérifications d'éléments techniques :
 - vérifications d'un calcul à la main ;
 - vérifications d'une preuve à la main ;
 - vérifications de la présence (ou non) des hypothèses d'un théorème ;
 - production d'exemples ou de contre-exemples pour mettre à l'épreuve un énoncé ;
2. les références aux résultats et méthodes du cours ou d'une source experte ;
3. les vérifications par changement de cadre (non observées en pratique).

Les arguments ou références ne nécessitant pas de raisonnement mathématique sont dits *externes*. Nous avons sélectionné *a priori* :

1. les références à l'énoncé de l'exercice ;
2. les références à une expérience personnelle de l'étudiant·e ;
3. les références aux éléments de corrections ;
4. les références à la mémoire du savoir enseigné (du type "je suis sûr·e de moi car on avait fait le même exercice en classe", "le prof a fait comme ça", "on fait toujours comme ça") ;
5. les références au contrat didactique (non observées en pratique).

Précisons que certaines sources, comme le texte du cours, sont externes à l'étudiant·es, mais sont des textes mathématiques. La citation d'un théorème du cours est donc considérée comme référence interne, car elle fait intervenir un savoir mathématique.

3.C.c. Retour sur nos choix expérimentaux

Rappelons que nous avons choisi, suite à l'expérimentation, de nous concentrer sur les évaluations du premier exercice. L'enseignant de la classe a confirmé que les étudiant·es avaient à disposition les connaissances pour résoudre l'exercice. L'analyse *a priori* de cet exercice montre la diversité des résolutions possibles, faisant appel aux domaines des suites numériques et des ensembles. Les étudiant·es peuvent donc mobiliser différentes notions, ce qui pourra produire des résolutions puis des évaluations diverses.

L'analyse *a priori* de l'évaluation nous a permis de construire une grille de codage, utilisée pour analyser le travail réalisé par les binômes. Sa version finale est présentée en annexe A. Cette grille est adaptée aux entretiens, durant lesquels les étudiant·es verbalisent leurs réflexions au cours de l'évaluation. En revanche, il paraît compliqué d'identifier toutes ces activités uniquement à partir des copies annotées. Cette grille est donc amenée à être adaptée et allégée pour être pertinente dans ce cadre. Par conséquent, dans la suite de ce texte, nous analyserons d'abord les entretiens, et nous chercherons ensuite des traces, dans les copies annotées, des activités que nous aurons identifiées préalablement.

4. Résultats des analyses de l'activité des étudiant·es

Avant de présenter les analyses des entretiens et des copies, nous donnons en première partie quelques éléments de repère concernant la résolution et l'évaluation des exercices par les étudiant·es.

4.A. Premières remarques sur la résolution et l'évaluation des exercices

Sur la résolution des exercices. Avant l'évaluation par les étudiant·es, nous avons consulté l'ensemble des copies afin de sélectionner celles que nous avons données à corriger en entretiens. L'analyse de ces copies nous permet de faire un état des lieux des stratégies utilisées par les étudiant·es et des erreurs les plus courantes.

La stratégie la plus fréquemment employée (34 étudiant·es) passe par l'étude de la croissance de la suite. Pour cette stratégie, les étudiant·es sont réparti·es en :

- calcul de la dérivée de la fonction associée : 25 étudiant-es ;
- étude de $s_{n+1} - s_n$: 7 étudiant-es ;
- étude de s_{n+1}/s_n : 1 étudiant-e ;
- décomposition en éléments simples du terme général : 1 étudiant-e (qui étudie ensuite $s_{n+1} - s_n$).

Parmi les étudiant-es qui n'ont pas utilisé cette stratégie, une partie argumente que le premier terme de la suite est la borne inférieure de l'ensemble et que la limite de la suite est la borne supérieure, sans plus de justifications. Deux étudiant-es tentent d'utiliser la caractérisation séquentielle de la borne supérieure, et un étudiant-e cite la caractérisation "avec des ε ", sans l'utiliser ensuite. Au final, aucun-e de ces étudiant-es n'a produit de preuve valide que la borne supérieure de l'ensemble est 1. Enfin, certain-es étudiant-es minorent la suite par 0 et montrent l'égalité $s_1 = 0$ pour conclure quant à la valeur de la borne inférieure.

Le premier examen des copies nous a également permis d'identifier les erreurs les plus fréquentes (présentes dans plus de 5 copies) dans le premier exercice. Il s'agit de :

- si $s_n \rightarrow 1$ lorsque $n \rightarrow +\infty$; alors (s_n) est majorée par 1 ;
- si $s_n \rightarrow 1$ lorsque $n \rightarrow +\infty$, alors $\sup(S) = 1$;
- si $s_1 = 0$, alors 0 minore S ;
- dérivation de la fonction $f : \mathbb{N}^* \rightarrow \mathbb{R}$ associée à la suite (s_n) .

Sur l'évaluation des copies. 30 étudiant-es ont chacun-e évalué 2 copies en travail à la maison. À l'examen de ces évaluations, on observe une grande variabilité en fonction de l'évaluateur-riche, tant au niveau de la validité de la correction que dans les annotations. Certain-es étudiant-es donnent de nombreux commentaires et conseils, tandis que d'autres se contentent de valider ou d'invalider les résultats. Des exemples de ces comportements sont présentés dans les figures 1 et 2. À la fin de l'expérimentation, dans une perspective formative, nous avons donc examiné toutes les évaluations afin de les compléter et les corriger si nécessaire.

Concentrons nous à présent sur les résultats des entretiens. Rappelons que lors de ces entretiens, des binômes d'étudiant-es ont évalué ensemble entre 2 et 4 copies, avec la consigne d'explicitement leurs discussions, réflexions et commentaires sur l'évaluation des copies.

4.B. Analyse des entretiens

Pour l'analyse, nous avons travaillé à partir de transcriptions des entretiens. Nous avons d'abord codé l'activité des étudiant-es de chaque binôme à l'aide de la grille d'analyse présentée ci-dessus. Sur cette base, nous avons réalisé une synthèse de l'activité de chaque binôme, puis essayé de repérer les aspects les plus caractéristiques de leurs activités. Nous présentons ici une synthèse de nos résultats.

Lors de l'analyse des entretiens, nous avons dégagé sept axes caractéristiques de l'activité des étudiant-es. Il s'agit de la compréhension de la preuve, l'évaluation de sa validité ou de sa valeur épistémique, l'attention portée au résultat de la preuve, l'attention portée à la démarche de résolution empruntée dans la copie, les corrections apportées par les évaluateur-rices, les remarques concernant la syntaxe et la rédaction et leurs liens avec l'invalidation de la preuve, et enfin les jugements de valeur. Pour illustrer nos résultats, nous choisissons de présenter dans la suite plus en détails trois de ces axes, qui nous semblent au cœur des enjeux de l'entrée dans les mathématiques du supérieur : compréhension de la preuve, validation et invalidation, corrections apportées. Nous reviendrons ensuite sur les arguments et ressources mobilisés par les étudiant-es.

Pour les exemples que nous donnons dans cette section, nous notons les binômes **B1**, **B2**, **B3**, **B4**, **B5**, **B6**.

Passer par une fonction est une bonne méthode.

(*) Soit $f(x) = \frac{xe - \frac{1}{xe}}{xe + \frac{1}{xe}}$ avec $xe \in \mathbb{R}^*$.

Ne pas confondre f le nom de la fonction et $f(x)$ désignant un ou des réels.

qu'est-ce que x ? → définir une bonne fois pour toute au moins vous devez écrire $\forall x \in \mathbb{R}^*$ au début de vos expressions.

Étudions $f(x)$.

Soit $x \in \mathbb{R}^*$

$$f(x) = \frac{1 - \frac{1}{x^2}}{1 + \frac{1}{x^2}}$$

$$f(x) = \left(x - \frac{1}{x}\right) \times \frac{1}{x + \frac{1}{x}}$$

$$= \frac{x^2 - 1}{x} \cdot \frac{xe}{x^2 + 1}$$

$$= (x^2 - 1) \times \frac{1}{1 + x^2}$$

OU ENCORE.

Soit une fonction f telle que $\forall x \in \mathbb{R}^*, f(x) = \frac{x - \frac{1}{x}}{x + \frac{1}{x}}$

Evitez "avec $x \in \mathbb{R}^*$ " en fin de "expression".

Correct. Vous auriez pu simplement multiplier tout par x dans l'expression de f .

Étudions la en \odot : OK, explicités vos calculs:

FIGURE 1 – Un exemple de correction très détaillée de l'exercice 1

La suite est strictement croissante pour $m \geq 1$ et $U_1 = 0$, ainsi: $\forall m \geq 1, U_m \geq 1$.

Un admet donc une borne inférieure qui ici est égale à 0 .

La suite étant donc comprise à l'intervalle $[0; 1[$ on peut trouver des minimum ou des maximum:

Exemple: - Un minimum de l'ensemble est -5

et un maximum de l'ensemble est 12 .

