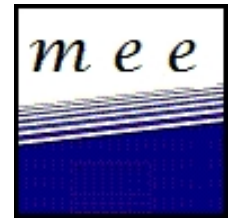


Mesure et évaluation en éducation



Évaluation et didactique des mathématiques : vers de nouvelles questions, de nouveaux travaux

Sylvie Coppé

Volume 41, numéro 1, 2018

URI : <https://id.erudit.org/iderudit/1055895ar>

DOI : <https://doi.org/10.7202/1055895ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Coppé, S. (2018). Évaluation et didactique des mathématiques : vers de nouvelles questions, de nouveaux travaux. *Mesure et évaluation en éducation*, 41 (1), 7–39. <https://doi.org/10.7202/1055895ar>

Résumé de l'article

Les travaux sur l'évaluation dans ses différents aspects sont nombreux et les entrées sont diverses (docimologie, évaluations internationales, évaluation formative, effets sur la motivation, liens avec l'orientation des élèves, etc.). Nous nous centrons sur les liens entre évaluations et apprentissages dans le cadre de l'enseignement primaire ou secondaire en prenant le point de vue de la didactique des mathématiques. Plus précisément, nous cherchons à montrer qu'une entrée par la didactique des mathématiques avec ses cadres théoriques, ses outils et ses méthodes peut éclairer d'un jour nouveau les résultats produits dans le cadre des recherches sur l'évaluation, mais aussi que des recherches sur l'évaluation peuvent être développées en didactique des mathématiques et ainsi renouveler certains questionnements didactiques.

Évaluation et didactique des mathématiques : vers de nouvelles questions, de nouveaux travaux

Sylvie Coppé

Université de Genève

MOTS CLÉS: didactique des mathématiques, évaluation sommative et formative, contrat didactique, milieu, évaluation entre pairs

Les travaux sur l'évaluation dans ses différents aspects sont nombreux et les entrées sont diverses (docimologie, évaluations internationales, évaluation formative, effets sur la motivation, liens avec l'orientation des élèves, etc.). Nous nous centrons sur les liens entre évaluations et apprentissages dans le cadre de l'enseignement primaire ou secondaire en prenant le point de vue de la didactique des mathématiques. Plus précisément, nous cherchons à montrer qu'une entrée par la didactique des mathématiques avec ses cadres théoriques, ses outils et ses méthodes peut éclairer d'un jour nouveau les résultats produits dans le cadre des recherches sur l'évaluation, mais aussi que des recherches sur l'évaluation peuvent être développées en didactique des mathématiques et ainsi renouveler certains questionnements didactiques.

KEY WORDS: mathematics didactics, summative and formative assessment, didactical contract, «milieu», peer assessment

The research on assessment in its various aspects is numerous, and the entries are varied (docimology, international assessments, formative assessment, effects on motivation, links with the orientation of the students, etc.). We focus on the links between assessment and learning in the context of primary or secondary education taking the point of view of mathematics didactics. More precisely, we try to determine how an entry by the mathematics didactics with its theoretical frameworks, its tools and its methods can clarify in a new way the results produced in the research on assessment. We also try to show that research on assessment can be developed in the field of mathematics didactics and thus can renew some didactical questions.

PALAVRAS-CHAVE: avaliação e didática das matemáticas, avaliação sumativa e formativa, contrato didático, “milieu”, avaliação pelos pares

Os trabalhos sobre avaliação nos seus vários aspectos são numerosos e as entradas são diversas (docimologia, avaliações internacionais, avaliação formativa, efeitos na motivação, ligações com a orientação dos alunos, etc.). Neste artigo pretendemos focar as ligações entre as avaliações e as aprendizagens na educação primária ou secundária, segundo o ponto de vista da didática das matemáticas. Mais especificamente, procuramos mostrar que uma entrada pela didática matemática com os seus quadros teóricos, os seus instrumentos e os seus métodos pode lançar uma nova luz sobre os resultados produzidos no contexto das investigações sobre avaliação, mas também que investigações sobre avaliação podem ser desenvolvidas na didática das matemáticas e, assim, renovar algumas questões didáticas.

Les travaux dans les recherches en éducation portant sur l'évaluation dans ses différents aspects sont nombreux et les entrées sont diverses (l'évaluation des élèves, des enseignants, des systèmes ; la docimologie ; les évaluations nationales et internationales ; les différents types ou fonctions, etc.). Des disciplines comme la sociologie ou la psychologie produisent également des études à partir des questions d'évaluation : par exemple, sur les effets de l'évaluation sur la motivation ou l'estime de soi ; sur les liens avec l'orientation des élèves, avec les inégalités sociales, avec le genre, etc. Dans ce texte, nous nous centrons sur les liens entre évaluations et apprentissages dans le cadre de l'enseignement primaire ou secondaire en prenant le point de vue de la didactique des mathématiques. Plus précisément, en analysant l'évolution de la place des recherches sur l'évaluation en didactique des mathématiques, nous cherchons à montrer qu'une entrée par la didactique des mathématiques avec ses cadres théoriques, ses outils et ses méthodes peut éclairer d'un jour nouveau les résultats produits dans le cadre des recherches sur l'évaluation, mais aussi que des recherches sur l'évaluation peuvent être développées en didactique des mathématiques et ainsi renouveler certains questionnements didactiques.

Dans la première partie, en parcourant les quatre grands cadres théoriques de la didactique des mathématiques, nous tenterons d'expliquer pourquoi les recherches en didactique en France ont tardé à s'intéresser aux questions d'évaluation. Puis, après avoir décrit quelques évolutions institutionnelles sur l'évaluation dans le contexte scolaire français, d'une part, et sur l'évaluation formative, d'autre part, nous terminerons par un état des lieux des travaux francophones actuels prenant en compte des questions liées à l'évaluation dans le cadre de la didactique des mathématiques et nous en développerons un exemple sur l'évaluation entre pairs.

Pour commencer, nous adoptons la définition générale proposée par De Ketele (1986), qui met en avant trois points essentiels relevés dans toutes les autres définitions : recueillir des informations et les traiter en vue de prendre une décision :

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

De plus, nous utilisons les six critères de Roegiers (2004) pour qualifier un système d'évaluation qui doit être 1) pertinent (base sur laquelle est certifiée la réussite); 2) valide (il doit évaluer ce qu'on veut évaluer); 3) fiable (réalisé dans de bonnes conditions); 4) efficace (pour aider à prendre les bonnes décisions); 5) réaliste (pas trop difficile à élaborer ni à corriger); et 6) visible (avec des critères explicites).

La didactique des mathématiques en France a tardé à s'intéresser à l'évaluation

Les travaux français en didactique des mathématiques ont tardé à s'intéresser aux questions d'évaluation tant sommative que formative. Bodin (1997), qui a travaillé sur les évaluations internationales, le souligne et le regrette. Il propose une définition, qu'il qualifie de « minimale », même si ce qu'il nomme « l'objet » de l'évaluation recouvre un ensemble assez large (du « savoir » d'un élève à un test, un manuel, etc.).

Évaluer suppose d'organiser et d'étudier des situations permettant de recueillir des informations, qui, après traitement, soient susceptibles de révéler quelque chose de fiable et de substantiel sur la « valeur » d'un « objet ». (Bodin, 1997, p. 60)

Dans cette définition est présente l'idée de recueillir des informations et de les traiter, comme nous l'avons vu précédemment, et aussi d'accorder une place centrale aux « situations », ce qui est cohérent avec une perspective didactique.

Nous allons maintenant expliciter des raisons de ce manque d'intérêt. Pour ce faire, nous partirons des quatre grands cadres théoriques de la didactique des mathématiques française : la théorie anthropologique du didactique, la théorie des situations didactiques, la théorie de l'action conjointe et la double approche. Nous tenterons ainsi de déterminer si et comment des questions liées à l'évaluation ont été prises en compte de façon explicite ou implicite en distinguant la perspective plutôt sommative ou formative.

La théorie anthropologique du didactique (TAD)

Commençons par la théorie anthropologique du didactique (TAD). Dans deux textes datant des années 1980, Chevallard (1986, 1989) aborde les questions d'évaluation en se situant dans une perspective d'évaluation sommative notée. En reprenant les résultats issus de la docimologie,

il critique l'idée de «déterminer LA note vraie». Il replace ces questions dans une perspective anthropologique, où les personnes sont vues comme sujets d'institution(s) et dans laquelle l'évaluation, par la note chiffrée, révèle un degré d'adéquation entre un rapport personnel et le rapport institutionnel à un objet de savoir dans une institution donnée. On pourrait y reconnaître la notion de référent ou les critères évoqués dans la définition proposée par De Ketele, mais Chevallard se situe dans une perspective critique en considérant que ce référent est fortement lié au contrat didactique et qu'alors ce serait des éléments de ce contrat qui seraient révélés davantage que des informations sur les connaissances des élèves sur les savoirs en jeu. La notion de contrat didactique étant définie comme «un système d'obligations réciproques entre l'élève et le professeur» (Brousseau, 1986), on peut penser que certains élèves, lors d'évaluations, produisent des réponses qui montrent davantage la conformité de leur rapport au savoir en jeu à ce moment-là que les connaissances acquises qu'ils pourraient mobiliser dans un autre contexte. De même, pour des élèves qui ne produisent pas les réponses attendues, peut-on en conclure pour autant une non-maîtrise des savoirs en jeu ou bien seulement la mobilisation d'un autre savoir parce que le problème posé peut être résolu d'une autre façon ?

