

Cultures et politiques de l'évaluation en éducation et en formation

COMPETENCES DE FIN D'ÉCOLE EN MATHÉMATIQUES EN FRANCE : ANALYSE DIDACTIQUE D'UNE ÉVALUATION EXTERNE

Nathalie Sayac*, Nadine Grapin*

* Université Paris Est, Laboratoire de Didactique André Revuz (LDAR-EA 4434)
nathalie.sayac@u-pec.fr

** Université Paris Est, Laboratoire de Didactique André Revuz (LDAR-EA 4434)
nadine.grapin@u-pec.fr

Mots-clés : *Compétences, mathématiques, QCM, évaluation externe.*

Résumé. *En France, depuis 2003 des bilans externes sont organisés en fin d'école et de collège et ont pour but de faire un état des lieux aussi objectif que possible des compétences et des connaissances des élèves. Dans les faits, si l'objectif affiché de ces évaluations est d'évaluer à la fois des compétences et des connaissances, tout se passe comme si évaluer des compétences se situait au même niveau qu'évaluer des connaissances, or nous avons estimé qu'il y avait là un raccourci problématique que nous avons souhaité lever. Nous avons donc conçu un outil d'analyse d'items, inspirés des travaux en didactique des mathématiques et prenant en compte plusieurs facteurs déclinés en termes de complexité du point de vue de la tâche mathématique et en termes de compétences, par niveaux. Cet outil nous a permis de constater que les items proposés dans cette évaluation ne pouvaient pas réellement témoigner du niveau de connaissances et de compétences des élèves de fin d'école en France, notamment parce qu'ils ne permettaient pas d'explorer réellement leur périmètre.*

L'évaluation des connaissances des élèves dans une discipline, par une institution nationale ou internationale à un moment donné du cursus scolaire se pratique depuis des décennies en France et contribue au pilotage du système éducatif. Comme la didactique des mathématiques s'intéresse aux relations entre les savoirs, les élèves et les enseignants, on peut considérer que le sous-système constitué des savoirs enseignés est évalué à l'occasion d'évaluations nationales, notamment à travers les productions des élèves, au regard des programmes scolaires (Roditi, 2012). Alors que Chevallard (1986) avait indiqué que son regard d'anthropologue didacticien lui faisait percevoir les « faits d'évaluation » comme des objets potentiels d'étude, une dizaine d'années plus tard, Bodin (1997) déplorait encore le « peu de recherches en didactique des mathématiques qui prennent en compte l'existence des faits d'évaluation [...] et le peu de recherches sur l'évaluation prennent en compte la spécificité des savoirs en jeu ». Actuellement, la question de l'évaluation des élèves, aussi bien en classe que par le biais d'évaluations externes commence à faire l'objet d'une attention plus importante de la part des chercheurs en didactique des mathématiques : par exemple, sur l'évaluation de compétences (Artigue & Winslow, 2010) ou à partir de l'évaluation internationale PISA (Deblois, Freiman & Rousseau, 2007).

Chercheuses en didactique des mathématiques, nous avons été sollicitées pour participer à la conception et l'analyse des évaluations bilans en mathématiques de fin d'école pour l'une, et de fin de collège pour l'autre, menées par la DEPP (Direction de l'évaluation, de la prospective et de la performance) de septembre 2005 à juin 2009. Il convient de préciser que nous n'avons pas été sollicitées en tant que chercheuses, mais en tant que formatrices pour compléter des équipes pluri-catégorielles d'experts composées de professeurs de terrain, de conseillers pédagogiques, d'Inspecteurs de l'éducation nationale (IEN) et d'Inspecteurs d'Académie-Inspecteurs Pédagogiques Régionaux (IA-IPR).

Cultures et politiques de l'évaluation en éducation et en formation

A la suite de ce travail, il nous a semblé pertinent d'interroger le contenu de ces évaluations et certains de leurs résultats à partir des outils et travaux développés en didactique des mathématiques. En effet, lors de la conception de ces évaluations, aucune analyse didactique des tâches n'est effectuée précisément ; or, un tel travail nous semble important pour apporter des éléments relatifs à la validation du contenu (Grégoire et Laveault, 2005) et expliciter ce qui est effectivement évalué pour mieux caractériser les connaissances des élèves en mathématiques lors de l'analyse des résultats.

C'est ce travail que nous vous présentons dans cet article. Nous exposerons dans un premier temps les conditions et le cadre de ces évaluations externes, puis nous préciserons notre questionnement, l'outil d'analyse que nous avons développé et les conclusions faites sur le contenu des évaluations bilans de la DEPP. Nous compléterons cette présentation par des résultats généraux que nous avons obtenus à partir d'une évaluation complémentaire portant sur les QCM. Nous concluons en proposant des outils d'aide à la conception d'évaluations plus équilibrées du point de vue des connaissances et des compétences.