General: $\forall m \geq 1$ et un ~~ensemble~~ $a < 0$ est minimum maximum

FIGURE 2 – Un exemple de correction (incorrecte) de l'exercice 1 avec très peu de détails

Compréhension de la preuve. Lors de l'évaluation d'une copie, les étudiant·es observé·es commencent par lire et essayer de comprendre les preuves proposées. Les activités de lecture et de validation d'une preuve s'effectuent la plupart du temps en parallèle. Il arrive qu'une phase uniquement d'explicitation et d'explication de la preuve précède toute validation ou invalidation, comme pour le binôme **B4**.

B4 : Je sais pas si c'est juste, mais c'est ça qu'il fait.

Lors de l'examen de la preuve, les étudiant·es identifient et détaillent les différentes étapes qui la compose, comme dans la citation suivante du binôme **B4**.

B4 : Après il a démontré qu'elle était croissante, du coup son minimum c'est sa première valeur vu qu'elle est strictement croissante.

Dans cette phase, les étudiant·es sont amené·es à identifier précisément les concepts mathématiques utilisés, qui parfois ne sont pas explicités dans la copie. Nous avons régulièrement observé cette situation dans les entretiens, lorsque les étudiant·es évalué·es utilisent le théorème de la limite monotone sans le citer, ou énoncent un théorème proche. Il arrive alors que les évaluateur·rices interprètent le raisonnement mené dans la copie à la lueur des théorèmes qu'il·elles connaissent.

B3 : En fait il a utilisé la limite monotone, mais dans l'autre sens.

Ceci est aussi une occasion pour les étudiant·es de se rappeler précisément des définitions ou théorèmes du cours, afin de comprendre si les concepts mobilisés le sont de la bonne manière. Par exemple, le binôme **B6** débat à deux reprises sur les différentes caractérisations de la borne supérieure vues en cours : définition, caractérisation séquentielle et caractérisation "avec des epsilons".

B6 : Non non, propriété des bornes sup c'est juste non vide, partie de \mathbb{R} , majoré donc borne sup.

B6-1 : Il existe u_n , si c'est ça.

B6-2 : Non, c'est pour tout epsilon.

B6-1 : Ça c'est caractérisation avec des epsilon, pas caractérisation séquentielle de la borne sup.

B6-2 : Ça c'est la caractérisation pour les suites. Là c'est une suite.

B6-1 : Si, parce que il caractérise l'ensemble des éléments de la suite par un ensemble.

B6-2 : Là l'existence de u_n mais il existe dans quoi u_n ?

B6-1 : C'est vrai que par contre là il manque de dire appartient à $A^{\mathbb{N}}$, mais il manque juste ça.

Notons également que contrairement aux autres binômes, les étudiant·es du binôme **B1** tentent peu d'explicitations et explications détaillées des raisonnements à l'oral. Il·elles se contentent le plupart du temps de lire la preuve à voix haute, et notent quand les démarches mises en œuvre dans la copie sont similaires aux leurs. On remarquera dans la suite que ce binôme utilise beaucoup de références externes pour les validations et les corrections des exercices.

Par ailleurs, s'il·elles n'ont pas réussi à comprendre un raisonnement ou un résultat annoncés dans la copie, les évaluateur·rices peuvent interpréter « ce que l'étudiant·e évalué·e a voulu faire dans la preuve ». Dans une copie corrigée par le binôme **B5**, il est écrit " $u(x) \in [0, 1[$ ", où u est la fonction associée à la suite introduite dans l'exercice 1. L'étudiant·e justifie correctement l'encadrement $u(x) \in [0, 1]$, mais sans prouver l'ouverture en 1, et sans faire le lien avec la borne supérieure de la suite. Le binôme a interprété que l'ouverture de l'intervalle a été citée pour justifier que 1 n'est pas le maximum de l'ensemble. Ces comportements font donc partie de la catégorie *Activités portant sur l'évaluation d'une preuve envisagée par l'élève corrigé, selon le·a correcteur·rice*.

On remarque enfin que les étudiant·es du binôme **B2** ont la spécificité de souvent anticiper les raisonnements à venir dans la copie. En conséquence, si la copie emprunte le chemin de résolution envisagé par les étudiant·es, la preuve est validée même si elle est incorrecte. Réciproquement, une preuve

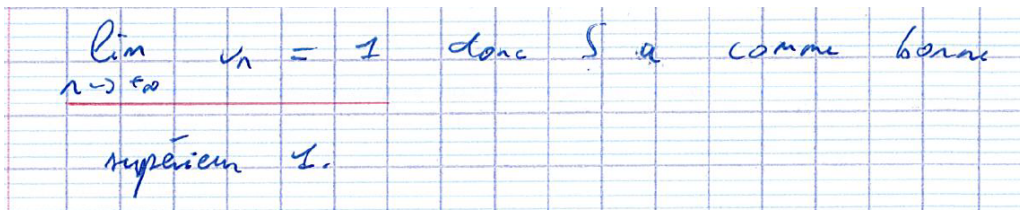


FIGURE 3 – Utilisation implicite de l'énoncé erroné "toute suite convergente converge vers la borne supérieure de l'ensemble de ses éléments"

correcte mais qui utilise une autre méthode peut être invalidée. Ces biais influent donc sur la validité des évaluations produites par ce binôme. On trouve ci-dessous deux exemples de ces anticipations.

Face à une copie dans laquelle il est seulement écrit " u_n croissante et converge vers 1".

B2 : Et du coup si elle converge vers 1, ça veut dire qu'en même temps 1 c'est le max et le sup ça voudrait dire. Est-ce qu'il l'a bien expliqué ?

B2 : Et est ce qu'il dit que 1 appartient à l'ensemble ? Ça vient peut-être après.

En conclusion, on observe toujours un essai de compréhension des preuves proposées dans la copie évaluée, qui est souvent parallèle à une première validation. Les étudiant·es lisent la preuve et identifient explicitement les étapes qui la constituent. Certains relèvent aussi les concepts et énoncés mathématiques en jeu.

Validation et invalidation. Au cours de l'évaluation, tous·tes les étudiant·es discutent de la validité des preuves, résultats ou démarches qu'il·elles rencontrent dans les copies. Nous présentons ici les caractéristiques de l'évaluation de la validité des preuves. Différentes situations peuvent amener les étudiant·es à invalider un élément d'une copie.

Premier cas : Lorsque, implicitement ou explicitement, une propriété avec quantification universelle invalide est utilisée dans la copie, les évaluateur·rices produisent des contre-exemples. La plupart du temps, cela se produit quand l'étudiant·e corrigé·e utilise un théorème du cours sans en vérifier toutes les hypothèses. Par exemple, dans la figure 3, l'étudiant·e semble utiliser l'énoncé "toute suite convergente converge vers la borne supérieure de l'ensemble de ses éléments".

Deux autres raisonnements invalides ont été relevés à plusieurs reprises par les binômes : l'utilisation de la propriété "toute partie non vide et majorée de \mathbb{R} admet une borne supérieure" sans expliciter que l'ensemble est non vide ; et l'utilisation du théorème de la limite monotone sans montrer que la suite est croissante ou qu'elle est majorée. Nous donnons ci-dessous une situation où le binôme **B3** donne un contre-exemple pour invalider la phrase "Pour $n = 1, s_1 = 0$ est la borne inférieure de S ".

B3 : Si elle était décroissante elle serait pas minorée par son premier terme.

Deuxième cas : les étudiant·es relèvent une erreur lorsqu'un résultat leur paraît trop peu justifié. Nous avons fréquemment observé cette situation dans le premier exercice, où des étudiant·es citent que "la limite 1 n'est pas atteinte, donc 1 n'est pas le maximum de l'ensemble" sans démonstration. Cette affirmation est alors invalidée par les évaluateur·rices la plupart du temps. Lorsqu'une preuve est peu justifiée, les étudiant·es sont amené·es à discuter de si elle est "assez justifiée" pour qu'ils la valident. Le binôme **B4** dit par exemple, à propos de la copie présentée en figure 4 :

B4-1 : La première partie de l'exercice 1 elle est bien écrite, mais la deuxième partie sur le max et la borne sup c'est un peu rapide, ça aurait pu être un peu plus détaillé.

B4-2 : Mais c'est la bonne réponse.

B4-1 : Par exemple là "par somme puis quotient", je trouve pas ça assez détaillé.

B4-2 : Après bon là il a pas donné le nom du théorème, mais il a dit qu'elle tendait vers sa borne sup, il manque le nom. Il faut dire pourquoi il dit ça.

