ciip : Conférence intercantonale instruction publique et culture Suisse romande et Tessin

Document de travail

irdp

25.1002
NOVEMBRE 2025

L'évaluation de la résolution de problèmes en mathématiques

Apports des résultats du test pilote 2023

Isaline Ruf et Géraldine Hoffer
Avec la collaboration d'Alina Matei et d'Éric Mäder

L'évaluation de la résolution de problèmes en mathématiques

Apports des résultats du test pilote 2023

Isaline Ruf et Géraldine Hoffer
Avec la collaboration d'Alina Matei et d'Éric Mäder

Les *Rapports de recherche «Orange»* sont des documents de travail permettant de présenter l'état d'avancement d'une recherche et/ou d'un mandat en cours: délimitation d'un objet ou d'une problématique, description d'une méthodologie de recherche, présentation de résultats intermédiaires. Ils constituent également l'instrument de publication privilégié pour la valorisation et la diffusion de travaux réalisés par des stagiaires ou des assistant.es.

Cet ouvrage applique les rectifications orthographiques de 1990.

Rédaction	Isaline Ruf et Géraldine Hoffer
Coordination scientifique	Viridiana Marc
Coordination éditoriale	Anne Bourgoz Froidevaux
Relecture	Mathilde Ceylan
Vérification des bibliographies	Isabelle Deschenaux
Mise en page et conception graphique	Doris Penot
Photo de couverture	© Adobe Stock
Édition et diffusion	CIIP – Conférence intercantonale de l'instruction publique et de la culture de la Suisse romande et du Tessin Institut de recherche et de documentation pédagogique Case postale 556 2002 Neuchâtel – Suisse www.ciip.ch documentation@ciip.ch | +41 32 889 86 18

Table des matières

Liste des abréviations

CECR	Cadre européen commun de référence pour les langues
CIIP	Conférence intercantonale instruction publique et culture Suisse romande et Tessin
CO	Compréhension de l’oral
COMEVO	Commission évaluation des objectifs du PER
EC	Épreuves cantonales
EpRoCom	Épreuves romandes communes
Gdid	Groupe de didacticiens et didacticiennes
GRés	Groupe de résonance
IRDP	Institut de recherche et de documentation pédagogique
MER	Moyens d’enseignement romands
MSN	Mathématiques et sciences de la nature
ONAÉ	Outil Numérique d’Apprentissage et d’Évaluation
PE	Production de l’écrit
PO	Production de l’oral
PER	Plan d’études romand
PistEval	Pistes pour l’évaluation

Résumé

Inscrits dans le mandat de l'Institut de recherche et de documentation pédagogique (IRDP), les travaux relatifs à l'évaluation des *objectifs d'apprentissage* du Plan d'études romand (PER), menés sur la période quadriennale 2020-2023, visent à créer une culture commune entre les cantons romands en matière d'évaluation des apprentissages des élèves. Dans cette perspective, ils permettent de se doter d'outils d'analyse et de mettre à disposition des enseignantes et enseignants des matériaux d'évaluation pertinents, validés et fiables. Ce texte fait partie d'une publication plus complète, qui rassemble les travaux réalisés dans plusieurs disciplines pour des élèves de 8e année (11-12 ans).

Ce rapport présente les principaux résultats concernant l'évaluation de la *résolution de problèmes* en Mathématiques. Il détaille les analyses effectuées sur les données récoltées auprès de 56 classes lors d'un test pilote déployé au printemps 2023, d'une part sur support papier-crayon et, d'autre part, sur support informatisé.

Les analyses se sont centrées sur les procédures de résolution mises en œuvre par les élèves, avec une attention particulière sur la possible influence de la mise à disposition de la calculatrice pour les problèmes effectués sur le support papier-crayon.

Les conclusions montrent des avantages concrets de l'utilisation du numérique dans l'évaluation, comme la possibilité de proposer des tâches adaptatives, différenciées suivant les actions des élèves, et de récolter de nombreuses données permettant de retracer leur raisonnement. Toutefois, ce type d'évaluation requiert une familiarisation préalable pour les élèves qui pourraient, en situation d'enseignement-apprentissage, s'habituer au format, aux fonctionnalités et aux outils proposés sur le support numérique. Quant à l'utilisation de la calculatrice, bien que sa mise à disposition lors de l'évaluation de *la résolution de problèmes* améliore la réussite des élèves à la tâche, elle ne favorise pas forcément la mise en œuvre de procédures expertes.

1. Cadre général

Les travaux présentés dans ce rapport s'inscrivent dans le cadre du projet EpRoCom, initié depuis 2016 et centré sur l'évaluation des apprentissages des élèves. Plus spécifiquement, ce rapport présente les résultats obtenus lors de la dernière période quadriennale (2020-2023), en ce qui concerne la mutualisation de ressources permettant d'**évaluer les compétences** inscrites dans le Plan d'études romand (PER) (CIIP, 2010/2024).

1.1 Objectif du projet

Depuis plusieurs années, des réflexions quant à l'évaluation des apprentissages des élèves au niveau romand ont été menées par les équipes de l'Institut de recherche et de documentation pédagogique (IRDP), structure scientifique rattachée à la Conférence intercantonale de l'instruction publique et de la culture de la Suisse romande et du Tessin (CIIP). Dès l'adoption du Plan d'études romand (PER) en 2010, l'enjeu majeur a été de définir les qualités et les caractéristiques d'une **évaluation de compétences**, celle-ci ne se résumant pas à évaluer une somme de connaissances, de savoir-faire et de savoir-être, bien que ces éléments participent au développement et à l'expression des compétences des élèves inscrites dans le PER. De ce fait, la mobilisation, la sélection et la combinaison de ces différents savoirs[1], considérés comme des ressources, sont nécessaires pour accomplir une tâche complexe et ainsi développer des compétences (Allal, 1999 ; Rey et al., 2003).

Pour faire face à cet enjeu ambitieux, le projet EpRoCom s'est attaché à mettre à disposition du corps enseignant de Suisse romande des exemples de ressources évaluatives **validées** (utilisables par tous les cantons), **pertinentes** (appropriées par rapport à l'objectif visé), **fiables** (qui renseignent les savoirs objets de l'évaluation), et **basées sur les *objectifs d'apprentissage* du PER**, lesquels sont formulés en termes de compétences à développer.

Pour le programme d'activités 2020-2023, les travaux se sont concentrés sur la 8^{e} année (élèves de 11-12 ans), dans les disciplines suivantes :

- le Français (production de l'écrit [PE]),
- l'Allemand langue étrangère (compréhension de l'oral [CO] et production de l'oral [PO]),
- les Mathématiques (résolution de problèmes).

Les concrétisations de ce projet (*cf.* chapitre 1.2) ont pour objectif de participer au développement d'une culture commune romande de **l'évaluation des *objectifs d'apprentissage* du PER**, et par conséquent de **compétences**.

[1] « Savoirs » est ici à prendre au sens large et comprend autant des savoirs encyclopédiques que des savoir-faire ou même des compétences.

1.2 Concrétisations du projet

1.2.1 Pistes pour l'évaluation (PistEval)

Dès la rentrée scolaire 2021, une des principales réalisations du projet a consisté en la création, sur des pages dédiées du site du PER, de *Pistes pour l'évaluation* (*PistEval*)[2]. Ces pages internet proposent des exemples de ressources évaluatives et des clarifications théoriques et didactiques permettant de soutenir le concept d'**évaluation de compétences** (Roth et al., 2021 ; Roth & Ruf, 2024). Elles sont destinées au corps enseignant romand, qui y accède sous identifiant.

Les pages *PistEval* contiennent actuellement, pour des élèves de 8e année, des exemples de ressources évaluatives pour la *compréhension de l'écrit* et la *production de l'écrit* en Français, ainsi que pour la *résolution de problèmes* en Mathématiques. Elles sont accompagnées d'étayages théorico-didactiques et méthodologiques, qui soutiennent d'une part le corps enseignant dans la prise en main des ressources et, d'autre part, mettent en lumière certaines réflexions essentielles à avoir lors du choix ou de l'élaboration de ressources pour évaluer les apprentissages réalisés par leurs élèves. Le statut illustratif des ressources proposées est particulièrement mis en avant, car, pour en permettre une utilisation adéquate, les enseignantes et enseignants doivent veiller à les ajuster afin qu'elles soient en adéquation avec les objectifs poursuivis, l'enseignement dispensé et les apprentissages réalisés dans leurs classes. Les réflexions mises en évidence par les étayages, idéalement réinvesties par le corps enseignant, non seulement en situation d'évaluation, mais aussi dans son enseignement, vont dans le sens d'une **évolution des pratiques évaluatives au service des apprentissages des élèves** (Roth & Ruf, 2024).

1.2.2 Outil Numérique d'Apprentissage et d'Évaluation (ONAÉ)

Au cours de 2022, le projet s'est enrichi d'un apport supplémentaire avec le développement d'une application dédiée à l'apprentissage et à l'évaluation des élèves sur support informatisé, appelée *ONAÉ* (*Outil Numérique d'Apprentissage et d'Évaluation*). Initialement conçue pour évaluer les compétences orales des élèves en Allemand langue étrangère (*compréhension de l'oral* et *production de l'oral*), *ONAÉ* présente la possibilité d'évoluer vers un outil plus polyvalent, destiné à servir le corps enseignant et les élèves. Ainsi, en complément des tâches prévues pour l'évaluation des compétences orales en Allemand, elle a aussi été utilisée, de manière exploratoire, pour évaluer la *résolution de problèmes* en Mathématiques.

ONAÉ est une application web, accessible directement en ligne via un navigateur. Son développement a été conçu pour un support numérique de type tablette, offrant ainsi de nombreux avantages (transport et mise en place facilités, usage intuitif, mobilité du dispositif, etc.). Cette application se compose de deux interfaces distinctes :

- Une **interface de gestion (*back-office*)**, qui permet d'enregistrer les informations relatives aux classes ou aux élèves, de programmer et de gérer en temps réel les passations, ainsi que d'exporter les données collectées.

[2] Les pages internet *PistEval* sont accessibles via la plateforme PER-MER :
https://bdper.plandetudes.ch/barome/francais-8/
https://bdper.plandetudes.ch/barome/maths-8/

- Une **interface élève (*frontend*)**, avec laquelle l'élève interagit. Cette interface propose plusieurs outils numériques, sélectionnés selon les besoins spécifiques de la tâche à accomplir.

Conçues pour être évolutives, ces interfaces pourront ainsi se développer progressivement, suivant les besoins des utilisatrices et utilisateurs.

1.3 Dispositif d'ensemble

Avant la mise à disposition des ressources évaluatives sur les pages *PistEval*, un **processus de validation** est indispensable pour garantir une prise en compte de la recherche et l'adéquation aux besoins de tous les cantons romands (Roth et al., 2021). Ce processus dure en moyenne quatre ans et implique différents groupes de travail qui partagent leur expertise. Ainsi, le groupe de conception de l'IRDP, chargé du projet, est soutenu par des groupes consultatifs, l'un composé d'enseignantes et d'enseignants issus de tous les cantons romands, et l'autre de didacticiennes et didacticiens pour chaque discipline concernée.

Comme convenu dès le début du projet par les organes décideurs, les travaux devaient être menés à partir des épreuves cantonales[3] (EC) produites par les cantons romands. Pour ce faire, certains axes du PER ont été priorisés, en accord avec les instances pilotant le projet. Pour la période 2020-2023, les axes priorisés ont été la *production de l'écrit* en Français, la *compréhension de l'oral* en Allemand[4], et la *résolution de problèmes* en Mathématiques. Suite à une première sélection parmi les EC, les ressources retenues ont fait l'objet d'une expertise tant scientifique, réalisée par les groupes de didacticiennes et didacticiens, que de terrain, par le groupe d'enseignantes et enseignants. Finalement, les ressources ont fait l'objet d'un processus de vérification de leur adéquation, à travers une démarche quantitative lors d'un test pilote réalisé en 2023 auprès des élèves, et/ou par une approche qualitative plus ciblée.

Pour le test pilote de 2023, afin de garantir une certaine représentativité de la population scolaire de la partie francophone de la Suisse, chacun des sept cantons romands a désigné huit classes de 8e année pour prendre part aux passations. Le test pilote s'est donc déployé auprès de **plus de 1000 élèves**, répartis dans 56 classes. Les passations se sont déroulées sur quatre périodes (de 45 ou 50 minutes selon les cantons), les deux premières étant consacrées à la partie informatisée (Allemand et Mathématiques) à l'aide du dispositif *ONAÉ* et les deux suivantes aux tâches papier-crayon (uniquement Mathématiques)[5].

Afin de soutenir la partie opérationnelle du dispositif d'ensemble, des prestataires externes ont également participé au projet :

- une équipe de développement informatique, pour maintenir et développer l'infrastructure technique des pages *PistEval* et de l'application numérique *ONAÉ* ;
- des administrateurs et administratrices de test pour superviser les passations dans les classes, ainsi que des codeurs et codeuses en charge de la correction et du codage des productions des élèves.

---

[3] La terminologie « épreuves cantonales (EC) » regroupe les différentes appellations utilisées dans les cantons pour désigner les évaluations externes : évaluation cantonale (FR et GE), épreuve commune (JU), épreuve cantonale de compétences (NE), épreuve cantonale de référence (VD) et examen cantonal (VS).

[4] Une partie expérimentale pour la *production de l'oral* (*PO*) a également été prévue, mais le manque de matériau cantonal initial a nécessité l'élaboration de tâches spécifiques par le groupe de conception.

[5] Les tâches de Français ont été testées dans les classes seulement en 2019 (*cf.* Roth et al. 2021).

Enfin, le dispositif a nécessité des analyses des résultats des élèves obtenus lors des tests pilotes ainsi que des processus de validation qualitatifs, permettant de déterminer si des tâches[6] d'origines cantonales différentes pouvaient s'adresser à l'ensemble des élèves de Suisse romande. L'introduction du support informatisé, par le biais de l'utilisation d'*ONAÉ*, a considérablement augmenté la quantité de données récoltées. Afin de convertir les informations brutes fournies par *ONAÉ* en données exploitables, un tri préalable suivi de plusieurs méthodes de traitement (*data process*) a été réalisé.

[6] Dans le contexte de l'évaluation, une tâche fait référence à la combinaison d'une consigne, d'une entrée et de réponses pour évaluer un objet (ALTE, 1998).

2. Évaluer la *résolution de problèmes* en Mathématiques

Comme évoqué dans les chapitres 1.1 et 1.3, les travaux de la période quadriennale 2020-2023 se sont concentrés sur l'évaluation de la *résolution de problèmes*. Tous axes et niveaux confondus, le PER (CIIP, 2010/2024) place la *résolution de problèmes* au centre de l'activité mathématique, en témoignent les visées prioritaires du domaine Mathématiques et sciences de la nature (MSN) :

> *Se représenter, problématiser et modéliser des situations et résoudre des problèmes en construisant et en mobilisant des notions, des concepts, des démarches et des raisonnements propres aux Mathématiques [...]. (CIIP, 2010, p. 5)*

Dans le PER, la *résolution de problèmes* apparait de manière transversale, à travers le premier chapitre de **chaque axe thématique** intitulé « Éléments pour la résolution de problèmes » (CIIP, 2010), présent pour **l'ensemble des années de la scolarité obligatoire**. S'y retrouvent notamment les progressions des apprentissages suivantes :

- tri et organisation des informations (liste, tableau, schéma, croquis…) ;
- mise en œuvre d'une démarche de résolution ;
- ajustement d'essais successifs ;
- pose d'une conjecture, puis validation ou réfutation ;
- déduction d'une ou plusieurs informations nouvelles à partir de celles qui sont connues ;
- vérification, puis communication d'une démarche et d'un résultat en utilisant un vocabulaire, une syntaxe ainsi que des symboles adéquats.

La priorisation de la *résolution de problèmes* dans le projet EpRoCom a pour objectif de soutenir l'**évaluation de compétences**. Le PER définissant une compétence comme la « possibilité, pour un individu, de mobiliser un ensemble intégré de ressources en vue d'exercer efficacement une activité considérée généralement comme complexe » (CIIP, 2010, p. 43), la *résolution de problèmes* s'inscrit effectivement dans cette forme d'activité complexe (Carulla et al., 2013), caractérisée par **la mobilisation, la sélection et la combinaison de différents savoirs et savoir-faire** (*cf*. chapitre 1.1).

2.1 Qu'est-ce qu'un problème en Mathématiques ?

S'il n'existe pas de définition universelle d'un « problème », un certain consensus existe pour considérer qu'un problème consiste en une question plus ou moins contextualisée, dont la démarche de résolution n'est pas immédiatement disponible pour l'élève, au regard de son âge et des apprentissages réalisés (Monaghan et al., 2009 ; Newell & Simon, 1972; Schoenfeld, 1985). En d'autres termes, lors de la résolution d'un problème, l'élève est amené à s'interroger, de manière autonome, sur la stratégie à mettre en œuvre, les outils et notions à convoquer, puis à utiliser de manière adéquate un ou plusieurs outils mathématiques pour arriver à la solution.

Au départ, un problème est relatif, comme le souligne Brun (1990, p. 2) : « Il n'y a problème que dans un rapport sujet/situation, où la solution n'est pas disponible d'emblée, mais possible à construire. C'est dire aussi qu'un problème pour un sujet donné peut ne pas être un problème pour un autre sujet, en fonction de leur niveau de développement intellectuel par exemple. » Par conséquent, au terme d'une séquence d'enseignement-apprentissage, selon ce qui a été automatisé en classe, un nombre important de tâches ne correspond plus à des problèmes, mais à des exercices pour l'élève. Ainsi, proposer des problèmes en évaluation revient à trouver l'équilibre entre une tâche qui soit effectivement un problème pour l'élève, mais dont tous les éléments nécessaires pour la résoudre ont été travaillés en classe (Ruf & Weiss, 2024).

Prenant appui sur ces éléments, une première étape du travail a consisté en l'analyse de l'ensemble des tâches constituant les EC produites par les cantons romands (*cf.* chapitre 1.3), afin de les catégoriser. En effet, ces tâches étant de diverses natures et évaluant des savoirs ou compétences différentes, il s'agissait de déterminer lesquelles permettaient d'évaluer la *résolution de problèmes* au niveau romand, en se basant sur les attentes du PER pour la 8e année.

Nous avons différencié deux types de problèmes, sur la base de la classification d'Houdement (2017) : les **« problèmes »** et les **« problèmes atypiques »**[7] (Ruf & Weiss, 2024). Le premier type correspond à des tâches qui permettent **d'évaluer la mobilisation, l'articulation et la maitrise d'outils ou notions mathématiques dans une situation complexe** (figure 1). Ces **« problèmes »** demandent à l'élève de se questionner quant à la démarche à mettre en œuvre et aux outils à mobiliser, voire à associer. Le plus souvent, ils requièrent la mise en œuvre d'une combinaison non automatisée de procédures se rattachant souvent à plusieurs notions ou outils mathématiques. Comme le relèvent Ruf et Weiss (2024, p. 11), « c'est la combinaison ou même l'imbrication des différentes étapes qui mettent l'élève en situation de résolution de problèmes ». En effet, la réponse ne peut généralement pas être trouvée en une seule étape et nécessite de rechercher des résultats intermédiaires, sans que ceux-ci soient guidés par des questions subsidiaires ; c'est donc à l'élève qu'est laissée la charge de « sélectionner et connecter des informations en qualifiant les résultats intermédiaires » (Académie de Créteil, 2018) de sous-tâches qu'elle ou il a construites. Ainsi, « pour parvenir à la solution, l’élève doit impérativement trier, rechercher, décomposer, organiser et mettre en relation les données de l'énoncé » (Académie de Créteil, 2018). Ces éléments se retrouvent d'ailleurs dans le chapitre « Éléments pour la résolution de problèmes » du PER, comme nous l'avons vu ci-avant (CIIP, 2010).

Voici deux exemples de « problèmes » considérés comme adéquats pour des élèves de 11-12 ans :

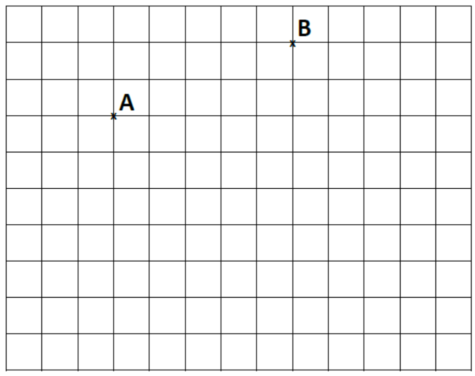
Figure 1 : Exemples de « problèmes » inspirés des EC, adéquats pour des élèves de 11-12 ans (Ruf & Weiss, 2024, p. 8)

Un boucher a préparé 60 saucissons. Il les vend 4 francs pièce. Cette semaine, il propose une action : si un client achète 4 saucissons, il ne paie que 14 francs. 	Dans ce système d'axes, les coordonnées du point A sont (-8 ; 3) et celles du point B sont (2 ; 9).

[7] Pour plus de détails, voir les catégories de tâches présentées sur les pages *PistEval* : https://bdper.plandetudes.ch/barome/maths-8/eclairages_categ/

À la fin de la journée, 20 clients ont acheté des saucissons. 8 d'entre eux ont profité de l'action et les autres se sont contentés d'un seul saucisson.

Combien d'argent le boucher a-t-il encaissé durant la journée grâce à la vente des saucissons ?



Place le point C qui a pour coordonnées (0 ; -6).

Le second type de tâches de *résolution de problèmes* regroupe les problèmes dits « **atypiques** », soit des tâches qui permettent d'évaluer des **compétences spécifiques à la *résolution de problèmes***. Elles demandent à l'élève d'imaginer puis de mettre en œuvre des stratégies de résolution comme l'ajustement d'essais successifs ou la mise en place d'une recherche de toutes les solutions car, la plupart du temps, l'élève ne dispose pas d'un outil lui permettant d'arriver directement à la solution (figure 2). Les « problèmes atypiques » requièrent souvent un temps de recherche considérable.

Figure 2 : Exemples de « problèmes atypiques » inspirés des EC, adéquats pour des élèves de 11-12 ans (Ruf & Weiss, 2024, p. 9)

Un abricotier magique produit des abricots de la manière suivante :

Jour	Nombre d'abricots
1^{er}	4
2^{e}	7
3^{e}	10
4^{e}	13
5^{e}	16

Combien d'abricots cet abricotier produira-t-il le 100^{e} jour ?

Colorie entièrement certains petits cercles de cette figure de manière à ce qu'elle n'ait aucun axe de symétrie.

2.2 L'évaluation de la *résolution de problèmes*

Lors de l'évaluation de la *résolution de problèmes*, ce n'est pas tant le résultat qui compte, mais le raisonnement de l'élève pour y parvenir (Julo, 1995). Ainsi, la procédure mise en œuvre par l'élève est centrale et l'analyse détaillée de sa production (en sus de la réponse soumise) permet d'identifier les outils et/ou connaissances mathématiques qu'elle ou il a mobilisés, la manière dont elle ou il les a agencés pour arriver à une solution et, si nécessaire, d'envisager des remédiations ciblées au regard du type d'erreurs identifiées.

