Document généré le 19 mai 2024 14:04

Éducation et francophonie

Évaluer les capacités des élèves à résoudre des problèmes dans le cadre d'une évaluation externe, en France
Les spécificités de la forme QCM

Assessing students' problem-solving abilities in the context of an external evaluation in France
Specific features of the MCQ format

Evaluar las capacidades de los alumnos en resolución de problemas en el cuadro de una evaluación externa, en Francia
Las especificidades de la forma QCM

Nathalie Sayac et Nadine Grapin

Volume 42, numéro 2, automne 2014

Résolution de problèmes en mathématiques : un outil pour enseigner et un objet d'apprentissage

URI : https://id.erudit.org/iderudit/1027906ar
DOI : https://doi.org/10.7202/1027906ar

Aller au sommaire du numéro

Éditeur(s)

Association canadienne d'éducation de langue française

ISSN

1916-8659 (numérique)

Découvrir la revue

Citer cet article

Sayac, N. & Grapin, N. (2014). Évaluer les capacités des élèves à résoudre des problèmes dans le cadre d'une évaluation externe, en France : les spécificités de la forme QCM. *Éducation et francophonie*, *42*(2), 64–83. https://doi.org/10.7202/1027906ar

Résumé de l'article

Dans le cadre d'une évaluation nationale en mathématiques en fin d'école primaire (élèves de 10-11 ans) en France nous proposons, dans cet article, d'étudier plus particulièrement les items relevant de la résolution de problèmes afin de déterminer quelles compétences et connaissances sont véritablement évaluées par un dispositif comportant une grande part de QCM (questions à choix multiple). À cette fin, nous serons amenées à utiliser un outil didactique décliné en facteurs de complexité et de compétence, mais aussi à considérer les stratégies que les élèves utilisent pour répondre à ce type d'évaluation. Dans une expérimentation menée parallèlement à ce bilan, nous examinerons l'activité de résolution de problèmes suivant deux modalités de questionnement (fermé ou plus classique) afin de déterminer si elles ont une incidence sur les résultats des élèves. Nous souhaitons ainsi appréhender l'activité de résolution de problèmes en mathématiques aussi bien d'un point de vue cognitif (notamment avec les rétroactions rendues possibles par les QCM) que d'un point de vue institutionnel, en tenant compte de l'influence de différentes modalités d'évaluation.

Tous droits réservés © Association canadienne d'éducation de langue française, 2014

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Évaluer les capacités des élèves à résoudre des problèmes dans le cadre d'une évaluation externe, en France Les spécificités de la forme QCM

Nathalie SAYAC
Université Paris-Est Créteil, France

Nadine GRAPIN
Université Paris-Est Créteil, France

RÉSUMÉ

Dans le cadre d'une évaluation nationale en mathématiques en fin d'école primaire (élèves de 10-11 ans) en France nous proposons, dans cet article, d'étudier plus particulièrement les items relevant de la résolution de problèmes afin de déterminer quelles compétences et connaissances sont véritablement évaluées par un dispositif comportant une grande part de QCM (questions à choix multiple). À cette fin, nous serons amenées à utiliser un outil didactique décliné en facteurs de complexité et de compétence, mais aussi à considérer les stratégies que les élèves utilisent pour répondre à ce type d'évaluation. Dans une expérimentation menée parallèlement à ce bilan, nous examinerons l'activité de résolution de problèmes suivant deux modalités de questionnement (fermé ou plus classique) afin de déterminer si elles

ont une incidence sur les résultats des élèves. Nous souhaitons ainsi appréhender l'activité de résolution de problèmes en mathématiques aussi bien d'un point de vue cognitif (notamment avec les rétroactions rendues possibles par les QCM) que d'un point de vue institutionnel, en tenant compte de l'influence de différentes modalités d'évaluation.

ABSTRACT

Assessing students' problem-solving abilities in the context of an external evaluation in France – specific features of the MCQ format

Nathalie SAYAC
University Paris-Est Créteil, France

Nadine GRAPIN
University Paris-Est Créteil, France

As part of a national mathematics assessment for students at the end of elementary school (10-11 years old) in France, in this article, we propose a more specific study of problem solving elements to determine which skills and knowledge are actually being assessed through a test made up in large part of MCQ (multiple choice questions). To this end, we will need to use an educational tool broken down into skill and complexity factors, but also to consider student strategies for dealing with this type of evaluation. In an experiment conducted in parallel with this assessment, we will compare the activity of problem solving using two methods of questioning (closed and more classical) to determine if they have an impact on student results. We thus hope to understand the activity of problem solving in mathematics both from a cognitive point of view (with feedback from the MCQ) and from an institutional point of view, taking into account the influence of different assessment methods.

