

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

ÉVALUER PAR QCM EN MATHÉMATIQUES EN FIN D'ÉCOLE : QUELLES STRATÉGIES SONT MISES EN ŒUVRE PAR LES ÉLÈVES ? QUELLES CONNAISSANCES SONT EFFECTIVEMENT ÉVALUÉES ?

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 : *Mathématiques, QCM, stratégies, degré de certitude.*

Résumé. *Le format de question QCM est utilisé régulièrement pour déterminer les connaissances des élèves en mathématiques à la fin de l'école primaire dans les évaluations à grande échelle, mais l'activité de l'élève face à cette forme spécifique de question est peu étudiée, en particulier pour des élèves de cet âge (10 – 11 ans). Notre recherche se situe dans le cadre de la didactique des mathématiques et vise à préciser les stratégies mises en œuvre par les élèves, les connaissances évaluées et le degré de certitude attribué à la réponse fournie. En éclairant les potentialités et les limites du format QCM, nous souhaitons apporter un élément de validité supplémentaire aux évaluations qui utilisent le format QCM afin de préciser les connaissances qui sont effectivement évaluées ; notre recherche porte principalement sur des connaissances mobilisées en résolution de problèmes et dans les transformations d'écriture de nombres entiers et décimaux.*

Le format de question QCM est utilisé régulièrement pour déterminer les connaissances des élèves en mathématiques à la fin de l'école primaire dans les évaluations nationales (par exemple, dans le bilan CEDRE¹ en France, les deux tiers des items ont ce format), mais les élèves de cet âge (10 - 11 ans) ne sont pas habitués à être évalués de cette façon, en classe. Il nous a alors semblé important de questionner ce format de questions pour préciser ce qu'il permet effectivement d'évaluer. Peu de recherches récentes ont été menées en France dans le cadre de la didactique des mathématiques sur ce sujet ; nous pouvons citer à titre d'exemple les recherches d'Adda (1976), Duval & 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...) ont été menées (notamment Choppin(1975), Leclercq (1986), Gilles (1996)).

Nous présenterons dans une première partie le cadre général de notre recherche, puis nous exposerons les deux expérimentations que nous avons menées et conclurons par des résultats et des perspectives sur l'utilisation des QCM pour évaluer les connaissances des élèves en fin d'école.

¹ Cycle des Évaluations Disciplinaires Réalisées sur Échantillon. Ce cycle d'évaluation mis en place en France par le Ministère de l'Éducation Nationale vise à évaluer les connaissances et les compétences des élèves de fin d'école et de fin de collège ; il fait l'objet de différentes publications, en particulier pour le bilan mathématiques fin d'école, de la Note d'Informations n°10-17 et d'une brochure (Lescure & Pastor (2008)).

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

1. Cadre et travaux de référence

1.1 Stratégies de réponse

On trouve dans les travaux relatifs aux QCM (Choppin (1975), Leclercq (1986, 1987)), différentes stratégies mais, toutes pensées pour des étudiants adultes. Voici, par exemple, les 3 modèles retenus par Choppin (1975) :

- Modèle 1 : quand l'étudiant « sait », il choisit la réponse correcte et quand il ne « sait pas », il choisit au hasard parmi les réponses proposées.

- Modèle 2 : commence comme le Modèle 1, mais au lieu de répondre au hasard quand il « ne sait pas », l'étudiant commence par éliminer les solutions qu'il sait être fausses et choisit au hasard parmi celles qui restent.

- Modèle 3 : l'étudiant commence par ranger les solutions possibles par ordre de plausibilité décroissante et, si la consigne l'oblige à ne fournir qu'une d'entre elles, alors il choisit celle dont la probabilité (subjective) est la plus élevée (à ses yeux).