Souvent, plusieurs procédures différentes sont possibles pour résoudre un problème. En fonction de la procédure mise en œuvre par l'élève, il est possible d'identifier son niveau de maitrise des compétences en jeu dans la situation proposée. Une procédure que nous avons qualifiée d'experte (*cf*. chapitre 3.2.2), qu'elle aboutisse à une réponse correcte ou non, est synonyme de compétence maitrisée, tandis que la mise en œuvre d'une procédure correcte montre que l'élève a saisi la situation mathématique, mais qu'elle ou il peut encore progresser et devenir plus efficace dans sa résolution. Quant à la mise en œuvre d'une procédure incorrecte, elle souligne une difficulté chez l'élève et il convient alors d'analyser son origine pour pouvoir y remédier.

2.3 Test pilote 2023 – partie Mathématiques

À partir de 2022, il a été décidé d'enrichir les exemples de tâches de *résolution de problèmes* disponibles en Mathématiques sur les pages *PistEval*, en particulier pour les chapitres du PER qui n'étaient que peu ou pas du tout illustrés. En effet, plusieurs chapitres n'étaient pas touchés par les tâches mises à disposition suite au test pilote de 2019, et il s'agissait de proposer des tâches permettant de les évaluer. Ainsi, afin de tester sur le terrain de nouvelles tâches, un test pilote a été mis sur pied entre avril et juin 2023, sur la base de l'expérience réalisée en 2019 (Roth et al., 2021). Ce test pilote fait partie intégrante du processus de validation des tâches de Mathématiques (*cf*. chapitre 1.3) en vue de leur intégration sur les pages *PistEval*, et visait à s'assurer de la bonne compréhension des tâches par l'ensemble des élèves romands, indépendamment de l'origine cantonale de la tâche.

Le test pilote ayant été réalisé sur deux supports pour les Mathématiques, papier-crayon et numérique (*cf.* chapitre 1.3), les deux types de productions récoltées ont nécessité une méthodologie d'analyse différente. Elles sont présentées dans les chapitres qui suivent.

3. Évaluer la résolution de problèmes sur support papier-crayon

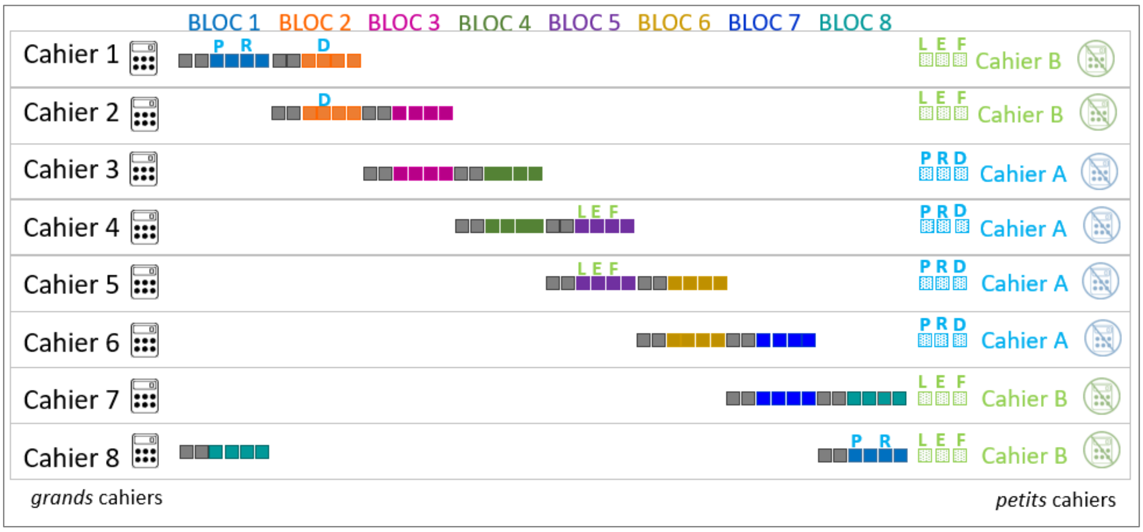
3.1 Méthodologie

Pour les passations dans les classes, les 36 problèmes de Mathématiques soumis aux élèves sur support papier-crayon ont été répartis dans différents cahiers selon la méthode des cahiers tournants : huit *grands* cahiers (numérotés de 1 à 8), composés de 12 problèmes pour lesquels la calculatrice était autorisée, et deux *petits* cahiers (A et B) contenant chacun trois problèmes[8] à effectuer sans cet outil de calcul (figure 3). Cette méthode, qui permet de tester un nombre important de tâches sans pour autant accroitre le temps de passation, consiste à répartir les problèmes dans des « blocs », autrement dit des ensembles de tâches. De ce fait, les problèmes n'ont pas tous été effectués par le même nombre d'élèves. Huit blocs différents, chacun constitué de six tâches, ont été formés. Chacun des huit *grands* cahiers comprenait des problèmes issus de deux blocs différents.

Un focus particulier de ce deuxième test pilote ayant porté sur l'impact que l'usage de **la calculatrice** peut avoir sur les procédures de résolution mises en œuvre par les élèves, six problèmes ont été soumis à une partie des élèves avec la calculatrice et à d'autres sans cet outil de calcul. Ce dispositif permet d'examiner comment la disponibilité de la machine peut influencer (ou non) les procédures de résolution. Lors de la passation, chaque élève a ainsi reçu, de manière aléatoire, un *petit* cahier à réaliser sans la machine. Une fois celui-ci terminé et rendu à l'administrateur ou l'administratrice du test, l'élève pouvait récupérer sa calculatrice et s'atteler aux 12 problèmes du *grand* cahier.

[8] À noter que les problèmes des petits cahiers se retrouvent également dans les grands cahiers. Lors de la répartition des cahiers, on a veillé à ce qu'aucun élève ne soit amené à résoudre un même problème une première fois sans calculatrice et une deuxième fois avec la machine.

Figure 3 : Dispositif des cahiers tournants déployé lors du test pilote de Mathématiques de 2023. Chaque carré représente une tâche. Les carrés de couleur grise correspondent aux problèmes d'ancrage, systématiquement présentés en position 1-2 et 7-8 de chaque grand cahier numéroté de 1 à 8. Les lettres en bleu et vert correspondent aux six problèmes ayant été réalisés avec la calculatrice par certains élèves et sans la calculatrice par d'autres, à savoir ***P*** *: Poterie,* ***R*** *: Récolte de coton,* ***D*** *: Le drapeau suisse,* ***L*** *: Long couloir,* ***E*** *: Entrées individuelles ou abonnement,* ***F*** *: Fête des Vignerons*



Chaque *grand* cahier comprenait quatre problèmes d'ancrage (signalés par leur couleur grise dans la figure 3), correspondant à des tâches présentes dans tous les cahiers (deux problèmes au début et deux au milieu de chaque cahier). Ces tâches ont donc été soumises à l'ensemble de l'échantillon romand, soit 1085 élèves, dans le but de vérifier l'équivalence des résultats des groupes d'élèves ayant répondu à des cahiers différents et donc de s'assurer, grâce à une analyse statistique, que la comparaison entre les résultats des problèmes des différents cahiers était possible. Quant aux 32 autres problèmes, en fonction de leur présence dans les cahiers, ils ont été résolus chacun par un groupe variant de 269 à 280 élèves issus de l'échantillon romand. Il convient de préciser que les cahiers ne sont pas conçus dans l'optique d'une épreuve dans laquelle les différents sujets de l'évaluation se doivent d'être équilibrés, mais bien de sorte à pouvoir soumettre aux élèves des classes sélectionnées un nombre adéquat de tâches. Cette manière de faire permet de profiter pleinement du dispositif mis en place, tout en évitant un nombre excessif de tâches pour ne pas décourager les élèves. Soulignons que, même si un problème donné était présent dans un cahier, il n'est pas certain que l'élève l'ait réalisé (par manque de temps par exemple), d'où des résultats manquants.

Préalablement à la passation du test pilote par les élèves, **un questionnaire a été soumis aux enseignantes et enseignants** de chacune des classes testées, afin de récolter des informations sur leurs élèves y prenant part[9] pour connaitre, par exemple, d'éventuelles mesures d'aide, dites « mesures d'accompagnement ». Des informations concernant le travail disciplinaire en classe ont également été collectées : pour les Mathématiques, le corps enseignant s'est notamment prononcé sur l'usage de la calculatrice dans leur classe (fréquence, fonctionnalités, type d'activités concernées).

[9] Les élèves au bénéfice d'objectifs adaptés dans l'une des disciplines testées n'ont pas pris part au test pilote de la discipline concernée.

Pour chaque tâche, les élèves se sont prononcés sur leur perception de son **niveau de difficulté** (facile/moyen/difficile), ainsi que sur leur **habitude à résoudre ce type de problèmes** (oui/non). En outre, à la fin du *grand* cahier, les élèves ont complété un questionnaire portant sur leur **utilisation de la calculatrice** en classe (fréquence et fonctionnalités utilisées).

À l'issue du test pilote, les productions papier-crayon des élèves ont été corrigées manuellement, puis codées. Le codage portait notamment sur **la réponse** (correcte, partielle, erronée, absence de réponse), **la procédure** (experte, correcte, incorrecte, incomplète, absence de procédure – *cf.* chapitre 3.2.2) et les **erreurs** (de calcul, d'unités, de copie, de dénombrement, etc.). Ces données ont été analysées et les résultats font l'objet du sous-chapitre qui suit.

3.2 Principaux résultats

3.2.1 Questionnaires relatifs à l'usage de la calculatrice

Afin de dresser un panorama de l'usage qui est fait de la calculatrice dans les classes de 8e année en Suisse romande, nous avons traité et analysé les réponses aux deux questionnaires : d'une part, celui soumis aux 1085 élèves et, d'autre part, celui envoyé aux 56 enseignantes et enseignants titulaires des classes ayant réalisé le test pilote, pour lequel 51 réponses nous sont parvenues en retour. Nous allons à présent exposer les principaux résultats issus de ces deux questionnaires.

Au niveau de l'**équipement dans les classes**, les données récoltées montrent que celui-ci varie d'un canton à l'autre, voire d'une classe à l'autre, que ce soit en termes de nombre de calculatrices à disposition des élèves (allant d'une calculatrice par élève à une calculatrice pour quatre à cinq élèves), de modèles de calculatrices utilisés ou même d'appartenance de cet outil (à savoir personnelle ou faisant partie du matériel de classe)[10].

Nous avons aussi interrogé les enseignantes et enseignants sur leur sentiment d'aisance personnelle face à l'utilisation de la calculatrice, ainsi que sur leur fréquence d'utilisation. Bien qu'une majorité du corps enseignant interrogé (90%) se sente à l'aise avec l'usage de la calculatrice (figure 4), elle est en moyenne peu autorisée en **situation d'enseignement-apprentissage**. En effet, la fréquence d’utilisation de la machine est (très) faible pour plus de la moitié des classes : 60% des enseignantes et enseignants de 8e année questionnés indiquent qu’elle ne l’est que rarement, et presque personne ne l’utilise à chaque leçon[11] (figure 5). Les réponses des élèves corroborent ces résultats : la calculatrice est utilisée en classe en moyenne une seule fois par mois ou encore moins par 42% des élèves (figure 6). Environ deux élèves sur cinq déclarent l'utiliser en classe une ou plusieurs fois par semaine.

[10] Pour plus de précisions, voir Ruf (2023).

[11] Par ailleurs, les données récoltées montrent qu’il y a peu de différence de recours à la calculatrice en fonction du thème : la fréquence moyenne de son utilisation est similaire, quel que soit le thème abordé, à savoir 1) les thèmes liés aux *Nombres et Opérations*, 2) le thème *Applications*, 3) les thèmes *Mesures* et *Aires et volumes*. Pour plus de détails, voir Ruf (2023).

Figure 4 : Répartition des réponses des enseignantes et enseignants à la question « Vous sentez-vous vous-même à l'aise avec l'utilisation de [la] calculatrice [utilisée par vos élèves] ? »

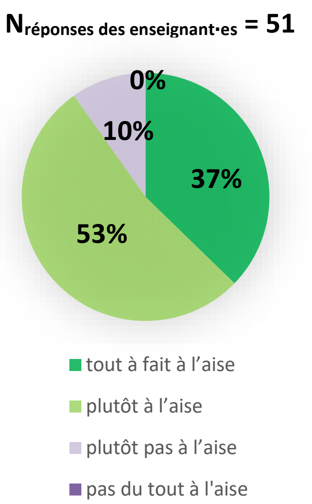


Figure 5 : Fréquence moyenne d'utilisation de la calculatrice[12]

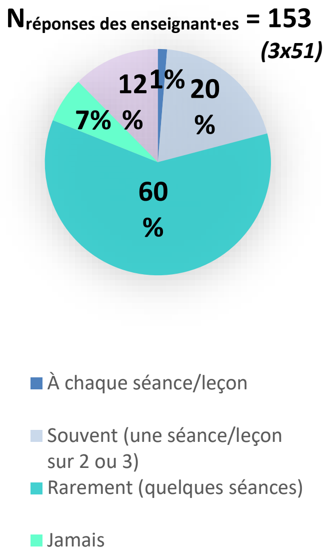
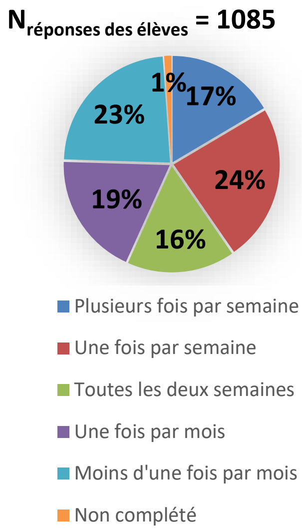


Figure 6 : Répartition des réponses des élèves à la question : « En classe, j'utilise en moyenne la calculatrice… »

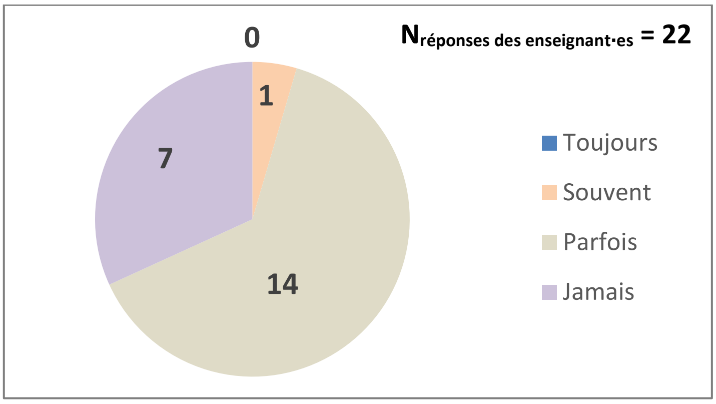


Les principales raisons évoquées par 29 enseignantes et enseignants autorisant « rarement » ou « jamais » l'utilisation de la calculatrice en classe sont la non-pertinence de la calculatrice au regard des tâches proposées aux élèves, la non-nécessité à ce niveau de scolarité et/ou que l'élève aura l'occasion de l'utiliser plus tard. Ces résultats rejoignent ceux de recherches antérieures soulignant la faible intégration de la calculatrice en classe de mathématiques (par exemple Trouche, 2002), que ce soit en Suisse romande ou ailleurs.

Quant aux 22 enseignantes et enseignants autorisant plus fréquemment l'utilisation de la calculatrice, il est intéressant de relever qu'elles et ils l'autorisent nettement moins en **situation d'évaluation** qu'en situation d'enseignement-apprentissage. En effet, la majorité des classes (64%) ont « parfois » la calculatrice à disposition, 32% des enseignantes et enseignants ne l'autorisent jamais, et personne ne l'autorise à chaque évaluation (figure 7). Une différence notable est ainsi constatée entre les situations d'enseignement-apprentissage et celles d'évaluation : quasiment personne ne laisse la calculatrice à disposition des élèves en évaluation (figure 7). Cette rupture nous questionne : pourquoi ne pas autoriser l'usage de la machine, au moins pour une partie de l'évaluation ? Les EC prévoient pourtant un tel dispositif : les élèves peuvent l'utiliser pour une partie de l'épreuve, plutôt axée sur la *résolution de problèmes*.

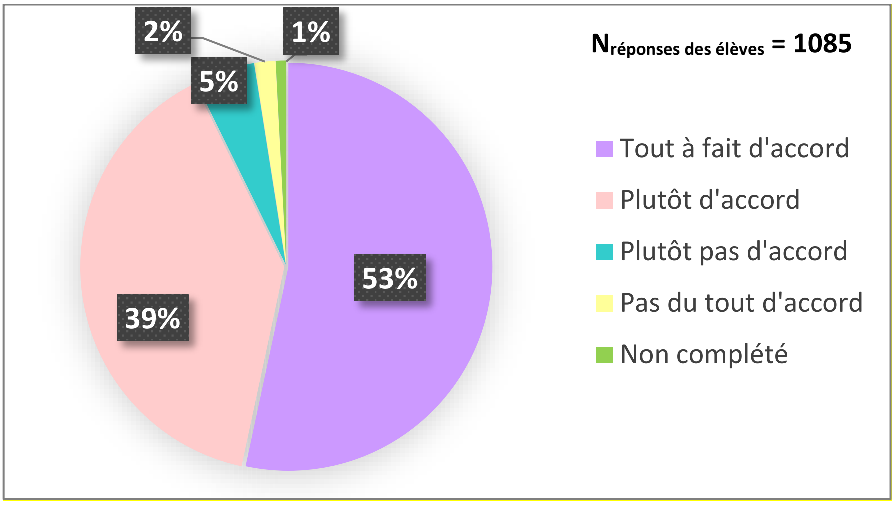
[12] La réponse « Je n'enseigne pas ce(s) thème(s) » s'explique par le fait que parfois les enseignantes et enseignants travaillent en duo dans une classe et se répartissent les thèmes.

Figure 7 : Répartition de la fréquence de l'accès à la calculatrice en situation d'évaluation, pour les 22 enseignantes et enseignants l'autorisant fréquemment



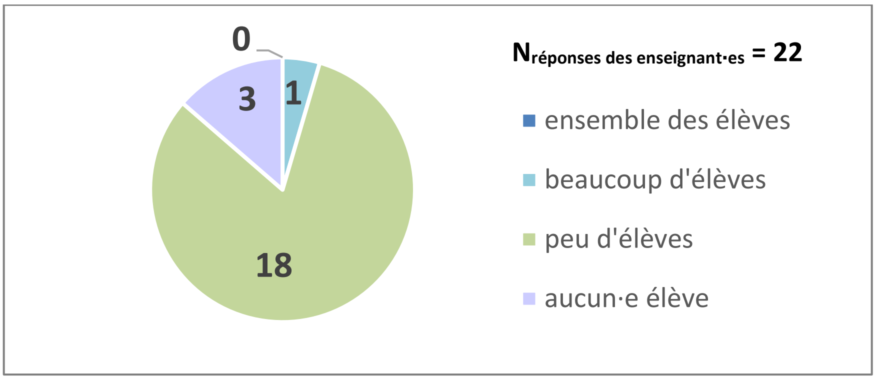
Nous avons également soumis un court questionnaire aux élèves concernant l'utilisation de la calculatrice. À l'affirmation « Je n'ai pas de difficulté à utiliser une calculatrice. », 92% des élèves indiquent être d'accord ou plutôt d'accord (*cf*. figure 8).

Figure 8 : Répartition des appréciations des élèves à l'affirmation : « Je n'ai pas de difficulté à utiliser une calculatrice. »



Les réponses des enseignantes et enseignants vont dans le même sens : selon eux, peu voire pas d'élèves rencontrent des difficultés. Seule une personne signale que c'est le cas de beaucoup d'élèves (figure 9). Nous nous attendions à ce que les difficultés rencontrées par les élèves soient rares. Ces résultats nous semblent cohérents avec le fait que la majorité des classes utilisent principalement les fonctionnalités basiques de la calculatrice (soit les quatre opérations), compétences faisant partie des attentes fondamentales du PER (CIIP, 2010 ; Ruf, 2023).

Figure 9 : Fréquence des difficultés rencontrées par les élèves quant à l'utilisation de la calculatrice en 8^e^ année, déclarées par les enseignantes et enseignants



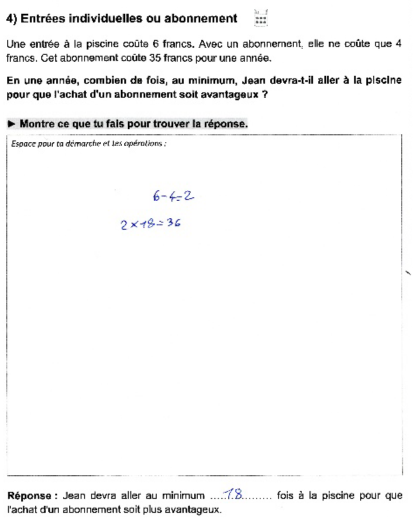
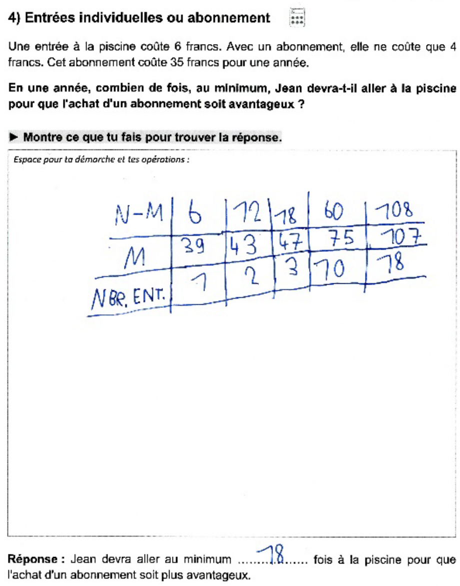
3.2.2 *Procédures mises en œuvre*

Dans cette partie, nous présentons les résultats issus des analyses réalisées à partir des 36 problèmes papier-crayon testés dans les classes lors du test pilote 2023. Les analyses effectuées se sont principalement focalisées sur les procédures de résolution mises en œuvre, celles-ci étant primordiales pour évaluer les compétences des élèves en *résolution de problèmes*.

En amont de la tenue du test pilote de 2023, nous avons différencié **cinq types de procédures**, décrites ci-dessous. Afin de les rendre plus concrètes, nous les avons illustrées avec de réelles productions d'élèves ayant résolu le problème intitulé « Entrées individuelles ou abonnement » :

- **Procédure experte** : procédure la plus directe (comprenant généralement le moins d'étapes) et la plus efficace pour arriver à la réponse.