1. Cadre des évaluations CEDRE

Les évaluations bilans de la DEPP menées en mathématiques à la fin de l'école et à la fin du collège ont pour objectif de mesurer les connaissances et compétences des élèves à ces deux moments de leur scolarité, en lien avec les programmes en cours (programmes scolaires et Socle commun de connaissance et de compétences). Elles sont menées dans le cadre d'un cycle (CEDRE : Cycle des évaluations disciplinaires réalisées sur échantillon) débuté en 2003 qui concerne d'autres disciplines que les mathématiques et qui se reproduit tous les 6 ans : l'évaluation bilan fin d'école menée en 2008 en mathématiques sera ainsi reconduite en 2014.

Ces évaluations bilans ont été analysées d'un point de vue institutionnel et ont donné lieu à la diffusion de notes d'information et d'une publication (Lescure & Pastor, 2012). A partir des productions des élèves et grâce à des outils statistiques, la DEPP a construit une échelle de performance permettant de catégoriser six groupes d'élèves, hiérarchisés selon leurs connaissances mathématiques et les compétences qu'ils maîtrisent. Pour le primaire, les items proposés en mathématiques ont permis de mesurer des connaissances relevant de six domaines (connaissances des nombres entiers naturels, fractions et décimaux, calcul, exploitation de données numériques, espace et géométrie, grandeurs et mesures) et de cinq compétences (identifier des notions, exécuter un calcul, traiter des données, produire en autonomie, contrôler - valider).

Comme pour de nombreuses évaluations institutionnelles, une forte proportion de questions du bilan CEDRE a été proposée sous forme de QCM afin de faciliter le traitement des réponses et la correction. Deux tiers des questions ont été posées sous cette forme, et le tiers restant, correspondant à des questions ouvertes, a porté soit sur du calcul mental soit sur les items relevant de la compétence « produire ».

2. Cadre de la recherche

Par delà la volonté que nous avons de croiser les travaux de la DEPP avec les travaux issus de recherche en didactique des mathématiques, il va sans dire que l'idée de travailler sur des données d'élèves de fin d'école, émanant d'un échantillon représentatif à l'échelle nationale d'une telle ampleur (3812 élèves), nous a grandement enthousiasmées. Par ailleurs, nous avons également la volonté d'utiliser cette recherche pour la formation des enseignants du premier degré en mathématiques, car il nous semble primordial de penser et de construire des liens entre la recherche et la formation.

Afin de circonscrire les domaines sur lesquels nous allions travailler, nous avons choisi de nous focaliser sur deux domaines spécifiques ayant fait l'objet de nombreux travaux en didactique des

Cultures et politiques de l'évaluation en éducation et en formation

mathématiques : les « fractions & décimaux » et « grandeurs & mesures ». Cette dernière notion a d'autant plus retenu notre attention que les résultats entre les évaluations-bilans école et collège y étaient très contrastés¹.

2.1 Problématique

Les évaluations bilans de la DEPP ont été conçues avec l'objectif de « faire un point aussi objectif que possible sur les compétences et les connaissances des élèves dans des domaines essentiels, à des moments clefs du cursus scolaire ». (Note d'information n°10 - 17, 2010). C'est un objectif que nous pouvons tout à fait partager dans l'absolu, mais en tant que didacticiennes des mathématiques, nous avons conscience qu'une évaluation externe ne peut en aucun cas rendre compte des apprentissages vraiment réalisés par des élèves, car comme le dit Winslow (2005) « on ne peut pas indiquer les résultats de l'enseignement de façon à la fois vérifiable dans le temps et par rapport aux individus et aux situations ». Il nous a donc semblé important de prouver cette réalité, à partir d'un travail approfondi sur les items constituant cette évaluation. Nous avons par ailleurs souhaité exploiter les résultats de ces évaluations bilans pour inférer des constats au niveau des apprentissages des élèves et pour déterminer quel reflet de la connaissance des élèves était produit par les items proposés, et quels apprentissages étaient réellement révélés.

La forme même de l'évaluation nous a également interpellées car il nous est apparu au cours de l'élaboration des items et l'analyse des résultats que la nature des items (QCM, questions ouvertes/fermées, etc.) n'était pas sans incidence sur les réponses données par les élèves. Nous avons donc souhaité vérifier cette hypothèse en analysant plus précisément les consignes et les tâches rattachées aux items.

Pour mener à bien cette exploration, nous avons décidé d'élaborer un outil permettant d'analyser finement les items d'évaluation proposés, en étant au plus proche de la réalité des connaissances des élèves ; nous avons cherché à rendre cet outil fonctionnel pour qu'il puisse être utilisé par des non spécialistes (en particulier les membres du groupe d'experts de la DEPP), mais aussi pour que des enseignants, en formation initiale et continue, puissent s'en emparer.

2.2 Construction d'un outil d'analyse pour les évaluations en mathématiques

2.2.1 Références théoriques

Pour analyser les items des évaluations bilans en vue de repérer les apprentissages qui s'y rattachaient, nous avons recherché des travaux menés en didactique des mathématiques et nous avons retenu les travaux relatifs à la notion de tâche en théorie de l'activité et à la taxonomie des opérations cognitives et les résultats de la didactique portant sur la notion mathématique traitée.