En complément de ces modèles, d'autres catégorisations existent, comme celle de Katz & al. (2000), qui distinguent des stratégies « *traditional* » et « *non traditional* ». Les premières reposent sur des procédures qui sont enseignées et que l'élève utiliserait pour répondre si la question était posée sous forme ouverte ; les autres, spécifiquement associées à des QCM, sont celles où l'élève s'appuie sur les réponses proposées pour faire un choix.

1.1.1 Liste des stratégies retenues

Nous avons donc commencé par lister les stratégies possibles pour des élèves de dix - onze ans car il nous a semblé indéniable qu'elles ne pouvaient être identiques à celles d'étudiants adultes. Cette liste s'est complétée au fur et à mesure de notre expérimentation, au plus près des stratégies réelles des élèves, pour arriver à la liste des stratégies suivante :

- S1 : l'élève effectue la tâche demandée mentalement ou explicitement puis trouve, parmi les propositions, celle qui correspond à la réponse trouvée ;

- S2 : l'élève reconnaît d'emblée la « bonne » réponse parmi celles proposées (connaissance intériorisée) ;

- S3 : l'élève commence par s'engager dans une procédure de résolution, mais sans aller jusqu'au bout (à la différence de S1) ; il utilise ensuite les différentes propositions de réponses pour conclure (en choisissant celle qui lui paraît la plus vraisemblable) ;

- S4 : l'élève élimine les propositions qui paraissent fausses, puis déduit de celle(s) qui reste(nt), la bonne réponse ;

- S5 : l'élève répond au hasard ;

- S6 : l'élève passe en revue superficiellement toutes les propositions, puis choisit celle qui lui paraît la plus vraisemblable ;

- S7 : l'élève ne sait pas expliquer sa procédure ;

- S8 : l'élève combine les nombres en présence de manière à trouver, parmi les choix possibles, une solution ;

- S9 : l'élève applique une règle simple intériorisée, correcte ou non (théorème en actes) ;

- S10 : l'élève teste les propositions de réponse une à une jusqu'à trouver celle qui convient.

1.1.2 Groupes de stratégies

Afin d'analyser plus efficacement le lien entre les différentes stratégies des élèves en fonction de leur niveau de connaissance, nous avons effectué des groupements de stratégies par type :

- Stratégies A (S1, S2, S9, S10) : *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), soit il teste les propositions de réponse et choisit celle qui peut convenir. Les stratégies S1, S2 et S9 peuvent s'apparenter aux « *traditional* » stratégies définies par Katz & al (2000), alors que la stratégie S10 est spécifique au format QCM.

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

- Stratégies B (S5, S6, S7, S8) : *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 (S3, S4) : *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.

1.2 Degrés de certitude

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).

En effet, même si des travaux de recherche antérieurs ont stipulé que la consigne ne devait pas être verbale, mais devait plutôt être présentée sous forme de pourcentage (0%, 20 %...) car la subjectivité liée aux expressions verbales était problématique (Gilles, 1996 et Leclercq dès 1983), nous n'avons pas suivi cette préconisation. Nous avons en effet considéré que, pour des élèves de cet âge qui ne maîtrisent pas les pourcentages, il était plus opportun de leur présenter une consigne verbale à quatre degrés de certitude. La présentation de l'échelle choisie permettait ainsi de doubler la consigne verbale d'une consigne numérique, tout en évitant le recours aux pourcentages.

1.3 Rétroactions

A la différence des questions ouvertes, les différents choix de réponse présents dans le QCM peuvent apporter une rétroaction et conduire l'élève à prendre une décision ; nous avons listé trois types de décision possibles :

- D1 : l'élève reste dans une stratégie de savoir et vérifie ses calculs et/ou modifie éventuellement sa procédure de résolution pour trouver une réponse figurant parmi les choix ;
- D2 : l'élève utilise une autre stratégie (de repli ou mixte). Plusieurs possibilités s'offrent alors à lui : il peut choisir le nombre le plus proche du résultat qu'il a obtenu par ses calculs, choisir une réponse au hasard ou celle qui lui paraît la plus « vraisemblable », etc.
- D1&D2 : même après avoir repris ses calculs, l'élève ne trouve pas une des réponses proposées, et de ce fait, utilise une stratégie autre qu'une stratégie de savoir.