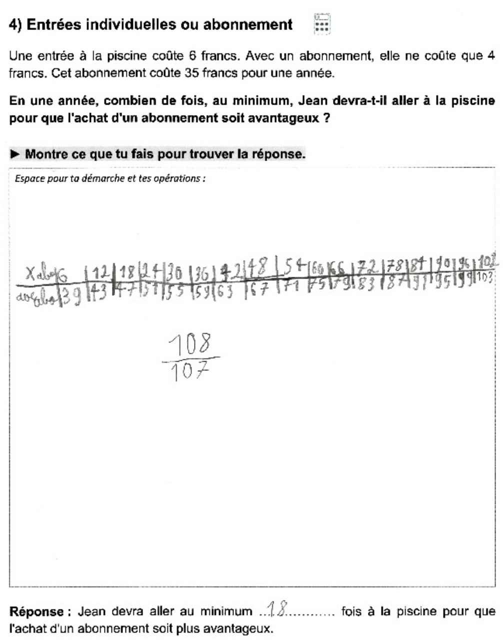
Figure 10 : Exemples de productions d'élèves issues du test pilote 2023, illustrant des procédures expertes pour le problème « Entrées individuelles ou abonnement »

<table>
<tr>
<td>Calcul de l'économie réalisée pour chaque entrée avec l'abonnement, puis calcul du nombre d'entrées nécessaires pour atteindre le coût de l'abonnement.

4) Entrées individuelles ou abonnement
Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année.
En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?
► Montre ce que tu fais pour trouver la réponse.
Espace pour ta démarche et tes opérations :
6-4=2
2×18=36
Réponse : Jean devra aller au minimum ...18......... fois à la piscine pour que l'achat d'un abonnement soit plus avantageux.</td>
<td>Calcul du coût sans et avec abonnement selon le nombre d'entrées (ajustement d'essais successifs).

4) Entrées individuelles ou abonnement
Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année.
En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?
► Montre ce que tu fais pour trouver la réponse.
Espace pour ta démarche et tes opérations :
N-M: 6, 12, 18, 60, 108
M: 39, 43, 47, 75, 107
NBR. ENT.: 1, 2, 3, 10, 18
Réponse : Jean devra aller au minimum ....18...... fois à la piscine pour que l'achat d'un abonnement soit plus avantageux.</td>
</tr>
</table>

- **Procédure correcte** : procédure permettant de résoudre le problème posé au prix d'un travail souvent plus laborieux et comportant plus d'étapes ; elle est parfois préférée par les élèves, car elles ou ils s'y sentent plus à l'aise.

Figure 11 : Exemple de production d'élève issue du test pilote 2023, illustrant la procédure correcte pour le problème « Entrées individuelles ou abonnement »

Recherche exhaustive du prix pour 1, 2, 3… 18 entrées, sans et avec abonnement.



4) Entrées individuelles ou abonnement

Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année.

En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?

► Montre ce que tu fais pour trouver la réponse.

Espace pour ta démarche et tes opérations :

X abo	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102
avec abo	39	43	47	51	55	59	63	67	71	75	79	83	87	91	95	99	103

108
107

Réponse : Jean devra aller au minimum ...18........... fois à la piscine pour que l'achat d'un abonnement soit plus avantageux.

- **Procédure incorrecte** : raisonnement incorrect de l'élève qui met en œuvre une procédure comportant par exemple des erreurs de mathématisation (autrement dit une traduction incorrecte de la situation en langage mathématique[13]), ou de représentation[14] du problème.

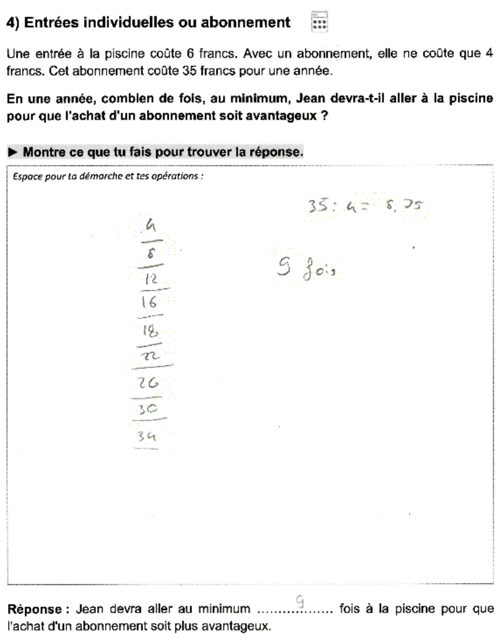
---

[13] Nous faisons référence ici à la progression des apprentissages du PER, pour l'objectif MSN23 : « Résolution de problèmes numériques en lien avec les ensembles de nombres travaillés, l'écriture de ces nombres et les opérations étudiées, notamment : […] traduction des données d'un problème en opérations arithmétiques, en utilisant au besoin des parenthèses : additions, soustractions, multiplications et divisions » (CIIP, 2010). Voir également Yvain-Prébiski (2021).

[14] Pour Julo (1995), la représentation d'un problème correspond à l'activité mentale qui permet de traiter les informations à disposition pour leur donner du sens et construire une compréhension de la situation.

Figure 12 : Exemples de productions d'élèves issues du test pilote 2023, illustrant des procédures incorrectes pour le problème « Entrées individuelles ou abonnement »

Division du prix de l'abonnement par le prix d'une entrée (avec ou sans abonnement).	Calcul du prix avec et sans abonnement pour 365 jours.



4) Entrées individuelles ou abonnement

Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année.

En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?

► Montre ce que tu fais pour trouver la réponse.

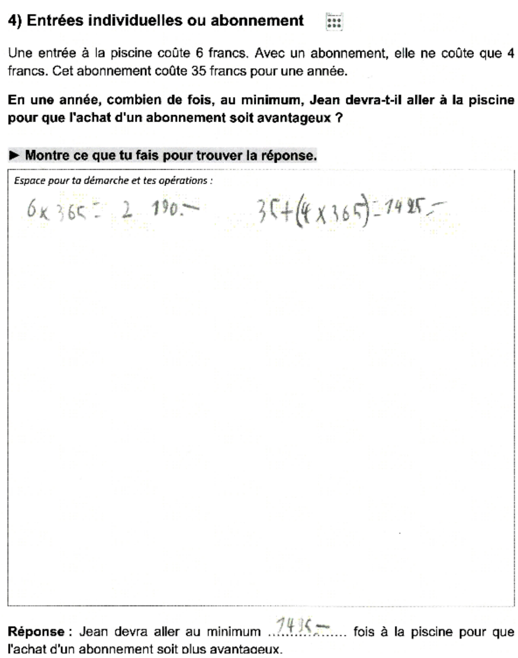
Espace pour ta démarche et tes opérations :

35 : 4 = 8,75

4
8
12
16
18
22
26
30
34

9 fois

Réponse : Jean devra aller au minimum ...9... fois à la piscine pour que l'achat d'un abonnement soit plus avantageux.



4) Entrées individuelles ou abonnement

Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année.

En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?

► Montre ce que tu fais pour trouver la réponse.

Espace pour ta démarche et tes opérations :

6x365 = 2 190.–   35+(4x365) = 1495.–

Réponse : Jean devra aller au minimum ...1495.–... fois à la piscine pour que l'achat d'un abonnement soit plus avantageux.

- **Procédure incomplète** : procédure non aboutie (mais début correct).

Figure 13 : Exemple de production d'élève issue du test pilote 2023, illustrant une procédure incomplète pour le problème « Entrées individuelles ou abonnement »

Calcul du coût sans et avec abonnement pour quelques entrées.

4) Entrées individuelles ou abonnement

Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année.

En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?

► Montre ce que tu fais pour trouver la réponse.

Espace pour ta démarche et tes opérations :

10 X avec abonement = 75
16 X Sans abonement = 60.-

15X avec abonoment = 15X4 = 60+35 = 95
15 X Sans abonement = 95x6 = 96

13x6 = 78

Réponse : Jean devra aller au minimum ..11............ fois à la piscine pour que l'achat d'un abonnement soit plus avantageux.

- **Procédure absente** : une réponse a été donnée par l'élève, mais il n'y a pas de trace pour la soutenir.

Figure 14 : Exemples de productions d'élèves issues du test pilote 2023, illustrant la procédure absente pour le problème « Entrées individuelles ou abonnement »

10) Entrées individuelles ou abonnement

Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année.

En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?

► Montre ce que tu fais pour trouver la réponse.

Espace pour ta démarche et tes opérations :

Réponse : Jean devra aller au minimum ....18......... fois à la piscine pour que l'achat d'un abonnement soit plus avantageux.

10) Entrées individuelles ou abonnement

Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année.

En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?

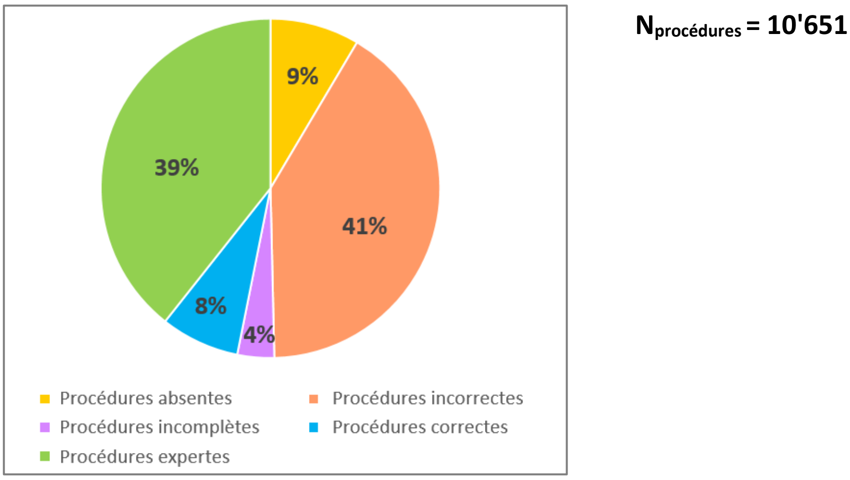
► Montre ce que tu fais pour trouver la réponse.

Espace pour ta démarche et tes opérations :

Réponse : Jean devra aller au minimum ...22.......... fois à la piscine pour que l'achat d'un abonnement soit plus avantageux.

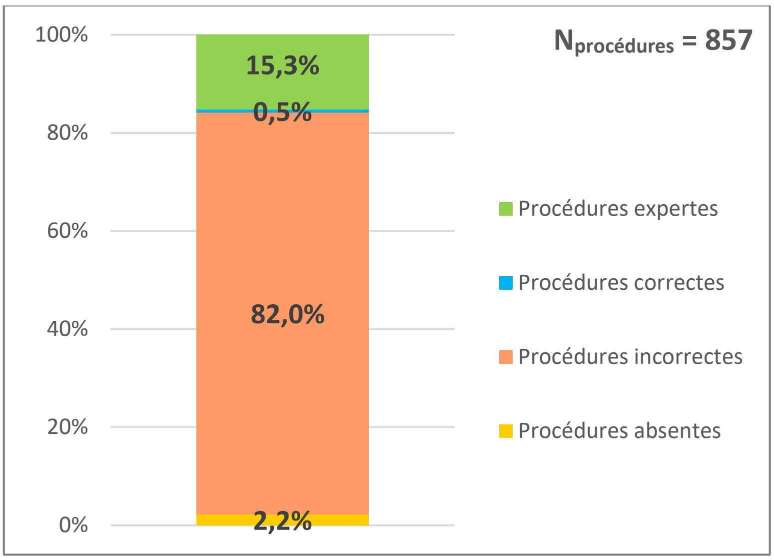
Dans le traitement des données mené suite au test pilote de 2023, nous avons analysé les productions de chaque élève, en nous intéressant à l'ensemble de la procédure de l'élève pour en identifier le type. La répartition de ces procédures, tous problèmes confondus (figure 15), révèle que celles qui s'observent le plus fréquemment sont les **procédures incorrectes** (41%), suivies par les **procédures expertes** (39%). Quant aux **procédures correctes**, elles ont été mises en œuvre dans 8% des cas, mais il convient de préciser que ce type de procédures ne s'appliquait pas à tous les problèmes. Seuls 9% des élèves ont soumis une réponse sans laisser de trace (**procédure absente**) et 4% des élèves ne sont pas arrivés au bout de leur raisonnement (**procédure incomplète**).

Figure 15 : Répartition globale des procédures mises en œuvre par les élèves sur l'ensemble des problèmes du test pilote pour lesquels la calculatrice était à disposition



Ce constat nous a amenées à questionner la qualité des tâches soumises aux élèves : le taux élevé de procédures incorrectes est-il la conséquence de problèmes inadéquats ? Dans le but d'identifier l'origine de ces nombreuses procédures incorrectes, nous avons analysé plus en détail les problèmes avec les plus hauts taux de ce type de procédures. Par exemple, la tâche d'ancrage intitulée « Parties de billes », réalisée par 857 élèves et pour laquelle les élèves avaient la calculatrice à disposition, présentait 82% de procédures incorrectes (figure 16).

Figure 16 : Répartition des procédures mises en œuvre pour le problème « Parties de billes »

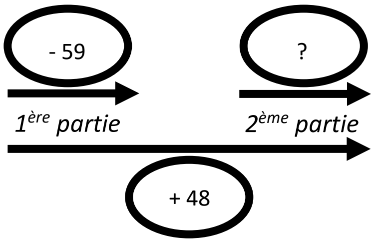


Il s'agissait du problème additif de type TTT (**t**ransformation, **t**ransformation, **t**ransformation)[15], selon la typologie de Vergnaud (1990), suivant :

> **Parties de billes**
>
> Des élèves jouent aux billes à l'école. Max a perdu 59 billes à la première partie. Il joue ensuite une deuxième partie. Après les deux parties, il a gagné 48 billes.
>
> **Que s'est-il passé lors de la deuxième partie ?**

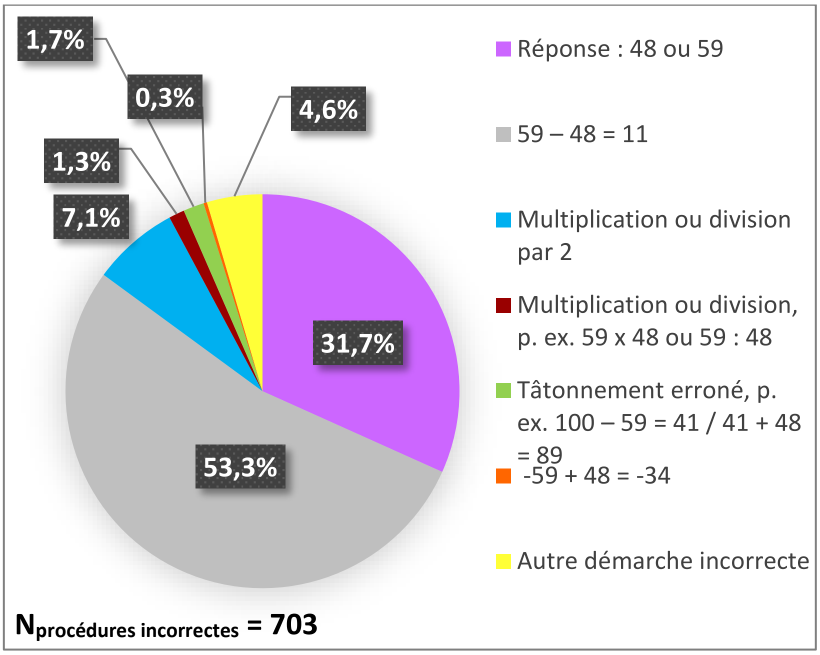
La situation en jeu peut être représentée au moyen d'un schéma (figure 17) et modélisée par le calcul : 59 + 48 = 107.

Figure 17 : Schématisation de la situation du problème « Parties de billes »



Intéressons-nous à présent plus en détail aux 703 procédures incorrectes mises en œuvre par les élèves et dont la répartition est présentée dans la figure ci-dessous.

Figure 18 : Répartition des procédures incorrectes mises en œuvre pour le problème « Parties de billes »



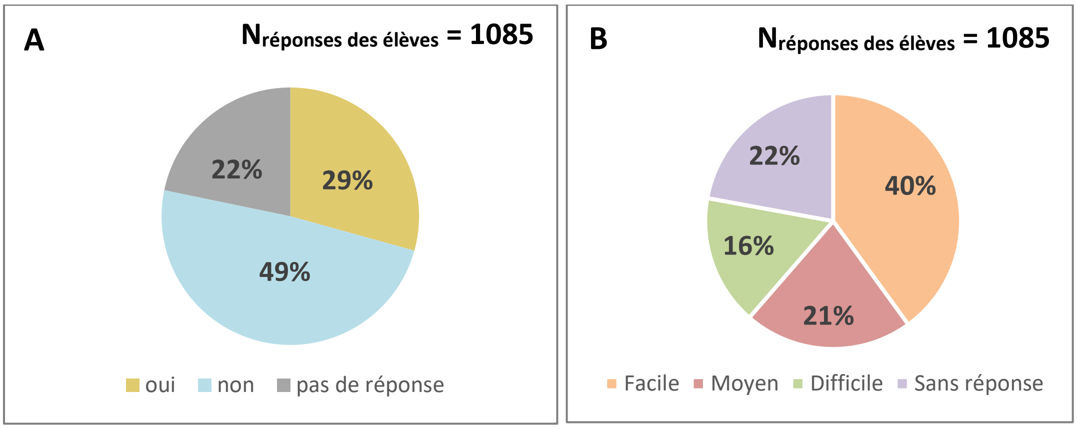
[15] Les problèmes de type TTT sont des problèmes additifs où « deux transformations se composent pour donner une transformation. Exemple : Séraphine a gagné 6 billes, puis elle en a perdu 9. En tout, elle en a perdu 3 » (*cf.* lexique du domaine MSN, CIIP, 2010/2024).

La majorité de ces procédures, soit plus de la moitié (53,3%), correspond aux cas où l'élève a réalisé la soustraction 59 – 48 = 11 et répondu que Max a perdu 11 billes. Autrement dit, il s'agit d'une mauvaise représentation de la situation, aboutissant par conséquent à une mathématisation incorrecte du problème. Il s'agit ici d'une erreur assez caractéristique des problèmes de type TTT. La deuxième procédure incorrecte la plus fréquente, mise en œuvre par près d'un tiers des élèves (31,7%), renvoie aux réponses « 48 » ou « 59 ». À nouveau, c'est une compréhension erronée de la situation, certainement due à un manque d'habitude à résoudre des problèmes de ce type, qui conduit à une telle procédure.

Ces deux premières catégories de procédures incorrectes, qui correspondent à 84% de l'ensemble des procédures de ce type, sont les plus caractéristiques de ce problème, et également des problèmes de type TTT considérés comme les plus difficiles parmi les problèmes additifs. Elles sont révélatrices des difficultés de représentation qu'engendrent ces situations auxquelles les élèves ne devraient être confrontés, selon le PER (2010), qu'à partir de la 7e année. Les autres procédures incorrectes sont nettement moins fréquentes. En font partie celles où l'élève réalise une multiplication ou une division par 2, par exemple 48 : 2 = 24 (7,1%), étant donné que deux parties sont jouées. Ici, c'est la mathématisation du problème qui est totalement erronée.

Ces résultats nous amènent à la conclusion que ce n'est pas le problème soumis aux élèves qui n'est pas adapté, les principales procédures incorrectes observées correspondant à des erreurs caractéristiques, anticipées à priori. Ils semblent montrer que les élèves sont potentiellement plus habitués aux autres types de problèmes additifs. En effet, le haut taux de procédures incorrectes ne semble pas à imputer à cette tâche spécifique, mais plutôt à un type de problèmes (TTT) potentiellement insuffisamment travaillés et entrainés en classe. D'ailleurs, près de la moitié des élèves (49%) dit ne pas être habituée à réaliser ce type de tâches (figure 19), alors qu'il s'agit clairement d'une compétence attendue par le PER (2010).

Figure 19 : Habitude déclarée (A) et difficulté perçue (B) des élèves pour le problème « Parties de billes »



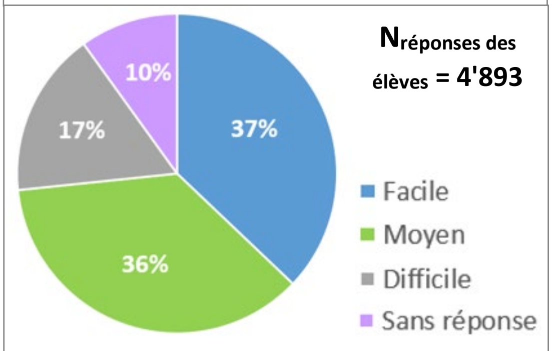
Nous nous sommes également intéressées à la perception de difficulté des élèves pour ce problème (figure 19). Les données montrent qu'une majorité le trouve facile (40%, soit deux élèves sur cinq). Plus d'un cinquième des élèves (21%) l'a estimé de difficulté moyenne et seul un élève sur six le perçoit comme étant difficile. Ces résultats semblent mettre en avant que les élèves n'appréhendent pas adéquatement la difficulté de ce problème au vu du taux très important de procédures incorrectes (plus de quatre élèves sur cinq), comme si, pour eux, leur raisonnement était correct.

3.2.3 Influence de la calculatrice

Le test pilote de 2023 visait également à étudier comment la manière de résoudre un problème est influencée par le fait d'avoir la calculatrice à disposition, soit l'impact de l'usage de cet outil de calcul sur le raisonnement des élèves. Dans cette partie, nous présentons les résultats obtenus pour les six problèmes (*cf.* annexe 6.1 pour les énoncés des problèmes) ayant été soumis à certains élèves avec la calculatrice et à d'autres sans.

La sélection des problèmes numériques choisis s'explique par le fait que, si les élèves sont tout à fait capables d'effectuer les opérations par écrit, l'aspect calculatoire n'est pas l'enjeu de ces six tâches. Ainsi, la calculatrice a pour rôle, en tant qu'outil de calcul, de décharger l'élève de cette partie de manière à ce qu'elle ou il puisse se focaliser sur l'aspect « raisonnement ». Comme le rappelle Tricot (2020, p. 26), en situation de *résolution de problèmes*, la mise à disposition de la calculatrice permet de « [déléguer] certains calculs à la machine, pas la résolution de problème elle-même ». Ainsi, la représentation de la situation, au sens de Julo (1995), et sa modélisation – soit le choix de l'opération – restent entièrement à la charge de l'élève.