Concernant l'analyse des tâches, nous nous sommes inspirées des travaux menés en didactique des mathématiques par Robert et Rogalski (2002) sur les différents niveaux de mise en fonctionnement des connaissances. En effet, il nous a semblé important de distinguer, les tâches qui amènent à des applications immédiates des connaissances, c'est-à-dire simples (sans adaptation) et isolées (sans mélanges), où seule une connaissance précise est mise en œuvre sans aucune adaptation, mis à part la contextualisation nécessaire (niveau de mise en fonctionnement technique), des tâches qui nécessitent des adaptations de connaissances qui sont en partie au moins indiquées (niveau de mise en fonctionnement mobilisable), de celles où l'élève doit reconnaître les connaissances à utiliser, sans qu'il y ait d'indication spécifique (niveau de mise en fonctionnement disponible).

Par ailleurs, les travaux de Duval (1993), repris par Gagatsis, Elia et Mougi (2002) sur les registres de représentation sémiotiques de concepts mathématiques nous ont paru également pertinents pour nous éclairer sur les difficultés que peuvent rencontrer les élèves confrontés à différents types de

¹ Pas d'inquiétude particulière en fin d'école alors que les résultats de fin de collège étaient très alarmants.

Cultures et politiques de l'évaluation en éducation et en formation

tâches. La distinction entre une tâche qui implique une reconnaissance (quand il s'agit de reconnaître ou d'identifier une représentation d'un objet mathématique dans un système déterminé), une tâche qui implique un traitement (quand il s'agit d'effectuer une tâche en restant dans le même registre de représentation) et une tâche qui implique une conversion (quand il s'agit de convertir quelque chose proposé dans un registre dans un autre registre) nous a semblé opportune pour analyser les items proposés dans une évaluation de fin d'école.

Du point de vue des connaissances en jeu dans les items relatifs aux « fractions-décimaux », nous avons exploité les travaux de Brousseau (1980), Douady et Perrin-Glorian (1986), Comiti et Neyert (1979), Roditi (2007), ainsi que les travaux de Moreira-Baltar (1996) et Perrin-Glorian (1990) pour les « grandeurs et mesures ». Ces travaux ont mis en évidence certaines difficultés rencontrées par les élèves notamment lors du passage d'une écriture fractionnaire à une écriture à virgule (et réciproquement) et des conceptions erronées telles que la conception du nombre décimal comme juxtaposition de deux entiers séparés par une virgule qui font obstacle à la réussite de certaines tâches. De même que la confusion entre aire et périmètre, les difficultés liées aux désignations usuelles des aires ou à la configuration des figures sont des éléments à prendre en compte pour analyser les tâches proposées en « grandeurs & mesures ».

Les travaux de Gras, repris par Bodin (2004) sur les taxonomies des énoncés de problèmes mathématiques permettent de classer par niveaux hiérarchisés de complexité cognitive, l'activité des élèves. En complément de cette taxonomie, les travaux canadiens réalisés dans le cadre de l'EIACA sur la « numératie des adultes et son évaluation » (2003) ont permis de dégager un organigramme de complexité, décliné en cinq facteurs de complexité (répartis en deux ensembles : deux facteurs qui concernent surtout les aspects textuels des tâches et trois facteurs qui concernent les aspects mathématiques) qui nous a semblé également intéressant pour analyser les items des évaluations bilans de la DEPP. Ces facteurs prennent en compte non seulement la complexité des informations ou des données mathématiques, le type et le nombre d'opérations ou de compétences à mettre en jeu pour répondre au problème posé, mais aussi le type d'appariement et la transparence du problème traité, la plausibilité des distracteurs : la prise en compte de ces derniers facteurs dans l'analyse de la complexité d'un exercice nous paraît nécessaire et complémentaire à une analyse de tâche menée uniquement dans un cadre didactique.

2.2.2 Définition de l'outil

Pour analyser les items dans une perspective de repérage des connaissances réellement en jeu pour les réussir, nous avons conçu un outil faisant la synthèse des différents travaux présentés ci-dessus et adapté à la forme et la nature des items. Pour chaque item, nous avons cherché ce qui pouvait avoir un impact sur la réponse des élèves, au-delà du savoir en jeu et des connaissances mathématiques nécessaires à sa réussite.

En effet, nous avons été surprises de constater que certains items, très proches du point de vue de la tâche mathématique à réaliser, avaient des taux de réussite très différents. Par exemple : le problème « *Monsieur Paul achète 9 rosiers à 4€ et 3 sapins à 17€ pièce. Quel est le montant de sa dépense ?* » a été réussi à 62,95 %. Alors que le problème « *Monsieur Jacques achète 8 cahiers et 5 stylos. Le prix d'un cahier est de 3€. Le prix d'un stylo est de 2€. Quel est le montant de sa dépense ?* » l'a été à 80,73 %.