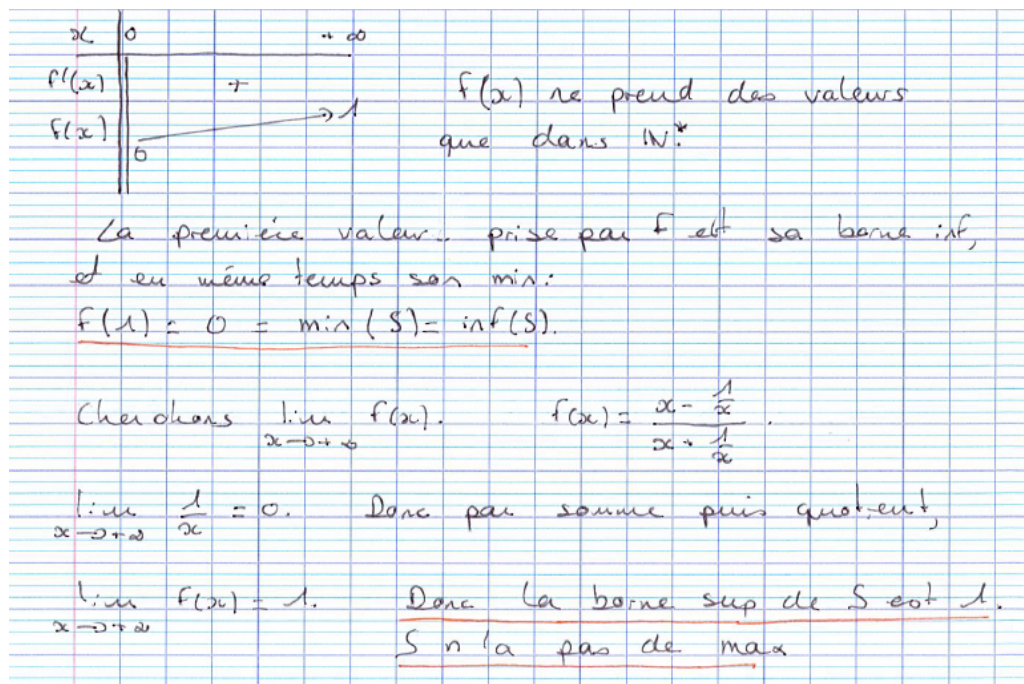


FIGURE 4 – Copie sur laquelle a travaillé le binôme B4

Les évaluateur·rices se posent donc la question des justifications nécessaires pour qu'une preuve soit considérée comme valide. Ils se demandent si une preuve utilisant un théorème sans le citer est valide, ou quels types de justifications sont nécessaires quand on mène des calculs.

Troisième cas : les étudiant·es discutent de la validité de propriétés des objets spécifiques à l'exercice. Par exemple, dans une copie corrigée par le binôme B4 (figure 5), il est affirmé que la fonction associée à la suite s_n est parfois décroissante, ce qui est invalidé par le binôme évaluateur.

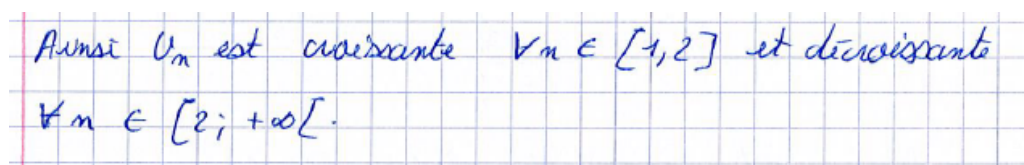


FIGURE 5 – Copie sur laquelle a travaillé le binôme B4

On observe donc plusieurs motifs d'invalidation au cours des entretiens. Elle peut porter sur les propriétés d'un objet utilisé dans le raisonnement, ou sur un théorème annoncé. Les évaluateur·rices peuvent alors justifier l'invalidation d'un énoncé en trouvant un contre-exemple. Les étudiant·es sont aussi amené·es à s'interroger sur les justifications nécessaires ou suffisantes pour valider une preuve ou un calcul. Notons aussi que certaines preuves sont évaluées comme invalides soit par rejet de la démarche entamée, soit parce que le résultat obtenu est incorrect.

Du côté de la validation d'une preuve, elle est le plus souvent directe et sans justification particulière : « Ça, c'est juste. ». La validation est parfois faite au regard des éléments de corrections qui ont été fournis aux étudiant·es. Pour les éléments techniques comme les calculs ou les tableaux de signe, la vérification à la main est systématique pour le binôme B2, et présente ponctuellement pour les binômes B3, B5 et B6. Il arrive aussi que les étudiant·es ne vérifient pas les calculs, mais les valident car il·elles se souviennent avoir obtenu le même résultat lors de leur résolution de l'exercice. Comme nous l'avons remarqué précédemment, les étudiant·es du binôme B1 font souvent référence à leur propre résolution des exercices pour valider les preuves qu'il·elles évaluent :

B1 : En gros c'est ce qu'il a fait, il a calculé le terme 1, ça fait 0. Moi après j'ai fait comme ça.

On observe donc que les validations sont moins argumentées que les invalidations, et font parfois référence aux expériences personnelles des étudiant·es et aux éléments de correction qui leur ont été fournis. Nous retrouvons également que le *sentiment de compréhension*, évoqué par Selden & Selden (2003), est un élément important dans la validation de preuves par les étudiant·es.

Corrections apportées. Après avoir repéré une erreur dans la copie qu'ils évaluent, les binômes observés proposent des éléments de correction, sous forme d'annotations.

Notons d'abord que la correction des erreurs de définition est systématique. La correction des calculs n'est quant à elle pas toujours présente. Certains binômes décortiquent les calculs pour chercher les éventuelles erreurs et les corriger, d'autres se contentent de noter quand le résultat est faux.

Nous avons identifié différents types de correction. Les étudiant·es peuvent compléter le raisonnement entrepris dans la copie, en ajoutant les arguments ou les justifications pour qu'il devienne valide. Cela amène les étudiant·es à réfléchir aux arguments nécessaires et aux arguments suffisants à donner en correction. Dans le dialogue suivant du binôme **B3**, un·e étudiant·e propose une certaine méthode pour montrer que la suite est minorée. L'autre étudiant·e corrige de manière plus générale, en affirmant qu'il faut montrer qu'elle est minorée par 0. Les étudiant·es discutent ensuite de s'il est nécessaire d'étudier la monotonie de la suite pour montrer que la borne inférieure est 0.

B3 : Il faudrait dire que si jamais cette suite elle est croissante, et que du coup elle est minorée par son premier terme. Du coup ça c'est une partie de ça.

La correction du binôme **B4** ci-dessous nous donne un second exemple. Les étudiant·es commencent par invalider l'argument fourni en proposant un contre-exemple : une suite stationnaire à 1 vérifierait les mêmes conditions de croissance, mais aurait un maximum. En regardant les éléments de correction, il·elles affirment ensuite qu'il faudrait encadrer la suite pour montrer que la borne supérieure n'est pas un maximum. La méthode proposée est donc suffisante. Cependant, pour conclure, il est seulement nécessaire de montrer que la borne supérieure n'est pas incluse dans l'ensemble.

B4-1 : Il a mis limite de $f(x) = 1$. Du coup ça a pas de max, mais c'est pas vraiment un max quand même.

B4-2 : Oui c'est vrai qu'elle peut être stationnaire.

B4-1 : Il fallait qu'il encadre la suite.

Ensuite, les binômes peuvent corriger une preuve en réorganisant les arguments, pour "remettre la preuve dans le bon ordre" comme dans l'exemple ci-dessous. La compréhension de l'ordre et de la bonne structure d'une preuve mathématique peut être une difficulté pour les étudiant·es en début de l'enseignement supérieur, et on retrouve en effet plusieurs erreurs de ce type dans les copies.

B5 : Genre dès le début il aurait pu commencer avec la limite, il aurait pu commencer avec le min et la borne inf, ensuite embrayer avec la croissance. Parce que ce qu'il a fait c'est pas mauvais, il a donné des arguments mais il a vendu la peau de l'ours avant de l'avoir tué en gros.

Enfin, les évaluateur·rices peuvent proposer une autre piste de résolution que celle envisagée dans la copie. Il arrive plus rarement que des étudiant·es proposent une autre méthode de résolution qu'ils jugent plus simple, après avoir validé celle présentée dans la copie. Dans l'exemple suivant, le binôme **B2** pense que le théorème invoqué dans la copie ne permet pas de résoudre l'exercice comme annoncé. Plutôt que de fournir une correction se basant sur la caractérisation séquentielle de la borne supérieure comme entamé dans la copie, les étudiant·es affirment qu'il faut utiliser le théorème de la limite monotone pour conclure.

B2 : Sup de S par caractérisation séquentielle ? Théorème des limites monotones ? Ouais c'est juste... C'est le bon résultat avec une enfin... Non non c'est juste le TLM (*Note : théorème de la limite monotone*).

Les corrections concernent donc plusieurs contenus mathématiques : définitions, calculs, preuves... Les étudiant-es peuvent compléter une preuve en rajoutant des éléments et des justifications pour la rendre valide. Notons que certains binômes cherchent à donner des arguments suffisants minimaux dans la correction, et on peut supposer que ces étudiant-es font preuve d'une compréhension avancée du raisonnement déductif et des concepts et énoncés mathématiques en jeu. D'autres, en revanche, complètent la preuve avec une condition suffisante non-nécessaire, comme un binôme qui propose de montrer que la suite est croissante pour prouver qu'elle est minorée. Enfin, les étudiant-es ont régulièrement tenté de remettre dans le bon ordre les arguments identifiés dans la copie, pour reconstruire une preuve valide.

Nous nous concentrons maintenant sur les arguments et ressources mobilisés par les étudiant-es pendant les différentes phases de l'évaluation.

Arguments internes et externes. Au cours de l'évaluation des copies, les étudiant-es font appel à certains arguments, à visée de validation de preuves ou de commentaire des copies. Rappelons qu'un argument ou une référence est dit *externe* s'il ne fait pas intervenir de savoir ou savoir-faire mathématique. Ceux qui requièrent un raisonnement mathématique sont dits *internes*.