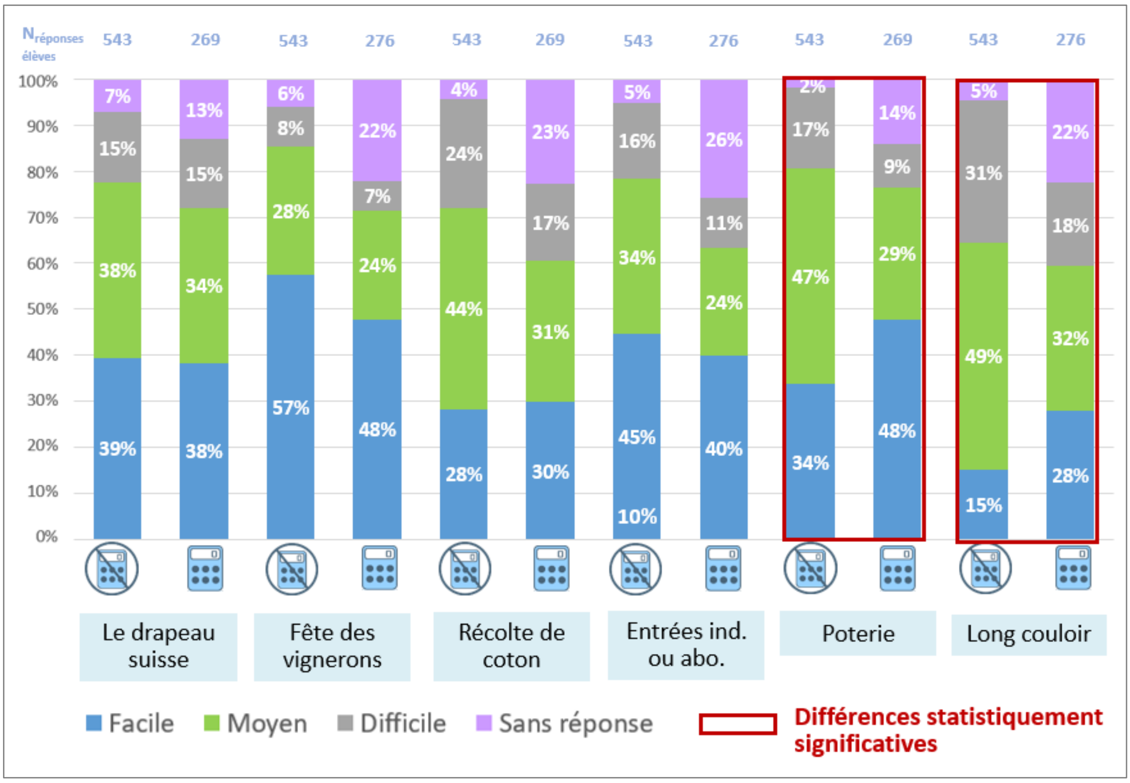
Figure 20 : Difficulté perçue par les élèves pour les six problèmes, indépendamment de la disponibilité de la calculatrice



Dans un premier temps, nous nous sommes penchées sur la perception de la difficulté des différents problèmes du point de vue des élèves. Tous problèmes confondus, plus de 70% des élèves les considèrent comme étant de difficulté plutôt facile à moyenne (figure 20).

Pour aller plus loin, nous avons analysé le lien entre la disponibilité de la calculatrice et la difficulté perçue, autrement dit si le fait d'avoir ou non la calculatrice à disposition impacte la perception de la difficulté. Nos résultats ont montré que pour seulement deux des six problèmes (« Poterie » et « Long couloir »), la différence en termes de difficulté perçue est statistiquement significative selon que la calculatrice est ou non disponible (figure 21). En d'autres termes, la plupart des problèmes ne semblent pas plus faciles pour les élèves si elles et ils ont la calculatrice à disposition.

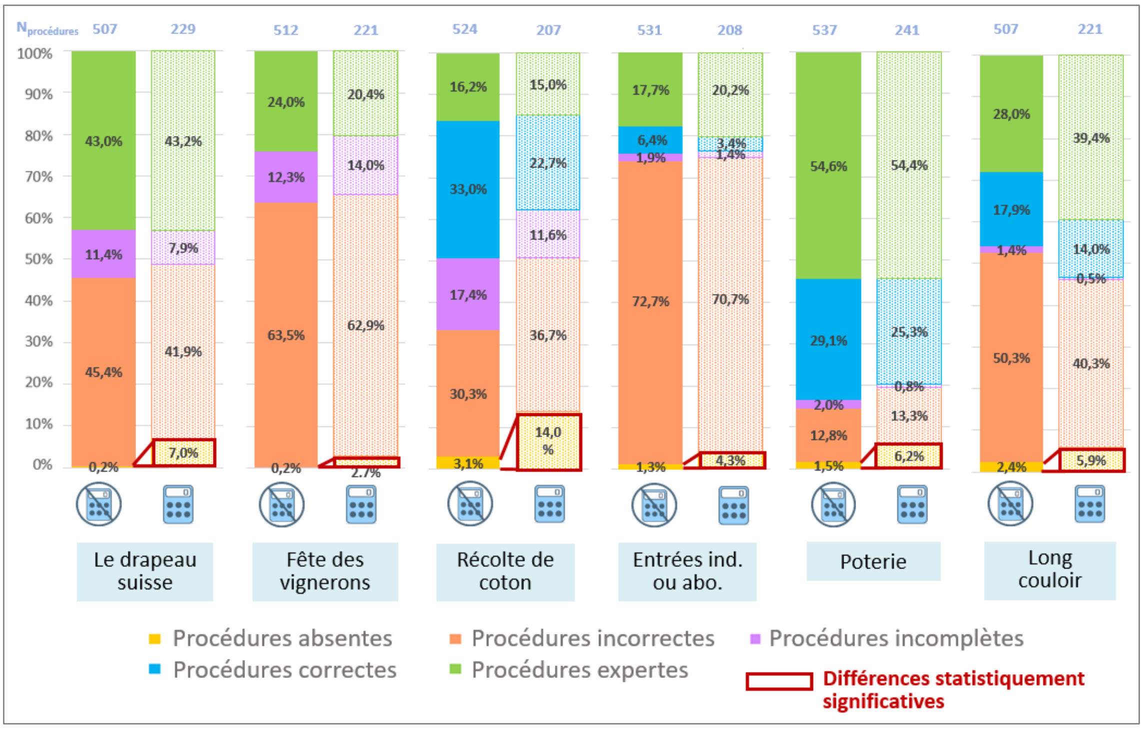
Figure 21 : Difficulté perçue par les élèves pour les six problèmes, en fonction de la disponibilité de la calculatrice



Dans un second temps, nous avons analysé les procédures mises en œuvre par les élèves selon que la calculatrice était ou non autorisée. Leur répartition pour chaque problème est présentée dans la figure 22 ci-dessous. Précisons que si les élèves avaient la calculatrice à disposition dans certains cas, nous ne pouvons être sures qu'elles et ils l'aient effectivement utilisée pour résoudre le problème. De fait, durant le codage des tâches, nous avons relevé à plusieurs reprises des traces de calculs posés et effectués à la main[16], alors que l'élève pouvait disposer de la machine.

[16] Dans les traces laissées par les élèves, les retenues inscrites sur les opérations posées en colonnes montrent bien que les calculs ont été réalisés par écrit et non à l'aide de la calculatrice.

Figure 22 : Répartition des procédures mises en œuvre pour chacun des six problèmes, en fonction de la disponibilité de la calculatrice (Ruf et al., en préparation). En rouge, des différences statistiquement significatives (seuil de 5%) entre les procédures en fonction de la disponibilité de la calculatrice.



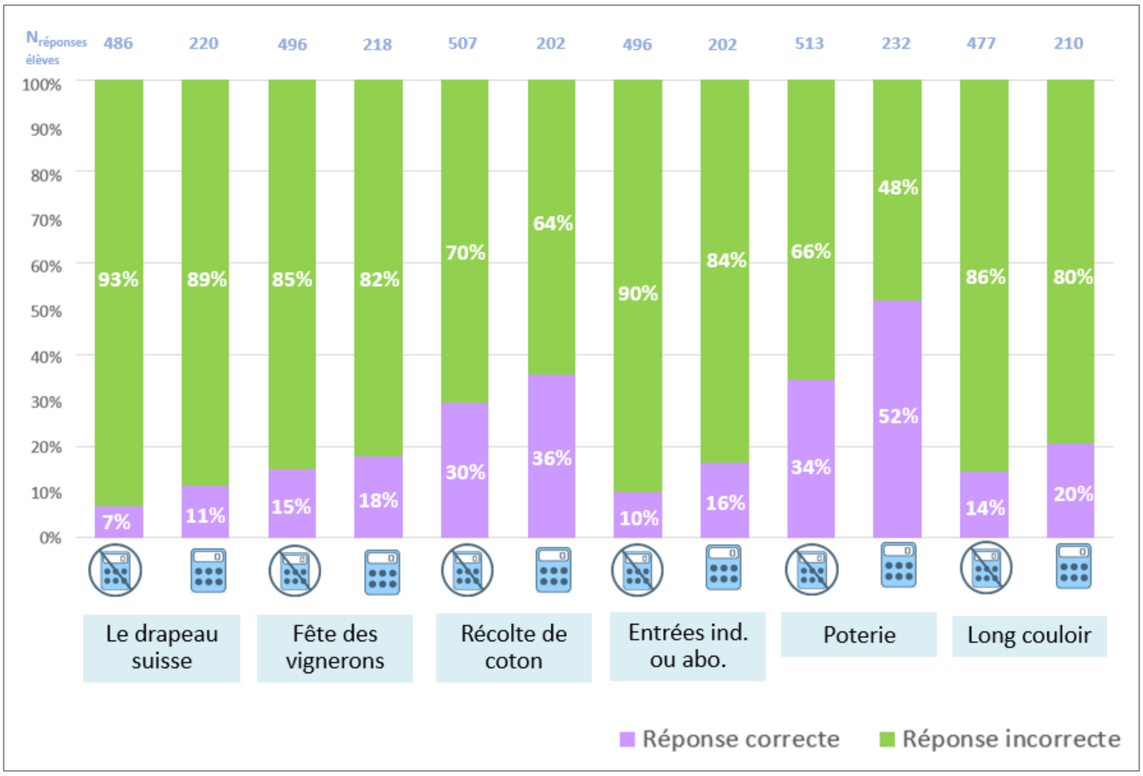
De manière générale, pour les six problèmes, les taux des différents types de procédures sont assez stables, que la calculatrice ait été autorisée ou non. La plus nette différence s'observe pour les procédures absentes, correspondant aux cas où une réponse a été donnée par l'élève, mais il n'y a pas de trace pour la soutenir. L'absence de trace augmente considérablement lorsque les élèves peuvent utiliser la machine, et la différence est statistiquement significative pour les six problèmes (figures 22 et 23). Cela n'est pas si surprenant et semble montrer que les élèves ne sont pas habitués à garder trace de leurs calculs, d'autant plus lorsqu'ils sont réalisés à la calculatrice (Ruf et al., en préparation). Si d'autres variations sont constatées en termes de procédures mises en œuvre selon la disponibilité de la machine, la différence n'est toutefois pas significative. La fréquence des procédures incomplètes a néanmoins tendance à diminuer avec la calculatrice pour cinq des six problèmes (figure 23). Une hypothèse revient à dire que le fait de décharger l'élève de la partie calculatoire l'aide à aller au bout de son raisonnement, que celui-ci soit correct ou non. Par ailleurs, la disponibilité de la machine tend à faire diminuer quelque peu le taux de procédures correctes, pour les problèmes pour lesquels ce type de procédure existe. Nous nous demandons si ce constat pourrait s'expliquer par le fait que l'élève, insuffisamment à l'aise avec la réalisation de l'algorithme « expert », opte pour une autre procédure permettant de contourner cette opération lorsque la calculatrice n'est pas autorisée.

Figure 23 : Différences du taux de procédures (avec et sans calculatrice) et de réussite pour chaque problème selon la disponibilité de la calculatrice. Les différences statistiquement significatives sont mises en évidence par les encadrés rouges.

	Procédures absentes	Procédures incorrectes	Procédures incomplètes	Procédures correctes	Procédures expertes	Réussite
Le drapeau suisse	↗ + 6,8%	↘ - 3,5%	↘ - 3,5%		↗ + 0,2%	↗ + 4%
Fête des vignerons	↗ +2,5%	↘ - 0,6%	↗ + 1,7%		↘ - 3,6%	↗ + 3%
Récolte de coton	↗ + 11,0%	↗ + 6,4%	↘ - 5,8%	↘ - 10,3%	↘ - 1,2%	↗ + 6%
Entrées ind. ou abo.	↗ +3,0%	↘ - 2,0%	↘ - 0,5%	↘ - 3,0%	↗ - 2,5%	↗ + 6%
Poterie	↗ + 4,7%	↗ + 0,5%	↘ - 1,2%	↘ - 3,8%	↘ - 0,2%	↗ + 18%
Long couloir	↗ +3,5%	↘ - 10,0%	↘ - 0,9%	↘ - 3,9%	↗ + 11,4%	↗ + 6%

Intéressons-nous à présent plus spécifiquement à la différence de réussite pour chaque problème (figures 23 et 24). Pour la tâche « **Le drapeau suisse** », bien que le taux de réussite augmente significativement lorsque l'outil de calcul est autorisé, la différence reste faible et semble montrer que l'effet « calculatrice » est négligeable pour ce problème. Il semblerait que les élèves s'égarent dans les calculs à réaliser même lorsque la calculatrice est disponible. Quant à la tâche « **Fête des Vignerons** », la différence en termes de réussite n'est pas significative selon que l'élève peut ou non disposer de la calculatrice (le taux de réussite est de 15% sans la calculatrice, respectivement de 18% avec). Ce constat nous amène à penser que ce problème devrait plutôt être proposé aux élèves sans la calculatrice, pour évaluer leurs compétences après un travail autour des additions et des soustractions, étant donné que ce ne sont visiblement pas les calculs qui leur posent problème. Pour ce qui est de la tâche « **Récolte de coton** », la mise à disposition de la calculatrice augmente le taux des réponses correctes : ce dernier passe de 30% à 36%. Cette différence n'est toutefois pas significative. Elle l'est en revanche pour le problème « **Entrées individuelles ou abonnement** », pour lequel le taux de réussite s'accroit de 6% avec la machine, passant de 10% à 16%. Cette augmentation plaide plutôt en faveur de l'utilisation de la calculatrice pour cette tâche. En ce qui concerne le problème « **Poterie** », les résultats montrent que la réussite est nettement meilleure et la différence est significative lorsque les élèves peuvent utiliser la machine : le taux de réponses correctes passe de 34% à 52%. Une hypothèse qui permettrait d'expliquer cette importante différence de taux de réussite est que la division est une opération complexe à effectuer, difficulté que la calculatrice permet d'effacer. Enfin, pour le problème « **Long couloir** », la différence des taux est également significativement plus grande lorsque la calculatrice est autorisée : 14% des élèves parviennent à la réponse attendue lorsqu'ils n'ont pas la machine à disposition, contre 20% lorsqu'elle est disponible.

Figure 24 : Taux de réussite des six tâches réalisées avec ou sans la calculatrice



Les différences observées au niveau des procédures expertes selon la disponibilité de la calculatrice ne sont pas constantes : leur taux de mise en œuvre augmente pour trois problèmes, et diminue pour les trois autres. Elles ne sont vraisemblablement pas en lien avec la possibilité d'utiliser ou non la calculatrice, mais plutôt avec le problème proposé.

3.3 Constats et perspectives

Les résultats obtenus montrent que, pour tous les problèmes papier-crayon confondus, qu'ils aient été effectués avec ou sans la calculatrice à disposition et indépendamment de la réponse soumise, nombreuses sont les **procédures incorrectes** mises en œuvre par les élèves romands de fin de 8e année. Ces résultats mettent en avant la difficulté de la *résolution de problèmes*. En effet, la mathématisation de certaines situations proposées a posé problème à une grande partie d'élèves, lesquels ont suivi un raisonnement erroné. L'important taux de procédures incorrectes observées sur l'ensemble du test pilote peut s'expliquer par le fait que la résolution de problèmes est une tâche difficile, par l'essence même de la nature du problème. Il en ressort qu'un travail et un entrainement considérables, en situation d'enseignement-apprentissage, sont nécessaires avant qu'une telle compétence soit maitrisée.

La calculatrice a également fait l'objet d'une étude dans nos travaux. En ce qui concerne son usage dans les classes, les enseignantes et enseignants disent se sentir en majorité à l'aise avec son utilisation. Néanmoins, les réponses soumises montrent que les élèves n'en disposent pas souvent en contexte d'enseignement-apprentissage, et encore moins en situation d'évaluation où elle n'est autorisée dans presque aucune classe. Quant à l'influence de la calculatrice sur le travail des élèves, les résultats obtenus à partir des six problèmes, testés avec et sans cet outil de calcul, montrent que sa mise à disposition a un impact positif sur la réussite de la tâche pour chacun des problèmes. La différence observée n'est toutefois pas significative dans tous les cas.

Au niveau des procédures mises en œuvre par les élèves, nos données montrent que, à l'exception des **procédures absentes**, l'influence de la machine sur le type de procédure mobilisée n'est pas univoque et semble plutôt dépendre du problème proposé (Ruf et al., en préparation).

Si la possibilité d'utiliser la calculatrice en *résolution de problèmes* se révèle avantageuse sur certains aspects, elle présente également des limites. C'est en ce sens que notre travail sur les productions des élèves issues du test pilote nous permettra de proposer un éclairage théorique spécifique à l'usage de cet outil de calcul sur les pages *PistEval* lors de l'évaluation de la *résolution de problèmes*. L'objectif de ce texte sera de rendre attentif le corps enseignant sur les avantages et les limites de la mise à disposition de la calculatrice, dans le but de l'aiguiller dans son choix d'autoriser ou non la machine à ses élèves pour résoudre des problèmes.

4. Évaluer la résolution de problèmes sur support informatisé

Dans le cadre du test pilote 2023, nous avons mené, en parallèle du support papier-crayon, une étude exploratoire concernant la résolution de problèmes mathématiques sur format numérique, au moyen de l'application *ONAÉ* (*cf*. chapitre 1.2.2).

4.1 Plus-values du support informatisé

L'utilisation du support informatisé nous a paru opportune pour évaluer la *résolution de problèmes*, étant donné qu'il permet, grâce à ses fonctionnalités embarquées, de récolter plus de traces des actions des élèves (en comparaison du support papier-crayon) et, ainsi, de mieux observer leur procédure de résolution. Outre la récolte automatique de nombreuses données (Coen, 2022 ; Wyatt-Smith et al., 2019), le support numérique offre d'autres avantages, comme des fonctionnalités embarquées facilitant le travail des élèves (Vermot et al., 2011), la possibilité de donner des *feedbacks* immédiats à l'élève (Tricot, 2020 ; Van Der Kleij et al., 2015) ou encore une correction automatique s'accompagnant d'un gain de temps pour le corps enseignant. De plus, d'autres études, à l'image de celles de Blumenthal et Blumenthal (2020) ou de Paleczek et al. (2021), ont mis en évidence la meilleure motivation des élèves face au support numérique, même si Tricot (2020), dans sa compilation de méta-analyses, arrive au constat que les outils numériques n'ont en moyenne aucun effet sur la motivation scolaire. En outre, dans un contexte de plus en plus connecté et suite à l'inscription de l'Éducation Numérique (ÉN) dans le PER depuis 2021, il existe de nouveaux outils qu'il convient d'exploiter. D'ailleurs, dans le référentiel de compétences pour la formation initiale et continue des enseignantes et enseignants pour le domaine de l'éducation numérique, publié en 2021 par la CIIP, l'évaluation sur support numérique est évoquée comme un potentiel d'innovation en la matière et relève des compétences attendues du corps enseignant (CIIP, 2021).

Si le support informatisé offre de nombreux avantages, des limites sont également à signaler, à l'instar des contraintes en ce qui concerne le format des tâches pouvant être proposées, ainsi que la guidance quant à la procédure pouvant être mise en œuvre à l'aide de ce support. En effet, des restrictions techniques liées au choix des outils numériques développés pour le test pilote de 2023 empêchent certaines procédures de résolution lorsque les tâches sont informatisées (Wang et al., 2007, 2008). La réalisation d'un arbre de classement en est un exemple. Se pose alors la question du statut de la tâche lorsque l'environnement numérique prend (trop) en charge une partie du cadrage : jusqu'où s'agit-il encore d'une *résolution de problèmes*, dans la mesure où une partie de la mobilisation des outils n'est plus à la charge de l'élève ? Lors de la sélection des tâches, nous avons été particulièrement attentives à ce que leur transposition sur support numérique ne les dénature pas de leur statut de *résolution de problèmes*, quand bien même ce support ne permet pas la mise en œuvre de toutes les procédures pouvant être restituées en format papier-crayon. En raison de ces procédures rendues impossibles par le support numérique, nous sommes également conscientes que l'évaluation de la compétence en jeu n'est que partielle.

Ci-après, nous présentons les principales plus-values du numérique exploitées dans le cadre du test pilote de 2023, à savoir les types de **données récoltées** ainsi que la possible intégration d'une **adaptativité** des tâches.

4.1.1 Récolte de données

Le support informatisé, par le biais de l'utilisation de l'application *ONAÉ* (*cf*. chapitre 1.2.2), a permis de recueillir de très nombreuses informations. Outre des données liées aux fonctionnalités numériques, telles que le nombre d'écoutes de l'énoncé ou le temps d'activité[17] passé sur une tâche, l'application a gardé trace de chaque action réalisée par l'élève au cours de son travail (Hoffer, en préparation). Pour les Mathématiques, enregistrer les actions des élèves se révèle spécialement instructif en situation de *résolution de problèmes*. En effet, pour ce type de tâches, la seule réponse fournie par l'élève est insuffisante pour renseigner son niveau de maitrise de la ou des compétences en jeu ; il est indispensable de pouvoir retracer son raisonnement à l'aide des traces laissées lors de son travail (*cf*. chapitre 2.2).

Dans le cadre du test pilote de 2023, le support numérique nous a, par exemple, permis de récolter de nombreuses données grâce à la fonctionnalité de la calculatrice intégrée. En effet, cette dernière permet de garder trace de l'ensemble des calculs effectués par l'élève, calculs qu'elle ou il n'aurait pas forcément transcrits si la tâche avait été proposée en format papier-crayon. Nous avons ensuite utilisé toutes ces données enregistrées par l'application pour automatiser la catégorisation des procédures de résolution mises en œuvre par les élèves lors d'un problème divisif (*cf*. chapitre 3.6.1). Par ailleurs, pour un problème de géométrie demandant de compléter une figure dans un quadrillage de manière à ce que le système d'axes corresponde à ses axes de symétrie, il a été possible, grâce à l'outil numérique, de garder trace des essais effectués par l'élève et, ainsi, de retracer la trajectoire suivie par l'élève au cours de son travail (*cf*. chapitre 3.6.2).

4.1.2 Adaptativité

Nous avons expérimenté une autre plus-value du support numérique, à savoir l'adaptativité permettant des réactions immédiates face aux actions de l'élève, par exemple en lui donnant quelques indications au cours de la réalisation d'une tâche ou en lui proposant une tâche d'approfondissement plus ou moins complexe. Ces messages, différenciés en fonction des difficultés et réussites rencontrées par l'élève, permettent de la ou le guider dans son travail. Selon Hattie (2009), les *feedbacks* immédiats donnés aux élèves au cours de leur travail sont un des principaux facteurs influençant grandement le succès de leurs apprentissages. Un tel retour est très utile en *résolution de problèmes*, où plusieurs compétences sont mobilisées, afin que l'élève ne soit pas bloqué par un obstacle qui n'est pas lié à la compétence de *résolution de problèmes* (non prise en compte d'une donnée de la consigne qui complique la résolution, difficulté calculatoire, etc.). Nous avons différencié deux formes d'adaptativité : les **relances (ou *feedbacks*)** et les **tâches adaptatives** :

- Une **relance (*feedback*)** correspond à un message d'alerte permettant, notamment, de recentrer l'élève sur son travail. Il peut s'agir du rappel d'une ou plusieurs contraintes de l'énoncé, pour rendre l'élève attentif à certains éléments éventuellement omis, afin d'éviter

[17] Le temps d'activité correspond au temps que l'élève a passé à travailler sur le problème. Il ne tient pas compte du temps d'appropriation, à savoir le temps passé par l'élève à prendre connaissance de la situation, mais commence au moment où elle ou il effectue sa première manipulation.

qu'elle ou il ne s'égare dans son travail. Cette aide prend la forme d'un message d'alerte intégré et préprogrammé, qui s'affiche à l'écran suite à une ou plusieurs actions anticipées lors de l'analyse à priori de la tâche. Ainsi, une relance a un ancrage didactique ; elle doit être minutieusement et didactiquement pensée pour représenter une aide pour l'élève. Les relances expérimentées à l'occasion du test pilote ne sont pas centrées sur la réponse donnée par l'élève, mais sur les actions qu'elle ou il a effectuées au cours de son travail. Elles sont présentées dans le chapitre 4.3.3.