2. Première expérimentation

2.1 Méthodologie

Notre première 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 nos hypothèses de corrélation entre les stratégies adoptées et le niveau scolaire des élèves. 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'élèves : « faibles », « moyens » et « forts » selon leur score de réussite à ces évaluations nationales, communes à tous les élèves.

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

Nous avons convenu d'élaborer un test comportant un nombre limité d'items pour que chaque élève puisse nous expliciter son raisonnement individuellement lors de la passation, et ainsi 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 (1976), Roditi (2007)). En choisissant des distracteurs correspondant à des erreurs récurrentes d'élèves, nous souhaitons savoir s'ils étaient attractifs pour certains élèves (selon leur niveau de connaissances) ou au contraire s'ils étaient d'emblée rejetés. L'utilisation des degrés de certitude, notamment pour les réponses correspondant à des conceptions erronées sur les fractions-décimaux nous permettait, au delà du score de réussite, de préciser si l'élève est réaliste ou non sur ses connaissances.

Chaque item était accompagné d'une échelle permettant à l'élève de choisir le degré de certitude qu'il accorde à sa réponse et un espace libre a été prévu afin que les élèves puissent éventuellement faire des calculs.

Une présentation à la classe entière a été réalisée avant la passation individuelle pour expliquer la nature du test et le déroulement de l'entretien, en précisant les consignes : « pour chaque exercice, il faut trouver la bonne réponse parmi celles proposées et indiquer son degré de certitude ». Nous avons choisi de faire des suivis individuels lors des passations pour que les élèves puissent nous expliquer précisément comment ils avaient choisi leurs réponses et leurs degrés de certitude. Nous avons noté toutes les explications ou remarques des élèves au fur et à mesure de la passation.

2.2 Résultats

Le premier constat que nous pouvons faire concerne l'aspect spécifique du fonctionnement cognitif des élèves en situation de test. Il s'avère en effet qu'il est parfois extrêmement difficile d'appréhender leur fonctionnement cognitif car il ne s'inscrit pas dans la même logique que celui d'un adulte. Il n'a d'ailleurs pas toujours été facile de comprendre comment un élève était arrivé à faire tel ou tel choix car il n'était pas capable de l'expliquer lui-même.

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 S5 (au hasard). 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 7 items du test, les élèves les plus faibles sont plus enclins à utiliser des stratégies de repli ou de substitution.

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, 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

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

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 fausse). 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 fausse conception dont l'élève est peu assuré (« ignorance reconnue », (Gilles, 1996)), que lorsqu'elle apparaît à l'élève comme certaine.

3. Seconde expérimentation

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 demandées quelles pouvaient en être les raisons, ce qui nous a amenées à investiguer la façon dont les élèves appréhendaient les QCM en comparaison des questions ouvertes. Pour cela, nous avons observé les stratégies utilisées par des élèves de fin d'école primaire (195 élèves au total) face à un QCM et étudié les procédures mises en œuvre pour résoudre une tâche mathématique selon le format de question (QCM ou ouvert) qui lui était assigné.

Il s'agissait, pour nous, d'apporter des éléments de validité « locaux » au sens où ils ne concernaient pas un test en entier (comme par exemple la validité de contenu (Grégoire & Laveault, 2005), mais des items pris indépendamment les uns des autres et pour les lesquels nous interrogeons la validité psycho-didactique, c'est à dire vérifier que « l'évalué va effectivement mobiliser la connaissance attendue » (Vantourout & Goasdoué, accepté, 2014).

3.1 Méthodologie

Nous avons conçu deux tests de six items, équivalents du point de vue des tâches mathématiques à réaliser, 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 ouvert et sous forme de 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.