Chevallard (1986, 1989) montre également les liens forts entre institutionnalisation et évaluation, ce que nous avons également mis en évidence par la suite (Coppé, 1993) en étudiant les processus de vérification mis en œuvre par les élèves lors d'évaluations sommatives. Nous avons notamment pointé le fait que certains élèves ne comprenaient qu'au moment de l'évaluation finale ce qu'il y avait à apprendre.

D'ailleurs, Chevallard (1999) reprend cette idée plus tard quand il développe la TAD, dans l'organisation didactique, à travers la définition des moments de l'étude :

Le sixième moment est celui de l'évaluation, qui s'articule au moment de l'institutionnalisation (dont il est à certains égards un sous-moment) : la supposition de rapports institutionnels transcendants aux personnes, en effet, fonde en raison le projet d'évaluer les rapports personnels en les référant à la norme que le moment de l'institutionnalisation aura ainsi hypostasiée. En pratique, il arrive un moment où l'on se doit de «faire le point» : car ce moment de réflexivité où, quels que soient le critère et le juge, on examine ce que vaut ce qui a été appris, ce moment de véridiction qui, malgré les souvenirs d'enfance, n'est nullement une invention de l'École, participe en fait de la «respiration» même de toute activité humaine. (Chevallard, 1999, p. 22)

Ces quelques textes montrent bien le souci de ce chercheur de ne pas ignorer ce qu'il nomme « les faits d'évaluation », voire comme il le dit « la tyrannie du processus d'évaluation » (Chevallard et Feldmann, 1986), mais les recherches françaises en didactique des mathématiques sur l'évaluation (au sens large) se sont peu développées avant les années 2000. Dans ce courant de travaux, c'est davantage la fonction sommative qui est prise en compte. Les notions de rapports personnels et institutionnels à des objets de savoir ainsi que celle de contrat didactique peuvent constituer des outils d'analyse pour mieux interpréter les productions des élèves. La détermination d'organisations mathématiques (ponctuelles, locales, globales) sur un thème donné peut servir de référence pour étudier la variété (voire la complétude) des tâches mathématiques proposées à l'étude et celles évaluées (travail sur la pertinence et la validité).

La théorie des situations didactiques

Considérons maintenant le cadre de la théorie des situations didactiques (Brousseau, 1986, 1998), dans lequel il s'agit de déterminer des situations fondamentales pour enseigner et apprendre une notion mathématique. L'importance est mise sur les objets mathématiques, l'élève est considéré comme un « élève générique » et, dans les années 1980, l'enseignant est peu modélisé. Les notions d'analyse a priori et d'analyse a posteriori sont centrales pour valider les ingénieries produites et pour justifier des possibilités d'apprentissage liées aux situations didactiques. Les questions d'évaluation ne sont pas centrales à ce moment du développement de la théorie. En revanche, les questions de validation le sont et c'est alors la notion de milieu didactique qui est mise en avant. Ainsi, dans les travaux de Margolinas (1992) est présente l'idée de distinguer, dans les phases de conclusion, des phases de validation mais aussi d'évaluation, comme « deux phases antagonistes » :

C'est parce que la validation a échoué que l'enseignante est contrainte, non seulement d'évaluer, mais encore de renforcer la stratégie qu'elle veut favoriser. (Margolinas, 1992, p. 89)

En fait, il nous semble que c'est une façon d'indiquer que, lorsque le milieu ne permet pas suffisamment de rétroactions, l'enseignant prend en charge la validation en donnant un jugement évaluatif. Ce travail n'a pas été poursuivi.

Une adaptation de la théorie est faite par Brousseau (1995), qui introduit les contrats de communication et la régulation didactique, ce qui lui permet de prendre davantage en compte le rôle de l'enseignant par les régulations qu'il effectue du fait des prises d'informations sur l'avancée du savoir dans la classe. Ainsi, dans ces études, les notions de contrat et de milieu sont centrales.

Dans les années 2000, des évolutions de ces concepts apparaissent grâce à l'étude de situations ordinaires (c.-à-d. des séances de classe non élaborées par le chercheur, où il est observateur), notamment dans les travaux de Perrin-Glorian et Hersant (2003), de Schubauer-Leoni, Leutenegger, Ligozat et Fluckiger (2007) ou d'Amade-Escot et Venturini (2009). Ainsi, Perrin-Glorian (1999) définit différents milieux : celui de l'enseignant et celui de l'élève ou des élèves, le milieu « potentiel » et le « milieu activé » :

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

Là encore est présente l'idée que, dans certaines situations, le milieu ne peut à lui seul donner des rétroactions à l'élève pour lui permettre de valider ses réponses et que, dans ce cas, cela amène une intervention de l'enseignant. Ainsi, on peut étudier comment et en quoi des rétroactions de l'enseignant peuvent ou non modifier et/ou enrichir le milieu pour provoquer des adaptations de connaissances chez les élèves.

La théorie de l'action conjointe

Cette idée est également présente dans le cadre de la théorie de l'action conjointe en didactique (Sensevy et Mercier, 2007), où la régulation constitue une des quatre actions de l'enseignant (définir, réguler, dévoluer et instituer) dans un processus dynamique en lien avec les notions de mésogénèse, topogénèse et chronogénèse.

Nous pensons que l'étude des régulations effectuées par l'enseignant ou déclenchées par les élèves permettrait d'aborder plus directement des questions qui relèvent plutôt de la fonction formative de l'évaluation. Les

notions de milieu et de contrat didactique nous semblent pertinentes pour les approfondir; nous y reviendrons dans la partie sur l'évaluation entre pairs.

La double approche

Enfin, à partir des années 2000, les travaux développés dans le cadre de la double approche (Robert et Rogalski, 2002), dans l'étude des pratiques ordinaires, s'intéressent aux liens entre pratiques des enseignants et apprentissages des élèves. Ainsi, ils comportent une forte prise en compte des pratiques de l'enseignant, qui sont alors modélisées avec les composantes de l'activité. Toutes ces recherches, essentiellement dans l'enseignement au collège¹, tentent de montrer des régularités et des variabilités dans les pratiques des enseignants. Certaines, comme celles de Horoks (2006) ou de Chesnais (2009), sans que ce soit toutefois leur axe d'étude principal, intègrent aussi l'étude des évaluations sommatives afin de déterminer des traces d'apprentissage qui vont au-delà de la réussite ou de l'échec à certaines tâches, qui sont alors mises en rapport avec celles travaillées dans la séquence. Il nous semble que ces travaux ouvrent la voie aux recherches sur les pratiques d'évaluation, sur lesquelles nous reviendrons.

À travers ce rapide tour d'horizon, nous avons voulu montrer que les objets d'étude tout comme les notions développées en didactique des mathématiques ont évolué depuis les travaux précurseurs. Les questions liées à l'évaluation des élèves, bien que considérées comme centrales dans les textes de Chevallard, n'ont pas été travaillées pour elles-mêmes ou sont restées des éléments peu documentés, alors que nous avons tenté de montrer en quoi les cadres ou outils théoriques le permettent.

Nous étudions maintenant ce qu'il en est du contexte actuel en faisant un rapide survol des évolutions institutionnelles sur l'évaluation dans le contexte de l'enseignement français, d'une part, et des recherches sur l'évaluation formative, d'autre part, pour montrer les évolutions des thèmes d'étude.

Évolutions institutionnelles

Après la période dite «des mathématiques modernes» (au début des années 1980), en France comme dans d'autres pays, la résolution de problèmes a été mise au centre des programmes de mathématiques, tant à

l'école primaire (Coppé et Houdement, 2009) que secondaire (Coppé, 2010). Depuis les années 2000, à travers les différents programmes, des évolutions institutionnelles sont apparues à la fois sur la place et le rôle des problèmes et sur l'évaluation, avec l'introduction du socle commun de connaissances et compétences, puis de l'évaluation par compétences². Par exemple, ce passage exprime le souci de l'institution d'indiquer que l'évaluation doit être un outil de régulation au-delà de sa fonction sommative ou certificative :

L'évaluation (qui ne se réduit pas au contrôle noté) n'est pas un à-côté des apprentissages. Elle doit y être intégrée et en être l'instrument de régulation, pour l'enseignant et pour l'élève. Elle permet d'établir un constat relatif aux acquis de l'élève, à ses difficultés. Dans cette optique, le travail sur les erreurs constitue souvent un moyen efficace de l'action pédagogique. L'évaluation ne doit pas se limiter à indiquer où en est l'élève ; elle doit aussi rendre compte de l'évolution de ses connaissances, en particulier de ses progrès. (MEN, 2004, 2008)

Dans les programmes actuels, il est davantage question de compétences et il est fait référence à une évaluation plus positive, comme le montre cet extrait du programme de maternelle (qui concerne toutes les disciplines) :