- Une **tâche adaptative** comprend une tâche initiale, suivie d'une tâche supplémentaire d'approfondissement plus ou moins complexe en fonction de la réponse donnée par l'élève à la tâche initiale. Il s'agit d'une réaction immédiate de l'application en fonction de la réponse validée par l'élève. Un tel dispositif, didactiquement ancré, vise à évaluer plus finement le niveau de maitrise de la ou des compétences en jeu, l'idée étant d'aller creuser à l'endroit supposé de la difficulté rencontrée par l'élève. Par exemple, à l'élève donnant une réponse incorrecte au problème initial, nous avons proposé une seconde tâche d'application portant sur le même concept mathématique, mais cette fois centrée sur la compétence technique, sans l'aspect « *résolution de problèmes* ». L'intention de cette seconde tâche est de voir si la non-réussite de la tâche initiale est liée à une lacune au niveau d'une notion technique (par exemple la notion d'axes de symétrie), ou si elle est plutôt à imputer à la part de *résolution de problèmes* comprise dans la première tâche. Nous en exposons les résultats dans le chapitre 4.3.4.

4.2 Méthodologie

Étant donné que notre corpus initial, issu des différents cantons (*cf*. chapitre 1.3), ne comprenait pas de tâches de *résolution de problèmes* sur support informatisé, nous avons été amenées à sélectionner des problèmes initialement destinés à un usage papier-crayon, puis à les transposer pour une passation sur un support numérique. Au vu des éléments présentés ci-avant, la sélection des problèmes a été opérée avec soin, toute tâche n'étant pas adaptable au format numérique. Au regard des contraintes liées aux fonctionnalités disponibles sur l'application *ONAÉ*, les problèmes nécessitant le recours à un dessin ou la qualification de résultats intermédiaires (car comportant plusieurs étapes dans leur résolution) ont par exemple été écartés. Par ailleurs, le recours au numérique avait notamment pour objectif de représenter une véritable plus-value ; les informations récoltées pour les problèmes retenus devaient ainsi se révéler plus riches qu'en format papier-crayon pour que la transposition numérique présente un réel intérêt.

Au total, ce sont cinq problèmes qui ont été sélectionnés et adaptés à un environnement informatisé avant d'être soumis aux élèves sur ce support. Étant donné que les problèmes mathématiques étaient proposés aux élèves en deuxième partie de test, à la suite des tâches d'Allemand, il était important de s'assurer que tous soient réalisés par un nombre suffisamment représentatif d'élèves (en fonction du temps restant à disposition). Un système rotatif de soumission des problèmes a donc été prévu. De la sorte, si l'ordre dans lequel les problèmes ont été présentés aux élèves était toujours le même (c'est-à-dire que le problème B succédait toujours au problème A, le problème C au problème B, etc.), le problème de départ variait (un ou une élève commençait par le problème A, un ou une autre par le problème B, etc.).

Par ailleurs, afin de familiariser les élèves avec certaines fonctionnalités numériques nécessaires à la résolution des tâches évaluatives présentes dans le test, la partie informatisée commençait par une courte « prise en main ». Il s'agissait de tâches avec des manipulations simples, comme ajuster le volume du son de la tablette, s'enregistrer, glisser-déposer une étiquette ou effectuer un calcul sur une calculatrice intégrée. Cette partie introductive permettait également de vérifier le niveau de maitrise de ces fonctionnalités par les élèves. L'analyse des données récoltées a montré que les élèves romands de 8e année ne rencontrent pas de difficulté lors de l'exécution de fonctionnalités simples sur une tablette. En effet, la grande majorité des élèves (81%) a passé moins de cinq minutes sur cette partie de prise en main, le temps d'activité effectif sur chacune des quatre tâches testant une fonctionnalité spécifique étant inférieur à une minute (Hoffer, en préparation).

La passation sur support informatisé a permis d'enregistrer l'ensemble des actions des élèves. Les données brutes des problèmes réalisés sur ce support ont été récupérées directement dans un fichier Excel. Ensuite, un tri préalable, suivi de plusieurs méthodes de traitement (*data process*), a été nécessaire afin de rendre ces données exploitables pour les analyses. Les résultats des analyses menées, dont certaines de manière exploratoire, sont présentés dans le chapitre suivant. Nous y exposons notamment le dispositif mis en place en vue d'une catégorisation automatisée des procédures mises en œuvre par les élèves, ainsi que les résultats relatifs à l'adaptativité des tâches.

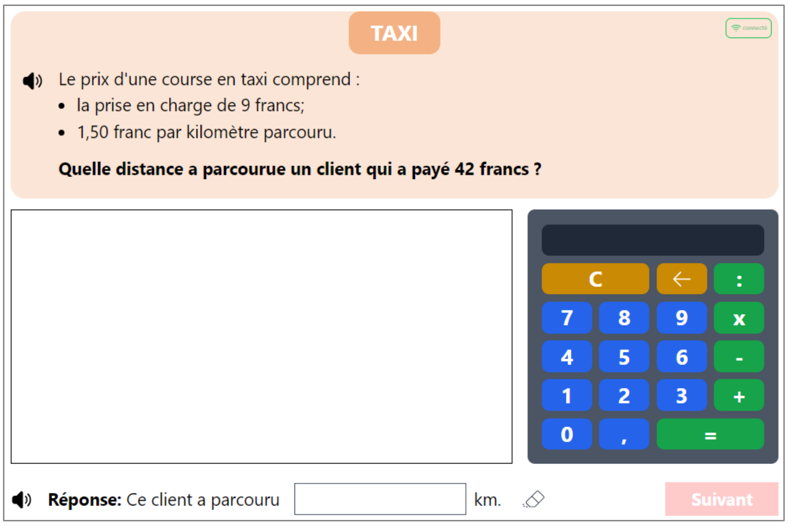
4.3 Principaux résultats

4.3.1 Catégorisation automatisée des procédures de résolution pour un problème mathématique divisif

Dans le cadre de l'expérimentation menée, nous avons exploité les traces laissées par les élèves pour reconstruire les procédures de résolution d'un problème divisif réalisé avec une calculatrice numérique intégrée (Hoffer & Ruf, 2025). Notre objectif était de « classer », de manière automatisée, les procédures mises en œuvre dans différentes catégories préétablies (*cf*. chapitre 3.2.2), lesquelles prenaient également en compte les éventuelles erreurs des élèves. L'utilisation d'une calculatrice intégrée, permettant de garder trace de l'ensemble des calculs effectués et/ou effacés par l'élève, s'est ainsi révélée particulièrement pertinente.

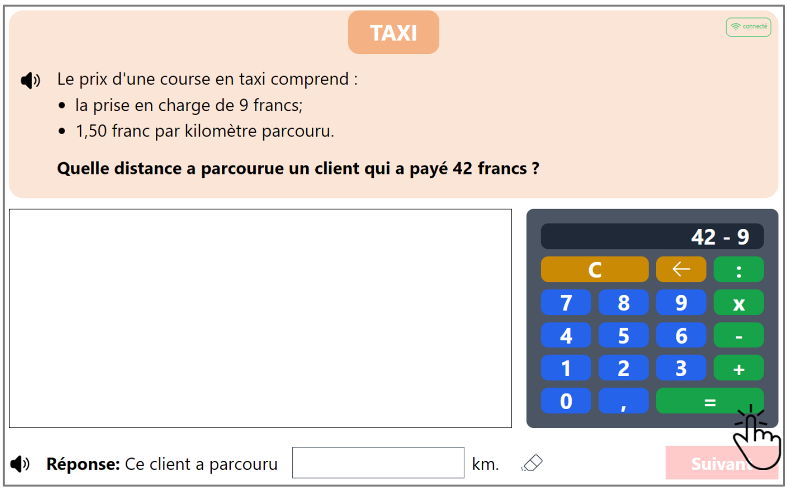
Nous nous sommes intéressées à un problème informatisé intitulé « Taxi ». Ce problème, dont les données récoltées pour les 1077 élèves l'ayant réalisé ont fait l'objet d'une analyse approfondie, se présente de la manière suivante :

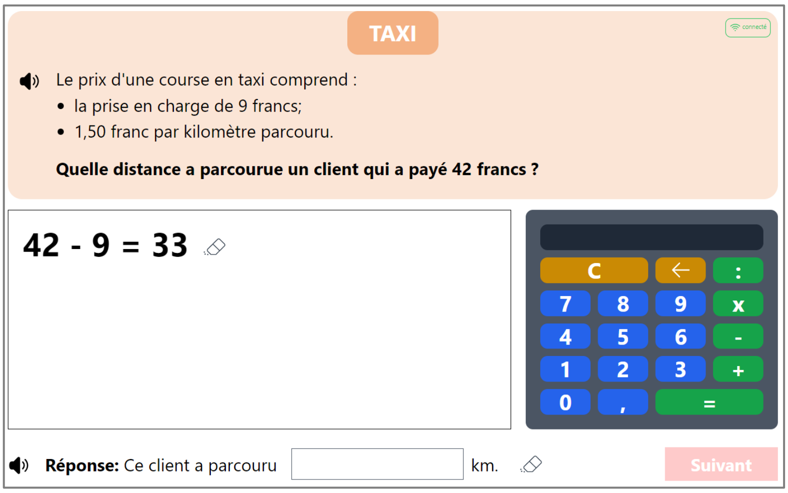
Figure 25 : Problème « Taxi », avec calculatrice embarquée (Hoffer & Ruf, 2025)



Lorsque l'élève inscrit un calcul sur la calculatrice, puis clique sur la touche « = », le calcul et son résultat s'affichent dans l'espace de travail à gauche (figure 26). Si souhaité, l'élève peut effacer le calcul à l'aide de la gomme. Chaque action de l'élève, à savoir « effectuer un calcul » et « effacer un calcul », est enregistrée par l'application, puis traduite sous forme de donnée brute dans un fichier Excel.

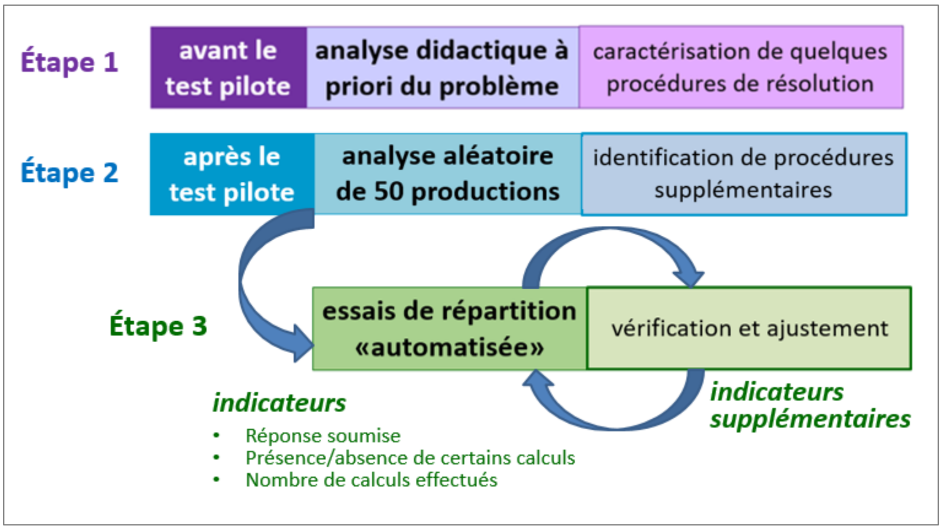
Figure 26 : Utilisation de calculatrice pour le problème Taxi (Hoffer & Ruf, 2025)





Pour analyser les données d'un point de vue didactique, nous avons appliqué une méthodologie en trois phases (Hoffer & Ruf, 2025), décrite ci-après.

Figure 27 : Méthodologie suivie pour catégoriser les procédures des élèves à partir de leurs actions (calculs et réponses), inspirée de Hoffer & Ruf (2025)



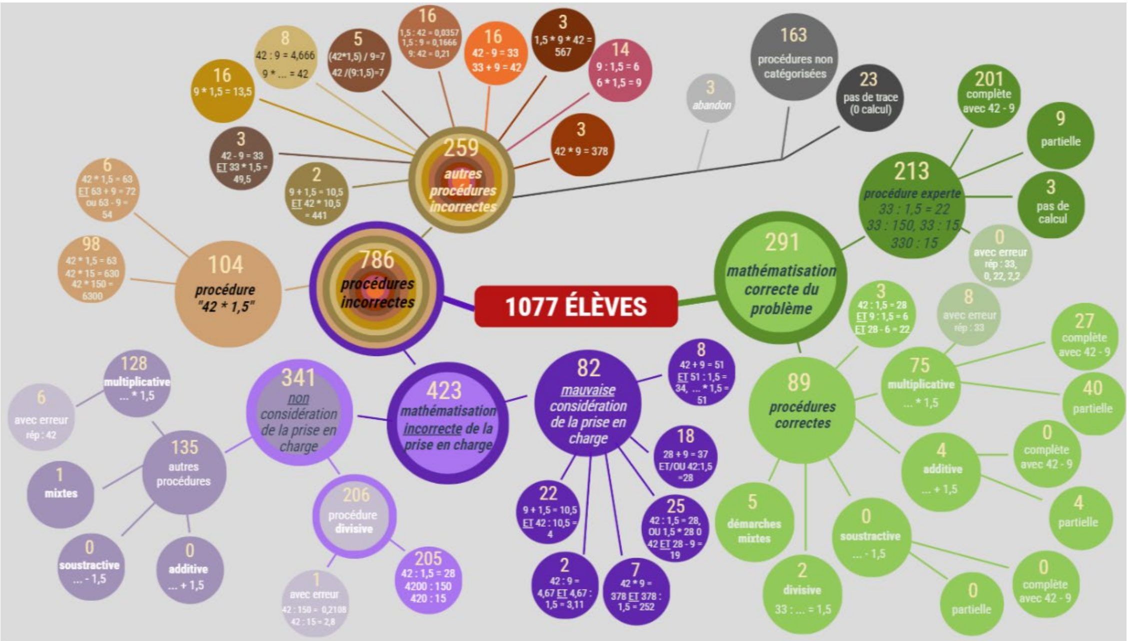
La première étape s'est déroulée en amont du test pilote et a consisté en une **analyse à priori du problème**. Elle a permis de caractériser six procédures de résolution, correctes ou non, auxquelles pouvaient s'ajouter une ou plusieurs erreurs (par exemple une erreur de saisie sur la calculatrice ou de report de la réponse).

Durant la deuxième étape, une **cinquantaine de productions d'élèves, choisies aléatoirement, ont été analysées manuellement**. La procédure mise en œuvre par chaque élève a été reconstruite, étape par étape, sur la base de la séquence de calculs et d'actions répertoriés dans le fichier Excel. L'analyse de ces productions a d'une part permis de retrouver les six procédures identifiées lors de la première étape, et d'autre part de mettre en évidence d'autres procédures qui n'avaient pas été anticipées. Le panorama des procédures établi lors de la première étape a ainsi pu être complété, dans le but qu'il soit aussi exhaustif que possible.

La troisième étape, qui s'est révélée particulièrement complexe, visait à **répartir de manière automatisée les procédures** mises en œuvre par les 1077 élèves ayant résolu ce problème, à partir des calculs enregistrés par l'application. En d'autres termes, il a fallu déterminer des **indicateurs** permettant de caractériser chaque procédure du panorama à l'aide de la réponse et d'un ou plusieurs calculs. Une fois les indicateurs définis, il s'agissait de vérifier qu'ils permettaient effectivement de répartir de manière fiable les productions des élèves dans les catégories procédurales correspondantes, autrement dit de s'assurer non seulement que les productions se retrouvent bien dans une (seule) catégorie, mais aussi que toutes les productions soient associées à la catégorie adéquate. La vérification de la catégorisation des différentes productions nous a amenées à ajuster ou compléter certains indicateurs, afin d'arriver à une catégorisation plus précise. Cette vérification a été réalisée selon un processus cyclique, jusqu'à arriver à une catégorisation automatisée stabilisée.

Grâce aux indicateurs définis manuellement à partir d'une analyse didactiquement ancrée, les données ont pu être agrégées en nouvelles variables et les fréquences de mise en œuvre de chacune des procédures déterminées de manière automatisée. La figure 28 ci-dessous répertorie l'ensemble des procédures identifiées, ainsi que le nombre de fois où elles ont été mises en œuvre (valeurs en jaune clair). Les résultats représentés avec la couleur violette permettent de constater que les frais de prise en charge du taxi ont causé des difficultés à plus du tiers des élèves (39%). Ce taux relativement élevé peut s'expliquer par le fait que la situation est peu familière aux élèves suisses âgés de 11 à 12 ans. Rares sont, selon nous, celles et ceux qui ont l'habitude de prendre un taxi, ce qui expliquerait que beaucoup d'élèves n'ont pas réussi à se représenter correctement la situation. D'autres procédures incorrectes ont également été identifiées (34%), elles sont signalées par les tons bruns. Néanmoins, plus d'un quart des élèves (27%) a résolu le problème en mobilisant une procédure correcte ou experte (secteur vert)[18].

Figure 28 : Carte présentant les fréquences (en jaune clair) de mise en œuvre de l'ensemble des procédures identifiées à priori et à partir des résultats du test pilote de 2023 (Hoffer & Ruf, 2025)



Suite à cette expérimentation, nous constatons que, bien que concluant, le travail réalisé pour automatiser le traitement des données est complexe et chronophage. Par ailleurs, les analyses réalisées sont valables seulement pour le problème concerné ; les indicateurs définis nous semblent intrinsèquement liés à la tâche elle-même et difficilement transférables à d'autres tâches complexes du même type (Hakem et al., 2005). Toutefois, nous constatons que ce travail en vaut la peine et qu'il permet d'observer si l'élève a reconnu la situation (additive et) multiplicative et à identifier l'origine des éventuelles difficultés rencontrées. S'il s'est avéré chronophage tel que nous l'avons mené, il peut être allégé en ajustant le niveau de finesse des analyses. Peut-être n'est-il pas nécessaire d'arriver au détail de la figure 28 ; identifier les plus grandes catégories de procédures de manière automatisée pourrait suffire et permettrait déjà un gain de temps considérable en termes d'analyse des productions. Le but serait d'obtenir, à terme, un outil efficace qui puisse soutenir le corps enseignant en termes de « diagnostic » ou d'évaluation formative du niveau de maitrise des compétences testées.

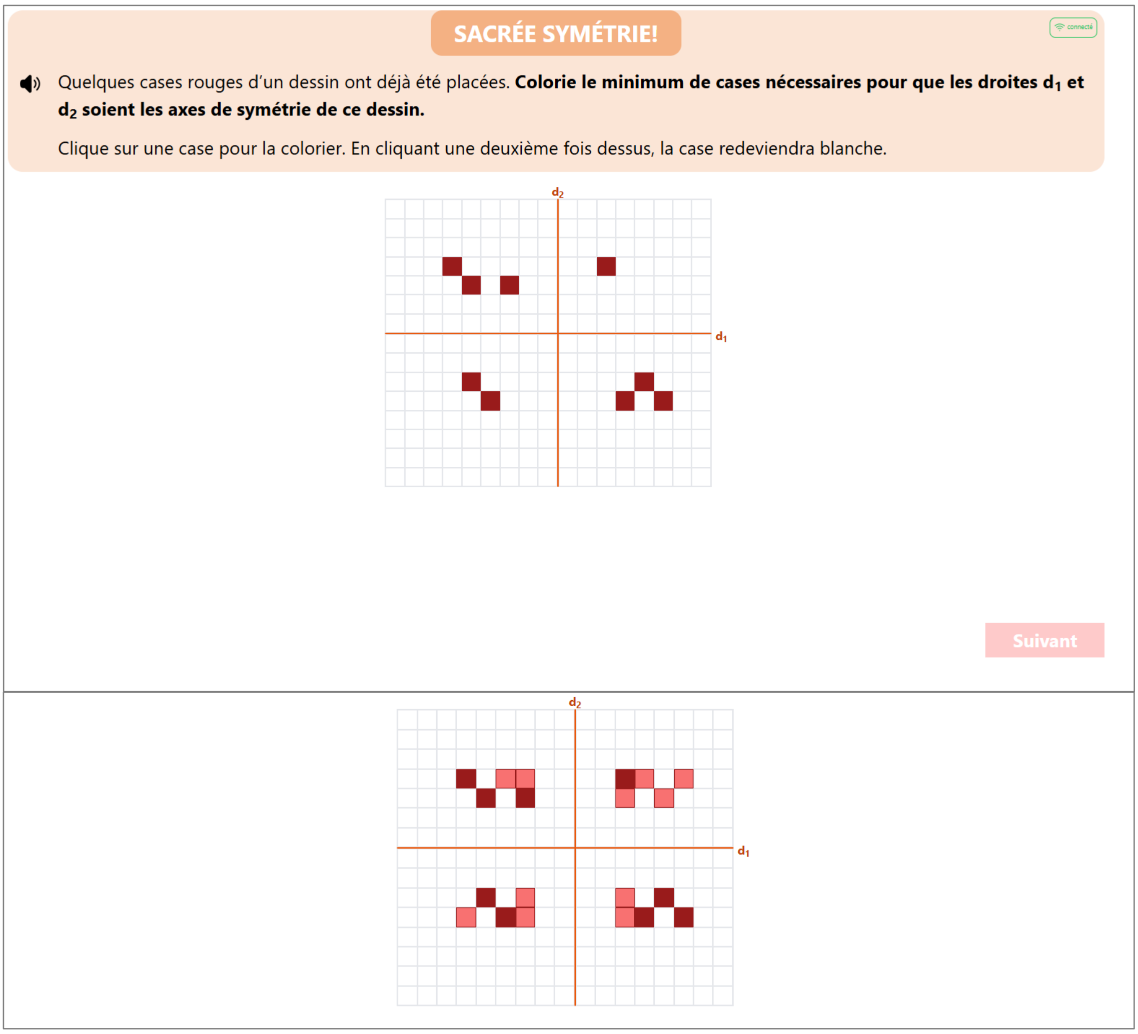
4.3.2 Répartition algorithmique des procédures de résolution vs répartition didactique

Un autre exemple de l'utilisation des données enregistrées par le support numérique lors de la *résolution de problèmes* est illustré par un problème de symétrie axiale, intitulé « Sacrée

[18] Des résultats plus détaillés sont disponibles dans l'article publié par Hoffer et Ruf (2025).

symétrie ! » (figure 29). Il se présente sous la forme d'un quadrillage comportant un système d'axes orthonormé, ainsi que, dans chaque quadrant, des parties d'une figure. Dans cette tâche, il est demandé à l'élève de compléter la figure de manière à ce que les deux axes (d_1 et d_2) correspondent aux axes de symétrie du dessin, et ce en coloriant le moins de cases possible. Cette deuxième contrainte a toute son importance, car l'élève pourrait colorier des cases « en trop », pour obtenir par exemple quatre rectangles symétriques horizontalement et verticalement qui respecteraient effectivement la première contrainte. Du point de vue de la réalisation de la tâche, lorsque l'élève clique sur une case du quadrillage, elle se colorie. En cliquant à nouveau sur la même case, la couleur disparait.