Les six items du test portaient sur différents domaines mathématiques (deux items « numération avec nombres entiers », deux items « fractions et décimaux » et deux items « résolution de problèmes ») et étaient de complexité variées. Nous avons veillé à ce que les questions restent identiques du point de vue de la tâche mathématique à réaliser, mais qu'elles varient par la forme (ouvert – fermé) sans changer de complexité ; nous avons pour cela utilisé l'outil d'analyse « facteurs de complexité et niveaux de compétences » (Sayac & Grapin, 2013) qui prend en compte le contexte de l'énoncé, la tâche mathématique et le niveau de compétences. Comme pour l'expérimentation précédente, le choix des distracteurs des questions portant sur les écritures de nombres (entiers et décimaux) a été effectué à partir des travaux en didactique des mathématiques cités précédemment ; pour les items relevant de la résolution de problèmes, nous avons proposé comme distracteurs des résultats d'étapes intermédiaires de la résolution du problème ou des résultats correspondant au choix d'une mauvaise opération.

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

3.2 Résultats

Dans ces actes, nous ne présentons que les résultats obtenus lors de l'étude comparative entre format ouvert et QCM, et ne détaillons pas les résultats spécifiques aux QCM pour cette expérimentation. Avant de faire un bilan global et d'interroger la pertinence du format QCM en évaluation en comparaison du format ouvert, nous donnons des résultats plus locaux, domaine par domaine.

3.2.1 Écriture de nombres entiers : passage d'une écriture en lettres à une écriture en chiffres

Dans le premier item, l'élève devait écrire en chiffres un nombre de l'ordre du million à partir de son écriture en lettres (sous forme de QCM, l'élève devait retrouver l'écriture en chiffres parmi quatre choix). La différence de format conduit, dans ce cas, à un changement important dans l'activité de l'élève, puisque dans le cas d'un QCM, il ne s'agit plus de produire une écriture, mais de la reconnaître. Nous observons pour cet item que la réussite au QCM est supérieure de 18 % à celle en ouvert : sous cette dernière forme, certaines erreurs qui relèvent *a priori* de « l'étourderie » sont évitées.

Dans le second item, l'élève devait donner (ou retrouver) le nombre de chiffres nécessaires pour écrire un nombre donné (par son écriture en lettres) dans quatre cas différents (le nombre de chiffres du nombre variant entre 6 et 9) ; les différentes possibilités offertes dans les QCM étaient ainsi de 6, 7, 8 ou 9 chiffres. Par exemple, il était demandé « *combien de chiffres sont nécessaires pour écrire vingt-cinq-millions-cinq-cent-mille* ». Comme pour le premier exercice nous constatons une meilleure réussite au QCM, mais avec des écarts de moindre importance avec le format ouvert, puisqu'ils varient selon les nombres de 5% à 15%. Nous expliquons ces écarts à la fois par une meilleure compréhension de la consigne sous format QCM (en ouvert, certains élèves n'ont pas compris la tâche et ont écrit les nombres demandés au lieu du nombre de chiffres) et par un effet rétroactif du QCM (les élèves qui se trompent en ouvert en écrivant un nombre de chiffres supérieur à 9 ou inférieur à 6 peuvent modifier leur réponse en QCM).

Pour ces deux items, nous constatons que pour répondre aux QCM la plupart des élèves utilisent des stratégies de savoir (96 % des élèves pour le premier exercice et 92 % pour le second) avec des procédures variées : écriture du nombre, puis comparaison avec les choix proposés, utilisation du nombre de chiffres par classe, etc.