RESUMEN

Evaluar las capacidades de los alumnos en resolución de problemas en el cuadro de una evaluación externa, en Francia: las especificidades de la forma QCM

Nathalie SAYAC
Universidad Paris Est Créteil, Francia

Nadine GRAPIN
Universidad Paris Est Créteil, Francia

En el contexto de una evaluación nacional en matemáticas al terminar la educación primaria (alumnos entre 10-11 años de edad) en Francia, proponemos en este artículo, estudiar específicamente los elementos concernientes a la resolución de problemas, con el fin de determinar cuáles son las competencias y los conocimientos que realmente se evalúan gracias al dispositivo que forma la mayor parte de QCM (preguntas de opción múltiple). Para ello, utilizaremos una herramienta didáctica enunciada en factores de complejidad y de competencia, pero también consideramos las estrategias que los alumnos emplean para responder este tipo de evaluación. En un experimento realizado paralelamente a este chequeo, confrontaremos la actividad de resolución de problemas siguiendo dos maneras de cuestionar (cerradas o de manera más clásica) con el fin de determinar si tienen alguna incidencia sobre los resultados de los alumnos. Tratamos así de cernir la actividad de resolución de problemas en matemáticas tanto desde un punto de vista cognitivo (especialmente con las retroacciones que ofrece las QCM) que de un punto de vista institucional, teniendo en cuenta la influencia de diferentes modalidades de evaluación.

Introduction

La résolution de problèmes est au cœur des apprentissages mathématiques et tient une place prépondérante dans les programmes scolaires dans de nombreux pays. En France, il est indiqué[1] que la résolution de problèmes en mathématiques est le moyen qui « permet d'approfondir la connaissance des nombres étudiés, de renforcer la maîtrise du sens et de la pratique des opérations, de développer la rigueur et le goût du raisonnement » et que « la résolution de problèmes joue un rôle essentiel dans l'activité mathématique ».

1. Dans le Bulletin officiel (BO), hors-série n° 3 du 19 juin 2008.

Nous proposons dans le cadre de cet article d'interroger une[2] des évaluations externes menées par le ministère de l'Éducation nationale autour des apprentissages mathématiques en fin d'école primaire (classe de cours moyen deuxième année [CM2], élèves âgés de 10-11 ans), de déterminer quelles connaissances et compétences des élèves sont réellement évaluées dans le domaine de la résolution de problèmes à partir de la nature et de la forme des items proposés : pour analyser l'ensemble des items de ce domaine, nous avons développé un outil que nous présentons dans une première partie. Nous interrogeons également, sur quelques items, les scores de réussite obtenus au regard des modalités d'évaluation choisies (questions à choix multiple [QCM] ou questions à réponse ouverte courte [QROC][3]), des stratégies adoptées par les élèves et des rétroactions rendues possibles par les différentes propositions de réponse dans le cas des QCM.

Les évaluations institutionnelles veulent évaluer les connaissances et compétences des élèves en résolution de problèmes : qu'en est-il exactement?

L'évaluation bilan menée en 2008 en fin d'école primaire en mathématiques avait pour objectif d'évaluer les connaissances et les compétences en mathématiques au regard des objectifs fixés par les programmes français sur un échantillon représentatif d'élèves (Ministère de l'Éducation nationale, 2010). Cette évaluation bilan comptait environ 450 items, dont 46 portaient sur des problèmes arithmétiques que nous avons analysés à l'aide d'un outil didactique décliné en facteurs de complexité et de compétence.

Forme des questions et classe de problèmes

Tous les problèmes proposés dans cette évaluation étaient des problèmes de réinvestissement, sans caractère inédit[4] (aussi bien dans le contenu mathématique que dans la forme de l'énoncé), mettant en jeu presque uniquement des nombres entiers (seuls deux problèmes impliquaient des nombres décimaux non entiers) et sans données numériques inutiles.

Par souci d'efficacité et de faisabilité, les évaluations menées nationalement comportent un grand nombre d'items en QCM (réponses corrigées mécaniquement par un logiciel) et un nombre plus restreint de QROC (réponses corrigées par des enseignants). La plupart des QCM sont proposées avec quatre propositions de réponse, trois distracteurs et une seule réponse correcte. Les distracteurs correspondent le plus souvent à des erreurs anticipées d'élèves, soit parce qu'elles se rattachent

2. Nous analysons l'évaluation bilan faite en mathématiques dans le cadre du Cycle des évaluations disciplinaires réalisées sur échantillon (CEDRE) en 2008 (Lescure et Pastor, 2012).
3. Les questions qui ne sont pas sous forme QCM demandent une réponse qui est dans tous les cas un nombre, et pour laquelle la présence des opérations ou du raisonnement n'est pas exigée.
4. Il s'agit de problèmes additifs ou multiplicatifs similaires à ceux qui sont présents en grande majorité dans les manuels scolaires.

à des conceptions erronées[5], soit parce qu'elles correspondent à des erreurs « types » dans la résolution de problèmes (mauvais choix de l'opération, erreur de calcul, résultat d'une étape intermédiaire, etc.). Le choix des distracteurs n'est pas sans incidence sur le choix de réponse que fera l'élève, et donc sur le score de réussite à l'item (Sayac et Grapin, accepté).