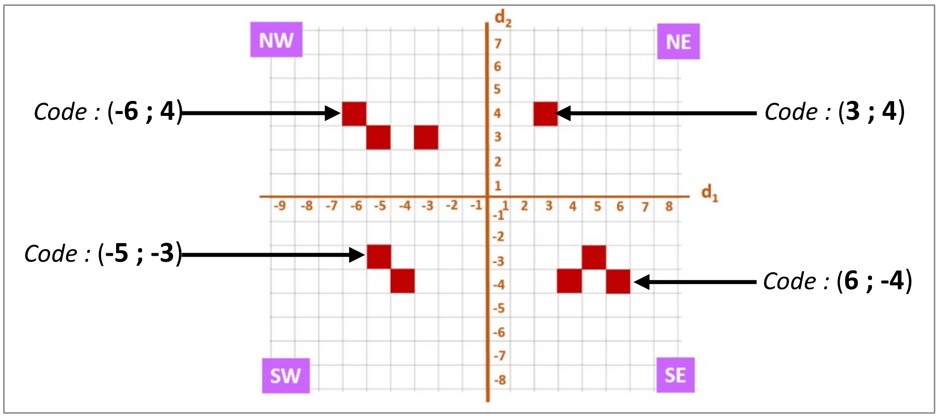
Figure 29 : Problème « Sacrée symétrie ! », tel qu'il était présenté aux élèves (en haut) et son corrigé (en bas)



1073 élèves ayant travaillé sur ce problème, une analyse manuelle de leurs productions se serait avérée fastidieuse et chronophage. De plus, la seule réponse soumise par l'élève ne rend pas forcément compte de son raisonnement préalable (restent dans la production finale uniquement les cases que l'élève a choisi de conserver coloriées, ses essais n'étant pas visibles). En ce sens, l'utilité de l'outil numérique est indéniable pour ce problème. Premièrement, il permet **de garder trace des essais effectués par l'élève de manière chronologique** en plus de la réponse soumise. La trajectoire qu'elle ou il a suivie peut ainsi être prise en compte lors de l'analyse. Deuxièmement, il permet une automatisation de **la catégorisation des productions, et ainsi d'alléger le travail de correction**.

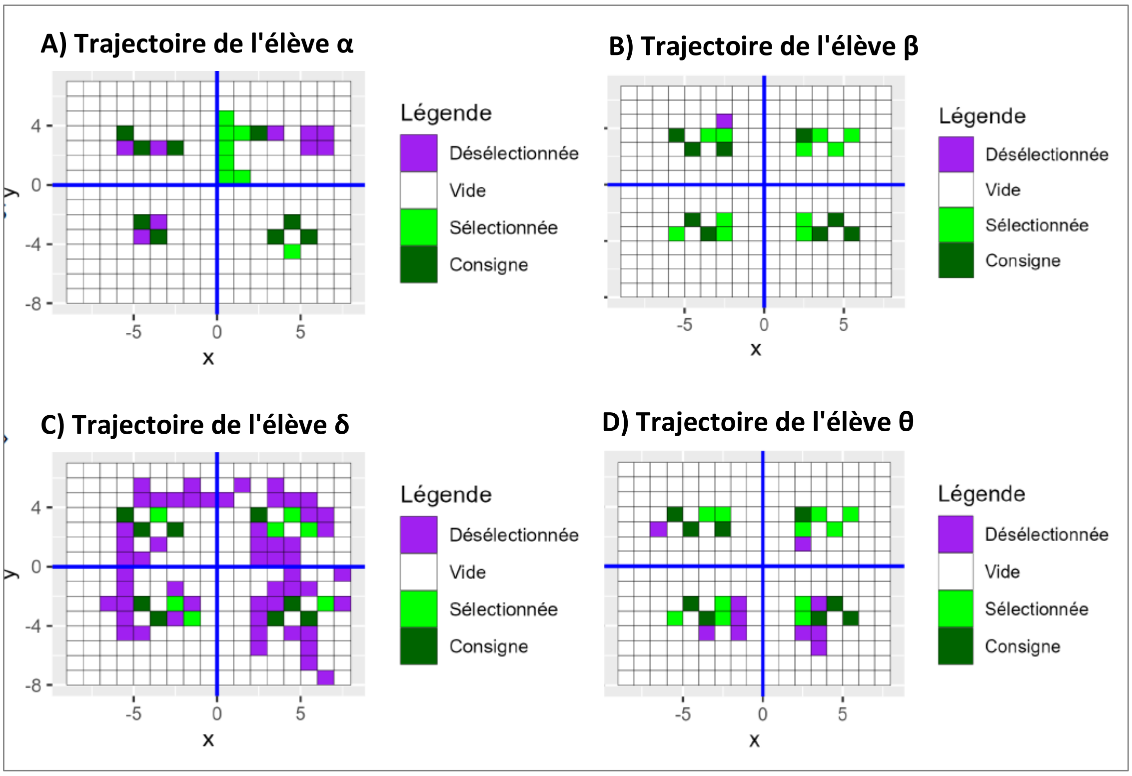
Afin de pouvoir tirer profit de ces avantages, il s'est avéré nécessaire de trouver un moyen pour identifier numériquement les cases (dé)coloriées par les élèves. Pour ce faire, chaque case du quadrillage a été associée à ses coordonnées du système d'axes, qui ont fait office de code (*cf*. exemples présentés dans la figure 30). Grâce à ce codage, la trajectoire du travail de l'élève, jusqu'au résultat final, a pu être retracée.

Figure 30 : Exemples de codes de différentes cases indiquées par les flèches



Ces trajectoires ont fait l'objet d'une analyse automatique, au moyen d'un algorithme, par nos collègues de l'IRDP, experts et expertes en statistique. Pour ce faire, l'équipe a tout d'abord incorporé les essais des élèves dans les images des réponses soumises, créant ainsi des « **images augmentées** », qui intègrent les cases décoloriées au résultat final (figure 31).

Figure 31 : Quatre exemples d'images augmentées montrant en vert foncé les cases coloriées initialement, en violet les cases coloriées puis décoloriées par l'élève (soit les essais non conservés) et en vert clair les cases coloriées dans la réponse finale (Maeder & Matei, soumis)

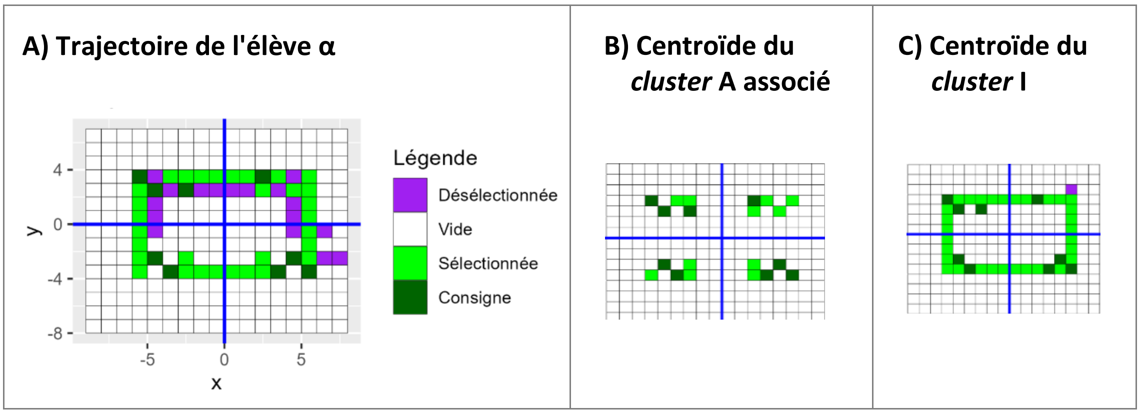


La suite de l'analyse a consisté en un travail de regroupement des productions par *clusters* à l'aide d'un algorithme de *clustering*, sur la base de ces images augmentées (Maeder & Matei, soumis). Il est intéressant de souligner que, contrairement à la méthode d'analyse présentée dans le chapitre 3.6.1, c'est l'algorithme appliqué qui a défini les catégories, sur la base d'une matrice construite spécialement pour cette application, en prenant en compte la proximité des zones du quadrillage dans lesquelles l'élève a travaillé (essais y compris). En d'autres termes, des élèves ayant travaillé dans les mêmes zones du quadrillage aboutissent à des procédures et des solutions qui se ressemblent et qui se retrouvent, par conséquent, regroupées dans la même catégorie. Chaque catégorie est illustrée par une production type (appelée centroïde) faisant office de représentant des tendances d'un groupe.

Ainsi, ce travail, réalisé sur plus de 1000 productions, a permis de définir, sur la base d'un indicateur statistique, **dix catégories** (ou *clusters*) de productions présentant des similitudes en termes de procédure de résolution et de solution, indépendamment de la justesse de la réponse (Maeder & Matei, soumis). La taille de ces *clusters* est variable, allant de trois à 230 productions. Le défi de ce travail a été d'obtenir des *clusters* interprétables, c'est-à-dire que l'on peut expliquer clairement la signification de chaque *cluster* obtenu.

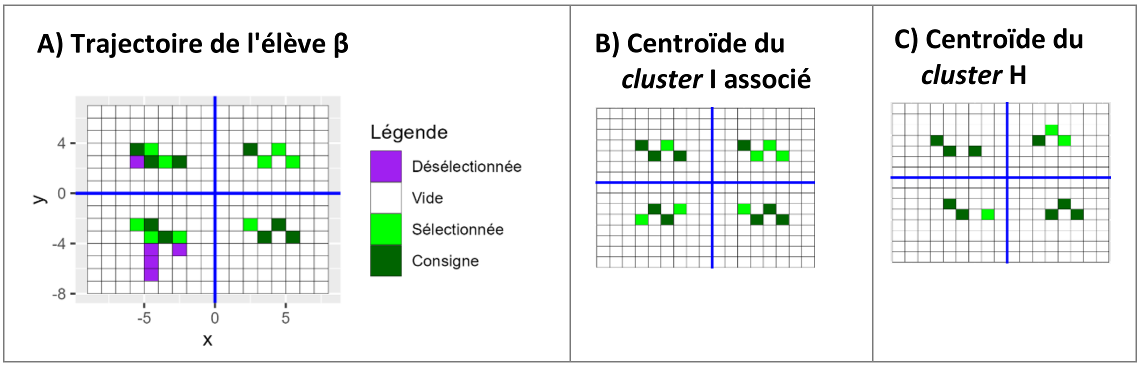
Nous avons regardé plus en détail ces *clusters*, afin d'avoir une idée des regroupements réalisés de manière automatisée par l'algorithme. L'impression générale issue de notre analyse est que le regroupement algorithmique s'est montré plutôt performant. En effet, la plupart des productions regroupées dans un *cluster* donné se ressemblent : soit elles sont identiques, soit elles correspondent à une légère variation du centroïde. Relevons tout de même que, pour quelques productions, peu nombreuses, le *cluster* auquel elles ont été associées est discutable d'un point de vue didactique, à l'image de la production de l'élève α, présentée dans la figure 32 ci-dessous. L'algorithme l'a associée au *cluster* A, lequel regroupe essentiellement des productions (presque) correctes, alors que la figure soumise par l'élève α est bien éloignée d'une telle réponse. Si nous avions été amenées à la catégoriser manuellement, nous l'aurions plutôt intégrée dans le *cluster* I.

Figure 32 : Production de l'élève α (A) et centroïdes du cluster A (B) et du cluster I (C)



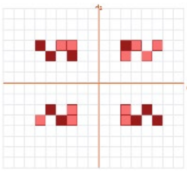
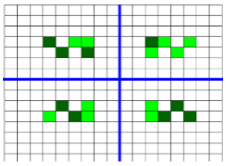
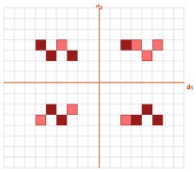
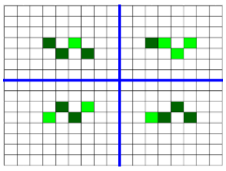
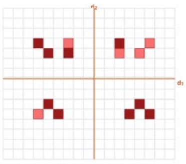
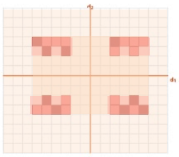
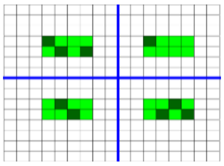
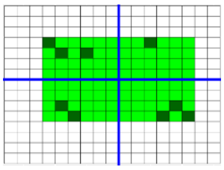
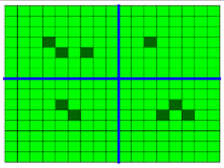
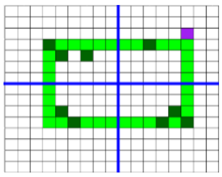
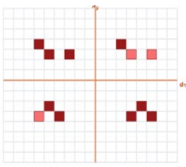
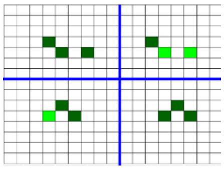
Un deuxième exemple est illustré par la production de l'élève β (figure 33). Cette dernière a été associée au *cluster* I, caractérisé par une translation d'une part, et associé à une symétrie d'autre part. En examinant cette production, nous remarquons toutefois qu'elle ne présente pas de symétrie, mais est le fruit d'une translation verticale des figures, correspondant au *cluster* H.

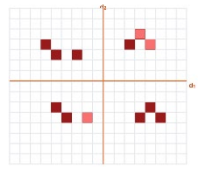
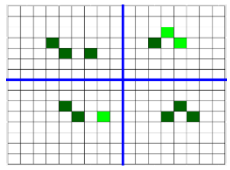
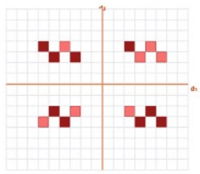
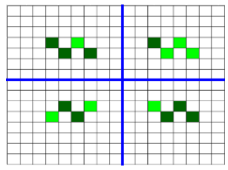
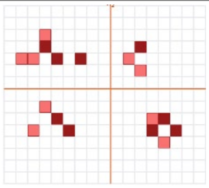
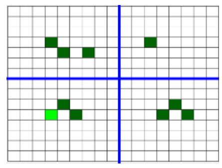
Figure 33 : Production de l'élève β (A) et centroïdes du cluster I (B) et du cluster H (C)



Ce travail d'identification puis de répartition des procédures au moyen d'un algorithme étant exploratoire, nous avons parallèlement analysé manuellement un échantillon de 50 productions choisies aléatoirement dans le but de comparer les procédés d'analyse et leurs résultats. Nous voulions observer s'il existait des similitudes entre les regroupements obtenus par les deux méthodes d'analyse. Une différence notable à souligner est que, contrairement à la méthode algorithmique, l'**analyse manuelle** s'est uniquement basée sur la réponse finale soumise par l'élève et a été réalisée sur nettement moins de productions que l'analyse par *clusters* (50 productions analysées manuellement contre 1073 algorithmiquement). Par ailleurs, son ancrage est didactique, alors que l'analyse algorithmique se base sur des similitudes des zones de travail. Cette **analyse didactique** non exhaustive a abouti à **huit catégories**. Le tableau ci-dessous présente parallèlement les regroupements obtenus **manuellement (didactiquement)** à gauche, et **algorithmiquement** à droite, chacun étant illustré par une production d'élève caractéristique.

Figure 34 : Présentation des résultats issus des méthodes d'analyse manuelle (didactique) et algorithmique, comparant dans une même ligne du tableau les regroupements des deux méthodes présentant des analogies

Regroupements manuels (didactiques)	N_{total} = 50	Regroupements algorithmiques	N_{total} = 1073	Caractéristiques des regroupements
Groupe 1 : N = 10		**Cluster A** N = 203 dont 157 réponses correctes		Réponse (presque) correcte (max. 1 case de plus coloriée dans chaque quadrant)
Groupe 2 : N = 4		**Cluster B** N = 83		Figure uniquement symétrique sur l'axe d_1
Groupe 3 : N = 3		Pas de cluster équivalent (répartition dans d'autres clusters)		Figure uniquement symétrique sur l'axe d_2
Groupe 4 : N = 5		**Cluster C** N = 78		Figure de forme rectangulaire, complétée de manière symétrique ou « géométrique » sur les axes d_1 et d_2, MAIS > 11 cases coloriées
		Cluster D N = 35		
		Cluster E N = 3		
		Cluster F N = 29		
Groupe 5 : N = 7		**Cluster G** N = 215		Figure translatée horizontalement

Groupe 6 : N = 7		**Cluster H** N = 86		Figure translatée verticalement
Groupe 7 : N = 5		**Cluster I** N = 111		Figure translatée d'une part et symétrique d'autre part
Groupe 8 : N = 9		**Cluster J** N = 230		Autres productions

La comparaison des regroupements issus des deux méthodes d'analyse nous a permis de constater que les catégories déterminées algorithmiquement et manuellement se ressemblent fortement. En effet, pour la plupart des groupes déterminés manuellement se trouve une équivalence parmi les centroïdes des *clusters* obtenus algorithmiquement. Un seul groupe (3), rassemblant des productions caractérisées par une symétrie axiale uniquement effectuée sur l'axe d_2, n'a pas de *cluster* équivalent. Le regroupement par *clustering* a réparti ces productions dans d'autres *clusters* selon une logique différente. Notons également que les deux méthodes ont défini un regroupement comprenant un nombre assez important de productions n'entrant dans aucune autre catégorie (groupe 8 et *cluster* J). Un examen plus fin des productions de ces regroupements pourrait être intéressant pour affiner la catégorisation, mais il s'avèrerait chronophage et ne permettrait pas forcément d'arriver à une compréhension plus fine du travail de l'élève.

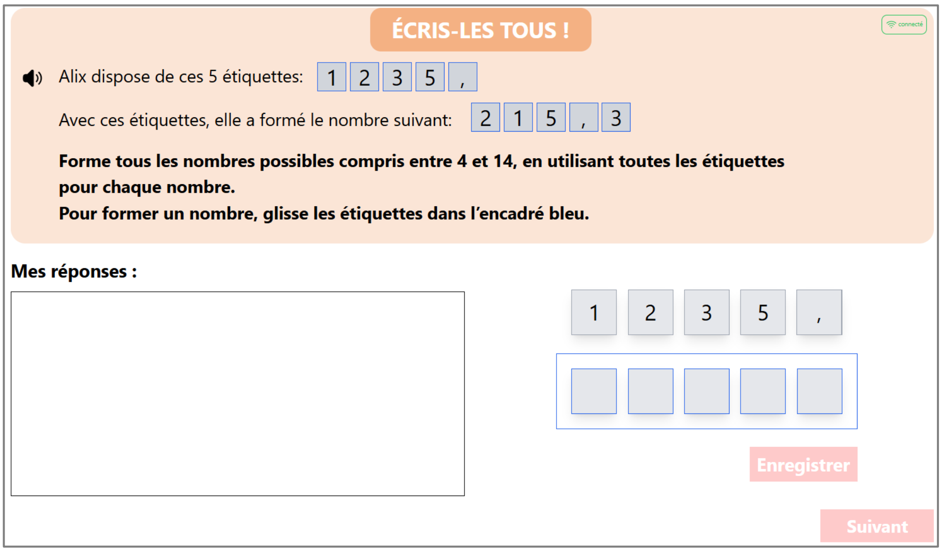
La similitude observée pour les catégories définies par les deux méthodes soulève la question suivante : l'intégration des essais dans les images augmentées traitées par l'algorithme est-elle nécessaire dans la catégorisation des productions ? En effet, l'analyse manuelle s'est basée uniquement sur la réponse finale (et ne prenait pas en compte les cases décoloriées par les élèves). Il faut cependant préciser que, pour être performant, l'algorithme a besoin d'énormément d'informations et qu'il a donc été nécessaire d'intégrer les cases décoloriées dans le processus d'analyse.

Somme toute, le travail comparatif mené nous a permis de constater que l'algorithme utilisé est assez efficace pour réaliser un premier tri automatique, sur un nombre très important de productions à analyser, et d'établir des *clusters* interprétables. Il s'agit ensuite de l'affiner manuellement, en intégrant des critères didactiques aux *clusters* déterminés. À terme, l'idée est de pouvoir proposer à l'enseignante ou l'enseignant une remédiation appropriée en fonction du *cluster* auquel la production de l'élève a été associée.

