

DES RECHERCHES AUTOUR DE « FAITS D'ÉVALUATION » EN MATHÉMATIQUES

Nathalie **SAYAC**

Université Paris Est Créteil - ESPE
Laboratoire de didactique André Revuz
nathalie.sayac@u-pec.fr

Nadine **GRAPIN**

Université Paris Est Créteil - ESPE
Laboratoire de didactique André Revuz
nadine.grapin@u-pec.fr

Résumé

Les deux recherches que nous présentons apportent deux regards complémentaires et différents sur une même évaluation externe, le bilan CEDRE fin d'école en 2008. En définissant deux facteurs de complexité et trois niveaux de compétence, nous appréhendons le contenu global de l'évaluation pour étudier la répartition des items et interroger la validité des résultats qui peuvent être produits. L'étude du premier facteur de complexité nous conduit par la suite à considérer les items d'un point de vue local et à observer les stratégies mises en œuvre par des élèves de CM2 pour répondre à des QCM : une première recherche porte principalement sur les stratégies des élèves en fonction de leur niveau de connaissance et la seconde recherche met en perspective les deux formats de question (QCM et question ouverte) pour analyser leur impact sur la réussite à l'exercice et sur les procédures mises en œuvre.

Mots clés

Évaluation, complexité, compétence, format de question, QCM.

UNE PREMIERE RECHERCHE AUTOUR D'UNE EVALUATION EXTERNE

Présentation et contexte de la recherche

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

à des évaluations menées par exemple en maîtrise de la langue et sur des compétences générales (2003, reprise en 2009), en langues vivantes (en 2004), etc. La reproductibilité des évaluations est rendue possible tant dans leur conception que dans leur passation, notamment en ne rendant pas public les items de ces évaluations.

A la différence des évaluations nationales menées en CE1 et en CM2, les évaluations bilans ne sont pas réalisées sur tous les élèves de fin d'école primaire, mais uniquement sur un échantillon représentatif de la population des élèves français. Ainsi, environ 3800 élèves de CM2 (soit un peu plus de 200 classes concernées) ont passé ces évaluations. Par conséquent, le modèle statistique¹ utilisé pour construire les échelles de scores permet de caractériser des groupes d'élèves par les compétences qu'ils maîtrisent ; il ne s'agit en aucune façon de déterminer les compétences maîtrisées par chacun des élèves ayant participé à l'évaluation.

Pour le primaire, les items proposés en mathématiques pour CEDRE 2008 ont ainsi permis de mesurer les connaissances relevant des six domaines (connaissances des nombres entiers naturels, fractions et décimaux, calcul, exploitation de données numériques, espace et géométrie, grandeurs et mesures) et différentes compétences (identifier des notions, exécuter un calcul, traiter des données, produire en autonomie, contrôler-valider).

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 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 d'enseignants en IUFM, pour compléter des équipes pluri catégorielles d'experts composées de professeurs de terrain, de conseillers pédagogiques, d'IEN (Inspecteur de l'Éducation Nationale) et d'IA-IPR (Inspecteur d'Académie - Inspecteur Pédagogique Régional).

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 indispensable pour expliciter ce qui est effectivement évalué et mieux caractériser les connaissances des élèves en mathématiques lors de l'analyse des résultats. Nous avons donc conçu une recherche permettant d'exploiter les résultats du bilan école et d'inférer des constats au niveau des apprentissages des élèves. Il s'agissait pour nous de considérer ce qui était réellement évalué dans les items proposés du point de vue des connaissances mathématiques et des compétences plus transversales, de déterminer quel reflet de la connaissance des élèves était produit par les items proposés, et quels apprentissages étaient révélés.

Nous avons rapidement pris conscience de l'ambiguïté de l'objectif affiché par le bilan de la DEPP qui consistait à évaluer à la fois des compétences et des connaissances. En effet, les items ont été élaborés pour vérifier l'acquisition de compétences, sans que le lien entre compétences et connaissances ne soit clairement établi. Tout s'est passé comme si évaluer des compétences se situait au même niveau qu'évaluer des connaissances, or il nous a semblé qu'il y avait là un raccourci problématique.

¹ Méthode de Réponse à l'Item ou MRI, fondé sur le postulat que la réponse d'un individu à un item (et notamment sa probabilité de fournir une réponse correcte) est déterminée - ou peut être expliquée - par certains attributs du sujet et par les propriétés de l'item lui-même. Ces modèles permettent d'exprimer sur une même échelle le niveau de difficulté d'un item et le niveau de compétences d'un individu.