L'utilisation de contre-exemples est présente pour tous les binômes, principalement pour invalider des arguments ou des résultats. Ces réflexions (arguments internes) peuvent amener à une compréhension plus profonde des théorèmes de leur cours. En effet, elles permettent de comprendre la raison d'être de certaines hypothèses des théorèmes que les étudiant-es utilisent dans leur pratique mathématique. Cependant, certain-es étudiant-es utilisent peu de contre-exemples pour invalider des preuves, au profit d'arguments et références externes.

Du côté des arguments externes, tous les binômes font des références à des expériences personnelles. La plupart du temps, elles visent à commenter la preuve qu'ils évaluent, et à la comparer avec leur propre résolution de l'exercice. Par exemple, les étudiant-es du binôme **B5** discutent de leur résolution de l'exercice quand il-elles essaient de comprendre la preuve engagée dans la copie. Il-elles disent alors :

B5 : Par contre là il a fait mieux que moi. Moi j'avais pas mis le "et", j'avais mélangé ensemble et... du coup en lisant sa copie j'ai appris de mon erreur aussi.

On voit déjà ici que l'évaluation de certaines copies, dont la stratégie de résolution est proche de celle engagée par les évaluateur-ices, peut leur permettre de prendre du recul sur leur propre pratique. Notons que les étudiant-es du premier binôme font de nombreuses références de ce type pour valider des preuves, sans plus d'argumentation. Ces étudiant-es n'utilisent que peu d'arguments internes.

Il nous a semblé que lors de la validation, l'invalidation ou la correction d'une preuve, les étudiant-es utilisent des arguments externes lorsqu'il-elles ne sont pas à l'aise avec les concepts en jeu, et ont des difficultés à mobiliser des arguments internes. Au niveau des mathématiques du supérieur, il paraît souhaitable que les étudiant-es apprennent à mobiliser des arguments internes pour évaluer les preuves, arguments qui peuvent aussi servir lors de la construction de nouvelles preuves.

Ressources utilisées. On observe dans ce recueil de données que les étudiant-es font aussi appel à des ressources extérieures lors de l'évaluation de copies de pairs. L'utilisation de ces ressources leur permet de valider, invalider, ou commenter les copies qu'il-elles évaluent. On trouve naturellement quelques références aux cours auxquels les étudiant-es ont assisté sur une notion, mais aussi aux supports de correction qui ont été fournis durant l'expérimentation.

On note d'abord que les étudiant-es font peu de références explicite à leur cours, et lorsqu'il-elles le font, les propriétés mentionnées sont parfois incorrectes. Par exemple, lors de la correction de l'exercice 2, le binôme **B1** affirme que le produit cartésien se développe sur l'union (ce qui est à la

fois problématique car c'est une propriété fautive, et parce que cette propriété est en fait l'objet même de la question). Seul·e un·e étudiant·e, dans le binôme **B2**, semble faire référence de manière précise à son cours. Les références au cours sont utilisées pour valider des preuves, mais aussi pour vérifier des définitions énoncées dans les copies. La situation semble donc différente de celle de recherche de problèmes, dans laquelle on s'attend à ce que le cours soit un support à la réflexion. Notons cependant que les étudiant·es ont déjà cherché l'exercice, et qu'il·elles ont pu à cette occasion se remémorer le cours sur le sujet. Nous avons aussi remarqué que la vérification de la présence des hypothèses d'un théorème n'est pas systématique chez les étudiant·es qui évaluent, ou n'est pas toujours aboutie. On peut supposer que cela est dû à une connaissance approximative des énoncés du cours, ou à un rapport aux théorèmes où l'importance des hypothèses est négligée. Cela peut amener les étudiant·es à valider des preuves qui sont incorrectes. Le cours ne semble donc pas être un support utilisé de manière explicite et précise.

Ensuite, quelques références à la feuille des éléments de correction sont faites. Les étudiant·es l'utilisent principalement pour valider une démarche, si celle-ci apparaît dans la feuille. Pour certain·es étudiant·es, cela est utile dans la phase de compréhension de la preuve qu'il·elles évaluent : ils s'en servent pour identifier précisément les méthodes engagées dans la copie.

Les étudiant·es mobilisent donc peu les ressources extérieures à leur disposition dans l'évaluation de leur copie.

Nous avons détaillé dans cette section trois principaux types d'activités observés dans le processus d'évaluation par les pairs : compréhension d'une preuve, validation et invalidation, et correction. Nous avons aussi remarqué que les étudiant·es observé·es portent une attention forte à la démarche mise en œuvre dans la preuve, et valorisent une bonne idée, même mal exploitée.

Dans la section suivante, nous complétons notre analyse à partir des évaluations réalisées individuellement en devoir maison, à la lumière de l'étude du travail des binômes observés.

4.C. Quelques résultats sur les évaluations réalisées à la maison

Comme nous l'avons vu dans la partie précédente, la méthodologie d'entretiens d'explicitation nous a permis d'identifier certaines activités d'étudiant·es lors d'une évaluation entre pairs. Cependant, nous ne savons pas à quel point ces activités apparaissent chez les étudiant·es ayant évalué seul·es. Nous cherchons donc ici des traces de ces activités dans les évaluations en devoir maison.

Il est plus difficile, seulement avec les copies évaluées, d'observer une invalidation : les étudiant·es barrent rarement les écrits de leur camarade, et n'indiquent pas toujours clairement quand une preuve leur paraît invalide. Pour les validations, on observe des symboles de coches (✓), et parfois le·a correcteur·rice commente la preuve validée. Ces données donnent peu de traces de l'activité de validation ou d'invalidation des étudiant·es, notamment les arguments et ressources utilisées.

Nous avons par contre accès aux corrections proposées par l'évaluateur·rice. On retrouve des motifs identifiés dans la partie précédente : correction de calculs ou de définitions (figure 6), complétion d'une preuve (figure 7), proposition d'une autre stratégie (figure 8), identification d'une justification manquante (figure 9).

Comme nous l'avons rapidement noté dans la section précédentes, les étudiant·es prêtent attention à la démarche entamée dans la copie qu'ils évaluent. Plusieurs commentaires dans les copies évaluées à la maison vont aussi dans ce sens (figure 10). Nous avons également remarqué de nombreux commentaires sur la propreté des copies, ainsi que des conseils pour rendre la rédaction plus fluide.

Notons enfin que certaines annotations des copies tendent à montrer que les étudiant·es ont utilisé des références externes, les éléments de correction à disposition, notamment dans le second exercice, moins maîtrisé que l'exercice 1.

FIGURE 6 – Exemple de correction d'un calcul

FIGURE 7 – Exemple de complétion d'une preuve

FIGURE 8 – Exemple de proposition d'une stratégie alternative

FIGURE 9 – Exemple d'identification de justification manquante

FIGURE 10 – Exemple d'un-e correcteur-riche qui souligne la pertinence des démarches

4.D. Synthèse des analyses de l'activité des étudiant·es

Les analyses des entretiens et des copies nous ont permis d'identifier les activités des étudiant·es en situation d'évaluer des pairs. Les trois axes principaux sont la compréhension des preuves, la validation ou l'invalidation de ces preuves, et leur correction. En plus d'évaluer le fond du contenu mathématique, les étudiant·es sont attentif·ves aux méthodes de résolution employées et au résultat final obtenu. Chez certain·es, ces observations sur la démarche peuvent servir à l'invalidation de la preuve évaluée, mais chez d'autres la démarche est seulement commentée.

Nous avons également observé que les étudiant·es relèvent les imprécisions de syntaxe et de rédaction, cependant il est difficile de comprendre quel est l'impact sur la validation ou l'invalidation des preuves.

Notons que la lecture, la compréhension et la validation des preuves s'effectuent en général en parallèle. Pour ce faire, les étudiant·es décomposent souvent la preuve en ses grands axes et éléments importants, pour essayer d'en comprendre la structure générale. Pendant les moments de validation, les étudiant·es s'interrogent régulièrement sur les justifications nécessaires à la validité d'une preuve ou d'un calcul. Ces réflexions pourront être utiles à une compréhension plus en profondeur de l'activité de preuve et des concepts en jeu dans les exercices.

On remarque que la vigilance des étudiant·es aux erreurs dans les preuves varie. Il est arrivé plusieurs fois que les étudiant·es, pensant reconnaître une méthode classique ou un théorème du cours, ne fassent pas attention à la présence de toutes les hypothèses. Ainsi, certaines copies qui utilisent le théorème de la limite monotone sans vérifier la majoration de la suite en jeu ne sont pas invalidées par les pairs. On trouve aussi des copies utilisant le théorème de la limite monotone sans le citer explicitement mais en vérifiant toutes les hypothèses, et qui sont invalidées. On constate ici un manque de recul de certain·es étudiant·es sur la validation de l'utilisation d'un théorème. Les étudiant·es semblent aussi plus critiques envers des méthodes qu'il·elles ne connaissent pas, ou n'ont pas eux-mêmes ou elles-mêmes mises en place.