Ces problèmes, se résolvent tous deux par des opérations similaires ; même si nous anticipions un score de réussite moindre dans le premier exercice parce que les nombres en présence conduisaient à des opérations plus difficiles pour les élèves, cet écart de presque 20 points nous a surprises. Pour dépasser ce simple constat, une analyse des productions des élèves sur ces deux exercices nous a permis de constater que leurs erreurs ne reposaient pas uniquement sur des difficultés de calcul, mais également sur les opérations à mettre en jeu : 17 % des élèves font la somme des prix unitaires pour le premier problème ($17 + 4$), alors qu'ils ne sont que 3,5 % à obtenir 5 par la même procédure dans le deuxième problème. On constate alors que pour cet exemple, en plus de la taille

Cultures et politiques de l'évaluation en éducation et en formation

des nombres, la structuration de l'énoncé, éventuellement le contexte, semblent intervenir dans la réussite à l'exercice.

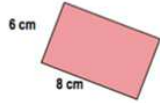
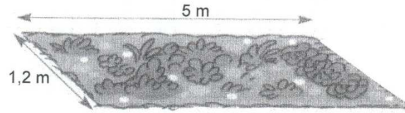
Ce constat nous a amenées à élaborer un outil d'analyse prenant en compte des paramètres extra mathématiques tels que les éléments relatifs au contexte, à l'énoncé, à la forme de la question, etc. Nous avons donc choisi d'établir un outil décliné en deux facteurs de complexité et un facteur de niveau de compétences.

Facteur de complexité 1 : le contexte de l'énoncé

Dans ce facteur, le niveau de langue de l'énoncé ainsi que la nature des informations à traiter (texte, graphique, schéma, etc.) nous semblent important à considérer. Nous estimons également que la nature de l'item et le type de réponses à produire ne sont pas sans incidence sur les taux de réussite : une question ouverte ou fermée, un QCM, un vrai/faux confrontent les élèves à des stratégies de réponse différentes (ce que nous étudions dans une recherche complémentaire).

Le contexte de l'énoncé, permettant ou non une représentation aisée du problème posé, peut également se révéler pertinent pour analyser le score de réussite des élèves.

Par exemple : A l'item 1 sous forme de QCM, le taux de réussite est de 53%, alors qu'à l'item 2, également sous forme de QCM, le taux de réussite est de 63%.

<p>Item 1 : Quel est le périmètre de la figure A ? Coche la bonne réponse.</p> <p><input type="checkbox"/> 14 cm <input type="checkbox"/> 24 cm <input type="checkbox"/> 28 cm <input type="checkbox"/> 48 cm</p>	 <p>Figure A</p>
<p>Item 2 : Quel est le périmètre de ce parterre de fleurs ? Coche la bonne réponse.</p> <p><input type="checkbox"/> 2,40 m <input type="checkbox"/> 6,20 m <input type="checkbox"/> 10 m <input type="checkbox"/> 12,40 m</p>	

Cet écart de 10 points nous paraît donc intéressant à approfondir, compte-tenu du fait que le savoir mathématique en jeu dans ces deux items est le même (déterminer le périmètre d'un rectangle), et que la complexité des calculs est plus élevée dans l'item 2 que dans l'item 1 puisque les nombres proposés ne sont pas tous entiers. La différence de réussite entre ces deux items peut alors s'expliquer par la forme même de l'énoncé (dimensions figurant explicitement sur le parterre de fleurs et non sur le rectangle de l'item 1, dessin figuratif contextualisant l'item 2 par exemple) ; ces éléments, qui nous semblent avoir une incidence sur la réussite à cet exercice, sont ainsi pris en compte dans ce premier facteur.

Facteur de complexité 2 : les savoirs mathématiques en jeu

Ce facteur est directement lié au savoir mathématique en jeu. De ce point de vue, la tâche à réaliser peut être plus ou moins simple, nous nous référons aux divers travaux effectués en didactique des mathématiques et aux travaux de Duval (1993) autour des changements de registres de représentation pour évaluer ce facteur de complexité. Dans ce facteur, seront également pris en compte les variables didactiques, ainsi que les distracteurs attachés aux items car ils peuvent avoir une influence non négligeable sur la réussite des élèves, dans un sens positif ou négatif.

Facteur de niveau de compétences

Les items ont été élaborés pour vérifier l'acquisition de compétences, sans que le lien entre compétences et connaissances n'ait été véritablement établi. La notion de compétence traverse davantage le champ des sciences de l'éducation que celui de la didactique des mathématiques. En effet, les chercheurs en didactique sont souvent critiques vis-à-vis de cette notion et préfèrent entrer par les situations pour aborder les apprentissages des élèves. Winslow (2005) évoque bien la notion de « compétences mathématiques spécifiques », mais ne semble pas convaincu par cette notion de compétence, même spécifiée. De même que Vergnaud (date non indiquée) ne conçoit

Cultures et politiques de l'évaluation en éducation et en formation

pas de travailler sur les compétences des élèves sans se donner les moyens de faire avancer les recherches sur l'activité des élèves et des enseignants.