L'évaluation constitue un outil de régulation dans l'activité professionnelle des enseignants ; elle n'est pas un instrument de prédiction ni de sélection. Elle repose sur une observation attentive et une interprétation de ce que chaque enfant dit ou fait. Chaque enseignant s'attache à mettre en valeur, au-delà du résultat obtenu, le cheminement de l'enfant et les progrès qu'il fait par rapport à lui-même. Il permet à chacun d'identifier ses réussites, d'en garder des traces, de percevoir leur évolution. Il est attentif à ce que l'enfant peut faire seul, avec son soutien (ce que l'enfant réalise alors anticipe souvent sur ce qu'il fera seul dans un avenir proche) ou avec celui des autres enfants. Il tient compte des différences d'âge et de maturité au sein d'une même classe. (MEN, 2015)

Ces évolutions vont de pair avec, d'une part, des travaux internationaux, comme ceux portant sur l'introduction des démarches d'investigation à la suite du rapport Rocard (Rocard et al., 2007), qui vise à lutter contre la désaffection des études scientifiques et, d'autre part, avec le développement des évaluations internationales (PISA³, TIMSS⁴). Ces évaluations pointent, pour certains pays, des difficultés des élèves à résoudre des problèmes dans lesquels il faut mobiliser différentes compétences qui ne se

réduisent pas à appliquer des procédures techniques. Depuis une dizaine d'années, plusieurs projets européens de recherche comme Fibonnaci⁵, Primas⁶, S-TEAM⁷ ou LEMA⁸, proposent des ressources pour les enseignants afin de favoriser une plus grande authenticité des questions posées, plus en lien avec celles de la vie courante, une recherche plus ouverte, un aspect expérimental fort avec des retours critiques et une dimension collaborative du travail des élèves. Ces évolutions institutionnelles souhaitées à la fois des pratiques enseignantes ainsi que des activités et apprentissages des élèves ne vont-elles pas de pair avec un renouvellement des questions sur l'évaluation, à la fois sur les différents types et les différentes fonctions ?

En effet, si l'évolution des pratiques des enseignants vise à rendre les élèves plus actifs en proposant des problèmes plus ouverts et moins guidés ainsi que des situations plus en lien avec celles de la vie courante, il nous semble nécessaire que les méthodes et fonctions de l'évaluation évoluent. Par exemple, dans le cas de problèmes d'investigation, évaluer seulement la production de l'élève en tant que produit fini ne donne que très peu d'informations sur le travail réalisé et sur l'avancement des connaissances. De plus, si la tâche est ouverte pour les élèves, des outils de régulation devraient être proposés pour soutenir leur travail.

Enfin, la présence d'injonctions institutionnelles fortes sur les méthodes de différenciation amène à penser des activités plus adaptées à chaque élève et, donc, à affiner les diagnostics et rétroactions.

Bref historique sur les recherches sur l'évaluation formative

Les thèmes des recherches en évaluation ont aussi évolué en intégrant de plus en plus largement l'évaluation formative. Des années 1930 à 1960, les études docimologiques ont été développées, basées sur des modèles de la mesure qui cherchaient à montrer les biais (essentiellement attribués aux correcteurs) dans les notes chiffrées lors d'évaluations sommatives ou certificatives. Le terme d'évaluation formative a été introduit tout d'abord par Scriven (1967), puis adapté par Bloom (1968) aux apprentissages des élèves dans le cadre de la pédagogie de la maîtrise et, enfin, précisé par Bloom, Hastings et Madaus (1971) en distinguant «évaluation formative et évaluation sommative». À cette époque et dans un courant néobehavioriste, l'évaluation formative est plutôt considérée comme un dispositif,

après une séquence d’enseignement et une évaluation sommative, dans une perspective de remédiation, le but étant de reprendre ce qui a été enseigné afin de réduire les écarts.

Nous reproduisons ci-dessous le schéma de Hadji (1989), dans lequel il indique les différents types et fonctions de l’évaluation en se situant dans une perspective temporelle avant/pendant/après la séquence de formation. L’évaluation formative a lieu dans le cours de la séquence de formation et elle vise à réguler les apprentissages en se centrant sur les processus, et non sur les produits.

Tableau 1
Différents types et fonctions de l’évaluation en fonction du temps (Hadji, 1989)

En (1) : AVANT L'ACTION DE FORMATION	En (2) : PENDANT L'ACTION	En (3) : APRÈS L'ACTION DE FORMATION
<p>Évaluation :</p> <ul style="list-style-type: none"> ● Diagnostique ● Pronostique ● Prédictive <p>Fonction :</p> <ul style="list-style-type: none"> ● Orienter ● Adapter <p>Centrée sur :</p> <ul style="list-style-type: none"> ● Le producteur et ses caractéristiques (Identification) 	<p>Évaluation :</p> <ul style="list-style-type: none"> ● Formative ● « Progressive » <p>Fonction :</p> <ul style="list-style-type: none"> ● Réguler ● Faciliter (l'apprentissage) <p>Centrée sur :</p> <ul style="list-style-type: none"> ● Les processus } de pro- ● L'activité } duction 	<p>Évaluation :</p> <ul style="list-style-type: none"> ● Sommative ● Terminale <p>Fonction :</p> <ul style="list-style-type: none"> ● Vérifier ● Certifier <p>Centrée sur :</p> <ul style="list-style-type: none"> ● Les produits

Les informations recueillies lors de l’évaluation, qui peut être formelle (des tests, des questions à choix multiples) ou informelle (dans les interactions), servent à envisager une régulation de l’enseignement en matière de planification ou de différenciation pour soutenir la progression de l’élève. Les régulations peuvent être diverses à la fois dans leurs modes, leurs temporalités ou leurs buts.

Dans les années 1980, un élargissement conceptuel de l’évaluation formative à la notion de régulation est survenu (Allal, 1988 ; Allal et Mottier Lopez, 2005 ; Mottier Lopez, 2015, 2012). Dans les travaux anglophones, la définition de Black et Wiliam (1998) met en avant la notion fondamentale de *feedback* pour le caractère formatif :

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

Pour Shavelson et ses collègues (2008), il y a un continuum entre des évaluations formelles et planifiées (voire des évaluations « prêtes à l'emploi » produites par des chercheurs ou *embedded-in-the-curriculum formative assessments*), celles moins formelles dans l'interaction et celles à la volée (*on the fly*), qui peuvent arriver de façon spontanée à l'initiative de l'enseignant ou d'un élève.

Enfin, Allal (1999b) considère l'importance de l'implication de l'élève dans des démarches d'autoévaluation ou d'évaluation entre pairs, ce qui indique que la gestion de l'évaluation formative ne relève plus seulement de l'enseignant, mais favorise une participation active des élèves, ce qui change les responsabilités dans la classe :

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

C'est aussi la position de Earl (2003), qui distingue l'évaluation des apprentissages (*assessment of learning*), l'évaluation au service de l'apprentissage (*assessment for learning*) et l'évaluation en tant qu'apprentissage (*assessment as learning*), par les buts ou fonctions (jugements sur la production de l'élève, informations pour l'enseignant, autositionnement) et par les acteurs (l'enseignant bien sûr, mais aussi l'élève). Cela permet d'envisager autrement les responsabilités dans la classe.

Tableau 2
Caractéristiques des trois types d'évaluation selon Earl (2003)
 (traduction libre de l'auteure)

Approche	But	Références	Principal évaluateur
Évaluation des apprentissages	Appréciation sur le positionnement, l'avancement, les diplômes, etc.	Autres élèves	Enseignant
Évaluation au service de l'apprentissage	Informations pour les prises de décision de l'enseignant	Normes externes ou attentes	Enseignant
Évaluation en tant qu'apprentissage	Autoévaluation, vérifications ou ajustements	Objectifs personnels et normes externes	Élève

En conclusion de cette partie, nous soulignons donc que des évolutions notables ont eu lieu dans les recherches sur l'évaluation formative, à la fois dans les travaux francophones avec la notion de régulation et dans les travaux anglophones avec le rôle central des *feedbacks*. Cela va de pair avec la prise en compte des théories de l'apprentissage pour l'évaluation, comme le souligne Allal (1999b):

J'ai tenté de montrer en quoi une conception constructiviste de l'évaluation formative exigerait une diversification des modalités de régulation. (Allal, 1999b, p. 37)

De plus, nous pensons que la place centrale accordée à l'élaboration et à l'étude de dispositifs d'évaluation formative à mettre en place dans les classes a été remplacée par celle des fonctions, ce qui permet d'élargir les types d'études. Citons, à titre d'exemple, l'étude des processus de régulation tout au long de séquences d'enseignement-apprentissage en fonction des différents moments de l'étude ou encore celle portant sur la nature et les effets des interactions entre l'enseignant et les élèves.

Des travaux actuels en didactique des mathématiques dans les pays francophones

Nous allons maintenant faire une revue des travaux récents en didactique des mathématiques portant explicitement sur des questions d'évaluation en les classant afin d'obtenir un panorama des recherches actuelles. Pour ce faire, nous avons étudié les actes de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE) depuis que le groupe EVADIDA⁹ a été créé en 2014; les actes du colloque Espace mathématique francophone (EMF) depuis 2009, première année qui comportait un projet spécial sur l'évaluation¹⁰; les quelques thèses en didactique dont le sujet est explicitement l'évaluation; et les deux projets récents *Assess Inquiry in Science, Technology and Mathematics Education (ASSIST-ME)*¹¹ et *Nouveaux outils pour de nouvelles pratiques d'évaluation et d'enseignement des mathématiques (NéOPRÆVAL)*¹², dans lesquels les didacticiens des mathématiques ont été impliqués¹³.