Cette catégorisation de l'activité des étudiant·es nous permet donc de préciser les éléments qu'il·elles prennent en compte lors de l'évaluation de pairs. Ils peuvent concerner les preuves en elles-mêmes, les résultats présentés, la démarche entreprise ou la rédaction et la syntaxe. Les étudiant·es sont aussi attentif·ves aux éléments non mathématiques, comme la propreté ou la clarté de la copie qu'il·elles évaluent. Certain·es s'intéressent également à la qualité et l'efficacité des méthodes empruntées. Tout ceci est sans doute lié au contrat didactique de la classe vis-à-vis de l'évaluation, très spécifique en CPGE.

Pour compléter nos analyses de l'activité des étudiant·es, nous avons mis en place un questionnaire qui a permis de recueillir la perception qu'ont les étudiant·es de leur activité. Nous synthétisons les résultats dans la section suivante.

5. Perception des étudiant·es sur leurs activités : retours du questionnaire

Comment les étudiant·es ont-il·elles perçu l'activité d'évaluation entre pairs à laquelle il·elles ont participé? Ont-il·elles pris conscience de certaines choses au cours de ce processus (relativement à la preuve, à l'évaluation, à la rédaction, au contrat didactique...)? Pour étudier cette question, nous avons fait passé un court questionnaire aux étudiant·es.

Il a été transmis à la fin de l'expérimentation, et a reçu 38 réponses. Les premières questions se concentrent sur le ressenti des étudiant·es quant aux apports de cette expérimentation, de manière générale et sur la compréhension des exercices. Nous questionnons ensuite les critères d'évaluation qui ont semblé importants aux étudiant·es. Les dernières questions interrogent les étudiant·es sur leurs impressions sur la posture d'évaluateur·rice.

Le détail des questions et un tableau résumant les statistiques des réponses aux questions fermées sont disponibles en annexe B. Nous résumons ci-dessous les observations qui ressortent des réponses au questionnaire, relativement à nos questions de recherche. Il faut noter que ces résultats émanent de déclarations des étudiant·es, et attestent donc de la perception qu'il·elles ont de l'activité proposée, qui peut différer de leur activité réelle. Il est néanmoins intéressant de savoir comment il·elles ont perçu leur activité et, sur le plan "méta", ce dont il·elles ont pu prendre conscience lors de l'évaluation entre pairs.

Critères d'évaluation jugés importants. Nous avons demandé aux étudiant·es les critères d'évaluations qu'il·elles estiment être les plus importants, qui sont synthétisés ci-dessous :

- la clarté de la rédaction (12 étudiant·es) ;
- la rigueur dans les justifications et l'utilisation d'un théorème (9 étudiant·es) ;
- la propreté de la copie (7 étudiant·es) ;
- la démarche de résolution (6 étudiant·es) ;
- le résultat final (5 étudiant·es) ;
- la connaissance précise du cours (3 étudiant·es) ;
- le fait de répondre aux questions posées (2 étudiant·es) ;
- le fait de ne pas bluffer (1 étudiant·e).

Notons que les aspects de communication du raisonnement sont les plus représentés dans les réponses (clarté de la rédaction et propreté de la copie), ce qui semble être une prise de conscience pour une partie des étudiant·es. Des critères de validité mathématique sont aussi évoqués par les étudiant·es, qui insistent sur l'importance de la rigueur dans les justifications et la démarche de résolution. Nous avons également relevé que 9 étudiant·es identifient le critère de "l'importance du raisonnement" comme essentiel, sans que l'on puisse tirer plus de détails quant aux caractéristiques du raisonnement auxquelles il·elles se réfèrent.

Perception sur les apprentissages mathématiques. Les réponses des étudiant·es semblent indiquer qu'évaluer des pairs est perçu comme contribuant effectivement à leur apprentissage. Cependant, dans leur perception, être soi-même évalué par des pairs s'avère moins utile à la compréhension des exercices, et est très dépendant de la personne qui évalue.

Des étudiant·es notent qu'évaluer des pairs nécessite une maîtrise avancée du cours. De plus, il·elles trouvent utile d'être confronté·es à plusieurs résolutions d'un même exercice, ce qui permet « d'échanger des savoirs et des savoir faire ». De plus, certain·es étudiant·es notent que l'évaluation d'autres copies les a amené·es à confronter leurs propres réponses à celles qu'il·elles ont évaluées, ce qui leur permet de « prendre du recul sur [leur] propre travail ». Nous remarquons ici que ces réponses font écho aux références aux expériences personnelles (utilisées comme références externes) mobilisées par les étudiant·es lors des entretiens.

Les étudiant·es ont également relevé l'intérêt de rencontrer plusieurs erreurs de résolution, en particulier des erreurs qu'il·elles auraient pu faire aussi.

Plus largement, certain·es affirment que l'évaluation de copies de pairs leur a permis de mieux saisir les "étapes clés" qui constituent un raisonnement. Cela leur permet donc de comprendre la structure générale d'une preuve qu'ils rencontrent. On peut supposer que les étudiant·es perçoivent un intérêt qui va au-delà de l'exercice spécifique traité, sur la preuve par exemple.

Perception des aspects méta-cognitifs. Comme nous l'avons cité ci-dessus, une partie des étudiant·es disent se rendre mieux compte de l'importance d'une rédaction claire et soignée, donc de l'importance de la communication de la preuve en plus de la preuve en elle-même. Par ailleurs dans le questionnaire, tout comme dans les entretiens, plusieurs étudiant·es ont remarqué que les "arnaques" peuvent être facilement identifiées par le correcteur·rice.

On peut interpréter tout cela comme relevant d'une prise de conscience de la dimension humaine de l'activité de preuve, et de la nécessité de tenir compte du point de vue de lecteur·rice lors de la production d'une preuve.

Les étudiant·es relèvent donc plusieurs raisons pour lesquels le processus d'évaluation par les pairs leur a semblé utile. Cela leur permet d'examiner d'autres rédactions pour leur propre stratégie de résolution, et de découvrir d'autres solutions à un même exercice. De plus, cela leur donne l'occasion de réfléchir à l'importance de la structuration d'une preuve pour sa validité. Enfin, il·elles développent un autre regard sur l'activité d'évaluation et sur la communication d'un contenu mathématique.

6. Conclusions

6.A. Résultats et perspectives

Nous avons expérimenté un processus d'évaluation entre pairs avec des étudiant·es du début de l'enseignement supérieur, qui s'est avéré fonctionnel (les étudiant·es sont bien entré·es dans la tâche proposée) et qui a engendré une activité riche chez les étudiant·es.

Nos questions de recherches portaient sur l'activité des étudiant·es évaluateur·rices pendant une évaluation de pairs (sur les plans mathématique et cognitif, mais aussi méta-mathématique et méta-cognitif), et plus particulièrement les éléments qu'il·elles prennent en compte lors des évaluations, les corrections qu'il·elles apportent, ainsi que les ressources auxquelles il·elles font appel.

Pour répondre à ces questions, après avoir sélectionné deux exercices portant sur les bornes supérieure et inférieure et sur les manipulations d'ensembles, nous avons réalisé une analyse *a priori* des différentes stratégies de résolution des exercices, puis une analyse *a priori* des évaluations possibles de ces exercices. Nous avons ensuite observé des étudiant·es évaluer des copies, afin d'identifier leur activité. Enfin, sur la base de ces analyses, nous avons examiné les traces écrites des évaluations réalisées en devoir maison.

Activités des étudiant·es lors de l'évaluation. Lors de l'évaluation d'une copie, les étudiant·es se trouvent dans des phases de lecture et de compréhension de preuves, des phases de validation et d'invalidation et des phases de correction.

Lors de la lecture et la compréhension de la copie, les évaluateur·rices cherchent à expliquer et expliciter le raisonnement entrepris dans la copie, notamment en identifiant les différentes étapes qui constituent la preuve. Pour cela, il·elles peuvent faire référence aux contenus de leur cours afin de préciser les concepts et énoncés mathématiques en jeu. Parfois, il·elles sont amené·es à interpréter ce que la personne évalué·e "a voulu faire" dans la copie.

Lors de la phase de validation et d'invalidation, les étudiant·es peuvent s'intéresser à la preuve dans son ensemble, à une étape du raisonnement, aux propriétés des objets manipulés, au résultat ou encore à la démarche de la preuve. Pour justifier une invalidation, il·elles peuvent être amené·es à trouver des contre-exemples à des énoncés mathématiques utilisés, implicitement ou explicitement, dans la copie. Nous avons observé ceci particulièrement lorsque l'étudiant·e évalué·e semble utiliser un théorème présent dans le cours (comme le théorème de la limite monotone), sans en vérifier toutes les hypothèses. Enfin, les évaluateur·rices discutent de la précision nécessaire dans les justifications permettant de valider une preuve, et identifient les résultats "pas assez justifiés".

Nous notons de plus qu'au cours de l'évaluation, une attention particulière est portée par les étudiant·es à la démarche de résolution de l'exercice empruntée dans la copie. Si ce n'est pas forcément un critère de validation de la preuve, de nombreux étudiant·es commentent lorsqu'il·elles estiment que la personne corrigée a eu "de bonnes idées" dans la preuve. Les étudiant·es prennent également en compte le résultat de la preuve, dont la vérité peut influencer leur opinion de correcteur·rices sur la validité de la preuve.