Méthodologie de la recherche

Pour mener à bien cette exploration, nous avons conçu un outil d'analyse d'items composé de facteurs de complexité et de compétences qui s'inspirent de divers travaux de chercheurs - taxonomie de Gras adaptée par Bodin (2004), niveaux de mise en fonctionnement des connaissances de Robert (2008), des registres de représentation sémiotiques de Duval (1993), notion de compétences de Perrenoud (1997), Rey, Carette, Defrance & Kahn 2006, Beckers (2002) - et qui prend également en compte les résultats des travaux en didactique des mathématiques qui identifient des difficultés liées à certains domaines mathématiques.

Cet outil se décline en trois facteurs d'analyse : deux facteurs de complexité FC1 et FC2 et un facteur 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 car il permet d'évaluer la complexité de la tâche du point de vue de la consigne et de la manière dont l'élève est amené à comprendre ce qu'il doit faire. La présence d'un exemple donné dans la consigne entraîne, du point de vue de la tâche que l'élève a à réaliser, une simplification de celle-ci qui n'est certainement pas sans incidence sur l'activité de l'élève. A contrario, une consigne² « brute » peut être mal comprise (Image 1) et amener l'élève à ne pas répondre à ce qui est attendu :

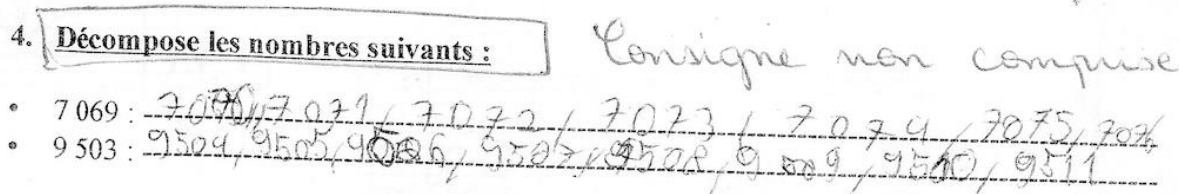


Image 1 - Production d'élève témoignant d'une consigne mal comprise

Facteur de complexité 2 : les savoirs mathématiques en jeu

Ce facteur est directement lié au savoir mathématique en jeu et nous convoquons des travaux développés en didactique des mathématiques dans le domaine de la numération des nombres entiers pour évaluer la complexité de la tâche proposée. Nous nous référons également aux travaux de Duval (1993) autour des changements de registres de représentation pour évaluer ce facteur de complexité. C'est dans ce facteur que sont prises en compte les variables didactiques propres au domaine étudié (taille des nombres, écriture des nombres, présence de zéros, etc.) et qui peuvent avoir une influence non négligeable sur la complexité de la tâche à réaliser. Par exemple, il n'est pas équivalent de demander aux élèves d'écrire en chiffres le nombre « quarante-trois-mille-cent-vingt-quatre » et le nombre « quarante-mille-vingt-quatre ».

Facteur de niveau de compétences

Pour ce facteur, la définition de compétence qui nous est apparue la plus adaptée est celle que Perrenoud (1997) a proposée, mais complétée par la prise en compte de l'aspect inédit de la tâche à réaliser retenu par certains auteurs (Beckers 2002, Rey, Carette, Defrance et Kahn,

² Par exemple : « décompose les nombres suivant le modèle : $6428 = 6 \times 1000 + 4 \times 100 + 2 \times 10 + 8 \times 1$ »

2002). Pour nous, une compétence se définit donc comme :

Une capacité d’agir de manière opérationnelle face à une tâche mathématique qui peut s’avérer inédite, en s’appuyant sur des connaissances que l’élève mobilise de façon autonome (Sayac, Grapin 2015).

Nous nous sommes inspirées des travaux relatifs aux différents niveaux de mises en fonctionnement des connaissances (Robert & Rogalski 2002) et des adaptations listées par Robert (2008) pour prendre en compte le caractère inédit³ des tâches et le degré d’autonomie développé par l’élève confronté aux différentes tâches.

Ainsi, pour ce facteur, trois niveaux de compétences permettent d’évaluer dans quelle mesure la tâche proposée en évaluation 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 (image 2).