Des travaux sur ou à partir des évaluations à grande échelle

Une première série de travaux portent sur les évaluations à grande échelle internationales (études amorcées par Bodin, 2008) ou nationales, par exemple les études faites par Roditi (2015) ou Roditi et Salles (2015) sur des items du PISA. Ces auteurs ont étudié les problèmes posés grâce aux niveaux de mise en fonctionnement des connaissances proposés par Robert (1998): mise en fonctionnement directe d'une procédure, mise en fonctionnement d'une procédure avec adaptation de l'énoncé et mise en fonctionnement d'une procédure avec introduction d'intermédiaires. Ils ont tenté de mieux évaluer le niveau d'acquisition des connaissances lié à certains items, qui va au-delà du seul taux de réussite, en donnant une interprétation didactique des résultats de ces items:

Un regard didactique porté sur l'évaluation de 2012 ne peut manquer de pointer que l'OCDE ne se donne les moyens ni de recenser précisément les connaissances acquises des élèves ni d'estimer le niveau d'acquisition de ces connaissances. Inversement, les didacticiens qui ont concentré leurs recherches sur les phénomènes d'enseignement et d'apprentissage des savoirs n'ont pas suffisamment développé d'outils théoriques et pratiques pour étudier la question de l'évaluation des connaissances des élèves. (Roditi, 2015, p. 1052)

Des travaux portent également sur les évaluations nationales en France et permettent encore une fois d'affiner les analyses. Par exemple, Grugeon-Allys et Grapin (2015) ont analysé les items des évaluations du Cycle des évaluations disciplinaires réalisées sur échantillon (CEDRE) de 2008 et 2014 pour les thèmes du numérique et de l'algébrique. Ces auteures ont développé des analyses croisées qui portent sur le niveau local (pour chaque tâche : les variables choisies, les techniques possibles, les registres) ou global (couverture du domaine par les différents types de tâches), sur des analyses de la complexité de la tâche déterminée a priori et sur l'indice de difficulté de l'item déterminé a posteriori. La participation de ces auteures ainsi que de Sayac et Grapin (2015) au travail de conception de l'évaluation du CEDRE a permis une évolution des items.

Chesné (2014) développe et analyse un scénario de formation continue des enseignants de mathématiques, toujours à partir des évaluations du CEDRE, montrant comment il est possible de prendre en compte des résultats des évaluations nationales dans la formation des enseignants. Ainsi, pour lui, l'entrée par les résultats des évaluations à grande échelle peut être un levier pour faire évoluer les pratiques des enseignants.

En Belgique, Demonty et ses collègues (2015) prennent appui sur un item inclus dans des ressources proposées par le ministère belge pour évaluer des compétences portant sur une tâche algébrique. Ces auteures remettent en cause le modèle des compétences en trois niveaux de Rey et ses collaborateurs (2003) grâce à une analyse du contenu mathématique en jeu et à une expérimentation sur 163 élèves, ce qui leur permet d'affirmer «l'importance de regards croisés sur la question de l'évaluation de compétences» (Demonty, Fagnant et Dupont, 2015, p. 1).

Ces exemples de travaux qui s'appuient sur des études à grande échelle à partir d'items conçus sans références explicites à des analyses didactiques montrent que les outils didactiques sont adaptés pour analyser les items proposés dans ces types d'évaluation à grande échelle puisqu'ils permettent d'éclairer d'un nouveau jour les analyses réalisées. En particulier, les analyses a priori des savoirs en jeu et l'étude des variables didactiques permettent d'analyser plus finement les procédures ou connaissances disponibles ou encore les difficultés prévisibles des élèves. Les outils de la TAD (Chevallard, 1999), en particulier les praxéologies mathématiques, permettent d'analyser la variété des tâches proposées et de réinterroger les résultats obtenus en termes de réussite et d'échec. Toutes ces études

devraient aider à l'analyse de processus différenciateurs et avoir un potentiel en formation des maîtres. Ainsi, Vantourout et Goasdoué développent l'idée «qu'une conjonction de travaux en didactiques et en psychologie permet de renouveler les débats sur la validité des évaluations» (2014, p. 140). C'est ce qu'ils nomment les approches et la validité psycho-didactiques en recentrant l'évaluation au service des apprentissages :

Tout se passe comme si les analyses a posteriori, qui portent sur des tâches que les chercheurs n'ont pas conçues mais qu'ils «récupèrent», ne pouvaient que déboucher sur des analyses très globales des processus cognitifs, par opposition aux travaux qui reposent sur l'élaboration de tâches et qui de ce fait peuvent faire des hypothèses plus précises sur le fonctionnement cognitif, notamment à travers l'analyse a priori des tâches. (Vantourout et Goasdoué, 2014, p. 144)

Des travaux sur les pratiques d'évaluation des enseignants

Un autre courant de travaux porte sur les pratiques évaluatives des enseignants, que ce soit du primaire (Sayac, 2016) ou du secondaire (Horoks et Pilet, 2015). La théorie des champs conceptuels (Vergnaud, 1981, 1991) ou les notions d'organisations mathématiques (ponctuelles, locales et globales) sont des outils pour penser l'organisation des savoirs enseignés et, pour les évaluations sommatives, les liens entre ce qui est travaillé et ce qui est évalué. Horoks et Pilet (2015) distinguent tâches enseignées et tâches évaluées. Ces auteures essaient de déterminer, dans les pratiques enseignantes, les liens ou les écarts entre ces deux types de tâches proposées aux élèves, ce qui les amène à définir «les pratiques d'évaluation» (Horoks, Kiwan, Pilet, Roditi et Haspekian, à paraître).

La recherche de cours de Chanudet (2015) en Suisse sur le dispositif «Développements en mathématiques», qui consiste à proposer un cours portant sur la résolution de problèmes en tant qu'objet d'enseignement dans lequel l'évaluation a une grande part, a pour but de mieux identifier les savoirs ou savoir-faire en jeu dans la résolution de problèmes. Cela lui permet de reprendre le questionnement suggéré par Georget (2009) ou Choquet-Pineau (2014) sur les contenus des institutionnalisations proposées par les enseignants lors de ces séances. En effet, Choquet-Pineau (2014) indique que les enseignants, lors de séances de résolution de problèmes de type ouvert en France, peinent à dégager ce qui peut être appris de façon décontextualisée, en dehors du problème particulier. Cela met en question, d'une part, ce qui peut ensuite être évalué en termes de savoirs,

de savoir-faire mathématiques ou de savoir-faire plus généraux et, d'autre part, les types de régulations qui peuvent être apportées par l'enseignant (à savoir si elles portent essentiellement sur le problème en jeu ou sur des savoirs plus décontextualisés).

À travers ces thèmes qui portent sur les liens entre ce qui est travaillé dans les séances de classe et les évaluations ou régulations proposées, nous constatons qu'on peut travailler sur la pertinence et la fiabilité des évaluations sommatives, sur les liens entre fonctions sommatives et formatives de l'évaluation, mais aussi sur le processus d'institutionnalisation.

Proposer et utiliser des outils d'évaluation

Les deux projets déjà cités, récemment terminés, NéOPRÆVAL (2014-2018) et ASSIST-ME (2013-2016), ont porté sur la proposition d'outils d'évaluation formative ou diagnostique. Notons que les deux s'appuient sur des travaux faits dans des groupes de recherche collaborative dans le cadre des lieux d'éducation associés (LEA) de l'Institut français de l'éducation (IFE), où les outils produits ont été élaborés lors d'une collaboration entre chercheurs et environ 40 enseignants de mathématiques ou de sciences.

Dans le premier projet, NéOPRÆVAL, une partie du travail porte sur l'adaptation d'évaluations diagnostiques en algèbre à différentes années d'études de collège (Grugeon-Allys, Pilet, Chenevotot et Delozanne, 2012) et sur l'élaboration de parcours différenciés en algèbre. Pour ce faire sont utilisés les travaux de Grugeon (1997), qui définit la compétence algébrique par ses différentes dimensions, et de Pilet (2012), qui a élaboré et implanté en ligne (Delozanne, Previt, Grugeon-Allys et Chenevotot, 2010) des parcours différenciés d'enseignement portant sur l'algèbre au collège en France. Ces évaluations ont été mises en ligne et peuvent être utilisées par les enseignants afin d'obtenir un diagnostic pour chaque élève sur la compétence algébrique, puis de proposer des activités différenciées.

Le second projet, ASSIST-ME, a comme objectif d'analyser l'influence de nouveaux dispositifs d'évaluations formatives, en lien avec les évaluations sommatives dans le cadre de démarches d'investigation, sur les apprentissages et les pratiques enseignantes en sciences (au lycée¹⁴) et en mathématiques (au collège). Une autre partie du travail a été de concevoir et de diffuser des méthodes d'évaluations formatives à mettre en œuvre dans les classes pour les disciplines concernées à différents ordres d'enseignement.

Un premier résultat commun aux deux recherches (même si le nombre d'enseignants concernés est faible) est que la participation des enseignants à de tels projets permet d'avoir des effets sur les outils et les méthodes d'évaluation utilisés dans les classes, notamment en ce qui concerne l'analyse des erreurs, voire l'amélioration des *feedbacks*. Autrement dit, une meilleure prise en compte des élèves dans leurs processus d'apprentissage est notée. Toutefois, l'entrée par l'évaluation permet aussi d'agir sur les pratiques des enseignants en ce qui concerne les contenus à enseigner/enseignés. Trois points sont ainsi mis en avant : une réflexion sur l'éventail des types de tâches proposées lors des séquences d'apprentissage, un changement de responsabilités dans la validation et une attention plus grande aux processus d'institutionnalisation (Chaussecourte, Coppé, Roditi et Trgalova, 2016; Horoks et al., à paraître).