4.3.3 Tâches adaptatives : effet des relances (recentrer l'élève sur la résolution du problème)

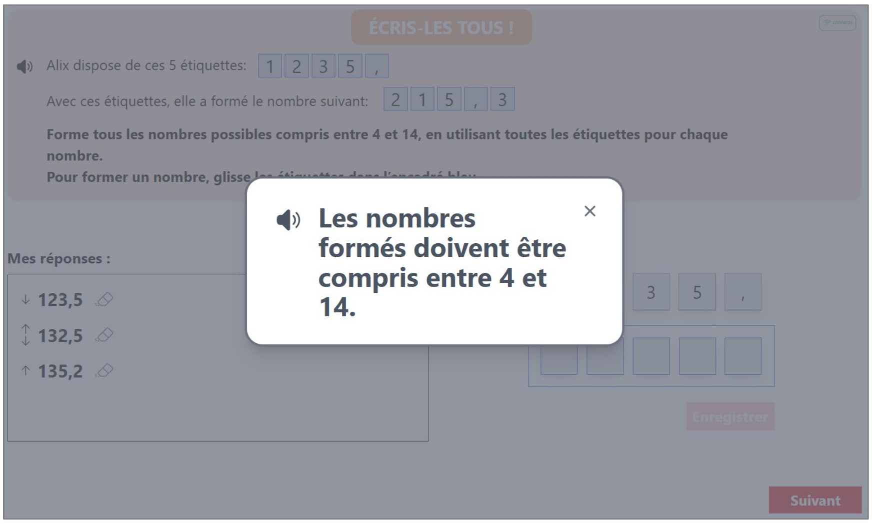
Deux problèmes de Mathématiques ont été conçus de manière à proposer **des relances** visant à rappeler certaines contraintes de l'énoncé, pour recentrer l'élève sur son travail. Ces *feedbacks* ont été anticipés sur la base de l'analyse à priori des difficultés caractéristiques que pourraient rencontrer les élèves ou d'erreurs typiques qu'elles ou ils pourraient commettre. Lors de l'analyse des données récoltées par l'application, nous nous sommes principalement intéressées à deux paramètres : le **nombre d'apparitions** de ces relances, ainsi que la ou les **actions effectuées par les élèves suite à l'apparition de ce message**. L'objectif était de déterminer l'effet de ces rétroactions sur le travail de l'élève. Voici un exemple de problème pour lequel une relance a été prévue :

Figure 35 : Problème « Écris-les tous ! », pour lequel une relance a été prévue



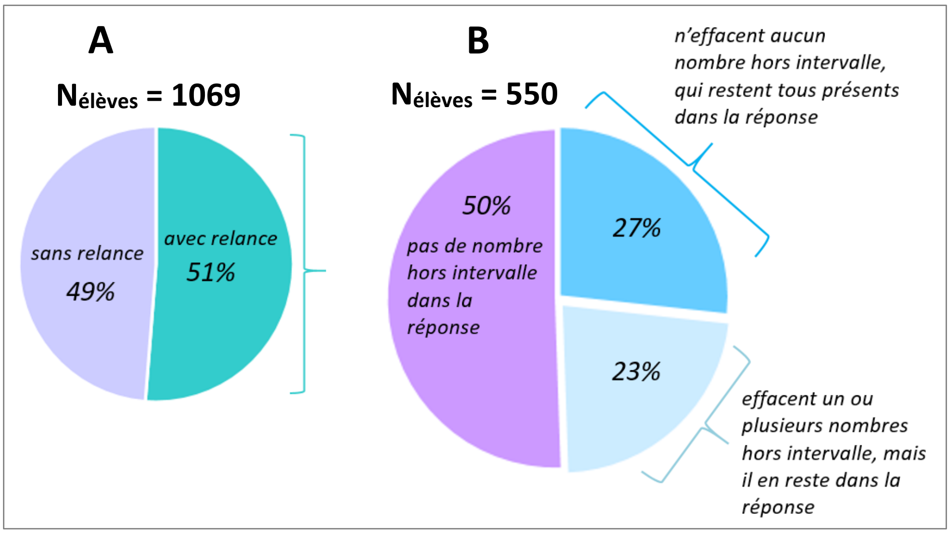
Pour réaliser ce problème, l'élève doit, à l'aide des étiquettes à disposition, former tous les nombres possibles compris dans l'intervalle donné. Une fois un nombre formé, elle ou il le valide, et le nombre s'affiche alors dans la zone de travail du problème. Une erreur anticipée à priori est la formation de nombres hors intervalle. Nous avons donc programmé une alerte, paramétrée pour se déclencher à partir du troisième nombre hors intervalle formé par l'élève, lui rappelant autant de fois que nécessaire que : « Les nombres formés doivent être compris entre 4 et 14 » (figure 36).

Figure 36 : Relance qui apparait au bout du 3e nombre formé hors intervalle



L'analyse des données a montré que cette erreur est effectivement fréquente : plus de la moitié (51%) des élèves a vu ce message s'afficher sur son écran (figure 37). Les actions effectuées par les élèves suite à l'apparition de la relance révèlent qu'elle a eu un effet positif pour beaucoup. En effet, 50% des élèves ayant reçu le *feedback* ont finalement effacé tous les nombres hors intervalle formés avant de soumettre leur réponse. De plus, 23% ont effacé un ou plusieurs nombres hors intervalle. Le fait qu'il en restait encore dans la réponse finale peut s'expliquer par le fait que la relance s'affichait seulement au bout du troisième nombre hors intervalle formé. Il se peut que des élèves n'aient effacé que le dernier nombre formé, pensant que la relance concernait uniquement leur dernière action. En revanche, pour 27% des élèves, l'alerte n'a eu aucun effet : ils ont conservé tous les nombres hors intervalle dans leur réponse.

Figure 37 : Pourcentage d'élèves A) ayant ou non reçu la relance et B) ayant effacé tous les nombres hors intervalle, ou ayant effacé une partie des nombres hors intervalle ou encore n'ayant effacé aucun nombre hors intervalle suite à l'apparition de la relance

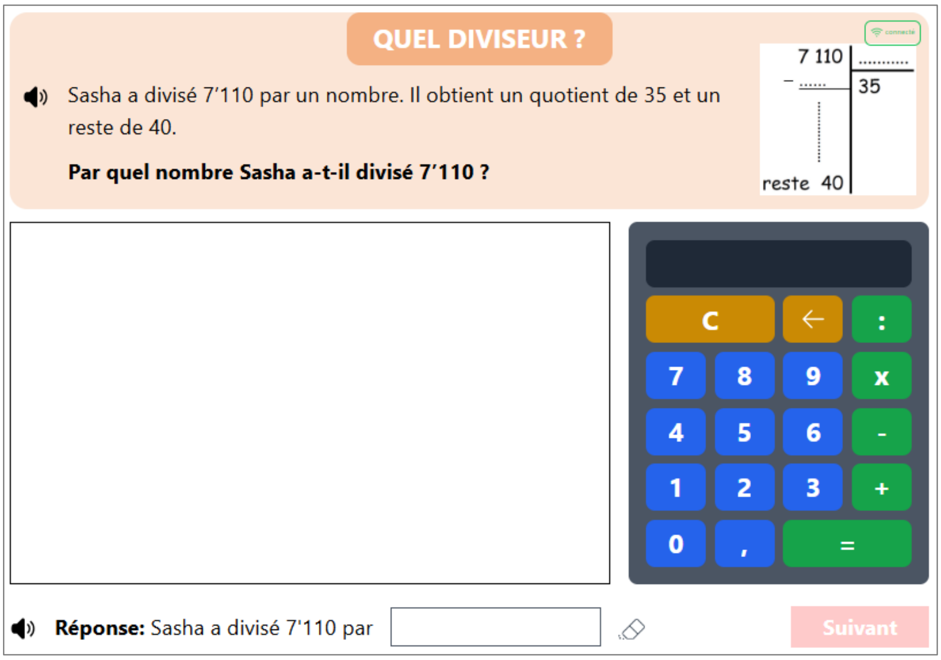


Face au constat que certains élèves ont vu cette relance apparaitre jusqu'à 24 fois (!) sans pour autant la prendre en compte, nous émettons deux hypothèses. La première est que la formulation d'une relance reprenant mot pour mot ceux de l'énoncé, en l'occurrence « Les nombres formés doivent être compris entre 4 et 14. », n'est pas efficace. Pour un *feedback* formatif (Lecoix, 2024) plus performant, il conviendrait de reformuler le message, par exemple : « Les nombres formés doivent être plus grands que 4 et plus petits que 14. » Dans le cadre de l'expérimentation menée, nous n'avons cependant pas pu soumettre la tâche à un nouvel échantillon d'élèves en modifiant la formulation de cette relance, afin de voir si son impact s'en verrait amélioré.

Notre seconde hypothèse est que la finalité de ces messages n'a pas été comprise par les élèves. En effet, il est probable que ces élèves ont fermé cette fenêtre s'ouvrant sur leur écran de travail à la manière d'une fenêtre popup publicitaire, probablement sans même tenir compte du message qui s'y trouvait. Cela suggère que les élèves ne sont pas habitués à une telle interactivité et qu'elles et ils doivent apprendre à l'utiliser. Il est important qu'elles et ils s'entrainent à utiliser des outils numériques et leurs fonctionnalités, par exemple à prendre en considération les *feedbacks* donnés automatiquement par la machine et à les traiter de manière efficace. Il semblerait que l'élève qui a vu la relance s'afficher 16 fois pour le premier problème l'ait compris par elle/lui-même : elle ou il a finalement effacé les 18 nombres hors intervalle formés, pour soumettre au final une réponse parfaitement correcte.

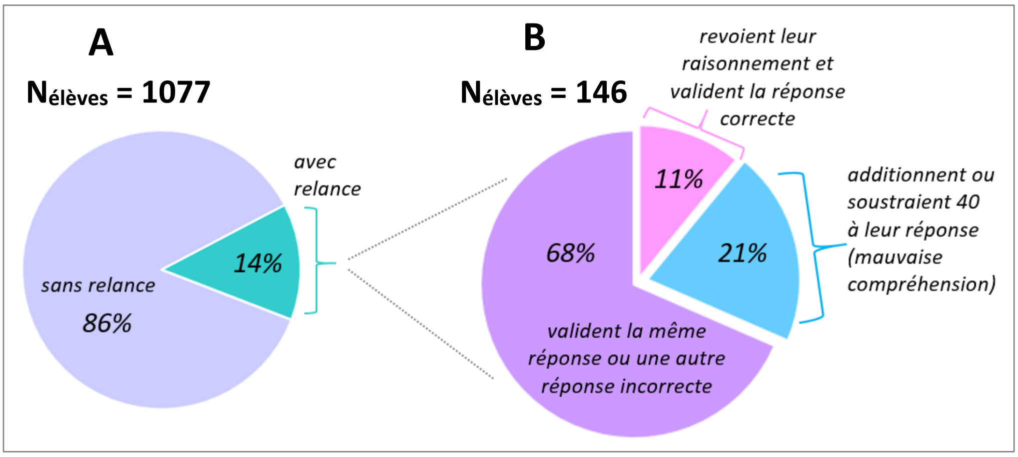
Un même type de **relance** a également été paramétré pour le problème divisif ci-dessous, dont l'objectif est de tester la compréhension de la division euclidienne. Dans ce problème, l'élève doit retrouver le diviseur d'une division, connaissant le dividende, le quotient et le reste.

Figure 38 : Problème « Quel diviseur ? », pour lequel une relance a été prévue



Une des difficultés de ce problème réside dans la prise en compte du reste, à savoir de commencer par le soustraire au dividende avant de réaliser la division du dividende par le quotient. Notre analyse à priori nous a permis d'émettre l'hypothèse que beaucoup d'élèves diviseraient le dividende par le quotient sans considérer le reste. Ainsi, les élèves qui ont effectué la division 7 100 : 35 et répondu 203 (ou 203,1, voire 203,12) recevaient le message suivant : « As-tu bien pris en compte le reste de 40 ? » Il est intéressant de constater que cette relance est apparue à seulement 14% des élèves (figure 39). Parmi ces élèves, un faible pourcentage (11%) a bien saisi le sens du *feedback*, revu son raisonnement et validé la réponse correcte. Pour ce groupe d'élèves, ce message d'alerte s'est révélé efficace. En revanche, il a été mal compris par 21% des élèves l'ayant reçu. En effet, ces élèves ont essayé de prendre en compte le reste, mais l'ont mathématisé de manière incorrecte. De plus, le message n'ayant eu aucun effet pour la majorité des élèves l'ayant reçu (68%), l'efficacité de cette relance n'a pas été celle attendue.

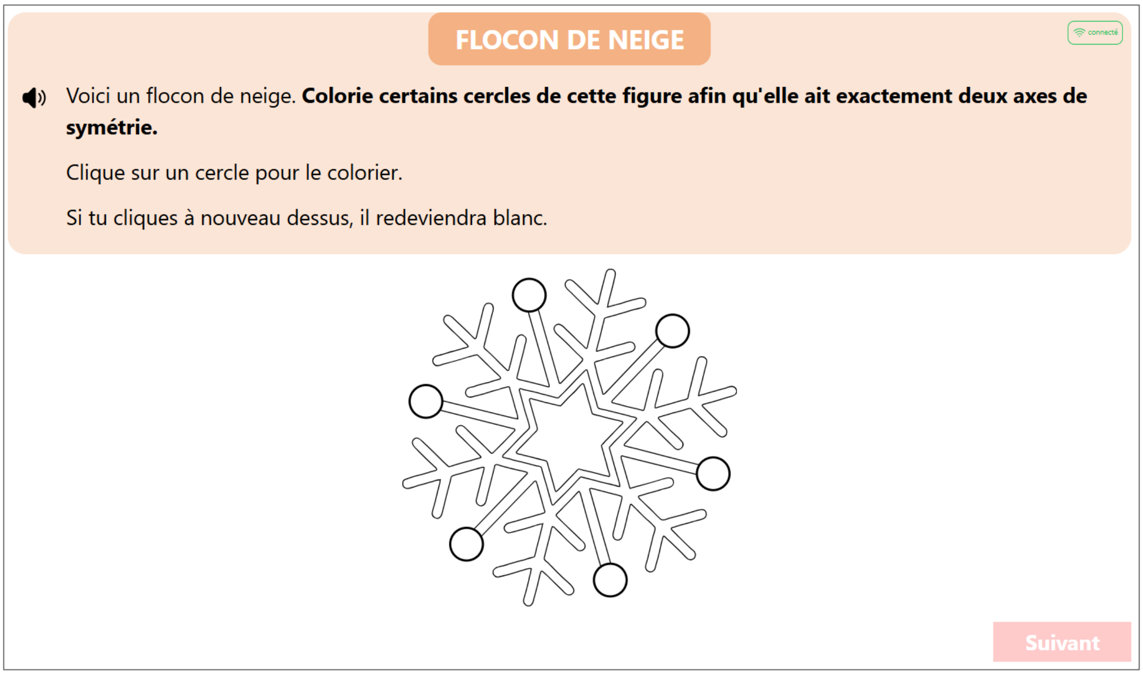
Figure 39 : Pourcentage A) d'élèves ayant ou non reçu la relance et B) de leurs actions suite à l'apparition de la relance



4.3.4 *Tâches adaptatives : évaluer plus finement la maitrise d'une compétence*

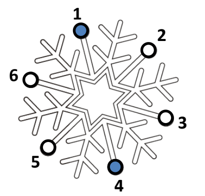
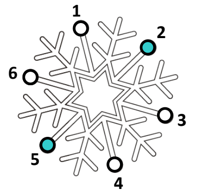
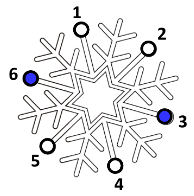
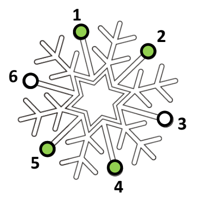
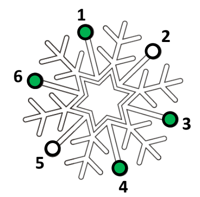
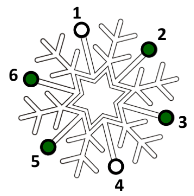
Outre les relances décrites dans le chapitre précédent, lesquelles ont pour objectif de recentrer l'élève sur son travail, un autre type d'adaptation a été éprouvé lors du test pilote. Il s'agit de proposer un problème initial à l'élève, puis de lui soumettre, en fonction de sa réponse, une seconde tâche plus ou moins complexe, permettant d'affiner l'évaluation de la ou des compétences en jeu. Pour ce faire, nous avons choisi un problème d'axes de symétrie d'une figure plane, en l'occurrence un flocon de neige (figure 40). Le choix de proposer ce problème sur support informatisé est motivé par le fait qu'il facilite les essais : il permet d'alléger la charge visuelle que les opérations précédentes laisseraient sur une feuille.

Figure 40 : Problème « Flocon de neige » présenté initialement à l'ensemble des 1076 élèves



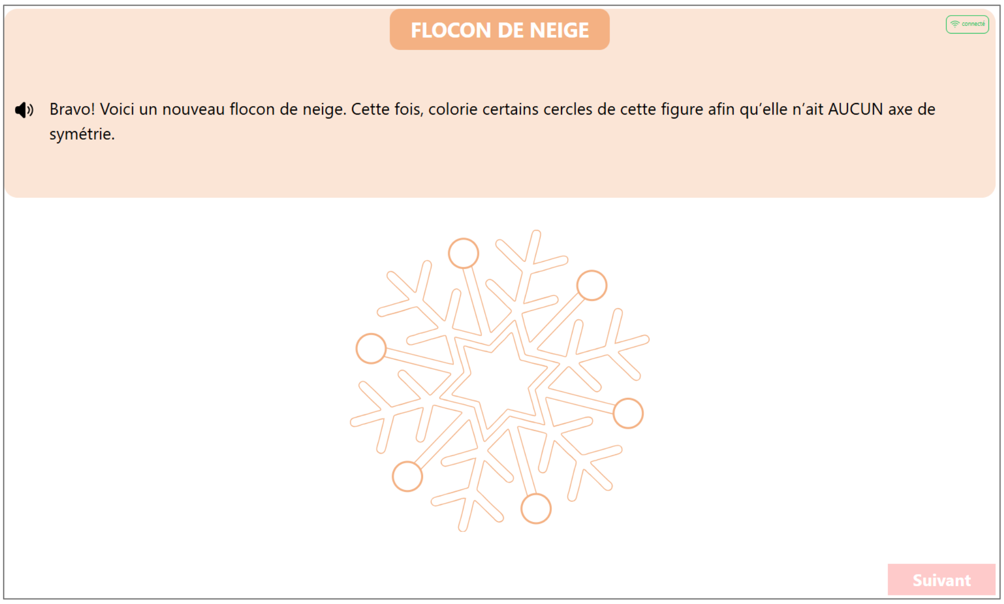
Dans ce problème initial, il est demandé à l'élève de colorier certains des six cercles de manière à ce que le flocon ait exactement deux axes de symétrie. L'élève peut colorier un cercle en cliquant dessus, et le décolorier en cliquant à nouveau sur le même cercle. Plusieurs réponses correctes sont possibles : outre l'orientation donnée au flocon, l'élève peut choisir de colorier deux ou quatre cercles opposés (figure 41).

Figure 41 : Réponses correctes possibles pour le problème initial « Flocon de neige »

2 cercles coloriés			
4 cercles coloriés			

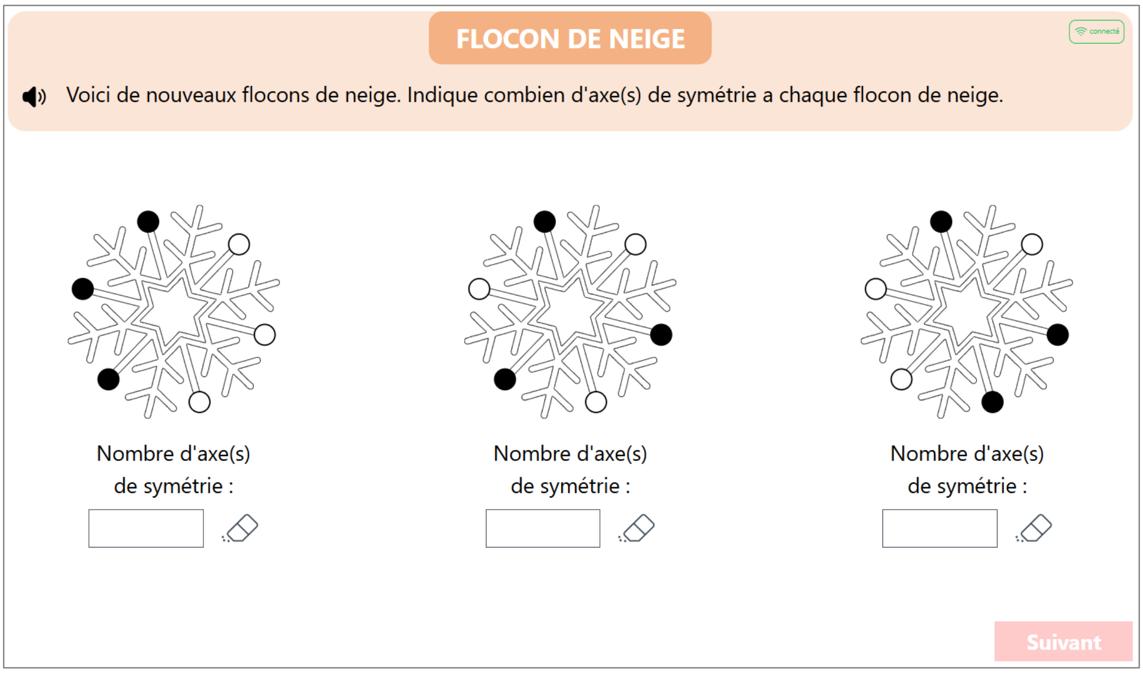
À l'élève qui donne une réponse correcte à cette tâche initiale, autrement dit qui valide une figure avec exactement deux axes de symétrie, est proposée ensuite une tâche similaire, mais plus complexe, où elle ou il doit cette fois colorier certaines parties de la figure pour que celle-ci n'ait aucun axe de symétrie (Adaptation 1, figure 42). Une telle tâche d'approfondissement permet de s'assurer de l'acquisition de la compétence chez l'élève, et que sa réponse à la tâche initiale n'est pas le fruit du hasard, mais de sa maitrise du repérage d'axes de symétrie en situation de résolution de problèmes.

Figure 42 : Adaptation 1 du problème « Flocon de neige », qui était soumise aux élèves ayant répondu correctement au problème initial



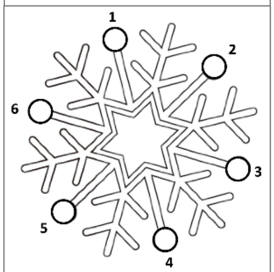
En revanche, une tâche d'application plus simple (adaptation 2, figure 43) est proposée à l'élève donnant une réponse incorrecte. Dans ce cas, il s'agit de déterminer le nombre d'axes de symétrie de trois flocons, dont des cercles sont déjà coloriés. Cette tâche d'application est exemptée de la dimension *résolution de problèmes* et seule la compétence de repérage d'axes de symétrie est évaluée. Cette manière de faire permet de déterminer si le fait que l'élève échoue à la tâche initiale est à imputer à l'aspect *résolution de problèmes* ou si les difficultés sont liées au repérage des axes de symétrie. Cette volonté d'aller creuser à l'endroit supposé de la difficulté rencontrée par l'élève s'explique par la possibilité d'affiner l'évaluation de la compétence en jeu.

Figure 43 : Adaptation 2 du problème « Flocon de neige », qui était soumise aux élèves ayant répondu de manière incorrecte au problème initial



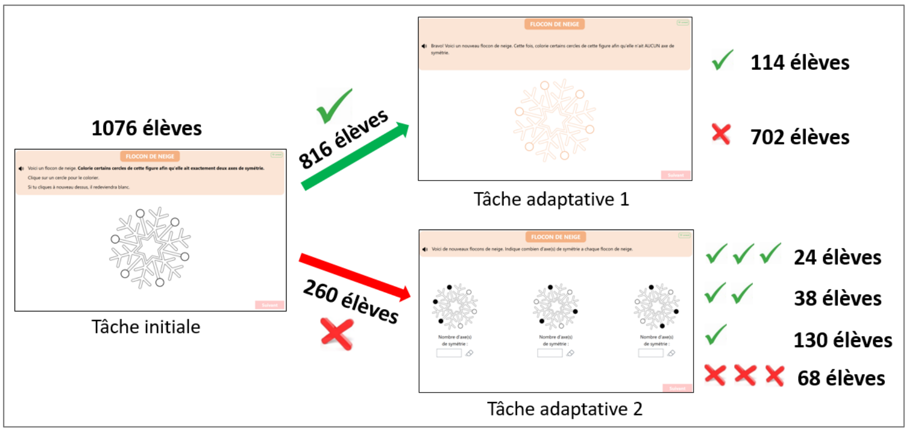
Lors de la récolte des données, l'application a enregistré chaque clic effectué par l'élève, à savoir chaque cercle colorié/décolorié. Chacun des six cercles étant codé par un numéro allant de 1 à 6 (figure 44), les productions des élèves ont pu être corrigées automatiquement. Grâce au support informatisé, la fréquence des réponses correctes a ainsi pu être obtenue très rapidement.