Dans l'évaluation, sur les 46 items rattachés à la résolution de problèmes, 36 se présentaient sous la forme de QCM, 1 sous forme de Vrai-Faux et 9 sous une forme de QROC. La spécificité du format QCM modifie la tâche proposée à l'élève. Il ne s'agit plus de « produire une réponse », mais de « reconnaître un résultat », ce qui n'implique pas nécessairement la même activité chez l'élève et qui pourrait ainsi justifier une correction « experte » pour les QROC (ce que nous approfondirons plus loin).

Pour préciser les types de problèmes proposés, nous avons utilisé les typologies de problèmes additifs et multiplicatifs décrites par Vergnaud (1990) et reprises par Fénichel et Pfaff (2005). Cette analyse (voir l'annexe 1) montre une répartition assez déséquilibrée des problèmes additifs avec une majorité de problèmes de transformation et de composition de mesures, mais une quasi-absence de problèmes de comparaison et de composition de transformation. Ces derniers étant considérés comme difficiles pour des élèves de l'école primaire (Fénichel et Pfaff, 2005, p. 108), ils sont peu présents dans les manuels scolaires et nous supposons qu'ils sont peu travaillés en classe. De ce fait, les concepteurs de l'évaluation ont choisi de ne pas en proposer. Pour les problèmes multiplicatifs, nous constatons une prédominance des problèmes de proportionnalité (pour lesquels la valeur de l'unité n'est pas donnée dans l'énoncé) menant à la recherche d'une quatrième proportionnelle dans les cas classiques, aucun problème de pourcentage n'étant proposé. Cette analyse pointe la question du contenu de l'évaluation et de la validité des résultats qui pourra en être faite : comment choisir les tâches de l'évaluation pour que celles-ci évaluent vraiment la capacité à résoudre des problèmes, relativement aux programmes de l'école? Nous constatons qu'entre redondance de certains types de problèmes et absence (ou presque) d'autres, un équilibre serait à trouver.

Décrire le contenu de l'évaluation à travers les problèmes qui y sont posés ne suffit pas à déterminer les connaissances et les compétences qui seront évaluées. Poser la question de l'évaluation de la compétence d'un élève à résoudre un problème demande que l'on interroge d'abord les connaissances et les capacités mises en jeu dans la résolution pour déterminer des indicateurs qui permettront de les évaluer. En effet, résoudre un problème demande, en plus de la mobilisation de connaissances mathématiques, la possibilité de se représenter le problème (Julo, 1995 et 2002) et de mettre en œuvre des compétences transversales, comme celles de « vérifier, justifier, valider » ou encore de « rédiger une solution » (Coppé et Houdement, 2002).

5. Par exemple, au niveau de la comparaison de décimaux, sont proposés des distracteurs qui correspondent à la représentation erronée « un nombre décimal = 2 entiers séparés par une virgule » et qui induisent à penser que, par exemple, 3,14 > 3,8 car 14 > 8.

Pour le bilan de 2008, l'évaluation des capacités des élèves à résoudre des problèmes a été faite sans tenir compte réellement de tels processus ou de telles capacités définies spécifiquement et préalablement lors de la conception. Seuls deux types de tâches ont été proposés à l'élève : résoudre un problème (42 items) et reconnaître l'opération à mettre en œuvre dans la résolution (4 items)[6]. Par ailleurs, pour les 9 QROC pour lesquelles l'élève devait résoudre le problème, seules les solutions conduisant à la réponse exacte ont été considérées comme justes, et aucun crédit partiel n'a été accordé du fait de la nature de l'évaluation.

Description des items à partir de facteurs de complexité et de niveaux de compétence

Pour décrire plus précisément le contenu de l'évaluation, nous avons développé un outil s'appuyant sur différents travaux en didactique des mathématiques (Sayac et Grapin, 2013) qui tient compte à la fois de l'énoncé de la question, des savoirs mathématiques en jeu et des compétences dont l'élève doit faire preuve pour répondre. Cet outil, qui se veut fonctionnel pour l'analyse du contenu d'évaluations externes contenant de nombreux items, se décline en deux facteurs de complexité et en trois niveaux de compétence.

Définition des facteurs de complexité et des niveaux de compétence

Facteur de complexité 1 (FC 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 importants à considérer. Le contexte sémantique de l'énoncé (Julo, 1995), la présence de mots inducteurs permettant ou non une représentation aisée du problème posé, peut également se révéler pertinent pour analyser le score de réussite des élèves. En ce qui a trait au contexte de l'énoncé, il est également important de considérer dans quelle mesure ce contexte est lié à la « vie courante » des élèves[7], car, même si l'efficacité de ces problèmes est loin d'être prouvée (Boaler, 1993; Beswick, 2011), nous rejoignons Van den Heuvel-Panhuizen (2003) qui estime que « la caractéristique principale de ces situations est que les étudiants puissent les imaginer, qu'elles soient ''réelles" dans l'esprit des étudiants[8] ». Ainsi, divers éléments sont pris en compte pour évaluer ce premier facteur de complexité, même si son appréciation générale est laissée à l'initiative du chercheur.