3.2.2 Écriture de nombres décimaux : passage d'une écriture à virgule à une écriture sous forme de fraction décimale et réciproquement

Les différents items relevant de l'évaluation de cette connaissance sont semblables dans leur forme : l'élève doit produire (ou reconnaître lors du QCM) l'écriture à virgule d'un nombre décimal à partir de son écriture en fraction décimale et réciproquement. A titre d'exemple, les items A et B sous forme de QCM, permettant d'illustrer notre propos sont donnés ci-dessous ; l'item A est représentatif du sens de transformation de l'écriture fractionnaire vers l'écriture décimale (sens Q ->D) alors que l'item B est représentatif de la transformation en sens inverse (sens D -> Q).

Item A : A quel nombre correspond la fraction $62/10$?			
<input type="checkbox"/> 6,2	<input type="checkbox"/> 0,62	<input type="checkbox"/> 62,10	<input type="checkbox"/> 620
Item B : Quelle est la fraction égale à 237,8 ?			
<input type="checkbox"/> 237/8	<input type="checkbox"/> 2378/100	<input type="checkbox"/> 2378/100	<input type="checkbox"/> 237/10

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 -> D (item A), 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 -> Q, item B), la différence de réussite est ici

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

significative (autour de 20 %) en faveur du QCM. Cette différence de réussite, selon le format, dans le sens Q-> 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 $237 + 8/10$). Notons que pour un élève de fin d'école primaire, il associe généralement une seule écriture à virgule à la fraction décimale $62/10$, alors qu'il peut associer différentes écritures avec une fraction décimale à une écriture à virgule : $237,8 = 2378/100 = 23780/1000 = 237 + 8/10\dots$

Au delà de la différence de réussite entre ouvert et QCM selon le sens de la transformation, une autre différence apparaît dans les stratégies employées par les élèves pour répondre à la question (tableau 1 ci-dessous). Alors que la plupart des élèves emploient une stratégie de savoir pour répondre à l'item A, ils ne sont plus que la moitié à utiliser une stratégie de ce type pour répondre à l'item B. La différence de réussite entre le format ouvert (qui met en jeu la plupart du temps une stratégie de savoir) et le format QCM peut s'expliquer justement par la réussite à partir d'une stratégie autre que celle de savoir : ce qui interroge de la validité de ce type de format pour évaluer cette connaissance.

	Stratégies A	Stratégies B	Stratégies C
$62/10 = ???$	91,3%	5,1%	3,6%
$237,8 = ???$	51,3%	33,3%	15,4%

Tableau 1 - Répartition des stratégies suivant les items A et B (sous forme de QCM)

Par ailleurs, les erreurs produites par les élèves en format ouvert sont « cohérentes » avec celles que l'on trouve dans le QCM, c'est à dire que, parmi les choix qui sont proposés au QCM, ils choisissent une réponse identique à celle qu'ils ont fournie en ouvert : parmi les choix de réponse, la « bonne » ne joue pas un rôle d'attracteur. Ces élèves conservent donc leur représentation erronée, quelle que soit la forme de la question posée, ce qui est en lien avec la certitude qu'ils accordent à leur réponse.

3.2.3 Résolution de problèmes

Comme pour les autres exercices nous avons veillé à ce que les deux problèmes proposés dans le test soient déclinés avec des niveaux de complexité similaires entre chacun des deux formats (QCM et ouvert).

Le premier problème était additif et de faible complexité (pour la fin de l'école primaire). A titre informatif, l'énoncé du problème QCM était le suivant : *Une course automobile de 2 879 kilomètres se déroule sur trois jours. Le premier jour, les concurrents parcourent 745 km. Le deuxième jour, les concurrents parcourent 1 123 km. Quelle est la distance parcourue le troisième jour ?* Les distracteurs proposés étaient : 1 868 km, 1 434 km, 1 756 km, 1 011 km.

Nous constatons que pour ce type de problème, les scores de réussite sont 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. Nous avons alors observé que pour le quart des élèves qui se échouent au QCM et réussissent en ouvert, la moitié d'entre eux choisissaient le distracteur correspondant au résultat de l'étape intermédiaire : ces élèves effectuent la première opération, constatent que leur résultat figure parmi les choix proposés et s'arrêtent à cette étape. Le choix de ce type de distracteur, correspondant au résultat d'une étape intermédiaire de calcul, est donc à interroger pour juger de la validité de l'item.