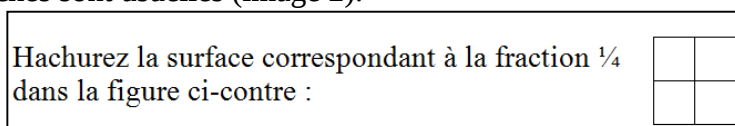


Image 2 : Tâche illustrant un niveau 1 de compétence

- Niveau 2 : pour les tâches qui nécessitent des adaptations de connaissances qui sont en partie au moins indiquées. Les tâches sont relativement usuelles (image 3).

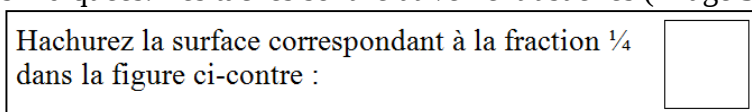


Image 3 : Tâche illustrant un niveau 2 de compétence

- Niveau 3 : pour les tâches qui nécessitent des adaptations de connaissances qui sont totalement à la charge de l’élève. Les tâches sont inédites (image 4).

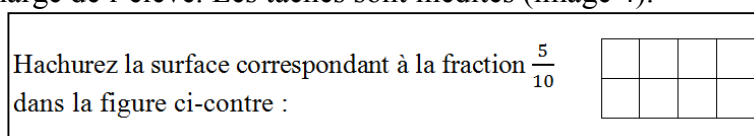


Image 4 : Tâche illustrant un niveau 3 de compétence

Chaque item du bilan CEDRE 2008 a donc été affecté de trois nombres selon son niveau de complexité et de compétence :

- le facteur de complexité 1 (FC 1) coté entre 0 et 2 points
- le facteur de complexité 1 (FC 2) coté entre 1 et 3 points
- le niveau de compétence de valeurs 1, 2 ou 3.

Résultats de la recherche

En appliquant notre outil à l’ensemble des items du bilan CEDRE 2008, nous avons ainsi pu constater qu’ils ne se répartissaient pas de façon homogène du point de vue des niveaux de

³ Nous entendons par inédit le fait que l’élève soit confronté à une tâche à laquelle il n’a pas l’habitude d’être confronté, par opposition aux tâches qu’il rencontre plus fréquemment (dans les manuels notamment).

complexité et de compétences comme l'illustre le tableau 1 :

	Facteur de complexité			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
Nombre d'items	128	266	33	296	112	19	333	86	8
Répartition en %	30 %	62 %	8 %	69 %	26 %	4 %	78 %	20 %	2 %

Tableau 1 : Répartition des items CEDRE 2008 selon les facteurs de complexité et les niveaux de compétence

Au-delà de ce résultat, nous avons également constaté que :

- Les items de niveau 3 de compétences sont souvent affectés d'un niveau élevé de complexité (supérieur ou égal à 3), ce qui ne permet pas d'évaluer pleinement la connaissance en jeu.

- Les élèves qui ne réussissent pas les items de faibles niveaux de complexité et de compétence semblent se heurter frontalement à un problème lié à la connaissance mathématique en jeu (ex : les décimaux).

- Il y a de nombreux items qui correspondent au niveau 1 de compétences et à un niveau faible de complexité. Ce qui ne permet pas d'augurer de la résistance des connaissances en jeu si on élevait le niveau de compétence et/ou de complexité.

De plus, 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 items relatifs aux 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 et de laisser entrevoir qu'il est possible de concevoir des évaluations externes plus acceptables d'un point de vue scientifique, en jouant sur les différents niveaux de complexité et de compétences.

DEUX RECHERCHES COMPLEMENTAIRES A LA PREMIERE

Le format QCM et les stratégies des élèves

Suite à cette première recherche, nous avons souhaité explorer le domaine des formats de questions, et plus particulièrement celui des QCM (Question à choix multiple) pour questionner la validité de ce format de question spécifique. Dans la réalité des classes de l'école primaire, en France, ce format est rarement, voire jamais, utilisé alors qu'il l'est régulièrement pour déterminer les connaissances des élèves de fin d'école primaire (10 - 11 ans) en mathématiques dans les évaluations nationales ou internationales standardisées.

Dans le cadre de la didactique des mathématiques en France, peu de recherches récentes ont été menées sur ce domaine ; nous pouvons citer à titre d'exemple les recherches d'Adda (1976), Duval et Pluvinage (1977) et Pluvinage (1979). Par contre, de nombreuses études portant sur l'utilisation des QCM pour des évaluations d'étudiants à visées différentes (formative, sommative, certificative, etc.) ont été menées (notamment Choppin 1975, Leclercq 1986, Gilles 1996).