Nous avons finalement retenu pour la notion de compétence la définition suivante, qui nous a semblée la plus acceptable, de notre point de vue : une capacité d'agir efficacement dans un type défini de situation, capacité qui s'appuie sur des connaissances, mais ne s'y réduit pas. (Perrenoud, 1997). Nous avons également choisi de prendre en compte l'aspect inédit d'une tâche à réaliser que l'on trouve souvent pour caractériser la notion de compétence car cela pouvait croiser, de manière intéressante, les niveaux de mise en fonctionnement des connaissances décrits par Robert et Rogalski (2002).

Ainsi, pour ce facteur, trois niveaux de compétences ont été retenus. Ils visent à évaluer dans quelle mesure l'item proposé est plus ou moins proche de la notion de compétence que nous avons retenue qui intègre aussi bien la restitution de connaissances qu'une capacité autonome de mobilisation de ces connaissances :

- **Niveau 1** : pour les tâches qui amènent à des applications immédiates des connaissances, c'est-à-dire simples (sans adaptation) et isolées (sans mélange), où seule une connaissance précise est mise en œuvre sans aucune adaptation, mis à part la contextualisation nécessaire. Les tâches sont usuelles.
- **Niveau 2** : pour les tâches qui nécessitent des adaptations de connaissances qui sont en partie au moins indiquées. Les tâches sont relativement usuelles.
- **Niveau 3** : pour les tâches qui nécessitent des adaptations de connaissances qui sont totalement à la charge de l'élève. Les tâches sont inédites.

Ces trois niveaux, construits en intégrant des éléments issus d'autres champs que les sciences de l'éducation rejoignent le découpage proposé par Rey, Carette, Defrance et Kahn (2003, p 26) autour de l'évaluation par compétences :

- **La compétence élémentaire** : savoir exécuter une opération en réponse à un signal ;
- **La compétence avec cadrage** : interpréter une situation inédite et choisir la compétence élémentaire qui convient ;
- **La compétence complexe** : choisir et combiner plusieurs compétences pour traiter une situation nouvelle et complexe.

Ces chercheurs considèrent que seuls, les deux derniers niveaux méritent d'être appelés compétence. Nous estimons pour notre part que seul le niveau 3 peut témoigner d'une réelle appropriation de connaissances.

Le niveau du facteur de compétences permet donc d'évaluer un item d'un point de vue complémentaire aux deux facteurs précédents puisqu'il distingue la nature des tâches proposées, au-delà du savoir en jeu. Le facteur de complexité 2 se réfère principalement au savoir mathématique en jeu alors que le niveau de compétences est davantage lié aux connaissances.

2.3 Utilisation de l'outil et méthodologie

Afin de pouvoir établir des comparaisons entre les niveaux de complexité des items, nous avons choisi d'attribuer un nombre de points à chaque catégorie de facteur de complexité (2 points pour le premier facteur, 3 points pour le second), ce qui aboutit à un total de 5 points. Plus le nombre de points est élevé, plus nous estimons que l'item est complexe. Pour chaque item, nous avons donc associé un niveau de complexité et un niveau de compétences qui nous permet de mieux lire et interpréter les résultats obtenus. Nous avons comptabilisé le nombre d'items correspondant à chaque niveau de compétences et regardé quel niveau de complexité (sur 5) était affecté à ces items. Cela nous a donné un panorama global de l'évaluation bilan de la DEPP qui nous a permis, comme nous le souhaitions, de déterminer quel reflet de la connaissance des élèves est réellement produit par les items proposés, et quels apprentissages sont révélés.

Cultures et politiques de l'évaluation en éducation et en formation

Afin de confronter le bilan CEDRE 2008 à l'outil que nous avons développé, nous allons présenter les résultats des deux travaux dans les deux domaines que nous avons retenus plus spécifiquement : « fractions/décimaux » & « grandeurs et mesures ».

2.3.1 Résultats pour les items relatifs aux fractions/décimaux

En appliquant les facteurs de complexité et de compétences aux 50 items de ce domaine, nous avons constaté que l'ensemble des items relatifs aux fractions/décimaux ne relevait pas de niveaux de complexité et de compétences très élevés (tableaux 1 et 2).