---

6. Exemple d'item dans lequel l'élève doit reconnaître l'opération :
Martin a 24 ans. Il a 5 ans de plus que son frère Jacques. Pour calculer l'âge de Jacques, on peut faire :
□ 19 + 5 □ 24 - 5 □ 24 + 5 □ 24 - 19
Le faible nombre d'items correspondant à ce type de tâche, dans lequel l'élève n'a ni à calculer ni à produire lui-même une démarche de résolution, mais où il doit seulement la reconnaître, ne permet pas de donner des éléments complémentaires d'analyse, notamment pour préciser les capacités des élèves des bas niveaux en résolution de problèmes.

7. Les programmes français de mathématiques (2008, cycle 3) précisent explicitement que « la résolution de problèmes liés à la vie courante permet d'approfondir la connaissance des nombres étudiés, de renforcer la maîtrise du sens et de la pratique des opérations, de développer la rigueur et le goût du raisonnement ».

8. Traduction libre.

Facteur de complexité 2 (FC 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 pour évaluer ce facteur de complexité (notamment la classification des problèmes additifs et multiplicatifs [Vergnaud, 1990]). Dans ce facteur, seront également pris en compte les variables didactiques ainsi que les distracteurs des situations proposées, car ils peuvent avoir une influence non négligeable sur la réussite des élèves, dans un sens positif ou négatif.

Niveau de compétence des items (NC)

Pour ce facteur, nous nous sommes inspirées des différents niveaux de mise en fonctionnement des connaissances décrits par Robert (1998) et par Robert et Rogalski (2002) ainsi que des trois niveaux de compétence de Rey (2003).

- 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.

Afin d'avoir une vision globale des items de l'évaluation à partir des facteurs et des niveaux de compétence, nous avons attribué pour chaque item : de 0 à 2 points pour le facteur de complexité 1, de 1 à 3 points pour le facteur de complexité 2 et un des trois niveaux de compétence tels que nous les avons définis. Nous proposons en annexe 2 un exemple de répartition des différents facteurs pour 4 items de résolution de problèmes.

Analyse de l'évaluation selon les facteurs de complexité et les niveaux de compétence

Pour chacun des 46 items, nous avons déterminé les deux facteurs de complexité et un niveau de compétence. Le tableau de répartition de ces items (voir le tableau 1) nous amène à deux constats majeurs. D'une part, aucun problème ne relève du niveau de compétence 3, c'est-à-dire qu'aucun n'est complètement inédit et ne demande à l'élève de faire preuve d'initiative pour le résoudre. D'autre part, la complexité ne relève pas a priori du contexte de l'énoncé (tous les items ont un facteur de complexité 1 inférieur ou égal à 1), ni véritablement du savoir mathématique en jeu : la plupart des nombres impliqués dans les problèmes sont entiers et conduisent à des opérations « simples » pour les résoudre.

Tableau 1. **Répartition des items de résolution de problèmes arithmétiques selon les facteurs de complexité et les niveaux de compétence**

	Facteur de complexité 1 Contexte de l'énoncé			Facteur de complexité 2 Savoir mathématique en jeu			**Nombre d'items total**
	FC 1 = 0	FC 1 = 1	FC 1 = 2	FC 2 = 1	FC 2 = 2	FC 2 = 3	
Niveau de compétence 1	9	25	1	24	9	2	**35**
Niveau de compétence 2	0	9	2	5	5	1	**11**
Niveau de compétence 3	0	0	0	0	0	0	**0**
Nombre d'items total	**9**	**34**	**3**	**29**	**14**	**3**	**46**

Ainsi, l'utilisation des facteurs de complexité et des niveaux de compétence vient en complément des éléments de description utilisés dans le paragraphe précédent. En effet, même si les problèmes proposés sont variés par rapport aux classes de problème et à leur forme (même si l'on peut déplorer la faible proportion de QROC par rapport aux QCM), leur complexité (aussi bien par rapport au contexte de l'énoncé qu'au savoir mathématique en jeu) et le niveau de compétence auquel ils se réfèrent est majoritairement de 1. De tels items ne permettent donc pas de savoir si les connaissances des élèves résisteraient à des situations plus complexes; par conséquent, ils ne sont pas suffisamment variés didactiquement pour pouvoir caractériser les capacités en résolution de problèmes des différents groupes élèves construits à partir de l'échelle des scores[9].

QCM ou QROC : des items instructifs

Trois problèmes assez proches relevant du champ conceptuel des structures additives et visant à évaluer la capacité d'un élève à résoudre un problème additif de composition de mesure (Vergnaud, 1990) ont été proposés dans l'évaluation à la fois sous forme de QCM et de QROC et ont produit des scores de réussite qui nous ont interpellées.