C'est aussi ce que soulignent Ash et Levitt (2003) avec la notion d'évaluation formative en tant que développement professionnel (*formative assessment as professional development*). Ces auteures concluent, grâce à des études réalisées avec des enseignants, que l'intégration des pratiques d'évaluation formative est un levier pour faire évoluer les pratiques enseignantes, notamment par une meilleure prise en compte des apprentissages des élèves de manière individuelle et collective, et ce, en développant un changement de regard sur l'élève et sur ses erreurs ; en rendant plus explicite pour l'enseignant et les élèves ce que doivent savoir et savoir faire ces derniers ; et en pointant les écarts entre ce qui est attendu et ce qui est fait. Autrement dit, nous pensons qu'il faut donc à la fois faire évoluer les pratiques en introduisant des dispositifs d'évaluation formative (cela revient à former les enseignants à davantage prendre en compte ce que font et savent les élèves) et former à l'évaluation formative pour changer les pratiques (sachant que l'entrée par les questions d'évaluation permet de réfléchir aux tâches proposées dans les séquences d'enseignement).

En ce qui concerne plus particulièrement le projet ASSIST-ME, nous avons noté que les enseignants qui ont participé aux expérimentations ont été très volontaires et intéressés par la mise en place, dans leurs classes, d'outils d'évaluations formatives comme les grilles d'autopositionnement (voir Annexe 1, un exemple issu de Gandit et Lepareur, 2017) ou par le dispositif d'évaluation entre pairs, qui a constitué pour eux une pratique vraiment nouvelle. Cependant, l'élaboration et l'utilisation de ces grilles ont nécessité des allers-retours importants entre les enseignants et les chercheurs ; cette pratique a donc besoin d'être soutenue.

En réfléchissant aux conditions de la diffusion, à destination des enseignants, de ressources portant sur l'évaluation formative (autres que des grilles critériées évaluant souvent des compétences générales qui ne prennent pas toujours en compte les savoirs enseignés et appris), nous notons que notre expérimentation a permis de voir la difficulté de cette tâche, notamment parce qu'il est essentiel que l'enseignant ait une bonne connaissance du savoir enseigné, des erreurs et des difficultés des élèves.

Les élèves se sont impliqués dans les activités proposées. Ceux de lycée se sont investis dans les évaluations entre pairs avec des difficultés pour se positionner sur les réponses de leurs camarades (avec quelquefois un positionnement des élèves calqué sur les remarques des enseignants lors des évaluations sommatives : « c'est bien », « pas clair », etc.) et, surtout, pour revenir ensuite sur leurs propres réponses. Nous avons constaté que les élèves considèrent que leur métier d'élève consiste à donner une réponse sans être responsable de sa validité. Ainsi, les responsabilités respectives des élèves et de l'enseignant peinent à évoluer. De plus, dans certaines classes, les élèves ne s'autorisent pas à formuler des jugements sur leurs productions ou sur celles de leurs camarades sans que ceux-ci soient validés par l'enseignant, ce qui rend sa tâche difficile, car il est sans cesse sollicité. Pour les élèves du collège, la difficulté de positionnement s'est révélée plus importante encore, car les arguments disponibles étaient assez pauvres (« c'est vrai parce que c'est vrai » ou « c'est faux parce que je n'ai pas trouvé ça »), ce qui nous amène à penser que les pratiques d'autoévaluation sont certainement peu développées actuellement dans les classes en France. Cela est à mettre en relation avec ce que nous avons déjà montré (Coppé, 1998), à savoir que les pratiques de vérification restaient de l'ordre du travail privé et n'étaient donc pas montrées à l'enseignant.

Ces premiers résultats généraux montrent que donner davantage de responsabilités ou d'autonomie aux élèves, ce qui est bien dans l'esprit des démarches d'investigation, ne va pas de soi ni pour les élèves ni pour les enseignants.

En conclusion, nous reprenons cette citation du rapport de l'Organisation de coopération et de développement économiques (OCDE, 2005) sur l'évaluation formative qui montre bien les enjeux de ce travail à partir de l'entrée par l'évaluation formative :

Le travail formatif appelle de nouvelles techniques d'amélioration du *feedback* entre les élèves et l'enseignant, lesquelles exigent de nouvelles formes de pédagogie et un bouleversement profond des pratiques en classe. (OCDE, 2005, p. 248)

Introduire des évaluations entre pairs

Pour terminer, nous allons décrire rapidement un travail que nous avons mené dans le cadre d'ASSIST-ME pour lequel nous avons choisi de travailler plus particulièrement sur l'évaluation entre pairs (Black, Harrison et Lee, 2004) afin de favoriser l'autoévaluation, dont Allal (1999a) indique qu'elle devrait avoir pour effet l'accroissement de l'autonomie de l'élève et une meilleure adaptation au monde. À travers ces activités d'évaluation des réponses des autres élèves, nous avons également favorisé l'argumentation de façon explicite par notre dispositif.

Méthodologie

Deux expérimentations ont été réalisées. La première en classe de 6^e (élèves de 11-12 ans) sur un problème complexe (c.-à-d. qui met en jeu plusieurs connaissances, plusieurs étapes et sans stratégie de résolution indiquée dans l'énoncé) portant sur les fractions (voir Annexe 2), travaillé sur une séquence de cinq à six séances et testé dans quatre classes de 6^e de la région lyonnaise en France, en 2014 et 2015 (Coppé et Moulin, 2017). La seconde expérimentation a été menée en classe de 4^e ou 3^e (élèves de 14-16 ans) sur un exercice d'algèbre classique (passage d'un programme de calcul à une expression littérale) pendant une séance de classe (voir Annexe 4) et testé dans sept classes de la région lyonnaise (Coppé, Moulin et Roubin, 2017).

Le scénario est identique : les élèves résolvent une partie du problème (ou l'exercice). Puis, l'enseignant collecte toutes les réponses de la classe et demande aux élèves de se prononcer sur la validité de celles-ci par écrit, avant et après un débat argumentatif sur les réponses, pendant lequel l'enseignant ne donne pas d'indications sur leur validité. Il organise les prises de parole et gère le temps. Dans certaines classes, nous avons aussi utilisé des votes à main levée. Nous avons recueilli les écrits des élèves avant et après, dans lesquels ils se positionnent en « vrai » ou en « faux » en argumentant. Nous avons filmé le débat, puis nous avons fait une analyse a priori des stratégies et réponses au problème ainsi qu'une analyse a priori des validations possibles pour chacune des réponses.

Notre recherche avait pour but de déterminer si et comment le positionnement de chaque élève sur chaque réponse évoluait grâce au débat. Nous faisons l'hypothèse selon laquelle l'enrichissement du milieu par l'introduction de toutes les réponses et le changement de contrat sur la responsabilité de la validation pouvaient être un levier pour cette évolution puisque, lorsqu'un élève doit évaluer une autre réponse que la sienne, il peut être amené à mobiliser d'autres connaissances que celles utilisées pour produire une réponse. Cela lui permet, à certaines conditions, de revenir sur sa propre réponse.

Résultats

Nos analyses montrent des résultats contrastés selon le problème. Pour le problème complexe, les élèves se sont impliqués dans sa résolution, dans les évaluations des réponses (avec des difficultés pour se positionner qui peuvent être dues à l'âge des élèves, à la difficulté du problème et au nombre important de réponses) et dans les débats. L'analyse a priori (voir Annexe 3) montrait qu'il était possible, par le biais de la validation des réponses des autres élèves, d'enrichir le milieu et de provoquer ainsi la mobilisation de connaissances autres que celles utilisées pour résoudre le problème. Toutefois, les réponses dépendent de la stratégie initiale (dans ce problème, il y avait deux grandes stratégies) et certaines étaient inaccessibles pour certains élèves, ce qui a entraîné des difficultés dans l'évaluation entre pairs. Nous avons constaté que les positionnements sur les réponses ont peu évolué, voire dans le mauvais sens, malgré des débats tout à fait riches du point de vue des mathématiques. Ainsi, nous pouvons dire que le milieu n'a pas toujours pu remplir sa fonction.

Pour l'exercice d'algèbre, plus simple, les résultats sont plus encourageants puisque les réponses ont évolué favorablement. Ainsi, dans chaque classe, quatre réponses (dont une seule correcte) étaient soumises à l'évaluation des élèves et au débat. Avant le débat, les élèves étaient répartis sur les quatre réponses, tandis qu'après le débat, seuls deux à trois élèves par classe se positionnaient encore sur une réponse erronée. De plus, lors du débat, nous avons observé l'émergence de deux types de thèmes de discussion (amorçés par l'enseignant ou les élèves) portant sur :

- des questions sur des objets de savoir anciens (au sens du temps didactique) explicitement enseignés, par exemple la multiplication des relatifs et les priorités opératoires, ce qui montre que les connaissances des élèves ne sont pas stabilisées, même sur des objets de savoir anciens;
- des demandes de précision sur des objets utilisés par les mathématiciens, par exemple le signe égal (« Il doit y avoir la même chose de chaque côté du signe = ? ») ou la lettre (« On n'est pas obligé de mettre x , on peut mettre une autre lettre ? »), mais aussi sur des propriétés ou savoir-faire (« Est-ce que les priorités opératoires existent en calcul littéral ? », « $5x$, c'est 5 fois x ou 5 plus x ») qui ne sont pas toujours explicités de façon aussi directe, ce qui montre bien tous les implicites qui peuvent faire obstacle. Castela (2008) et Pilet (2012) ont qualifié ces questions de besoins d'apprentissage implicites.