Le deuxième problème était un problème de proportionnalité complexe ; comme pour le problème précédent, nous présentons à titre informatif le problème proposé en QCM : « Pour faire une

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

bonne confiture, il faut mettre 300 g de sucre pour 400g de fruits. Combien faut-il mettre de sucre pour 1 kg de fruits ? ». Les distracteurs étaient : 600 g, 700 g, 750 g, 800 g.

A la différence du problème précédent, les scores de réussite sont ici très différents et en faveur du QCM (19 % de réussite en ouvert et 30 % de réussite en QCM) et cette fois, contrairement au problème précédent, 90 % des élèves qui ont résolu le problème en ouvert choisissent la bonne réponse au QCM. Nous constatons aussi une cohérence globale des réponses fausses entre les deux formats, c'est à dire que les élèves qui ont fait une erreur de raisonnement en ouvert, font souvent la même en QCM : l'effet rétroactif potentiel du QCM n'a pas provoqué la reprise de calculs (D1), mais plutôt des décisions de type D2 (choix de la réponse la plus proche pour la moitié des élèves ayant trouvé 900 g).

Si nous nous attachons plus spécifiquement aux stratégies mises en œuvre pour répondre à ces deux problèmes, nous observons que, pour le problème additif, davantage de stratégies de savoirs sont utilisées, alors que pour le problème de proportionnalité, ce sont majoritairement des stratégies de repli qui sont employées (tableau 2 ci-dessous).

	Stratégies A	Stratégies B	Stratégies C
Problème additif	73,8 %	10,3 %	15,9 %
Problème de proportionnalité	51,3 %	33,3 %	15,4%

Tableau 2 - Répartition des stratégies suivant les deux problèmes (sous forme de QCM)

La question de la validité d'un tel item (sous forme de QCM) pour évaluer les connaissances des élèves en résolution de problème de proportionnalité « complexe » se pose ici : même si la taille de notre échantillon et le faible nombre d'items que nous avons étudiés ne nous permettent pas d'établir des conclusions générales, il s'avère apparemment difficile de pouvoir évaluer les connaissances des élèves à l'aide de QCM dans le cas de la résolution de problèmes complexes.

Pour les deux problèmes, le nombre de rétroactions a été identique et a concerné environ 18 % des élèves. Dans le cas du problème additif, peu complexe, les rétroactions permises par le QCM 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 (D1) 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, les décisions prises ont été de type D2 pour 85 % des élèves, c'est à dire qu'ils 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.

4. Conclusion

Dans les deux expérimentations que nous avons proposées, 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 Pluvinage, (1979) que « pointer cette erreur et faire remarquer son caractère attractif à un élève qui l'a commise est un service à lui rendre ».

Enfin, même si nous n'avons à notre disposition qu'un faible nombre d'items, nous avons pu constater, dans la deuxième expérimentation, l'aspect potentiellement formateur du QCM notamment en situation de résolution de problèmes à travers l'étude des rétroactions fournies par

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

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 auto-contrôle constructif sur ses résultats.

5. Références et bibliographie

Adda J. (1976). Difficultés liées à la présentation des questions en mathématiques. *Educational Studies in Mathematics*, 7, 3-22.

Choppin (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., Pluvinaige 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 FA.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. (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 21^{ème} conférence internationale de l'AIPU*, Marrakech.

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.

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

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.

Sayac N., & Grapin N. (2013). Facteurs de complexité et de compétences en mathématiques : un outil au service de la formation des enseignants. *Actes du colloque ADMEE 2013*, Freiburg.

Vantourout M. & Goasdoué R. (accepté). Approches et validité psycho-didactiques des évaluations, *Éducation et Formation*.

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.