Ces problèmes avaient pour contexte les distances entre trois villes (annexe 3). Un premier constat peut être fait sur le score de non-réponse, beaucoup plus élevé lorsque l'exercice est proposé en ouvert, et porte sur la recherche d'une des mesures (17,7 % contre 7,4 % pour le problème 2 et 19,8 % contre 7,6 % pour le problème 3). On peut alors supposer que les élèves s'engageraient moins dans la résolution d'un problème « difficile » lorsque ce dernier est proposé en ouvert ou alors que les différentes propositions de réponse « sécurisent » les élèves ou sont utilisées directement par eux pour répondre à la question posée.

9. Le bilan CEDRE conduit à la construction d'une échelle de score et à la définition de six niveaux de performance; chacun de ces groupes est caractérisé par les items que maîtrisent les élèves du groupe (Lescure et Pastor, 2012).

Le deuxième constat porte sur le score de réussite global à chacun de ces exercices : la forme de la question (QCM ou QROC) a peu d'impact sur la réussite dans les deux premiers problèmes (autour de 85 % pour le problème 1, autour de 70 % pour le problème 2), alors que la différence est plus grande dans le cas du problème 3 : les scores de réussite diffèrent de 10 % (62,8 % pour la forme QROC, contre 72,7 % pour la forme QCM).

Ces constats nous ont amenées à investiguer de manière plus approfondie les scores de réussite des élèves et leurs stratégies de réponse aux QCM, avec d'éventuelles rétroactions.

Exploration des QCM

Une expérimentation complémentaire a donc été menée pour explorer ces dimensions. Elle s'est déroulée en juin 2013 et a concerné 195 élèves de CM2, de 9 classes différentes.

Les modalités d'évaluation : QCM ou QROC

Nous avons élaboré deux tests, composés de sept items équivalents du point de vue des tâches mathématiques à effectuer[10], l'un sous forme de QROC, l'autre sous forme de QCM; nous nous intéressons dans cet article uniquement aux items de résolution de problèmes. Dans un premier temps, les élèves ont réalisé le premier test (ouvert) avec leur enseignant, sans qu'aucune indication ou aide leur soit fournie. La trace de leurs calculs nous permettait donc d'avoir accès à leur procédure et à leurs erreurs éventuelles. Dans une deuxième phase (le lendemain), nous sommes intervenues dans les classes[11] pour faire passer le deuxième test (QCM) à tous les élèves, individuellement, afin d'observer leurs stratégies et de les interroger sur leur choix de réponses.

Les premiers problèmes que nous comparons (Problème 1 sous forme de QROC et de QCM, annexe 4) relevaient du champ conceptuel des structures additives et de la composition de mesures. Quelles que soient les procédures utilisées, les opérations à effectuer ne comprenaient pas de retenue pour focaliser l'attention sur le choix des opérations et non sur leur complexité. Du point de vue de notre outil d'analyse, ces problèmes sont affectés de faibles facteurs de complexité et de niveaux de compétence moyennement élevés (FC1 = 1, FC2 = 1 et NC = 2).

Les scores de réussite à ces deux problèmes sont extrêmement proches, puisque 70,8 % des élèves ont réussi le problème « ouvert » et 69,2 % le problème QCM, mais nous verrons par la suite que, du point de vue de l'activité de l'élève, ce résultat est à nuancer.

10. Seuls le contexte et les nombres (entre le problème sous forme de QROC et le problème sous forme de QCM) variaient, sans que cela ait un impact sur les procédures et les opérations en jeu.
11. La plupart du temps, nous nous sommes placées dans le couloir attenant à la classe ou dans un espace jouxtant la classe afin que les élèves puissent, à tour de rôle, venir réaliser leur test sans perturber le déroulement de la classe.

Si l'on regarde plus finement combien d'élèves ayant répondu correctement au problème ouvert ont coché la bonne réponse dans la version QCM, il s'avère que l'on ne retrouve pas la totalité des élèves, mais seulement 76 %. Parmi ceux qui ne répondent pas correctement, plus de la moitié (51 %) cochent le distracteur correspondant à une étape intermédiaire du problème. La question des stratégies adoptées par les élèves confrontés à des QCM que nous développerons par la suite est également un élément à prendre en compte dans cette analyse.

Les seconds problèmes relevaient du champ conceptuel des structures multiplicatives, et plus particulièrement d'un problème de proportion simple, avec une présence fictive[12] de l'unité (Problème 2, annexe 4). Du point de vue de notre outil d'analyse, ces problèmes sont affectés de facteurs de complexité et de niveaux de compétence plus élevés que le premier problème (FC1 = 1, FC2 = 2, NC = 3).