Ces deux expérimentations, certes contrastées, nous semblent prometteuses, d'abord pour développer des travaux sur l'introduction de l'évaluation entre pairs dans les classes : pour quels types de problèmes, à quel moment dans le processus d'enseignement, avec quels outils pour les enseignants, comment les développer ? Ou encore pour effectuer de l'autoévaluation : comment agir sur les responsabilités respectives de l'élève et de l'enseignant, quels types de soutien, quelles régulations au travail de l'élève et de l'enseignant lors des problèmes de recherche ?

Conclusion

Nous avons voulu montrer, à travers ces différents tours d'horizon, que des questions portant sur l'évaluation (à la fois sur les dispositifs et sur les fonctions) peuvent être étudiées en didactique des mathématiques (ou de façon conjointe) avec ses cadres et ses outils, ce qui devrait permettre d'apporter un autre éclairage ou d'affiner des résultats déjà produits. Cela devrait amener, selon nous, des études plus fines sur les apprentissages des élèves prenant en compte la nature des tâches qui leur sont proposées et, certainement, dans une perspective différenciée.

L'entrée par des questions d'évaluation permet de revoir certains concepts, comme ceux de contrat et de milieu. Par exemple, l'entrée par l'évaluation formative permet de reprendre les questions du partage des

responsabilités entre l'enseignant et les élèves (topos de l'élève et contrat didactique). Le concept de milieu peut également être revu dans une perspective dynamique. Aussi, la prise en compte des processus d'évaluation formative ou de régulation renforcerait le rôle de l'enseignant et celui des interactions sur le savoir en jeu. Une piste de recherche est de déterminer comment analyser les interactions entre enseignant et élèves ou entre les élèves dans une perspective didactique.

Réception : 1 décembre 2017

Version finale : 26 mars 2018

Acceptation : 28 avril 2018

NOTES

1. Dans le système éducatif français, « le collège » est l'appellation du 1^{er} cycle des études secondaires. Il dure quatre ans et fait suite à l'école primaire. La fourchette d'âge est généralement de 11-12 ans (début de la 6^e) à 14-15 ans (fin de la 3^e).
2. Il est difficile de dire de façon certaine que c'est sous l'impulsion des travaux sur l'évaluation, mais il est raisonnable de le penser.
3. Programme international pour le suivi des acquis des élèves.
4. *Trends in International Mathematics and Science Study*.
5. www.fibonacci-project.eu
6. www.primas-project.eu
7. http://cordis.europa.eu/result/rcn/56347_en.html
8. www.lema-project.org/web/lemaproject/web/eu/tout.php
9. Le réseau EVADIDA (Évaluations et didactiques) de l'ADMEE a pour objectif premier de viser un rapprochement entre évaluations et didactiques disciplinaires.
10. En 2009, un projet spécial avait été proposé: *Évaluations internationales: impacts politiques, curriculaires et place des pays francophones*. Il avait surtout pour but de présenter les évaluations internationales et d'informer les participants sur le sujet.
11. <http://assistme.ku.dk>
12. www.ldar.website/neopraeval
13. À noter que l'École d'été de didactique des mathématiques d'août 2017 comportait également un thème sur l'évaluation. Puisque les textes sont en cours d'écriture, nous ne les avons pas pris en compte, sauf des cas particuliers d'auteurs déjà cités.
14. Dans le système éducatif français, le lycée correspond au 2^e cycle des études secondaires. Ce 2^e cycle peut être général, technologique ou professionnel. Il correspond aux trois dernières années de l'enseignement secondaire (seconde, première et terminale), pour des adolescents âgés de 15 à 17, voire 18 ans (du début de la seconde à la fin de la terminale).

BIBLIOGRAPHIE

- Allal, L. (1988). Vers un élargissement de la pédagogie de maîtrise : processus de régulation interactive, rétroactive et proactive. Dans A.-M. Huberman (dir.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise* (pp. 86-126). Lonay, Suisse : Delachaux et Niestlé.
- Allal, L. (1999a). Acquisition et évaluation des compétences en situation scolaire. Dans J. Dolz et E. Ollagnier (dir.), *L'énigme de la compétence en éducation* (pp. 77-94). Bruxelles : De Boeck.
- Allal, L. (1999b). Impliquer l'apprenant dans le processus d'évaluation : promesses et pièges de l'autoévaluation. Dans C. Depover et B. Noël (dir.), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes* (pp. 35-56). Bruxelles : De Boeck.
- Allal, L. et Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en langue française. Dans Centre pour la recherche et l'innovation dans l'enseignement (dir.), *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires* (pp. 265-290). Paris : OCDE.
- Amade-Escot, C. et Venturini, P. (2009). Le milieu didactique : d'une étude empirique en contexte difficile à une réflexion sur le concept. *Éducation et didactique*, 3(1), 7-43. doi: 10.4000/éducation didactique.407
- Ash, D., & Levitt, K. (2003). Working within the zone of proximal development: Formative assessment as professional development. *Journal of Science Teacher Education*, 14(1). doi: 10.1023/A:1022999406564
- Black, P., Harrison, C., & Lee, C. (2004). Working inside the black box: Assessment for learning in the classroom. London: Granada Learning.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74. doi: 10.1080/0969595980050102
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1-12. Retrieved from <https://eric.ed.gov/?id=ED053419>
- Bloom, B. S., Hastings, J. T., & Madaus, G. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Bodin, A. (1997). L'évaluation du savoir mathématique : questions et méthodes. *Recherches en didactique des mathématiques*, 17(1), 49-96.
- Bodin, A. (2008). Lecture et utilisation de PISA pour les enseignants. *Petit x*, 78, 53-78. Repéré à www-irem.ujf-grenoble.fr/revues/revue_x/fic/78/78x4.pdf
- Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en didactique des mathématiques*, 7(2), 33-116.
- Brousseau, G. (1995, août). L'enseignant dans la théorie des situations didactiques. Dans R. Noirfalise et M.-J. Perrin Glorian (dir.), *Actes de la 6^e École d'été de didactique des mathématiques Saint-Sauves d'Auvergne*, 3-46.
- Brousseau, G. (1998). *La théorie des situations didactiques*. Grenoble : La Pensée sauvage.

- Castela, C. (2008). Travailler avec, travailler sur la notion de praxéologie mathématique pour décrire les besoins d'apprentissage ignorés par les institutions. *Recherches en didactique des mathématiques*, 28(2), 135-82.
- Chanudet, M. (2015, octobre). Questions posées par l'évaluation d'activités de résolution de problèmes : le cas particulier du cours de « développements en mathématiques » au cycle d'orientation à Genève. Dans L. Theis (dir.), *Pluralités culturelles et universalité des mathématiques : enjeux et perspectives pour leur enseignement et leur apprentissage* (pp. 837-846). Actes du colloque EMF 2015, Alger.
- Chaussecourte, P., Coppé, S., Roditi, E. et Trgalova, J. (2016). Collectifs de chercheurs ou collectifs d'enseignants engagés dans une recherche : effets sur leurs pratiques respectives. Dans Y. Matheron, G. Gueudet, V. Celi, C. Derouet, D. Forest, M. Krysinska et S. Besnier (dir.), *Enjeux et débats en didactique des mathématiques* (p. 529-550). Grenoble : La Pensée sauvage.
- Chesnais, A. (2009). *L'enseignement de la symétrie axiale en sixième dans des contextes différents : les pratiques de deux enseignants et les activités des élèves* (Thèse de doctorat). Université Paris-Diderot, Paris. Repéré à <https://halshs.archives-ouvertes.fr/tel-00450402/>
- Chesnè, J.-F. (2014). *D'une évaluation à l'autre : des acquis des élèves sur les nombres en sixième à l'élaboration et à l'analyse d'une formation d'enseignants centrée sur le calcul mental* (Thèse de doctorat). Université Paris-Diderot, Paris. Repéré à <https://tel.archives-ouvertes.fr/tel-01081505>
- Chevallard, Y. (1986). Vers une analyse didactique des faits d'évaluation. Dans J.-M. De Ketele (dir.), *L'évaluation : approche descriptive ou prescriptive?* (pp. 31-59). Bruxelles : De Boeck.
- Chevallard, Y. (1989). Évaluation, vérification, objectivation : la relation didactique comme caprice et miniature. Dans J. Colomb et J. Marsenach (dir.), *L'évaluateur en révolution* (pp. 13-36). Paris : INRP.
- Chevallard, Y. (1999). L'analyse des pratiques enseignantes en théorie anthropologique du didactique. *Recherches en didactique des mathématiques*, 19(2), 221-265.
- Chevallard, Y. et Feldmann, S. (1986). *Pour une analyse didactique de l'évaluation*. Marseille : IREM d'Aix-Marseille.
- Choquet-Pineau, C. (2014). *Une caractérisation des pratiques de professeurs des écoles lors de séances de mathématiques dédiées à l'étude de problèmes ouverts au cycle* (Thèse de doctorat). Université de Nantes, Nantes.
- Coppé, S. (1993). *Processus de vérification en mathématiques chez les élèves de première scientifique en situation de devoir surveillé* (Thèse de doctorat). Université Claude-Bernard Lyon I, Lyon.
- Coppé, S. (1998). Composantes privées et publiques du travail de l'élève en situation de devoir surveillé de mathématiques. *Educational Studies in Mathematics*, 35(2), 129-151. doi: 10.1023/A:1003086628291
- Coppé, S. (2010, juin). Les problèmes dans les programmes depuis la contre réforme : quelles évolutions ? Dans B. Grugeon-Allys (dir.), *Actes du 17^e colloque de la COR-FEM*, Besançon, 9-16. Repéré à <https://halshs.archives-ouvertes.fr/halshs-00960313/document>