Nous avons donc conçu une expérimentation permettant d'explorer tout à la fois la validité d'un QCM et les stratégies des élèves de l'école primaire confrontés à ce type de question. Nous avons fait l'hypothèse d'une corrélation entre le niveau de connaissances des élèves et les stratégies qu'ils utilisaient pour répondre aux QCM.

L'expérimentation a été menée en juin 2012 dans 6 classes de CM2 (soit 157 élèves au total), de Paris et de sa banlieue, variées en termes de population et de niveau scolaire afin de pouvoir tester notre hypothèse. Cette expérimentation a eu lieu après les évaluations nationales en mathématiques (début juin) afin de pouvoir bénéficier des résultats des élèves et ainsi déterminer leur niveau de connaissances de manière objective et permettant des comparaisons ; nous avons ainsi déterminé trois groupes d'élèves⁴ : « faibles », « moyens » et « forts » selon leur score de réussite à ces évaluations nationales, communes à tous les élèves.

Nous avons élaboré un test comportant un nombre limité d'items pour que chaque élève puisse nous expliciter son raisonnement individuellement lors de la passation, et ainsi nous permettre de déterminer au plus juste la stratégie employée par l'élève pour choisir sa réponse. Le test a été conçu à partir de sept items, de niveaux variés du point de vue de la difficulté (l'item 7 étant particulièrement difficile) et pour qu'ils relèvent de deux catégories : ceux pour lesquels l'élève doit mobiliser des connaissances relatives aux fractions et aux nombres décimaux et ceux pour lesquels l'élève doit résoudre un problème dans le cadre des grandeurs et mesures (problèmes additifs, de proportionnalité et de calcul de périmètre).

Pour les items relevant des fractions-décimaux, les différents choix de réponses ont été effectués à partir d'erreurs reconnues comme fréquentes par différents travaux en didactique des mathématiques, Douady et Perrin-Glorian (1986), Comiti et Neyert (1979), Roditi (2007). En choisissant des distracteurs correspondant à des erreurs récurrentes d'élèves, nous souhaitions savoir s'ils étaient attractifs pour certains élèves (selon leur niveau de connaissances) ou au contraire s'ils étaient d'emblée rejetés.

La question des stratégies que développent les élèves lorsqu'ils sont confrontés à des QCM a été étudiée par de nombreux chercheurs (Choppin 1975, Leclercq 1986, 1987), mais elle a toujours été traitée pour des étudiants adultes. Choppin (1975) a retenu 3 modèles, combinant de manière variée connaissances des élèves et opportunités liées à la forme spécifique d'un QCM alors que Katz, Bennet, et Berger (2000) ont distingué les stratégies « traditionnel » des « non traditionnel ». Les premières reposant sur des procédures qui sont enseignées et que l'élève utiliserait pour répondre si la question était posée sous forme ouverte alors que les autres, spécifiquement associées à des QCM, sont celles où l'élève s'appuie sur les réponses proposées pour faire un choix.

⁴ « Fort » correspond à un % de réussite de plus de 75% aux évaluations nationales, « moyen » entre 40% et 75%, « faible » moins de 40%.

De notre côté, nous avons dans un premier temps listé, de manière aussi exhaustive que possible (Annexe 1), les différentes stratégies que pouvaient utiliser des élèves de fin d'école pour répondre à des QCM, puis nous les avons regroupées suivant trois types selon qu'elles intégraient des niveaux différents de connaissances :

- Stratégies A : stratégies de savoirs. L'élève active des connaissances ou des savoir-faire (techniques – raisonnement) pour choisir la réponse qu'il pense être la bonne : soit il résout complètement la tâche (par la procédure de son choix, juste ou fausse : ce qui s'apparente aux stratégies « traditionnal » définies par Katz, Bennet & Berger (2000)), soit il teste les propositions de réponse et choisit celle qui peut convenir.
- Stratégies B : stratégies de substitution ou de repli. L'élève n'utilise pas ses connaissances mathématiques de façon explicite pour faire un choix : son choix ne repose pas de façon assurée sur ses connaissances.
- Stratégies C : stratégies mixtes. L'élève a initié un raisonnement pour répondre à la question posée, mais il se sert des différentes propositions de réponse pour finaliser son choix.