Enfin, les étudiant·es proposent des corrections des éléments qu'il·elles ont invalidés. Il peut s'agir d'erreurs de définition, qui sont systématiquement corrigées, ou d'erreurs de calcul, pour lesquelles les

étudiant-es se contentent parfois de noter que le résultat est erroné. Pour la correction des éléments relatifs à la preuve, les évaluateur-rices peuvent compléter le raisonnement entamé dans la copie ; nous avons alors observé des discussions portant sur le caractère nécessaire de certains arguments. Certain-es étudiant-es peuvent aussi changer la structure d'une preuve en modifiant l'ordre des différentes étapes de raisonnement, pour la rendre valide. Enfin, il arrive que des étudiant-es proposent une stratégie de résolution différente de celle entamée dans la copie.

Ainsi, on voit que le processus d'évaluation entre pairs peut amener les étudiant-es à une activité très riche autour de l'analyse de solutions et de preuves, impliquant certaines réflexions que l'on peut considérer d'ordre méta-mathématique.

Utilisation des arguments et références internes et externes. Les arguments et références mobilisées par les étudiant-es lors de l'évaluation n'ont pas seulement une visée de validation : ils peuvent aussi servir à éclairer des preuves qui n'ont pas été initialement comprises par les évaluateur-rices, ou sont utilisés pour commenter les raisonnements proposés dans la copie. Le principal argument interne que nous avons relevé est la production de contre-exemple pour invalider un énoncé affirmé dans la copie. D'autres arguments internes, références au cours et vérification d'éléments techniques (formules, calculs...) ont également été observés. Nous avons noté peu de références explicites au cours ou à des sources expertes lors des entretiens. De nombreux étudiant-es ont fait référence à leur propre résolution des exercices, soit pour commenter les preuves évaluées, soit pour valider ou invalider un calcul ou le résultat d'une étape du raisonnement. Enfin, la feuille de correction que nous avons fournie a été exploitée à plusieurs reprises lors des entretiens, et semble avoir été également utilisée pour la correction de copies en devoir maison.

En conclusion, il semble que les étudiant-es mobilisent à la fois des arguments internes et externes, avec une variabilité selon les individus. La dialectique entre ces deux dimensions dans les processus d'évaluation de pairs, et le développement des arguments internes chez les étudiant-es seraient certainement intéressants à étudier sur un temps plus long.

Apports de l'évaluation entre pairs à l'apprentissage. Les analyses des entretiens, des copies et du questionnaire réalisées nous ont permis de dégager des aspects de l'évaluation entre pairs pouvant participer à l'apprentissage des étudiant-es. Ceci va dans le sens des résultats des recherches existantes mentionnées en section 2.B..

Notre étude montre que ce processus peut permettre de développer un travail sur les contenus mathématiques en jeu. En effet, l'évaluation de copies de pairs met les étudiant-es face à d'autres stratégies de résolution que celles qu'il-elles auraient entrepris. Cela peut permettre de se rendre compte de la diversité de moyens possibles pour résoudre un problème mathématique. À l'inverse, les étudiant-es peuvent évaluer des preuves similaires à celles qu'il-elles ont mises en place. Il-elles ont alors l'occasion de considérer à nouveau à la validité de leur méthode, et d'observer d'autres erreurs qui peuvent survenir.

De plus, l'évaluation et la correction d'éléments techniques peut permettre aux étudiant-es de travailler le cours. Par exemple, lorsqu'il-elles cherchent des contre-exemples à certaines étapes de déduction, il-elles sont amené-es à réfléchir à l'importance des différentes hypothèses des théorèmes qui sont utilisés. On peut s'attendre à ce que cela permette une compréhension plus profonde du cours et des théorèmes mis en jeu, qui pourrait mener, plus généralement, à saisir l'importance des hypothèses dans les théorèmes.

L'évaluation par les pairs est également utile pour prendre du recul sur l'activité de preuve. Lors des différentes phases de l'évaluation décrites ci-dessus, les étudiant-es peuvent s'interroger sur la structure des preuves qu'il-elles évaluent, mettant en lumière l'importance des étapes d'une preuve et de leur organisation.

Enfin, l'évaluation entre pairs permet d'explicitier certains enjeux de l'évaluation, en rendant explicite le contrat didactique en jeu. Nous avons noté que les étudiant-es semblent mieux appréhender les évaluations d'un-e correcteur-riche. En particulier, ce processus semble mettre au jour des enjeux

de communication des contenus mathématiques, autant dans la présentation d'un écrit que dans la rédaction des raisonnements.

En conclusion, nous observons que les étudiant-es dans un processus d'évaluation entre pairs développent une activité mathématique qui diffère de celle que l'on peut imaginer ordinairement dans la résolution des mêmes exercices. Nous observons une activité d'analyse de preuve, d'interprétation des raisonnements, d'évaluation de la validité d'étapes ou de l'entièreté d'une preuve, ainsi que le développement des réflexions sur les plans méta-mathématiques et méta-cognitifs. Ceci nous semble porteur d'un potentiel pour l'apprentissage mathématiques, en particulier au début de l'enseignement supérieur.

Nous identifions certaines limites à ce travail : sur le plan de la recherche, notre travail exploratoire montre un potentiel pour l'apprentissage des mathématiques, mais il reste à comprendre comment peut se réaliser ce potentiel et le confirmer dans d'autres contextes (scolaires et mathématiques), et à étudier sur le long terme les effets d'un tel dispositif sur les apprentissages. Sur le plan du dispositif lui-même, l'expérience peu fructueuse du second exercice montre que les conditions de fonctionnement d'une évaluation entre pairs sont importantes et à bien contrôler. Cet exercice nous donne des indices quant aux critères de choix des exercices, nous y revenons dans la section suivante. Enfin, le contexte spécifique de la CPGE est aussi à prendre en considération, comme nous l'avons mentionné dans la présentation du dispositif expérimental : un contrat didactique particulier est en jeu, notamment en ce qui concerne l'évaluation, et ce contrat serait peut-être différent dans d'autres contextes du supérieur, à l'université par exemple, ou bien dans d'autres pays. De manière générale, le contrat didactique dans le cadre de l'évaluation par les pairs mériterait d'être étudié pour comprendre les ruptures que peut introduire une évaluation de preuve entre pairs mais aussi comprendre la nature du nouveau contrat qui se met en place dans ces évaluations.

Ce travail ouvre des perspectives de recherches didactiques sur les mathématiques dans le supérieur. Une étude plus approfondie serait à développer sur l'activité des étudiant-es en situation d'évaluation de pairs, avec un corpus contenant d'autres tâches et dans des contextes divers, ainsi que sur des mises en œuvres plus longues et répétées. Un approfondissement des activités possibles des étudiant-es en situation d'évaluation de pairs, en particulier en croisant avec les recherches sur l'apprentissage de la preuve serait certainement fructueux. Des recherches sur la viabilité en classes de mathématiques de dispositifs d'évaluation entre pairs seraient aussi intéressantes à déployer, afin de comprendre les conditions et contraintes de fonctionnement de ces dispositifs.

6.B. Quelques conseils et pistes de mise en place

Sur la base de nos lectures de recherche, de conseils donnés dans la littérature (Falchikov & Goldfinch, 2000; Topping, 2009) et de notre expérience pour cet article, nous formulons quelques idées, pistes, conseils pour les enseignant-es qui souhaiteraient mettre en place une évaluation entre pairs en mathématiques, en particulier au début de l'enseignement supérieur. Ces recommandations et points de vigilance ne sont pas tous à prendre comme des résultats étayés, mais il nous a semblé intéressant, dans cette section distincte des résultats de notre étude, de proposer quelques points qui peuvent aider à développer l'évaluation entre pairs en mathématiques.

Au préalable de l'expérimentation, lors du choix du ou des exercices à faire évaluer, il nous semble important :

- que les étudiant-es soient à l'aise avec le type d'exercice, voire que celui-ci porte sur des connaissances suffisamment anciennes et stabilisées ;
- qu'une partie de l'exercice porte sur la production d'une preuve (pas seulement d'un calcul) ;
- que l'exercice puisse être résolu avec plusieurs stratégies, et que plusieurs points de vue soient possibles sur les objets en jeu (en ce sens, un énoncé "ouvert" qui ne donne pas la réponse, c'est-à-dire qui ne soit pas de la forme "montrer que", nous semble assez adapté).

Pour la mise en place du processus d'évaluation entre pairs lui-même, nous recommandons de :

- clarifier les attentes et les objectifs aux étudiant-es (par exemple faire un lien avec l'activité d'auto-évaluation), et préciser si ce processus permettra de donner une note en tant qu'évaluateur·rice et/ou en tant que qu'évalué·e ;
- développer et clarifier les critères d'évaluation en impliquant les étudiant-es, et en particulier discuter des critères de validation d'une preuve possibles (références internes et externes) – On peut à cette occasion donner des exemples d'évaluation ;
- au cours des premières sessions, guider, encadrer et modérer les corrections et feedback donnés par les évaluateur·rices.

Ces trois points relèvent de l'explicitation et de l'accompagnement d'un changement de contrat didactique relativement à l'évaluation : quel est son statut ? À qui en incombe la responsabilité ? Cela nécessite pour l'enseignant·d'avoir bien identifié les spécificités de ce nouveau contrat, et notamment les responsabilités des étudiant-es et de l'enseignant-e, à l'image de ce que proposent Bloch & Gibel (2022) dans la mise en place d'un dispositif innovant en mathématiques en licence 1.