Facteur de complexité 1			Facteur de complexité 2			Niveau de compétences		
FC1 = 0	FC1 = 1	FC1 = 2	FC2 = 1	FC2 = 2	FC2 = 3	NC = 1	NC = 2	NC = 3
12 %	76 %	12%	44%	42%	14 %	82 %	18%	0%

Tableau 1 Répartition selon les facteurs de complexité (FC) et les niveaux de compétences (NC) des items relatifs aux « fractions/décimaux »

	Niveau de compétences = 1	Niveau de compétences = 2	Niveau de compétences = 3
Facteurs de Complexité = 1	4%	0%	0%
Facteurs de Complexité = 2	44%	0%	0%
Facteurs de Complexité = 3	24%	10%	0%
Facteurs de Complexité = 4	8%	8%	0%
Facteurs de Complexité = 5	2%	0%	0%

Tableau 2 Répartition des items relatifs aux « fractions-décimaux » en croisant les facteurs de complexité et de compétences

Il s'avère plus précisément que les items mettant en jeu des fractions (autres que les fractions décimales) portent essentiellement sur deux types de tâche, reconnaître ou représenter une fraction à partir d'un partage d'unité, et qu'ils relèvent tous du niveau de compétences 1. Ces items se distinguent par les fractions choisies (fractions supérieures à l'unité), la nature de l'unité à partager (segments - disques...) et le lien entre la fraction et le partage existant de l'unité (unité partagée en huitième et fraction à représenter en quart) ; ainsi pour une même tâche, la complexité émane, non pas de la situation, mais des adaptations que l'élève doit réaliser pour effectuer la tâche, sans pour autant que ces adaptations aient été réellement pensées lors de l'élaboration des items. L'évaluation des connaissances des élèves sur les nombres décimaux a porté, elle aussi, sur un nombre restreint de tâches : comparer des nombres décimaux, transformer ou reconnaître une écriture (passage écriture décimale à fraction décimale et réciproquement), placer un nombre décimal sur une droite graduée, intercaler un décimal entre deux nombres, poser des opérations avec des nombres décimaux. Par ailleurs, presque aucun de ces items n'est contextualisé.

2.3.2 Résultats pour les items relatifs aux grandeurs & mesures

Nous observons que les 42 items relevant de ce domaine ne permettent pas d'évaluer l'ensemble des connaissances relatives aux notions de grandeurs et de mesures. En employant la même méthodologie que précédemment, nous constatons à nouveau que l'ensemble des items de domaine ne relevait pas de niveaux de complexité et de compétences très élevés (tableaux 3 et 4).

Facteur de complexité 1			Facteur de complexité 2			Niveau de compétences		
FC1 = 0	FC1 = 1	FC1 = 2	FC2 = 1	FC2 = 2	FC2 = 3	NC = 1	NC = 2	NC = 3
19 %	71 %	10%	74%	19%	7%	69 %	19%	12%

Tableau 3 Répartition selon les facteurs de complexité et de compétences des 42 items relatifs aux grandeurs et mesures de CEDRE 2008

Cultures et politiques de l'évaluation en éducation et en formation

	Niveau Compétences = 1	Niveau Compétences = 2	Niveau Compétences = 3
Facteurs de Complexité = 1	21%	0%	0%
Facteurs de Complexité = 2	41%	7%	0%
Facteurs de Complexité = 3	7%	10%	2%
Facteurs de Complexité = 4	0%	2%	10%
Facteurs de Complexité = 5	0%	0%	0%

Tableau 4 Répartition des 42 items relatifs aux grandeurs et mesures de CEDRE 2008 en croisant les facteurs de complexité et de compétences

Concernant ce domaine, nous avons constaté qu'il y avait très peu d'items relatifs à des conversions d'unités ainsi qu'une surreprésentation des items relatifs à la lecture d'heures (12 sur 42). Il convient également de noter que la nature de l'évaluation CEDRE ne permet pas d'évaluer la capacité des élèves à utiliser correctement des instruments de mesure (une balance, par exemple). Les résultats globaux dans ce domaine sont donc à relativiser.

2.3.3 Bilan sur le contenu de l'évaluation 2008 dans les domaines « fractions & décimaux » et « grandeurs et mesures »

Il est donc apparu que les items des domaines « fractions/décimaux » et « grandeurs & mesures » ne permettaient pas réellement de rendre compte du niveau de connaissances et de compétences des élèves de fin d'école primaire, compte-tenu du fait qu'ils étaient inégalement répartis dans un espace défini à l'aide de niveaux de compétences et de complexité extrêmes. Globalement, pour l'évaluation CEDRE 2008 de fin d'école, nous avons pu constater qu'un peu plus des trois quarts des items relevaient d'un bas niveau de compétences, c'est à dire qu'ils correspondent à des applications immédiates de connaissances, et qu'ils ne permettent pas d'évaluer la résistance de ces connaissances dans des situations plus complexes. Par ailleurs, les items relevant de niveaux de compétences plus élevés sont également, pour plus de la moitié d'entre eux, d'un niveau de complexité élevé : en termes d'évaluation, il est donc difficile de déterminer si l'échec de l'élève relève d'un manque de compréhension de l'énoncé, d'un manque de connaissance mathématique ou d'un manque de compétences adaptées (ou les trois).