Les scores de réussite ont été, pour ces problèmes, très distincts, puisque 19 % des élèves ont réussi le problème « ouvert », alors qu'ils sont 30,3 % à avoir coché la bonne réponse pour le problème QCM. Parmi les élèves ayant donné la bonne réponse au problème ouvert, près de 90 % ont répondu dans le même sens au problème QCM. On peut donc évoquer une cohérence des réponses correctes pour ces problèmes. Néanmoins, on note que de nombreux élèves (39%) ont trouvé la valeur 900 g pour le problème ouvert, alors que cette réponse n'était pas proposée en distracteur dans le problème QCM. Ces élèves ont certainement été déroutés par le choix des réponses proposées[13] et ils ont, pour près de la moitié, coché la case 800 g qui correspondait à la valeur la plus proche de leur réponse en ouvert. Les autres choix se sont portés sur les autres réponses proposées, dont environ 20 % sur la bonne réponse.

Ces deux exemples, contradictoires dans leurs résultats, ne nous permettent évidemment pas d'établir un constat général sur l'impact de la forme de la question sur les scores de réussite des élèves, mais ils illustrent la complexité de la problématique de l'évaluation des élèves en résolution de problèmes qui ne peut être réduite à forme QROC/forme QCM. Au-delà des scores de réussite aux différents items, nous avons donc voulu explorer quelles stratégies les élèves développaient pour répondre aux questions posées.

Les stratégies de réponses

Comme nous l'avons relevé dans un problème de proportionnalité de notre expérimentation, le taux de non-réponse peut être plus faible lorsque la question est posée sous forme de QCM. On peut alors supposer que cet écart est dû au fait que les élèves ont la possibilité de choisir la réponse « au hasard ». Mais on peut aussi penser que la présence des distracteurs les « sécurise » et les conduit à mettre en place des stratégies autres que celles d'évitement (Focant, 2003). Pour les notions de stratégie et de procédure dans le cadre de la résolution de problèmes, nous nous référons à Julo (1995) chez qui « la notion de stratégie correspond à un niveau de description

12. Le 1 kg proposé devait être converti en 1000 g, ce qui aboutissait à un problème de proportion simple, sans présence de l'unité.
13. La notion de rétroaction que nous évoquerons plus loin nous permettra de revenir sur cette affirmation.

assez global et qui cherche à rendre compte de l'orientation générale que prend l'activité de résolution, alors que celle de procédure correspond à l'ensemble des opérations élémentaires que l'on met en œuvre pour atteindre le but proposé».

Groupes de stratégies

Dans une précédente recherche (Sayac et Grapin, accepté), nous avions cherché à savoir si les stratégies qu'élaborent les élèves pour répondre aux QCM étaient toujours les mêmes et si elles étaient en lien avec leur niveau de connaissances en mathématiques. Pour cela, nous avons répertorié les stratégies possibles pour des élèves de 11 ans, car il nous a semblé indéniable qu'elles ne pouvaient être identiques à celles d'étudiants adultes (Choppin, 1975; Leclercq, 1987). Nous en avons retenu dix (annexe 5), que nous avons finalement regroupées en trois groupes:

Stratégies A (S1, S2, S9, S10): stratégies de savoirs

Quand 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.

Stratégies B (S5, S6, S7, S8): stratégies de substitution ou de repli

Quand 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

Quand l'élève a entrepris un raisonnement pour répondre à la question posée, mais qu'il se sert des différentes propositions de réponse pour faire un choix.

Stratégies utilisées par les élèves

Concernant les différentes stratégies utilisées par les élèves pour répondre à nos deux problèmes QCM, on peut noter des différences assez marquées, puisque les stratégies A sont utilisées à 73% pour le problème 1 et à 51% pour le problème 2, alors que les stratégies B le sont à 10% pour le problème 1 et à 33% pour le problème 2. Notre outil d'analyse d'items nous a permis de distinguer ces deux problèmes au niveau des facteurs de complexité et de compétence, le problème 2 ayant des niveaux plus élevés que le problème 1, ce qui pourrait expliquer pourquoi les élèves ont davantage recours à des stratégies de substitution pour le problème le plus difficile. Nous pouvons également préciser que, parmi les élèves qui adoptent des stratégies A, les scores de réussite sont de 85% pour le problème 1 et de seulement 46% pour le problème 2. Concernant les stratégies de substitution, elles amènent à des scores de réussite également variés, puisque 35% des élèves qui les adoptent répondent correctement au problème 1, contre 17% pour le problème 2.

Si l'on veut affiner cette analyse, on peut relever que, parmi les différentes stratégies retenues, celle qui correspond à une étape intermédiaire de la résolution du problème[14] est la plus déployée, après celle qui amène l'élève à effectuer la tâche

14. 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).

demandée jusqu'au bout. Par ailleurs, la stratégie de hasard n'est pas une stratégie très utilisée par les élèves de fin d'école primaire (respectivement 1,54 % et 8,72 % pour les problèmes 1 et 2).