Le tableau 2 donne la répartition des stratégies utilisées suivant le niveau de connaissance des élèves :

	Stratégies A	Stratégies B	Stratégies C
Élèves faibles	64,9 %	25,4 %	9,7 %
Élèves moyens	63,5 %	24,7 %	11,8 %
Élève forts	76,1 %	12,8 %	11,1 %

Tableau 2 : Répartition des stratégies utilisées selon le profil des élèves

La recherche que nous avons menée tend à montrer que finalement, peu d'élèves répondent « au hasard » dans ce type de dispositif. De façon globale, les élèves ne semblent pas avoir perçu les avantages que ce dispositif pouvait leur procurer et ont été peu nombreux à utiliser la stratégie B1 (« au hasard », voir annexe 1). Il semblerait qu'un effet de contrat didactique empêche les élèves de fin d'école primaire à considérer que cette stratégie puisse être envisageable pour répondre, même dans les rares cas où nous leur avons suggéré face à leur désarroi.

Les élèves de l'école primaire n'adoptent pas un type de stratégies uniforme pour résoudre les tâches auxquelles ils sont confrontés, mais ils en changent suivant la nature de ces tâches et suivant leur niveau de connaissance. Alors que les élèves les plus forts n'utilisent quasiment que des stratégies de savoir pour répondre aux sept items du test, les élèves les plus faibles sont plus enclins à utiliser des stratégies de repli ou de substitution.

En complément des stratégies utilisées, des procédures mises en œuvre et de la réponse choisie, il nous a semblé complémentaire de pouvoir demander à l'élève d'estimer la qualité de sa réponse (bonne ou non) en lui demandant quel degré de certitude il lui attribuait. Définis par Leclercq et Poumay (2004, p. 3), les degrés de certitude sont « l'ensemble des jugements, des analyses, des régulations, conscientes ou non, (mais qu'il importe de rendre explicites, observables et conscientes) effectués par l'apprenant sur ses propres performances ».

En nous inspirant des travaux de Leclercq (1987, 2006), nous avons proposé une échelle de quatre degrés de certitude, adaptée à des élèves de fin d'école, variant de « pas sûr du tout » à « sûr et certain » : pas sûr du tout (1), pas très sûr (2), sûr (3), sûr et certain (4). Cette échelle, figurant sur la feuille de l'élève, était accompagnée d'une échelle numérique (de 1 à 4), lui laissant la possibilité de positionner leur degré de certitude entre deux degrés « entiers » (c'est-à-dire choisir un degré de certitude de 3,5 par exemple). L'utilisation des degrés de certitude, notamment pour les réponses correspondant à des conceptions erronées sur les fractions-décimales nous permettait, au-delà du score de réussite, de préciser si l'élève est réaliste ou non sur ses connaissances.

Les difficultés relevées sur les connaissances des nombres décimaux par de nombreuses recherches en didactique des mathématiques sont confirmées, à nouveau, par les scores de réussite aux trois items correspondants.

Logiquement, nous avons pu constater que les élèves les plus performants sont plus sûrs de leurs réponses que les élèves qui le sont moins (en moyenne, 2,88 pour les plus « faibles », 2,97 pour les « moyens », 3,31 pour les plus « forts »). Néanmoins, les élèves du groupe « faibles » ne sont pas en manque de certitude, même si leurs résultats ne sont pas à la hauteur de leur espérance. Plus précisément, sur les items portant sur l'écriture des nombres décimaux, environ 20% des élèves considèrent que l'écriture à virgule a,b est équivalente à celle de la fraction a/b . A ce stade de l'apprentissage, cette erreur considérée comme classique ne serait pas si préoccupante si les élèves qui la commettent n'accordaient pas un degré de certitude élevé à leur réponse (environ un cinquième des élèves accordent un degré de certitude supérieur à 3 à une réponse qui est fautive). L'« ignorance ignorée » (Gilles, 1996) de ces élèves est particulièrement importante à réaliser, notamment pour les enseignants : en effet, il sera plus simple de déconstruire une fautive conception dont l'élève est peu assuré (« ignorance reconnue », (Gilles, 1996)), que lorsqu'elle apparaît à l'élève comme certaine.