Par ailleurs, il s'est avéré que l'ensemble des items proposés pour évaluer les connaissances et compétences des élèves de fin d'école ne recouvrait pas toutes les tâches relatives aux domaines mathématiques traités. Concernant les décimaux, de nombreuses tâches relevant d'une même nature (comparer des nombres décimaux ou transformer une écriture décimale en une écriture fractionnaire et réciproquement) ont été proposées aux élèves, sans pour autant différencier leurs niveaux de complexité ou de compétences, ce qui est regrettable et préjudiciable à l'objectif visé par l'évaluation. Finalement, des constats ont été émis pour rendre compte du niveau de connaissances et compétences des élèves à partir d'un ensemble d'items peu représentatif de l'ensemble des tâches réalisables dans ce domaine. Néanmoins, l'outil qui nous a permis d'analyser plus finement les items de cette évaluation bilan et de réaliser quels en étaient les défauts, laisse entrevoir qu'il est possible de concevoir des évaluations externes avec un contenu plus équilibré, à la fois en termes de niveaux de compétences, mais aussi relativement aux programmes scolaires qui sont sensés être évalués à travers ces évaluations.

Nous pouvons donc constater que le découpage arbitraire des groupes qui précisent « le niveau global de maîtrise des connaissances et des compétences des élèves » (Lescure & Pastor, 2012, p 64) ne reflète pas réellement le niveau de connaissance des élèves dans un domaine donné (les fractions/décimaux notamment) et qu'il ne témoigne que de performances locales, pas forcément révélatrices de ce que savent réellement les élèves. En effet, le fait que les items proposés dans ce domaine relèvent majoritairement de niveaux de complexité et de compétences assez bas ne permet pas d'augurer de la résistance des connaissances en jeu si on élevait ces niveaux et relativise l'interprétation des scores de réussite des élèves suivant les différents groupes.

Cultures et politiques de l'évaluation en éducation et en formation

Finalement, il nous faut acter du fait que les élèves qui ne réussissent pas les items de niveau 1 de compétences semblent se heurter frontalement à un problème lié au savoir mathématique en jeu. Toutes les études et les évaluations sur les décimaux ont montré que cette notion mathématique est mal maîtrisée par les élèves qui ne comprennent pas toujours leur fondement et leur écriture. Ce constat est hélas encore aujourd'hui, d'une funeste actualité.

2.4 Recherche complémentaire sur la spécificité du format QCM

En analysant l'ensemble des items proposés dans l'évaluation bilan CEDRE 2008, il s'est avéré que certains parmi ceux qui étaient proposés sous forme de QCM avaient des résultats très différents de ceux proposés sous forme de question ouverte, alors que d'un point de vue mathématique, la tâche à réaliser était équivalente. Nous nous sommes donc demandé quelles pouvaient en être les raisons, ce qui nous a amenées à investiguer la façon dont les élèves appréhendaient les QCM à ce niveau de leur scolarité.

Nous avons exploré en termes de stratégies et de degrés de certitude (Leclercq, 2006) des élèves la modalité QCM utilisée majoritairement dans ces bilans afin de tester la validité de ces évaluations ; ce test contenait des items relevant des domaines « grandeurs et mesures » et « fractions et décimaux ». Nous avons ainsi constaté qu'en France, les élèves de l'école primaire n'adoptent pas un type de stratégies uniforme pour résoudre les tâches auxquelles ils sont confrontés, mais qu'ils en changent suivant la nature de ces tâches (Sayac & Grapin, soumis). Les stratégies qu'ils utilisent dépendent également de leur niveau de connaissances en mathématiques : les élèves « forts » utilisent principalement des stratégies de savoir, alors que les « faibles » sont plus enclins à utiliser des stratégies de repli ou de substitution, même s'ils rechignent à choisir leur réponse totalement au hasard. Il semblerait qu'un contrat didactique s'opère dans le cadre de ce type d'évaluation, ne permettant pas aux élèves d'envisager cette stratégie comme scolairement acceptable, à moins qu'ils n'aient pas suffisamment de recul pour utiliser le hasard comme recours possible.

Les items relatifs au lien entre écriture décimale et fraction de ce test ont produit des résultats assez inquiétants car, au-delà des scores de réussite assez faibles témoignant d'une mauvaise compréhension des décimaux, ils nous ont alerté sur le fait que certains élèves utilisent des règles erronées qu'ils appliquent avec un degré de certitude élevé (ignorance ignorée, Gilles 1996). Ainsi, l'intégration du degré de certitude au choix d'une réponse nous a permis, à l'échelle de notre test et de la taille de l'échantillon, de mieux mesurer le degré de connaissance des élèves et l'ancrage de leurs conceptions erronées, surtout pour les fractions - décimaux. Cette recherche, menée en 2012, a été complétée par une recherche en 2013 portant sur l'impact du format de question (ouvert et fermé) dans l'activité de l'élève². La prise en compte des résultats de ces deux recherches permet alors de mieux qualifier le facteur de complexité 1 pour les items relevant des domaines considérés.