Les rétroactions possibles

Rétroactions liées à la spécificité du format QCM

Résoudre un problème avec une question posée sous forme de QCM apporte à l'élève des éléments de contrôle complémentaires à ceux présents lorsqu'il résout le problème sous une forme ouverte. En effet, s'il obtient un résultat qui ne figure pas parmi les quatre choix de réponse, il sait que ce qu'il a trouvé est faux. Même si dans le cadre d'une évaluation bilan l'élève n'est pas dans une situation d'apprentissage et que l'enseignant n'intervient à aucun moment, nous considérons, dans le cas des QCM, que l'énoncé et les différents choix de réponse constituent un milieu (Brousseau, 1998) qui provoque des rétroactions et amène l'élève à agir sur la situation et à prendre des décisions (changer de stratégie ou de procédure de résolution si la réponse ne figure pas parmi les choix).

Ce type de rétroaction, pour la forme spécifique du QCM, apporte à l'élève « un moyen adéquat [...] pour répondre à ses doutes quant à sa résolution en voie d'élaboration » (Burgermeister et Coray, 2008) et, en ce sens, lui fournit un élément rétrospectif relevant d'un processus de contrôle. Il vient en complément d'autres éléments de contrôle entrant habituellement dans la résolution de problèmes, comme ceux listés par Houdement (2011) qui s'exercent dans le choix d'un modèle (contrôles pragmatiques, sémantiques et syntaxiques).

Les éléments de contrôle prospectifs et rétroactifs (Burgermeister et Coray, 2008) ne sont pas sans influence sur la façon dont l'élève oriente son choix de réponse in fine. La question ne se pose guère lorsque le résultat qu'il trouve coïncide avec un choix de réponse ou lorsque l'élève ne s'est pas engagé dans une stratégie de savoir (s'il a répondu au hasard par exemple). Par contre, lorsque ce n'est pas le cas, l'élève doit prendre une décision.

- D1 : L'élève reste dans une stratégie de savoir et il 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 ou choisir le nombre pour lequel l'ordre de grandeur correspond à ce qu'il avait anticipé prospectivement (ou qu'il détermine rétrospectivement) ou, encore, combiner les nombres en présence pour retrouver une des propositions de réponse (Stratégie S8) ou répondre au hasard.
- D1 et D2 : Même après avoir repris ses calculs, l'élève ne trouve aucune des réponses proposées et, de ce fait, il utilise une stratégie autre qu'une stratégie de savoir.

Effet des rétroactions sur les stratégies utilisées et la réussite

Dans le cadre de notre expérimentation, pour le problème 1 comme pour le problème 2, l'effet rétroactif du QCM a été observé pour un nombre semblable d'élèves : 36 élèves (18 % des élèves) pour le problème 1 et 35 élèves pour le problème 2. Par contre, les décisions qui ont été prises par les élèves sont très différentes d'un problème à l'autre : ils prennent majoritairement une décision de type 1 (D1) pour le problème 1 (26 élèves sur 36) et majoritairement une décision de type 2 (D2[15]) dans le problème 2 (30 élèves sur 35).

Si l'on observe ensuite la qualité de la réponse choisie après cette décision, on constate que, pour le problème 1, 16 élèves sur 36, soit plus de la moitié, trouvent ensuite la bonne réponse (15 d'entre eux avaient pris une décision D1); alors que, pour le problème 2, seuls 7 élèves sur 35 trouvent ensuite la bonne réponse.

Nous observons ainsi, sur la résolution de ces problèmes, un effet « QCM » lié aux éléments de contrôle apportés par les choix de réponse. On peut alors s'interroger sur le choix des problèmes à proposer sous la forme de QCM plutôt que sous forme ouverte dans les évaluations pour ne pas biaiser les résultats; notre expérimentation tendrait à montrer que, du point de vue des scores de réussite, l'effet « QCM » est finalement très limité pour les problèmes complexes (comme pour le problème 2 de notre expérimentation) dans le sens où l'élève ne sait pas comment modifier sa procédure pour trouver la bonne réponse. Néanmoins, le QCM ne conduit pas forcément l'élève à une même activité que pour une question sous forme ouverte et, de ce fait, ne permet pas d'évaluer les mêmes capacités.

Conclusion

Dans l'élaboration d'un dispositif d'évaluation, le choix des tâches est une étape importante pour garantir la validité des résultats : celles-ci doivent être représentatives du domaine étudié, mais elles ne peuvent pas être exhaustives. Une analyse didactique des tâches permettrait ainsi d'éviter certaines redondances et l'utilisation des facteurs de complexité et de compétence permettrait de garantir des problèmes de complexité variée. Comme nous l'avons montré dans la deuxième partie, le choix des items doit aussi s'accompagner d'une réflexion sur le format de la question (QCM ou non) et sur les modalités de correction afin de pouvoir évaluer différentes compétences qui interviennent dans la résolution d'un problème. Dans le champ de la psychométrie, des modèles (modèles diagnostiques) prenant en compte les différentes compétences qui interviennent dans la résolution d'un problème se développent (Loye, 2011); ils devraient ainsi permettre de prendre en compte des compétences transversales qui interviennent dans la résolution de problèmes et, pour les évaluations bilans, de mieux caractériser les compétences des élèves dans ce domaine.