Les items relatifs au lien entre écriture décimale et fraction 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). Même si l'utilisation de QCM en cours d'apprentissage n'est pas préconisée (Grégoire & Laveault, 2005), on mesure à travers ce travail et ce type de dispositif, l'intérêt que présentent de telles informations pour l'enseignant et pour l'élève. L'utilisation en classe par l'enseignant de tels indicateurs semble une voie intéressante pour mieux appréhender le niveau de ses élèves et pour réguler son enseignement. Pour les erreurs fréquentes repérées par les didacticiens, nous convenons avec Pluvineau (1979) que « pointer cette erreur et faire remarquer son caractère attractif à un élève qui l'a commise est un service à lui rendre ». Par ailleurs, une différenciation « active » des élèves peut, dans les faits, amener à renforcer les inégalités scolaires et constituer un leurre préjudiciable aux élèves les plus en difficulté (Crinon, Rochex 2011).

QCM vs QROC

Nous avons voulu compléter cette première recherche sur les QCM et les stratégies de réponse par un questionnaire plus ciblé sur la différence de format de question qui a déjà été étudié notamment par Bennett, Rock et Wang (1991), Frary (1985) ou Bridgeman (1992). Nous avons donc réalisé une seconde expérimentation confrontant deux tests équivalents du point de vue des tâches mathématiques, mais différents par leur format : l'un avec des QCM extraits de CEDRE 2008 ou 2014, l'autre avec des Questions à Réponse Ouverte Courte (QROC).

Nous avons réalisé cette expérimentation en juin 2013 auprès de 195 élèves de CM2, de neuf classes différentes, qui ont été soumis successivement à deux tests, l'un exclusivement sous format de QCM, l'autre reprenant les mêmes tâches mathématiques, mais présentées sous format QROC.

Pour cette deuxième expérimentation, nous avons conçu deux tests de six items, équivalents du point de vue des tâches mathématiques, l'un sous forme de questions ouvertes, l'autre, sous forme de QCM. Le premier test a été effectué individuellement par les élèves dans leur classe et le deuxième test, sous forme de QCM, a été passé quelques jours après. Pour ce dernier, nous

avons observé tour à tour, individuellement, chacun des élèves afin de le questionner sur la façon il a procédé pour choisir sa réponse et dans quelle mesure il était assuré de sa réponse ; la méthodologie employée pour l'étude de l'activité de l'élève en train de répondre à un QCM a été la même que celle employée pour la première expérimentation et nous a aussi permis d'étudier la stratégie mise en œuvre, la réponse choisie (réussite, échec/erreur significative), le degré de certitude qui lui a été accordé et la procédure développée pour répondre. En complément à ce travail spécifique aux QCM du test, une étude comparative entre les deux tests (en QROC et sous en QCM) nous a conduites à analyser les différences entre les procédures développées, les réponses fournies et à prendre en compte les rétroactions permises par le format spécifique QCM.

Le tableau 3 indique les pourcentages de réussite suivant le format des questions et la nature des tâches proposées dans les items du test :

	Pourcentage de réussite QROC	Pourcentage de réussite QCM
Écrire un nombre de l'ordre des millions	70,3%	88,2%
Donner (ou reconnaître) l'écriture à virgule d'un nombre écrit sous forme de fraction décimale	78,5%	74,9%
Donner (ou reconnaître) l'écriture sous forme de fraction décimale d'un nombre écrit avec une virgule.	51,8%	70,3%
Résolution d'un problème additif (simple)	70,8%	69,2%
Résolution d'un problème de proportionnalité (complexe)	19,0%	30,3%

Tableau 3 : Score de réussite selon le format de question

Nous pouvons d'emblée constater que la réussite à l'item QCM relatif à l'écriture des nombres est de 88%, alors qu'elle n'est que de 70 % au QROC, mais certaines réponses erronées au QROC semblent relever d'un manque d'attention plutôt que d'une méconnaissance majeure (écriture d'un chiffre à la place d'un autre, comme par exemple 5 542 006 ou lieu de 5 552 006). Même si des stratégies de savoir ont majoritairement été employées pour répondre au QCM, elles ne mettent pas en jeu les mêmes procédures qu'en QROC : si l'élève doit écrire le nombre en QROC, dans le QCM il doit reconnaître une écriture. On perçoit à travers les résultats des élèves à cet item la difficulté à laquelle on peut se confronter pour déterminer si l'élève a acquis ou non le savoir en jeu dans ces items relatifs à l'écriture des nombres entiers.