- Coppé, S. et Houdement, C. (2009, juin). Résolution de problèmes à l'école primaire : perspectives curriculaire et didactique. Dans C. Houdement, C. Buff, P. Danos, P. Masselot, C. Ouvrier-Bufferet, M.-J. Perrin-Glorian et A. Simard (dir.), *L'enseignement des mathématiques à l'école : où est le problème* (pp. 48-71). Actes du 37^e colloque de COPIRELEM, Auch. Repéré à www.arpeme.fr/documents/1060A7EB0E663362251D.pdf
- Coppé, S. et Moulin, M. (2017). Évaluation entre pairs et débat argumenté dans le cadre d'un problème complexe en mathématiques. *Canadian Journal of Science, Mathematics and Technology Education*, 17(4), 308-327. doi: 10.1080/14926156.2017.1378832
- Coppé, S., Moulin, M. et Roubin, S. (2017, janvier). Analyses de l'évolution de réponses d'élèves lors d'évaluations formatives dans le cadre de la résolution d'un problème complexe. Dans *L'évaluation : levier pour l'enseignement et la formation – Résumé* (pp. 32-34). Actes du 29^e colloque international de l'ADMEE-Europe, Dijon.
- De Ketele, J.-M. (1986). *L'évaluation, approche descriptive ou prescriptive?* Bruxelles : De Boeck.
- Delozanne, E., Previt, D., Grugeon-Allys, B. et Chenevotot, F. (2010). Vers un modèle de diagnostic de compétence. *Techniques et sciences informatiques*, 29(8-9), 899-938. doi: 10.3166/tsi.29.899-938
- Demonty, I., Fagnant, A. et Dupont, V. (2015). Analyse d'un outil d'évaluation en mathématiques : entre une logique de compétences et une logique de contenu. *Mesure et évaluation en éducation*, 38(2), 1-29. doi:10.7202/1036761ar
- Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Gandit, M. et Lepareur, C. (2017, janvier). Effets de différentes modalités d'évaluation formative sur l'évolution du milieu de l'élève. Dans *L'évaluation : levier pour la formation et l'enseignement* (pp. 725-731). Actes du 29^e colloque international de l'ADMEE-Europe, Dijon.
- Georget, J.-P. (2009). *Activités de recherche et de preuve entre pairs à l'école élémentaire : perspectives ouvertes par les communautés de pratique d'enseignants* (Thèse de doctorat). Université Paris-Diderot, Paris.
- Grugeon, B. (1997). Conception et exploitation d'une structure d'analyse multidimensionnelle en algèbre élémentaire. *Recherches en didactique des mathématiques*, 17(2), 167-210.
- Grugeon-Allys, B., Pilet, J., Chenevotot, F. et Delozanne, E. (2012). Diagnostic et parcours différenciés d'enseignement en algèbre élémentaire. *Recherches en didactique de mathématiques*, « Enseignement de l'algèbre, bilan et perspectives », hors série, 137-162.
- Grugeon-Allys, B. et Grapin, N. (2015, octobre). Méthodologie d'analyse d'évaluations externes. Dans L. Theis (dir.), *Pluralités culturelles et universalité des mathématiques : enjeux et perspectives pour leur enseignement et leur apprentissage* (pp. 1017-1027). Actes du colloque EMF 2015, Alger.
- Hadji, C. (1989). *L'évaluation, règles du jeu*. Paris : ESF.
- Horoks, J. (2006). *Les triangles semblables en classe de 2^{nde} : des enseignements aux apprentissages – Étude de cas* (Thèse de doctorat). Université Paris-Diderot, Paris. Repéré à <https://hal.archives-ouvertes.fr/tel-01136889/document>

- Horoks, J., Kiwan, M., Pilet, J., Roditi, E. et Haspekian, M. (à paraître). Régulations des apprentissages et évaluation formative : quels regards didactiques ? *Actes de la 19^e École d'été de didactique des mathématiques*, Paris.
- Horoks, J. et Pilet, J. (2015, octobre). Étudier et faire évoluer les pratiques d'évaluation des enseignants de mathématiques en algèbre au collège dans le cadre d'un LEA. Dans L. Theis (dir.), *Pluralités culturelles et universalité des mathématiques : enjeux et perspectives pour leur enseignement et leur apprentissage* (pp. 791-804). Actes du colloque EMF 2015, Alger.
- Margolinas, C. (1992). Éléments pour l'analyse du rôle du maître : les phases de conclusion. *Recherches en didactique des mathématiques*, 12(1), 113-158.
- Ministère de l'Éducation nationale [MEN] (2004). *Programmes des collèges : mathématiques – Introduction générale pour le collège*. Bulletin officiel du ministère de l'Éducation nationale, hors série, n° 5.
- Ministère de l'Éducation nationale [MEN] (2008). *Programmes des collèges : programme de l'enseignement de mathématiques*. Bulletin officiel spécial du ministère de l'Éducation nationale, n° 6.
- Ministère de l'Éducation nationale [MEN] (2015). *Annexe: programme d'enseignement de l'école maternelle*. Bulletin officiel spécial du Ministère de l'Éducation nationale, n° 2.
- Mottier Lopez, L. (2012). *La régulation des apprentissages en classe*. Bruxelles : De Boeck.
- Mottier Lopez, L. (2015). *Évaluations formative et certificative des apprentissages : enjeux pour l'enseignement*. Bruxelles : De Boeck Université.
- Organisation de coopération et de développement économiques [OCDE] (2005). *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires*. Paris : OCDE.
- Perrin-Glorian, M.-J. (1999). Problèmes d'articulation de cadres théoriques : l'exemple du concept de milieu. *Recherches en didactique des mathématiques*, 19(3), 279-321.
- Perrin-Glorian, M.-J. et Hersant, M. (2003). Milieu et contrat didactique, outils pour l'analyse de séquences ordinaires. *Recherches en didactique des mathématiques*, 23(2), 217-276.
- Pilet, J. (2012). *Parcours d'enseignement différencié en algèbre élémentaire* (Thèse de doctorat). Université Paris-Diderot, Paris.
- Rey, B., Carette, V., Defrance, A. et Kahn, S. (2003). *Les compétences à l'école : apprentissage et évaluation*. Bruxelles : De Boeck.
- Robert, A. (1998). Outil d'analyse des contenus mathématiques à enseigner au lycée et à l'université. *Recherches en didactique des mathématiques*, 18(2), 139-190.
- Robert, A. et Rogalski, J. (2002). Le système complexe et cohérent des pratiques des enseignants de mathématiques : une double approche. *Canadian Journal of Science, Mathematics and Technology Education*, 2(4), 505-528. doi: 10.1080/14926150.209556538
- Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science education now: A renewed pedagogy for the future of Europe*. Brussels: European Commission.

- Roditi, E. (2015, octobre). Didactique et évaluation : un nouveau regard sur le PISA 2012. Dans L. Theis (dir.), *Pluralités culturelles et universalité des mathématiques : enjeux et perspectives pour leur enseignement et leur apprentissage* (pp. 1042-1053). Actes du colloque EMF 2015, Alger.
- Roditi, E. et Salles, F. (2015). Nouvelles analyses de l'enquête PISA 2012 en mathématiques : un autre regard sur les résultats. *Éducation et formations*, 86-87, 235-255. Repéré à http://cache.media.education.gouv.fr/file/revue_86-87/56/9/depp-2015-EF-86-87-nouvelles-analyses-enquete-PISA-2012-mathematiques_424569.pdf
- Roegiers, X. (2004). *L'école et l'évaluation : des situations pour évaluer les compétences des élèves*. Bruxelles : De Boeck.
- Sayac, N. (2016, janvier). Nature et diversité des tâches mathématiques proposées en évaluation sommative par des professeurs des écoles, en France : résumé (pp. 609-611). *Actes du 28^e colloque de l'ADMEE-Europe*, Lisbonne.
- Sayac, N. et Grapin, N. (2015). Évaluation externe et didactique des mathématiques : un regard croisé nécessaire et constructif. *Recherches en didactique des mathématiques*, 35(1), 101-126.
- Schubauer-Leoni, M. L., Leutenegger, F., Ligozat, F. et Fluckiger, A. (2007). Un modèle de l'action conjointe professeur-élèves : les phénomènes didactiques qu'il peut/doit traiter. Dans G. Sensevy et A. Mercier (dir.), *Agir ensemble : l'action didactique conjointe du professeur et des élèves* (pp. 51-91). Rennes : PUR.
- Scriven, M. (1967). The methodology of evaluation. In R. Mills Gagné, R. W. Tyler, & M. Scriven, *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Sensevy, G. et Mercier, A. (dir.). (2007). *Agir ensemble : l'action didactique conjointe du professeur et des élèves*. Rennes : PUR.
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., ... Yin, Y. (2008). On the role and impact of formative assessment on science inquiry teaching and learning. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing science learning: Perspectives from research and practice* (pp. 295-314). Arlington, VA: NSTA Press.
- Vantourout, M. et Goasdoué, R. (2014). Approches et validité psycho-didactiques des évaluations. *Éducation et formation*, 302, 139-156. Repéré à <https://halshs.archives-ouvertes.fr/halshs-01239551/document>
- Vergnaud, G. (1981). *L'enfant, la mathématique et la réalité : problèmes de l'enseignement des mathématiques à l'école élémentaire*. Berne : P. Lang.
- Vergnaud, G. (1991). Théorie des champs conceptuels. *Recherches en didactique des mathématiques*, 10(23), 133-170.