Enfin, Nous suggérons également que tous·tes les étudiant-es cherchent au préalable l'exercice qu'il·elles vont évaluer, et soient tous·tes évalué·es par leurs pairs sur cet exercice. Il peut être intéressant de mettre en place un "cycle d'évaluation" (assessment cycle) tel que le propose Reinholz (2015) (figure 11). Il a été créé pour décrire comment l'évaluation entre pairs peut être un appui à l'auto-évaluation.

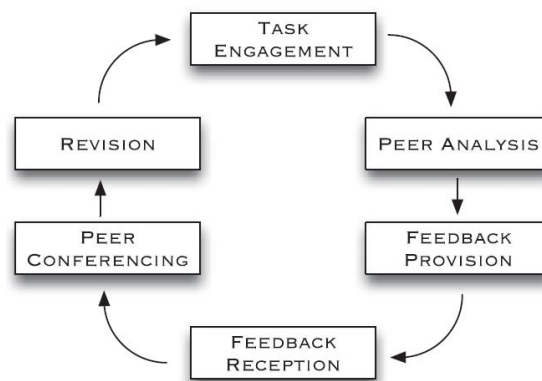


FIGURE 11 – Le cycle d'évaluation de Reinholz (2015)

Les activités suivantes sont présentes dans le cycle :

- la réalisation de la tâche (*task engagement*) ;
- l'évaluation par un ou des pairs (*peer analysis*) ;
- le retour des évaluateur·rices sur le travail (*feedback provision* et *feedback reception*), qui peut être écrit ou oral ;
- la discussion entre évaluateur·rices et évalué·e (*peer conferencing*) ;
- la révision du travail par l'évalué·e.

L'auteur précise qu'une évaluation entre pairs peut inclure toutes les phases du cycle ou seulement certaines d'entre elles, et leur ordre pourrait être modifié.

Nous pointons enfin le fait que ce processus d'évaluation nécessite que les étudiant-es aient déjà une idée du contrat didactique concernant l'évaluation dans la formation concernée. Ainsi, il faut certainement un temps d'appropriation d'un nouveau contrat didactique d'évaluation avant d'initier des évaluations entre pairs. Les classes préparatoires aux grandes écoles nous ont semblé être appropriées pour l'expérimentation, et nous nous questionnons sur les conditions de mise en place de ce processus à l'université, certainement plus complexes.

Il est souvent recommandé de développer l'évaluation entre pairs sur plusieurs sessions au fil d'un semestre ou d'une année. Cependant, selon notre expérience, même une seule session d'évaluation entre pairs semble initier une prise de conscience de certains enjeux aux étudiant·es (enjeux de preuve, enjeux d'évaluation...) et leur faire se poser des questions formatrices. Nous encourageons donc nos collègues à expérimenter, même ponctuellement, un dispositif d'évaluation entre pairs.

A Grille d'analyse

<i>Évaluations portant sur le fond du raisonnement mathématique</i>		
1	Activités portant sur l'évaluation d'une preuve proposée dans la copie	
1Z	Lire la preuve	
1A	Essayer de comprendre la preuve	Identifier des étapes du raisonnement mathématique Reconnaître les concepts mathématiques utilisés Expliquer une étape du raisonnement Anticiper un raisonnement
1B	Évaluer la validité de la preuve	Invalidier / Valider la démarche Invalidier / Valider une étape du raisonnement / d'un calcul Invalidier / Valider la preuve
		Noter l'absence de la réponse à une question Pointer un manque de justification ou une bonne justification Pointer une information non nécessaire au raisonnement
1C	Évaluer la valeur épistémique de la preuve	Évaluer la vraisemblance de la démarche Évaluer la valeur épistémique d'une étape / d'un calcul Évaluer la valeur épistémique de la preuve
1D	Corriger la preuve	Corriger une étape du raisonnement Corriger la structure d'une preuve Compléter un raisonnement Proposer une autre méthode de résolution Corriger un détail technique
		Apporter des précisions sur une étape du raisonnement Expliquer les raisons d'une erreur
2	Activités portant sur l'évaluation d'une preuve envisagée par l'élève corrigé	
2A	Essayer de comprendre la preuve	
2B	Évaluer la validité de la preuve	
2C	Évaluer la valeur épistémique de la preuve	
2D	Corriger la preuve	
3	Activités portant sur l'évaluation du résultat d'une preuve	
3B	Évaluer la validité du résultat	
3D	Corriger le résultat	

<i>Évaluations portant sur la forme du raisonnement mathématique</i>		
4	Activités pourtant sur l'évaluation de la rédaction et de la syntaxe d'une preuve	
4B ₁	Évaluer la validité de la rédaction et de la syntaxe	Valider / Invalider une définition Valider / Invalider la syntaxe d'un énoncé
4B ₂	Évaluer la clarté de la rédaction	
4D	Corriger la rédaction ou la syntaxe	
<i>Évaluations ne portant pas sur le raisonnement mathématique</i>		
5	Activités ne portant pas sur l'évaluation du raisonnement mathématique	
5A	Juger de la propreté de la copie	
5B	Essayer de comprendre le comportement de l'élève	
5C	Juger de la qualité ou de la légitimité d'une méthode	
5D	Juger de la compréhension d'un élève	

B Questionnaire

Le questionnaire a été partagé aux étudiant-es via la plateforme Framiforms, avec les questions :

1. Nom
2. Prénom
3. Étiez-vous volontaire pour faire les corrections sous ma supervision ?
4. Le fait de corriger des copies d'autres élèves vous a-t-il semblé utile ? Pourquoi ?
5. Le fait d'être corrigé par d'autres élèves vous a-t-il été utile ? Pourquoi ?
6. Est ce que le fait de corriger des copies d'autres élèves vous a permis de mieux comprendre les exercices ? Pourquoi ?
7. Est ce que le fait d'être corrigé par d'autres élèves vous a permis de mieux comprendre les exercices ? Pourquoi ?
8. Quels sont les critères d'évaluation les plus importants que vous retenez ?
9. Qu'est ce qui a été le plus facile pour vous dans le fait de corriger des copies ?
10. Qu'est ce qui a été le plus difficile pour vous dans le fait de corriger des copies ?
11. Vous êtes-vous senti à l'aise en corrigeant les copies ? Pourquoi ?
12. Est-ce une expérience que vous aimeriez refaire ? Pourquoi ?

Le tableau ci-dessous synthétise les réponses aux questions fermées. 38 étudiant-es ont répondu à ce questionnaire.

Question	Oui et plutôt oui Utile et plutôt utile	Non et plutôt non Inutile et plutôt inutile
Le fait de corriger des copies d'autres élèves vous a-t-il semblé utile ?	36	2
Le fait d'être corrigé par d'autres élèves vous a-t-il été utile ?	27	11
Est ce que le fait de corriger des copies d'autres élèves vous a permis de mieux comprendre les exercices ?	27	11
Est ce que le fait d'être corrigé par d'autres élèves vous a permis de mieux comprendre les exercices ?	17	21
Vous êtes-vous senti à l'aise en corrigeant les copies ?	26	12
Est-ce une expérience que vous aimeriez refaire ?	34	4

C Éléments de correction

Éléments de correction, projet d'évaluation par les pairs

Lycée

Consignes :

- Écrivez "Correcteur / Correctrice :" suivi de votre nom en haut des copies, et corrigez dans une couleur bien visible.
- Évaluez les copies, en commentant et faisant des annotations.
- Portez attention à la validité des réponses, ainsi que la rédaction et l'argumentation des copies.

Exercice 1 On considère l'ensemble $S = \left\{ \frac{n-\frac{1}{n}}{n+\frac{1}{n}} , n \in \mathbb{N}^* \right\}$.

S admet-il une borne supérieure ? une borne inférieure ? un minimum ? un maximum ? Si oui, les déterminer.

Solution. Pour faciliter la rédaction, on note $s_n = \frac{n-\frac{1}{n}}{n+\frac{1}{n}}$ si $n \in \mathbb{N}^*$.

Point 1 : Existence des bornes inf et sup

- *Idée 1* : Montrer que $\forall n \in \mathbb{N}^*, 0 \leq \frac{n-\frac{1}{n}}{n+\frac{1}{n}} \leq 1$.
- *Idée 2* : La suite (s_n) converge, elle est donc bornée.
- *Idée 3 (Pour la minoration uniquement)* : La suite (s_n) est croissante, donc $\forall n \in \mathbb{N}^*, s_n \geq s_1$.

Point 2 : Calcul de la borne inf et du minimum

- *Idée 1* : $s_1 = 0$ est un minorant de S et $s_1 \in S$, c'est donc la borne inf et le minimum de S .
- *Idée 2* : La suite (s_n) est croissante, donc le premier terme de cette suite est le minimum de l'ensemble S .
- *Idée 3* : Utiliser la caractérisation séquentielle de la borne inf et le fait que $s_1 = 0 \in S$.

Point 3 : Calcul de la borne sup

- *Idée 1* : Utiliser la caractérisation séquentielle de la borne sup.
- *Idée 2* : Utiliser la théorème de la limite monotone.