Concernant les items relevant du passage d'une écriture à virgule (D) à une écriture sous forme de fraction décimale (Q) et réciproquement nous constatons que la différence de format (ouvert-QCM) a un impact sur la réussite des élèves dans un des sens de transformation, mais pas dans l'autre : dans le sens $Q \Rightarrow D$, il n'y a pas de différence significative dans la réussite entre l'item posé sous forme ouverte et sous forme de QCM, alors que dans le sens inverse $D \Rightarrow Q$, la différence de réussite est ici significative (autour de 20 %) en faveur du QCM. Cette différence de réussite, selon le format, dans le sens $Q \Rightarrow D$, s'explique par un score important de non réponse en ouvert (environ 11 %) et par un effet rétroactif du QCM sur certaines réponses, qui ont été produites en ouvert, mais que les élèves ne retrouvent pas dans les choix (en particulier, des réponses justes, du type $237 + 8/10$).

Concernant les deux items relevant de la résolution de problèmes (un problème additif, un problème de proportionnalité), nous avons également constaté des différences selon la forme (ouvert/QCM) et la complexité des problèmes. Pour le premier problème, assez simple, nous

avons constaté que les scores de réussite étaient extrêmement proches entre les deux modalités (ouvert-QCM), mais il convient de noter que seulement les trois quarts des élèves qui ont réussi le problème en ouvert, le réussissent en QCM, alors que nous pensions que la réussite à un problème en ouvert entraînerait la réussite à son équivalent en QCM assez logiquement. A la différence du problème précédent, les scores de réussite du deuxième problème qui était assez complexe ont été très différents et en faveur du QCM (19 % de réussite en ouvert et 30 % de réussite en QCM) et cette fois, contrairement au problème additif, 90 % des élèves qui ont résolu le problème en ouvert choisissent la bonne réponse au QCM.

Pour les deux problèmes, le nombre de rétroactions permises par la forme QCM a été identique et a concerné environ 18 % des élèves. Dans le cas du problème additif, les rétroactions ont conduit la moitié des élèves à trouver, à la suite de la décision prise, la bonne réponse (à titre indicatif, 75 % ont repris leurs calculs ou les a vérifiés à la suite de la rétroaction). Sur le problème de proportionnalité, seulement 20 % des élèves a pu trouver la bonne réponse à la suite de la rétroaction ; à la différence du problème additif, 85 % des élèves n'ont pas cherché à reprendre leurs calculs, mais ont mis en place des stratégies autres que celles « de savoir ». Ainsi, dans le cas de problèmes complexes, l'effet rétroactif du QCM semble sans influence réelle sur la mobilisation de connaissances *a priori* évaluées.

Même si nous n'avons à notre disposition qu'un faible nombre d'items, nous avons pu constater, dans cette deuxième expérimentation, l'aspect potentiellement formateur du QCM notamment en situation de résolution de problèmes à travers l'étude des rétroactions fournies par les différents choix de réponse. Au-delà d'un apprentissage possible de stratégies propres aux QCM (écarter les réponses invraisemblables, utiliser les ordres de grandeur, tester les différentes propositions, etc.), apprendre avec les QCM pourrait permettre d'aider à la résolution de problèmes et développer chez l'élève une certaine autonomie : d'une part les différents choix de réponse « sécurisent » l'élève, mais d'autre part les choix de réponses lui permettent d'exercer un autocontrôle constructif sur ses résultats.

CONCLUSION

Ces deux études, menées à partir d'évaluation externes, nous ont permis de réaliser à quel point l'évaluation des élèves était un acte complexe, nécessitant la prise en compte de nombreux paramètres ne relevant pas exclusivement de connaissances didactiques.

Suite à ces recherches, nous pouvons conclure que l'étude des bilans évaluatifs dans une approche didactique permet de réaliser ce qui est réellement évalué du point de vue des connaissances en jeu, mais aussi de mieux comprendre comment les élèves s'emparent des questions posées selon les différents formats adoptés. Au-delà de ces constats, nous pensons que l'évaluation des apprentissages mathématiques des élèves doit faire l'objet d'une formation spécifique pour permettre aux professeurs de mieux exploiter les différentes modalités d'évaluation qui se trouvent à leur disposition (ce qui inclut les QCM) en fonction de leurs objectifs d'enseignement et ainsi en permettre un usage adapté selon les visées diagnostique, formative ou finale des évaluations qu'ils sont régulièrement amenés à concevoir dans leur classe.

REFERENCES BIBLIOGRAPHIQUES

ADDA J. (1976) Difficultés liées à la présentation des questions en mathématiques, *Educational*

Studies in Mathematics 7 3–22.

BECKERS J. (2002) *Développer et évaluer des compétences à l'école : vers plus d'efficacité et d'équité*. Bruxelles : Labor.