Figure 44 : Codage des cercles du flocon



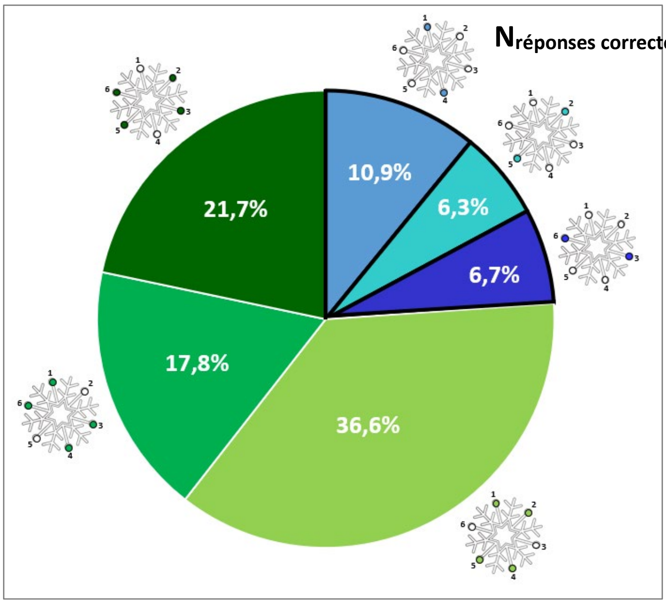
L'analyse des données (figure 45) a montré que le problème initial était réussi par de nombreux élèves (76%). Parmi ces réponses correctes, les résultats révèlent que plus de 75% des élèves ont choisi de colorier quatre cercles et que seul environ un quart des élèves en a colorié deux.

Figure 45 : Fréquence des réponses correctes/incorrectes pour la tâche initiale et les deux tâches adaptatives



Par ailleurs, parmi les différentes orientations possibles, celle représentée par les cercles coloriés en violet (figure 46) est nettement préférée des élèves, certainement car c'est celle qui se rapproche le plus d'une orientation horizontale. Précisons que l'application a non seulement enregistré le nombre de coloriages/décoloriages effectués par chaque élève, mais aussi l'ordre dans lequel les cercles ont été coloriés/décoloriés. Cependant, ces données n'ont pas été exploitées pour le moment.

Figure 46 : Répartition des réponses correctes à la tâche initiale



Quant aux résultats relatifs à l'adaptativité de la tâche, si plus des trois quarts des élèves réussissent le problème initial, seuls 14% d'entre eux ont soumis une réponse correcte au second problème (figure 45). Dès lors, on peut se demander si le repérage d'axes de symétrie est réellement maitrisé par les 702 élèves ayant répondu correctement à la tâche initiale, mais pas à la tâche d'approfondissement. Nos données ne permettant pas de connaitre l'origine de l'échec à la seconde tâche, il serait nécessaire de les compléter. Au niveau de la réussite à la deuxième tâche adaptative, 90% des élèves ayant répondu de manière incorrecte à la tâche initiale rencontrent également des difficultés à résoudre la tâche d'application suivante. Pour les 68 élèves ayant donné un nombre incorrect d'axes de symétrie pour les trois flocons proposés, le repérage d'axes de symétrie pour une figure plane donnée n'est clairement pas maitrisé. Une déduction similaire peut être réalisée pour les 130 élèves ayant donné uniquement une bonne réponse à la tâche d'application. Dans les autres cas de figure, le niveau de maitrise de la compétence évaluée est plus nuancé, et il est difficile d'identifier l'origine des difficultés rencontrées. Une troisième tâche nous paraitrait nécessaire pour approfondir encore l'évaluation du repérage des axes de symétrie d'une figure plane, par exemple en proposant des figures moins complexes que ce flocon.

4.4 Constats et perspectives

L'analyse des données a montré que l'évaluation de la *résolution de problèmes* est non seulement possible sur format informatisé, mais aussi que ce support présente des plus-values par rapport au papier-crayon, bien que le choix des outils soit contraint par le support numérique (*cf*. chapitre 4.1). En font notamment partie la **récolte de traces** par le biais des actions réalisées par les élèves durant leur travail, **les relances immédiates** qui leur sont proposées et la possibilité de créer des **tâches adaptatives** permettant d'affiner l'évaluation des compétences en jeu.

Les **nombreuses données** qu'il est possible de recueillir favorisent la reconstitution du raisonnement de l'élève et permettent d'identifier plus précisément la procédure de résolution. Par ailleurs, grâce à l'enregistrement numérique des données, il est possible d'automatiser, du moins en partie, la catégorisation des procédures mises en œuvre par chaque élève, ce qui permet un gain de temps considérable en termes d'analyse de productions. Néanmoins, au regard des deux dispositifs exploratoires présentés dans ce travail (*cf*. chapitres 3.6.1 et 3.6.2), force est de constater que le regard « humain » demeure nécessaire pour vérifier le travail réalisé par la machine. En effet, les résultats donnés par un algorithme doivent être complétés par ou combinés à une analyse didactique, et ce afin que les catégories obtenues reflètent le niveau de maitrise des compétences testées. Par conséquent, l'outil numérique ne remplace pas l'expertise humaine, mais la facilite en prenant en charge une partie du traitement des données (Delozanne et al., 2010 ; Hakem et al., 2005 ; Stacey et al., 2018). En conclusion, l'analyse numérique doit être associée à une approche « humaine » prenant en charge l'aspect didactique. Cela révèle également le côté difficilement transposable d'un tel procédé d'analyse à d'autres types de problèmes que celui pour lequel il a été programmé (Hakem et al., 2005). Une perspective en termes de gestion des données serait de travailler par famille de problèmes.

Une autre plus-value du support informatisé est qu'il permet de donner des ***feedbacks*** conditionnés par les actions que les élèves ont réalisées. Selon Tricot (2020), l'efficacité d'un *feedback* est renforcée s'il est immédiat et élaboré, c'est-à-dire s'il fournit une explication et pas seulement une appréciation de la réponse. C'est en ce sens que les rétroactions proposées lors du test pilote de 2023 ont été conçues : elles ne visaient pas de simples indications sur la justesse de la réponse, mais possédaient un ancrage didactique basé sur une analyse à priori des problèmes. Nous avons cependant remarqué que leur élaboration reste délicate, qu'elles sont encore perfectibles et qu'à l'avenir, il s'avèrerait judicieux de pouvoir les tester en amont. N'oublions pas que la qualité du *feedback* est primordiale pour aider l'élève à progresser. Ainsi, afin qu'il ait un réel impact sur les apprentissages, la formulation de chaque *feedback* doit être examinée attentivement pour qu'il englobe les recommandations suivantes (Sadler, 1989 ; Van Der Kleij et al., 2015) :

- le *feedback* précise les attentes ;
- il met en évidence l'écart entre la production et ce qui est attendu ;
- il fournit des conseils pour réduire l'écart entre la performance actuelle et celle exigée.

La dernière plus-value examinée dans ce travail concerne la possibilité de proposer une série de **problèmes adaptatifs**. Cet avantage mériterait d'être approfondi en testant d'autres problèmes. En effet, les essais réalisés dans ce sens lors du test pilote de 2023 ont montré que plusieurs tâches successives sont nécessaires afin d'évaluer plus finement la maitrise de la ou des compétences en jeu.

5. Conclusion

Le travail réalisé durant la période quadriennale 2020-2023, centré sur l'évaluation de la *résolution de problèmes* en 8e année, a révélé que :

- La *résolution de problèmes* est une **compétence difficile** à maitriser pour beaucoup d'élèves. En effet, nos résultats ont montré qu'environ la moitié des élèves ayant pris part au test pilote mettent en œuvre des procédures incorrectes pour résoudre les problèmes que nous leur avons soumis. Il est ainsi important que cette compétence soit travaillée et entrainée régulièrement en situation d'enseignement-apprentissage.
- Bien que la mise à disposition de la calculatrice lors de l'évaluation de la *résolution de problèmes* **améliore la réussite** des élèves à la tâche, elle ne favorise pas forcément la mise en œuvre de procédures expertes. Nos résultats ont également montré que le fait d'autoriser la calculatrice augmente de manière significative l'absence de traces, c'est-à-dire que, souvent, les élèves ne donnent pas à voir les calculs qu'ils ont effectués à l'aide de la machine. L'utilisation d'un support numérique comprenant la fonctionnalité de calculatrice intégrée se révèle très intéressante ici, tous les calculs réalisés – même ceux non pertinents à l'image d'essais – sont automatiquement enregistrés par le support/l'application.
- Le recours au support informatisé lors de l'évaluation de la *résolution de problèmes* présente de nombreux avantages, notamment en termes de **traces récoltées** et de possible **adaptativité**. Des limites sont à signaler, comme le guidage des procédures pouvant être mises en œuvre par les élèves ou le temps nécessaire à la programmation des tâches en amont de la passation. Relevons également que, si le support informatisé permet le traitement automatisé des données récoltées à partir d'un paramétrage établi sur la base d'une analyse, celui-ci est généralement valable uniquement pour le problème spécifique en question. En effet, bien que chronophage, ce traitement se révèle difficilement transposable à d'autres problèmes que celui pour lequel il a été déterminé.
- L'**expertise humaine**, notamment en termes de didactique, reste indispensable lors de l'analyse des productions d'élèves, même si une partie du traitement peut être automatisée et déléguée à la machine.
- L'évaluation de la *résolution de problèmes* au moyen d'un support numérique requiert une familiarisation préalable pour les élèves : elles et ils doivent tout d'abord être habitués au format, aux fonctionnalités et aux outils proposés sur ce support, en situation d'enseignement-apprentissage.

La plupart des problèmes papier-crayon proposés aux élèves romands lors du test pilote de 2023 ont pu être validés par le dispositif mis en place pour être intégrés sur les pages *PistEval*, en vue d'une mise à disposition pour le corps enseignant romand. Quant aux problèmes informatisés, testés de manière exploratoire dans les travaux menés, d'autres développements seront nécessaires avant de pouvoir proposer de telles tâches aux enseignantes et enseignants, notamment en termes de retour qui leur sera transmis sur les compétences de leurs élèves. À terme, l'utilisation du numérique vise à proposer un outil d'assistance au diagnostic, afin de soutenir les enseignantes et enseignants dans leur travail d'évaluation des compétences, grâce à un retour automatisé et différencié pour chaque élève.

6. Références bibliographiques

Académie de Créteil. (2018). *La résolution de problèmes et la classification Vergnaud*. https://www.dsden94.accreteil.fr/IMG/pdf/la_typo.pdf

Allal, L. (1999). Acquisition et évaluation des compétences en situation scolaire. In J. Dolz & E. Ollagnier (éds), *L'énigme de la compétence en éducation* (pp. 75-94). De Boeck Supérieur.

Association of language testers in Europe (ALTE). (1998). *Multilingual glossary of language testing terms* (Local Examinations Syndicate). Cambridge University Press.

Blumenthal, S., & Blumenthal, Y. (2020). Tablet or paper and pen? : examining mode effects on German elementary school students' computation skills with curriculum-based measurements. *International Journal of Educational Methodology*, *6*(4), 669-680. https://doi.org/10.12973/ijem.6.4.669

Brun, J. (1990). La résolution de problèmes arithmétiques : bilan et perspectives. *Math-Ecole, 141*, 2-15.

Carulla, C., Corti, D., de Pietro, J.-F., Kassam, S., Roth, M., Abchi Sánchez, V., & Singh, L. (2013). *Quoi et comment évaluer en référence au Plan d'études romand (PER) : premières pistes : rapport scientifique intermédiaire*. Institut de recherche et de documentation pédagogique (IRDP).

Coen, P.-F. (2022). Exploiter les traces numériques à l'école. *Éducateur, 9*, 5-6.

Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP). (2021). *Référentiel de compétences pour la formation initiale et continue des enseignant.es dans le domaine de l'éducation numérique.* CIIP. https://www.ciip.ch/files/1050/Documents/IRDP/Tableau-r%C3%A9sum%C3%A9-RC-NUM-SO.pdf

Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP). (2010/2024). *Plan d'études romand (PER)*. CIIP. https://www.plandetudes.ch

Delozanne, É., Prévit, D., Grugeon-Allys, B., & Chenovotot-Quentin, F. (2010). Vers un modèle de diagnostic de compétence. *Techniques et sciences informatiques, 29*(8-9), 899-938. https://doi.org/10.3166/tsi.29.899-938

Hakem, K., Sander, E., & Labat, J.-M. (2005). DIANE (Diagnostic informatique sur l'arithmétique au niveau élémentaire). In L. Trouche, M. Joab & P. Tchounikine (éds), *Environnements informatiques pour l'apprentissage humain : actes de la conférence EIAH 2005, Montpellier, 25, 26 et 27 mai* (pp. 81-92). Université Montpellier II.

Hattie, J. (2009, reprinted). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hoffer, G. (en préparation). *ONAÉ : un outil numérique au service des apprentissages et de l'évaluation : analyse scientifique*. IRDP.

Hoffer, G., & Ruf, I. (2025). L'évaluation de la résolution de problèmes sur support informatisé : processus de catégorisation automatisée des procédures des élèves. *Revue de mathématiques pour l'école, 242*, 41-61.

Houdement, C. (2017). Résolution de problèmes arithmétiques à l'école. *Grand N, 100*, 59-78.

Julo, J. (1995). *Représentation des problèmes et réussite en mathématiques : un apport de la psychologie cognitive à l'enseignement*. Presses Universitaires de Rennes.

Lecoix, É. (2024, 12 janvier). *Quelques éléments de réflexion autour des « feedbacks » : les différents types d'étayages pédagogiques au service des apprentissages*. Académie d'Aix-Marseille. https://www.pedagogie.ac-aix-marseille.fr/jcms/c_11191662/fr/quelques-elements-de-reflexion-autour-des-feedbacks-les-differents-types-d-etayages-pedagogiques-au-service-des-apprentissages

Maeder, É., & Matei, A. (soumis). Identification of student navigational patterns in a mathematical task through image clustering.

Monaghan, J., Pool, P., Roper, T., & Threlfall, J. (2009, 30 January). Open-start mathematics problems: an approach to assessing problem solving. *Teaching Mathematics and its Applications: an International Journal of the IMA, 28*(1), 21-31. https://doi.org/10.1093/teamat/hrn023

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Paleczek, L., Seifert, S., & Schöfl, M. (2021, 14 September). Comparing digital to print assessment of receptive vocabulary with GraWo-KiGa in Austrian kindergarten. *British Journal of Educational Technology, 52*(6), 2145-2161. https://doi.org/10.1111/bjet.13163

Rey, B., Carette, V., Defrance, A., & Kahn, S. (2003). *Les compétences à l'école : apprentissage et évaluation*. De Boeck.

Roth, M., Ruf, I., Abchi Sanchéz, V., Soussi, A., & Weiss, L. (2021). EpRoCom : dispositif romand de mutualisation de tâches évaluatives : premiers constats : résumé. *irdp FOCUS, 08.2021*.

Roth, M., & Ruf, I. (2024, 5 mars). Ressources évaluatives pour les enseignant.es romand.es : une démarche intercantonale. *La Revue LEeE, 8*. https://doi.org/10.48325/rleee.008.08

Ruf, I. (2023). *La calculatrice dans les classes romandes de 8e année : pratiques déclarées des enseignant.es et impact sur les procédures des élèves en résolution de problèmes divisifs* [Mémoire de Master]. Haute école pédagogique du canton de Vaud (HEP VD).

Ruf, I., & Weiss, L. (2024). Évaluer la résolution de problèmes de manière certificative : une question complexe. *Revue des mathématiques pour l'école, 241*, 3-14.

Ruf, I., Hoffer, G., & Matei, A. (en préparation). *Mise à disposition de la calculatrice en résolution de problèmes : influence sur les procédures mises en œuvre par les élèves*. IRDP.

Sadler, D. R. (1989, June). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144. https://doi.org/10.1007/BF00117714

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press. https://doi.org/10.1016/C2013-0-05012-8

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007, April). A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement, 67*(2), 219-238. https://doi.org/10.1177/0013164406288166

Wang, S., Jiao, H, Young, M. J., Brooks, T., & Olson, J. (2008, 12 September). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments: a meta-analysis of testing mode effects. *Educational and Psychological Measurement, 68*(1), 5-24. https://doi.org/10.1177/0013164407305592

Stacey, K., Steinle, V., Price, B., & Gvozdenko, E. (2018, 11 November). Specific mathematics assessments that reveal thinking: an online tool to build teachers' diagnostic competence and support teaching. In T. Leuders, K. Philipp & J. Leuders (eds), *Diagnostic competence of mathematics teachers* (pp. 241-261). Springer International Publishing. https://doi.org/10.1007/978-3-319-66327-2_13

Tricot, A. (2020). *Quelles fonctions pédagogiques bénéficient des apports du numérique ?* Centre national d'étude des systèmes scolaires (Cnesco).

Trouche, L. (2002). Les calculatrices dans l'enseignement des mathématiques : une évolution rapide des matériels, des effets différenciés. In D. Guin & L. Trouche (éds), *Calculatrices symboliques : transformer un outil en un instrument du travail mathématique : un problème didactique* (pp. 21-53). La Pensée sauvage.

Van Der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015, 1 December). Effects of feedback in a computer-based learning environment on students' learning outcomes: a meta-analysis. *Review of Educational Research, 85*(4), 475-511. https://doi.org/10.3102/0034654314564881

Vergnaud, G. (1990). Psychologie du développement cognitif et didactique des mathématiques, un exemple : les structures additives. *Petit x, 22*, 51-69.

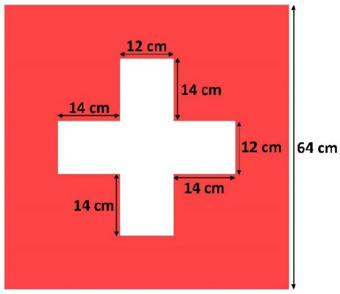
Vermot, B., Behrens, M., & Marc, V. (2011). Étude comparative de deux tests assistés par ordinateur en vue de l'élaboration d'épreuves communes en Suisse romande. In J.-G. Blais & J.-L. Gilles (dirs), *Évaluation des apprentissages et technologies de l'information et de la communication : le futur est à notre porte*. (pp. 225-244). Presses de l'Université Laval.

Wyatt-Smith, C., Lingard, B., & Heck, E. (2019). Évaluations numériques des apprentissages et les mégadonnées : conséquences sur le professionnalisme des enseignants. *Recherche et prospective en éducation : réflexions thématiques, 2*5. https://unesdoc.unesco.org/ark:/48223/pf0000370940_fre

Yvain-Prébiski, S. (2021). Didactical adaptation of professional practice of modelling: a case study. In F. K. S. Leung, G. A. Stillman, G. Kaiser & K. L. Wong (eds), *Mathematical modelling education in east and west* (pp. 305-315). Springer International Publishing. https://doi.org/10.1007/978-3-030-66996-6_26

7. Annexes

7.1 Problèmes testés avec et sans calculatrice

Problème de calcul d'aire	**<u>Le drapeau suisse</u>** Voici une représentation d'un drapeau suisse. Le drapeau suisse est particulier : il est carré.  **Calcule l'aire de la partie rouge de ce drapeau suisse.**
Problème additif	**<u>Fête des Vignerons</u>** À Vevey, sur la place du Marché, plusieurs stands proposent des sandwichs pour les visiteurs de la Fête des Vignerons. Le 1^{er} stand avait 1'481 sandwichs à disposition et en a déjà vendu 213. Le 2^{e} en avait 1'648 à disposition et en a déjà vendu 354. Le 3^{e} en a déjà vendu 189. Actuellement, il reste 3'200 sandwichs à vendre en tout. **Combien de sandwichs le 3^{e} stand avait-il à disposition ?**
Problèmes multiplicatifs et divisifs	**<u>Récolte de coton</u>** En Inde, une agricultrice a engagé 15 personnes pour récolter 7,5 tonnes de coton. Une personne ramasse 40 kg de coton par jour. **Combien de jours ces personnes vont-elles travailler ?**
	<u>Entrées individuelles ou abonnement</u> Une entrée à la piscine coûte 6 francs. Avec un abonnement, elle ne coûte que 4 francs. Cet abonnement coûte 35 francs pour une année. **En une année, combien de fois, au minimum, Jean devra-t-il aller à la piscine pour que l'achat d'un abonnement soit avantageux ?**
	<u>Poterie</u> Pour confectionner des pots de fleurs, on a acheté de l'argile. Il faut 2,2 kg d'argile pour un pot. **Combien de pots de fleurs peut-on confectionner avec 39 kg d'argile ?**
	<u>Long couloir</u> On souhaite recouvrir entièrement le sol d'un couloir rectangulaire de 47 m de long et 1,30 m de large avec du carrelage. On peut acheter des carreaux de carrelage de forme carrée mesurant 1,30 m de côté. **Combien de carreaux doit-on acheter pour recouvrir entièrement le couloir ?**

Inscrits dans le mandat de l'Institut de recherche et de documentation pédagogique (IRDP), les travaux relatifs à l'évaluation des *objectifs d'apprentissage* du Plan d'études romand (PER), menés sur la période quadriennale 2020-2023, visent à créer une culture commune entre les cantons romands en matière d'évaluation des apprentissages des élèves. Dans cette perspective, ils permettent de se doter d'outils d'analyse et de mettre à disposition des enseignantes et enseignants des matériaux d'évaluation pertinents, validés et fiables. Ce texte fait partie d'une publication plus complète, qui rassemble les travaux réalisés dans plusieurs disciplines pour des élèves de 8e année (11-12 ans).

Ce rapport présente les principaux résultats concernant l'évaluation de la *résolution de problèmes* en Mathématiques. Il détaille les analyses effectuées sur les données récoltées auprès de 56 classes lors d'un test pilote déployé au printemps 2023, d'une part sur support papier-crayon et, d'autre part, sur support informatisé. Les analyses se sont centrées sur les procédures de résolution mises en œuvre par les élèves, avec une attention particulière sur la possible influence de la mise à disposition de la calculatrice pour les problèmes effectués sur le support papier-crayon.

Les conclusions montrent des avantages concrets de l'utilisation du numérique dans l'évaluation, comme la possibilité de proposer des tâches adaptatives, différenciées suivant les actions des élèves, et de récolter de nombreuses données permettant de retracer leur raisonnement. Toutefois, ce type d'évaluation requiert une familiarisation préalable pour les élèves qui pourraient, en situation d'enseignement-apprentissage, s'habituer au format, aux fonctionnalités et aux outils proposés sur le support numérique. Quant à l'utilisation de la calculatrice, bien que sa mise à disposition lors de l'évaluation de la *résolution de problèmes* améliore la réussite des élèves à la tâche, elle ne favorise pas forcément la mise en œuvre de procédures expertes.