Annexe 1 : Tableau de positionnement sur la résolution de problème
(Gandit et Lepareur, 2017)

	1	2	3	4
	J'ai bien compris ce qu'on cherchait, j'ai fait des essais, j'ai eu des idées.	J'ai compris ce qu'on cherchait, mais j'ai fait des erreurs dans mes recherches, j'ai eu des idées.	J'ai compris ce qu'on cherchait, mais je n'ai pas su comment démarrer. Je n'avais pas d'idée.	Je n'ai pas compris ce qu'on cherchait, je n'ai pas su démarrer.
Comprendre le problème, faire des recherches, faire preuve d'initiative, être original	J'ai donné de bonnes idées pour l'affiche, j'ai reconnu des erreurs dans les autres affiches.	J'ai donné de bonnes idées pour l'affiche, je n'ai pas vu d'erreurs dans les autres affiches.	Je n'ai pas vraiment trouvé d'idées pour les affiches, ni de questions à poser. J'ai voté au hasard.	Je n'ai pas eu d'idées et je n'ai pas voté.
Donner des réponses, émettre des hypothèses ou conjectures, formuler des questions	J'ai trouvé, plusieurs fois, des raisonnements corrects pour prouver qu'une réponse était vraie ou bien était fausse.	J'ai trouvé, une seule fois, un raisonnement correct pour prouver qu'une réponse était vraie ou bien fausse.	Je n'ai trouvé aucun raisonnement correct pour prouver qu'une réponse était vraie ou bien fausse.	Je n'ai pas cherché de raisonnement.
Prouver que c'est vrai, prouver que c'est faux	Ce que j'ai écrit dans l'affiche était correct, je suis intervenu(e) dans le débat pour répondre aux autres.	J'ai participé à la rédaction de l'affiche, il y avait des erreurs, je suis intervenu(e) dans le débat pour répondre aux autres.	J'ai participé à la rédaction de l'affiche, mais c'était faux, je ne suis pas intervenu(e) dans le débat pour répondre aux autres.	Je n'ai pas participé à la rédaction de l'affiche, je ne suis pas intervenu(e) dans le débat.
Communiquer par écrit ou par oral				

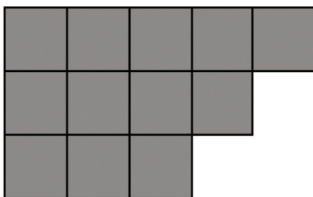
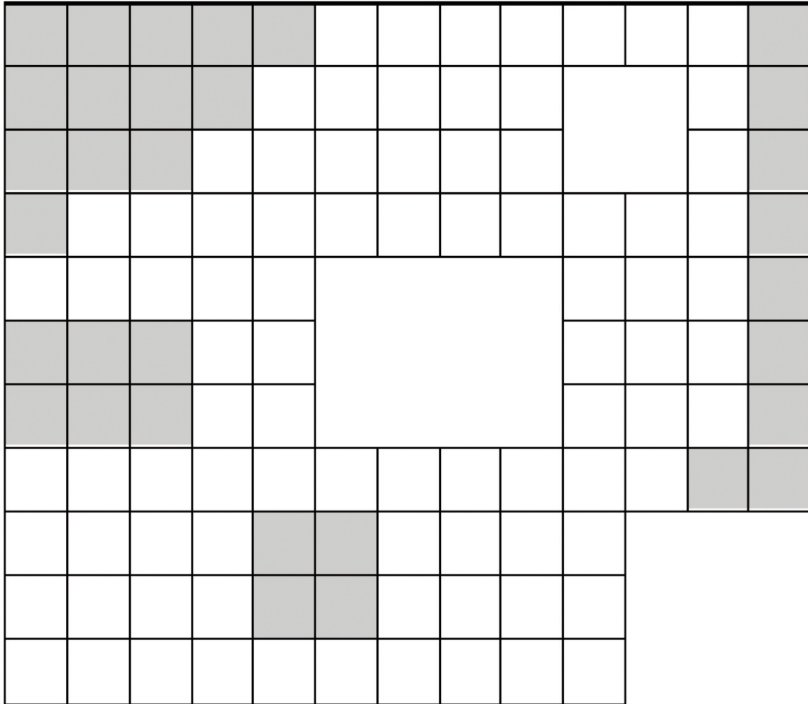
Annexe 2: Problème portant sur les fractions posé en 6^e

Le parc

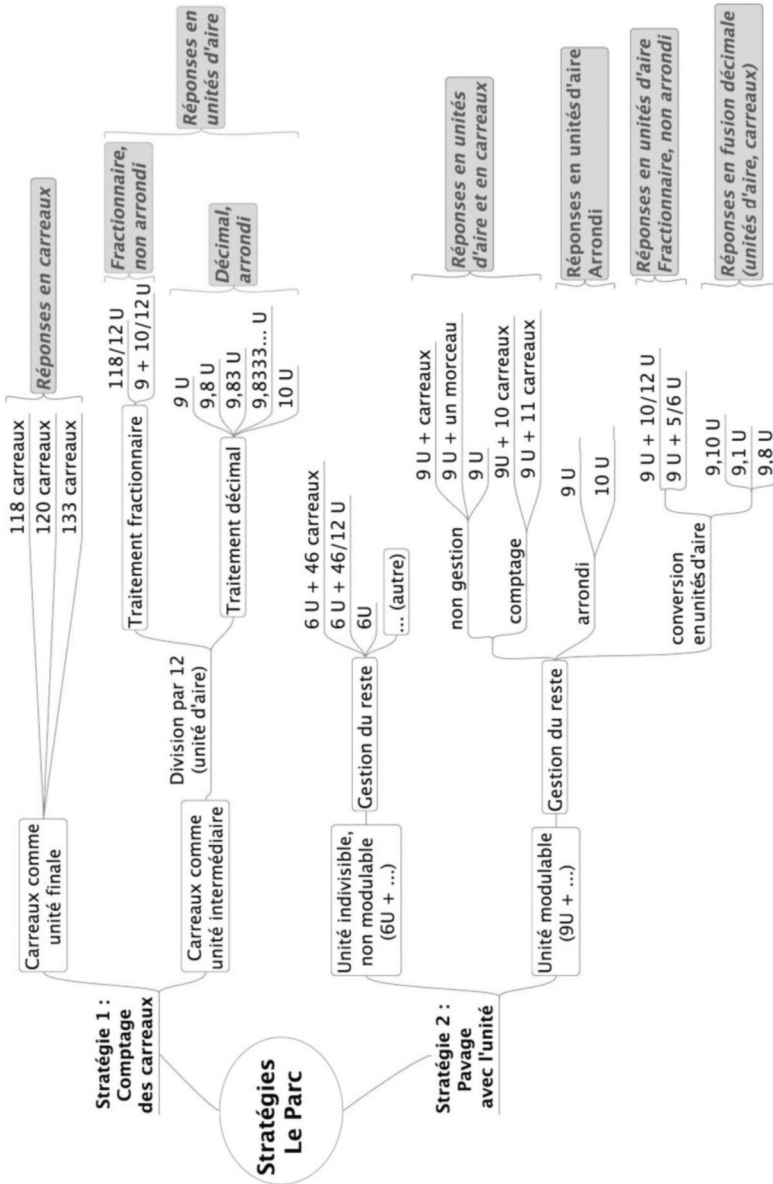
La grille représente un parc, avec des massifs de fleurs, dans lequel se trouvent des bâtiments :

- Les massifs de fleurs sont représentés en gris clair.
- Les trous dans la grille représentent les bâtiments.

On veut mesurer l'aire du parc (sans les bâtiments) avec une certaine unité d'aire. Cette unité d'aire est l'aire de la surface en gris foncé qui se trouve en dessous de la grille.



Annexe 3: Problème portant sur les fractions posé en 6^e



Annexe 4: Problème d'algèbre classique posé en classes de 4^e et 3^e

Voici un programme de calcul.

Choisir un nombre.

Ajouter 4.

Multiplier la somme par 5.

Soustraire 8 au résultat.

Quelle expression littérale décrit ce programme de calcul?