BENNETT R E., ROCK D A., WANG M. (1991) Equivalence of Free-response and Multiple-choice Items, *Journal of Educational Measurement*, 28-1, 77-92

BODIN A. (2004) Taxonomie des énoncés mathématiques, classement par niveaux hiérarchisés de complexité cognitive. <http://www.apmep.asso.fr/07-Documents-et-articles>.

BRIDGEMAN B. (1992) A comparison of Quantitative Questions in Open-Ended and Multiple-choice Formats, *Journal of Educational Measurement*, 29-3, 253-271

CHOPPIN B.H. (1975) Guessing the answer on objective tests, *British Journal of Educational Psychology* 45 206-213.

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

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 37–35.

DUVAL R., PLUVINAGE F. (1977) Démarches individuelles de réponse en mathématique, *Educational Studies in Mathematics* 8.1 51– 116.

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"*.

GREGOIRE J., LAVEAULT D. (1997) *Introduction aux théories des tests en sciences humaines*. Bruxelles : De Boeck.

FRARY R B. (1985) Multiple-choice versus free-response: a simulation study. *Journal of Educational Measurement*, 28-1, 21-31

KATZ, I., BENNETT, R. E., & BERGER, A. (2000) Effects of Response Format on Difficulty of SAT Mathematics Items: It's Not the Strategy. *Journal of Educational Measurement* 37.1 39–57.

LECLERCQ D. (1986) *La conception des questions à choix multiple*. Bruxelles : Labor.

LECLERCQ D. (1987) *Qualité des questions et signification des scores*. Bruxelles : Labor.

LECLERCQ D. & POUMAY M. (2004) Une définition opérationnelle de la métacognition et ses mises en œuvre. *Communication présentée à la 21ème conférence internationale de l'AIPU*, Marrakech.

LECLERCQ D. (2006) L'évolution des QCM. In Figari & Mottier-Lopez (Eds) *Recherche sur l'évaluation en éducation* 139– 146, L'Harmattan, Paris.

PERRENOUD P. (1997) *Construire des compétences dès l'école*. Paris : ESF.

PLUVINAGE F. (1979) Loto-questionnaires (pour l'évaluation et l'autocontrôle en mathématiques), *Educational Studies in Mathematics* 10 443–485.

REY B., CARETTE V., DEFRANCE A., KAHN S. (2002) Création d'épreuves étalonnées en relation avec les nouveaux socles de compétences pour l'enseignement fondamental. *Le point sur la*

Recherche en Éducation 23 51–63.

ROBERT A. (2008) Problématique et méthodologie communes aux analyses des activités mathématiques des élèves en classe et des pratiques des enseignants de mathématiques. In Vandebrouck F. (Ed) (pp. 33–57). Toulouse : Octares.

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.

ROCHEX J., CRINON J. (2011) *La construction des inégalités scolaires. Au cœur des pratiques et des dispositifs d'enseignement*. Rennes : Presses Universitaires de Rennes.

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.

ANNEXE 1

LISTE DES DIX STRATEGIES UTILISEES PAR DES ELEVES DE FIN D'ECOLE POUR REPENDRE A UN QCM

Stratégies A : stratégies de savoirs.

A1 : l'élève effectue la tâche demandée mentalement ou explicitement puis trouve, parmi les propositions, celle qui correspond à la réponse trouvée

A2 : l'élève reconnaît d'emblée la « bonne » réponse parmi celles proposées (connaissance intériorisée)

A3 : l'élève applique une règle simple intériorisée, correcte ou non (théorème en actes)

A4 : l'élève teste les propositions de réponse une à une jusqu'à trouver celle qui convient.

Stratégies B : stratégies de substitution ou de repli.

B1 : l'élève répond au hasard

B2 : l'élève passe en revue superficiellement toutes les propositions, puis choisit celle qui lui paraît la plus vraisemblable

B3 : l'élève ne sait pas expliquer sa procédure

B4 : l'élève combine les nombres en présence de manière à trouver, parmi les choix possibles, une solution.

Stratégies C : stratégies mixtes.

C1 : l'élève commence par s'engager dans une procédure de résolution, mais sans aller jusqu'au bout (à la différence de A1) ; il utilise ensuite les différentes propositions de réponses pour conclure (en choisissant celle qui lui paraît la plus vraisemblable)

C2 : l'élève élimine les propositions qui paraissent fausses, puis déduit de celle(s) qui reste(nt), la bonne réponse.