Point 4 : Existence du maximum :

- *Idée 1* : Montrer $1 \notin S$ par l'absurde.
- *Idée 2* : Montrer que $\forall n \in \mathbb{N}^*, \frac{n-\frac{1}{n}}{n+\frac{1}{n}} < 1$.

Références

- Allal, L., & Laveault, D. (2009). Assessment for Learning : Évaluation-soutien d'Apprentissage. *Mesure et évaluation en éducation*, 32(2), 99–106. doi: 10.7202/1024956ar
- Allal, L., & Mottier Lopez, L. (2005). Formative Assessment of Learning : A Review of Publications in French. In *Formative Assessment - Improving Learning in Secondary Classrooms* (OECD éd., pp. 241–264). Paris.
- Artigue, M. (1988). Ingénierie didactique. *Recherches en didactique des mathématiques*, 9(3), 281–308. Consulté sur <https://revue-rdm.com/1988/ingenierie-didactique-2/>
- Baiduri. (2022). Effect of self and peer assessments on mathematics learning achievement. *Al-Jabar : Jurnal Pendidikan Matematika*, 13(1), 13–21. doi: 10.24042/ajpm.v13i1.10731
- Biehler, R., Durand-Guerrier, V., & Trigueros, M. (2024). New trends in didactic research in university mathematics education. *ZDM – Mathematics Education*, 56(7), 1345–1360. doi: 10.1007/s11858-024-01643-2
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education : Principles, Policy & Practice*, 5(1), 7–74. doi: 10.1080/0969595980050102
- Bloch, I., & Gibel, P. (2022). Situations de recherche pour l'accès aux concepts mathématiques à l'entrée à l'université. *epiDEMES*, 1 / 2022. (Publisher : Episciences.org) doi: 10.46298/epidemes-9886
- Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education : learning for the longer term*. London ; New York : Routledge.
- Cnesco. (2022). *Conférence de consensus - L'évaluation en classe, au service de l'apprentissage des élèves*. Consulté le 2025-01-08, sur <https://www.cnesco.fr/fr/evaluation-en-classe/>
- Coppé, S. (1993). *Processus de vérification en mathématiques chez les élèves de première scientifique en situation de devoir surveillé* (Thèse de doctorat non publiée). Université Claude Bernard - Lyon I.
- Coppé, S. (2018). Évaluation et didactique des mathématiques : vers de nouvelles questions, de nouveaux travaux. *Mesure et évaluation en éducation*, 41(1), 7–39. (Publisher : ADMEE-Canada - Université Laval) doi: 10.7202/1055895ar
- Coppé, S., & Moulin, M. (2017). Évaluation entre pairs et débat argumenté dans le cadre d'un problème complexe en mathématiques. *Canadian Journal of Science, Mathematics and Technology Education*, 17(4), 308–327. doi: 10.1080/14926156.2017.1378832
- De Ketele, J.-M. (1986). En guise de conclusion. In *L'évaluation : approche descriptive ou prescriptive ?* (pp. 248–271). Bruxelles : De Boeck.
- Detroz, P., Crahay, M., & Fagnant, A. (2017). *L'évaluation à la lumière des contextes et des disciplines*. Louvain-la-Neuve : De Boeck Supérieur. doi: 10.3917/dbu.detro.2017.01
- Dreyfus, T., Nardi, E., & Leikin, R. (2012). Forms of Proof and Proving in the Classroom. In G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education : The 19th ICMI Study* (pp. 191–213). Dordrecht : Springer Netherlands. doi: 10.1007/978-94-007-2129-6_8
- Durand-Guerrier, V. (2008). Truth versus validity in mathematical proof. *ZDM*, 40(3), 373–384. doi: 10.1007/s11858-008-0098-8
- Durand-Guerrier, V., Boero, P., Douek, N., Epp, S. S., & Tanguay, D. (2012). Examining the Role of Logic in Teaching Proof. In G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education : The 19th ICMI Study* (pp. 369–389). Dordrecht : Springer Netherlands. doi: 10.1007/978-94-007-2129-6_16
- Durand-Guerrier, V., Meyer, A., & Modeste, S. (2019). Didactical Issues at the Interface of Mathematics and Computer Science. In G. Hanna, D. A. Reid, & M. De Villiers (Eds.), *Proof Technology in Mathematics Research and Teaching* (Vol. 14, pp. 115–138). Cham : Springer International Publishing. (Series Title : Mathematics Education in the Digital Era) doi: 10.1007/978-3-030-28483-1_6

- Duval, R., & Egret, M.-A. (1993). Introduction à la démonstration et apprentissage du raisonnement déductif. *Repères-IREM*, 12, 114–140.
- Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education : A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, 70(3), 287–322. doi: 10.3102/00346543070003287
- Grugeon-Allys, B. (2017). Chapitre 2. Évaluer en mathématiques : une approche didactique et épistémologique. In P. Detroz, M. Crahay, & Annick Fagnant (Eds.), *L'évaluation à la lumière des contextes et des disciplines* (pp. 63–90). Louvain-la-Neuve : De Boeck Supérieur. doi: 10.3917/dbu.detro.2017.01.0063
- Guedet, G., & Vandebrouck, F. (2022). Transition secondaire-supérieur : Ce que nous apprend la recherche en didactique des mathématiques. *epiDEMES*, 1 | 2022. (Publisher : Episciences.org) doi: 10.46298/epidemes-7486
- Hanna, G., & De Villiers, M. (Eds.). (2012). *Proof and Proving in Mathematics Education : The 19th ICMI Study* (Vol. 15). Dordrecht : Springer Netherlands. doi: 10.1007/978-94-007-2129-6
- Horoks, J., & Pilet, J. (2023). Quels apports de la didactique pour penser l'évaluation ? L'exemple des mathématiques. In Cnesco (Ed.), *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 182–192). Cnesco-Cnam. Consulté sur https://www.cnesco.fr/wp-content/uploads/2023/09/Cnesco-CC-Eval_Note-Julie-HOROKS-Julie-PILET.pdf
- Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, 39(10), 1774–1787. doi: 10.1080/03075079.2013.821974
- Morrisette, J. (2014). Un panorama de la recherche sur l'évaluation formative des apprentissages. *Mesure et évaluation en éducation*, 33(3), 1–27. doi: 10.7202/1024889ar
- Niss, M. (Ed.). (1993). *Investigations into Assessment in Mathematics Education : An ICMI Study* (Vol. 2). Dordrecht : Springer Netherlands. doi: 10.1007/978-94-017-1974-2
- Reinholz, D. (2015). The assessment cycle : a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 41(2), 301–315. doi: 10.1080/02602938.2015.1008982
- Robert, A., & Vandebrouck, F. (2017). Activités mathématiques des élèves avec les technologies numériques. *Recherches en didactique des mathématiques*, 37(2-3), 333–382. Consulté sur <https://revue-rdm.com/2017/activites-mathematiques-des-eleves/>
- Sadler, P., & Good, E. (2006). The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment*, 11(1), 1–31. doi: 10.1207/s15326977ea1101_1
- Selden, A. (2012). Transitions and Proof and Proving at Tertiary Level. In G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education : The 19th ICMI Study* (pp. 391–420). Dordrecht : Springer Netherlands. doi: 10.1007/978-94-007-2129-6_17
- Selden, A., & Selden, J. (2003). Validations of Proofs Considered as Texts : Can Undergraduates Tell Whether an Argument Proves a Theorem ? *Journal for Research in Mathematics Education*, 34(1), 4–36. (Publisher : National Council of Teachers of Mathematics) doi: 10.2307/30034698
- Sommerhoff, D., & Ufer, S. (2019). Acceptance criteria for validating mathematical proofs used by school students, university students, and mathematicians in the context of teaching. *ZDM*, 51(5), 717–730. doi: 10.1007/s11858-019-01039-7
- Stylianides, G. J., Stylianides, A. J., & Moutsios-Rentzos, A. (2024). Proof and proving in school and university mathematics education research : a systematic review. , 56(1), 47–59. doi: 10.1007/s11858-023-01518-y
- Topping, K. (2003). Self and Peer Assessment in School and University : Reliability, Validity and Utility. In *Optimising New Modes of Assessment : In Search of Qualities and Standards* (Vol. 1, pp. 55–87). Dordrecht : Kluwer Academic Publishers. (Series Title : Innovation and Change in Professional Education) doi: 10.1007/0-306-48125-1_4

Topping, K. (2009, janvier). Peer Assessment. *Theory Into Practice*, 48(1), 20–27. doi: 10.1080/00405840802577569

Weber, K. (2008, juillet). How Mathematicians Determine if an Argument Is a Valid Proof. *Journal for Research in Mathematics Education*, 39(4), 431–459. doi: 10.5951/jresmetheduc.39.4.0431

Weber, K., & Alcock, L. (2004). Semantic and Syntactic Proof Productions. *Educational Studies in Mathematics*, 56(2), 209–234. doi: 10.1023/B:EDUC.0000040410.57253.a1

Juliette Veuillez--Mainard

Institut Montpellierain Alexander Grothendieck, Université de Montpellier, CNRS

e-mail: juliette.veuillez-mainard@umontpellier.fr

Simon Modeste

Institut Montpellierain Alexander Grothendieck, Université de Montpellier, CNRS

e-mail: simon.modeste@umontpellier.fr