3. Conclusion

Le dernier constat que nous pouvons faire concerne l'intérêt certain que nous avons eu à porter notre regard de didacticienne sur une évaluation externe, commanditée par une direction du ministère de l'éducation nationale. Cette collaboration fructueuse entre chercheurs et évaluateurs institutionnels gagnerait à être renouvelée parce qu'elle permet d'enrichir, de questionner et de mieux connaître le travail des uns et des autres. Ainsi, l'outil que nous avons développé autour des facteurs de complexité et de compétences et que nous continuons d'enrichir par des travaux spécifiques à l'évaluation par QCM a été utilisé comme support pour la conception de l'évaluation

² Cette recherche a fait l'objet de la communication « Évaluer par QCM en mathématiques en fin d'école : quelles stratégies sont mises en œuvre par les élèves ? Quelles connaissances sont effectivement évaluées ? » présentée lors de ce même colloque et détaillée dans ces actes.

Cultures et politiques de l'évaluation en éducation et en formation

CEDRE 2014 en mathématiques : à la suite de notre travail, un rééquilibrage du contenu a été alors envisagé pour qu'il y ait davantage d'items niveaux de complexité élevés.

4. Références et bibliographie

- Artigue, M., & Winslow, C. (2010). International comparative studies on mathematics education: a view from the anthropological theory of didactics. *Recherches En Didactique Des Mathématiques*, 1(30), 47–82.
- Bodin, A. (2004). *Taxonomie des énoncés mathématiques, classement par niveaux hiérarchisés de complexité cognitive*. Rapport EVAPM.
- Brousseau, G. (1980). Problèmes de l'enseignement des décimaux. *Recherches en Didactique des Mathématiques*, 1 (1), 11-59.
- Comiti, C., Neyret, R. (1979). A propos des problèmes rencontrés lors de l'enseignement des décimaux en classe de CM. *Grand N*, 18, 5-20.
- Deblois, L., Freiman, V., & Rousseau, M. (2007). Les résultats des élèves aux tests internationaux et leur possible influence sur les thèmes de recherche. In *Actes du colloque du Groupe des Didacticiens des Mathématiques du Québec* (pp. 135 – 146). UQAR.
- Douady, R. & Perrin-Glorian M.-J. (1986). *Liaison école - collège : Nombres décimaux*. Brochure n°62. Paris : IREM de Paris 7.
- Duval, R. (1993). Registres de représentations sémiotiques et fonctionnement cognitif de la pensée. *Annales de Didactique et de Sciences Cognitives*, 5, 37-65.
- Gagatsis, A., Elia, I., Mougi, A. (2002) The nature of multiple representations in developing mathematical relations. *Scientia Paedagogica Experimentalis* 39 (1), 9–24.
- Gilles, J.-L. (1996). Utilisation des degrés de certitude et normes de réalisme en situation d'examen et d'auto-estimation à F.A.P.S.E. - ULG, *Colloque de l'ADMEE-EUROPE "Dix années de travaux de recherche en évaluation"*
- Grégoire, G. & Laveault D. (2005). *Introduction aux théories des tests en sciences humaines*. Bruxelles : De Boeck.
- Leclercq, D. (2006). L'évolution des QCM. In G. Figari et L. Mottier-Lopez (Ed.), *Recherche sur l'évaluation en éducation* (pp. 139-146). Paris : L'Harmattan.
- Lescure, S., Pastor, J.-M. (2012). *Mathématiques en fin d'école primaire. Le bilan des compétences*. Paris : Scéren.
- Moreira-Baltar, P. (1996). À propos du concept d'aire. *Petit x*, 43, 43-68.
- Perrenoud, P. (1997). *Construire des compétences dès l'école*. Paris : ESF.
- Perrin-Glorian, M.-J. (1990) L'aire et la mesure. *Petit x*, 24, 5-36.
- Rey, B., Carette, V., Defrance, A. & Kahn, S. (2003). *Les compétences à l'école - Apprentissage et évaluation*. Bruxelles : De Boeck.
- Robert, A. & Rogalski, M. (2002) Comment peuvent varier les activités mathématiques des élèves sur des exercices ? le double travail de l'enseignant sur les énoncés et sur la gestion en classe. *Petit x*, 60, 6-25.
- Roditi E. (2007) La comparaison des nombres décimaux, conception et expérimentation d'une aide aux élèves en difficulté. *Annales de didactique et de sciences cognitives*, 12, 55-81.
- Vergnaud, G. (date non indiquée) Les compétences, Bravo ! Mais encore ? Réflexions critiques pour avancer. Site Pedagopsy.eu, http://www.pedagopsy.eu/competences_vergnaud.htm
- Winslow C. (2005) Définir les objectifs de l'enseignement mathématique : la dialectique matières - compétences. *Annales de didactique et de sciences cognitives*, 10, 131-155.
- Enquête internationale sur l'alphabétisation et les compétences des adultes* (2003), Ministère de l'Éducation, du Loisir et du Sport, Canada.
- Note d'information de la DEPP, 10-17 octobre 2010 : les compétences en mathématiques des élèves de fin d'école primaire*, Ministère de l'Éducation Nationale, France.