15. Comme expliqué précédemment, certains élèves trouvaient pour réponse 900 g pour le problème 2. Ce choix n'étant pas proposé, ils ont alors majoritairement choisi 800 g, qui était la masse la plus proche de celle qu'ils auraient voulu choisir.

En ce qui concerne les QCM, même si leur usage en évaluation bilan modifie l'activité de l'élève et par conséquent influe sur les résultats, nous constatons qu'ils amènent l'élève à exercer un autocontrôle sur ses résultats; en ce sens, ils pourraient être exploités de façon constructive en classe pour apprendre à l'élève à reprendre sa démarche, à vérifier ses calculs en toute autonomie, et, par conséquent, l'amener à développer des compétences transversales nécessaires en résolution de problèmes. Cette exploitation, encore à la marge en France, mériterait d'être développée.

Pour conclure, nous dirons que l'évaluation des élèves en résolution de problèmes mathématiques est une question cruciale pour tous les acteurs institutionnels de l'école, mais aussi pour les chercheurs en éducation et en didactique car elle est multidimensionnelle. Elle pose aussi bien la question de la nature de l'évaluation, de son contenu que des résultats qu'elle donne à voir. À travers une évaluation externe menée à grande échelle en France, en fin d'école primaire, nous avons pu montrer comment le choix des problèmes proposés (facteurs de complexité et niveaux de compétence), la nature des questions posées (QCM ou forme ouverte) et la façon dont les élèves répondaient (stratégies, rétroactions) n'étaient pas sans incidence sur ce qui était évalué (capacités, connaissances), ni sur les résultats obtenus. À l'heure où les évaluations nationales et internationales prennent une place de plus en plus grande dans les choix politiques des pays, il convient donc de rester vigilant et de se donner les moyens d'avoir un regard objectif et constructif sur ces évaluations.

Références bibliographiques

BESWICK, K. (2011). Putting context in context. An examination of the evidence for benefits of « contextualised » tasks. *International Journal of Science and Mathematics Education*, *9*, 367-390.

BOALER, J. (1993). Encouraging transfer of 'school' mathematics to the 'real world' through the integration of process and content; context and culture. *Educational Studies in Mathematics*, *25*, 341-373.

BROUSSEAU, G. (1998). *Théorie des situations didactiques.* Grenoble : La pensée sauvage.

BURGERMEISTER, P. F. et CORAY, M. (2008). Processus de contrôle en résolution de problèmes dans le cadre de la proportionnalité des grandeurs : une analyse descriptive. *Recherches en didactique des mathématiques*, *28*(1), 63-106.

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

COPPÉ, S. et HOUDEMENT, C. (2002). Réflexion sur les activités concernant la résolution de problèmes à l'école primaire. *Grand N*, *69*, 53-62.

FOCANT, J. (2003) Impact des capacités d'autorégulation en résolution de problèmes chez les enfants de 10 ans. *Éducation et francophonie, 31*(2), 45-64.

HOUDEMENT, C. (2011). Connaissances cachées en résolution de problèmes arithmétiques ordinaires à l'école. *Annales de didactique et de sciences cognitives, 16,* 67-96.

JULO, J. (1995). *Représentation des problèmes et réussite en mathématiques.* Rennes: Presses universitaires de Rennes.

JULO, J. (2002). Des apprentissages spécifiques pour la résolution de problèmes? *Grand N, 69,* 31-52.

FÉNICHEL, M. et PFAFF, N. (2005). *Donner du sens aux mathématiques. Nombres, opérations et grandeurs.* Tome 2. Paris: Bordas.

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

LESCURE, S., PASTOR, J.-M. (2012). *Mathématiques en fin d'école primaire. Le bilan des compétences.* Paris: Scéren.

LOYE, N. (2011). Validité du diagnostic issu d'un mariage entre didactique et mesure sur un test existant. Dans G. Raîche, K. Paquette-Côté et D. Magis (dir.), *Des mécanismes pour assurer la validité de l'interprétation de la mesure en éducation* (p. 11-29). Québec: Presses de l'Université du Québec.

MINISTÈRE DE L'ÉDUCATION NATIONALE (2010). *Les compétences en mathématiques des élèves de fin d'école primaire.* Note d'information de la Direction de l'évaluation, de la prospective et de la performance (DEPP), nº 10-17.

REY, B., CARETTE, V., DEFRANCE, A. et KAHN, S. (2003). *Les compétences à l'école: apprentissage et évaluation.* Bruxelles: De Boeck.

ROBERT, A. (1998). Outil d'analyse des contenus mathématiques à enseigner au lycée et à l'université. *Recherches en didactique des mathématiques, 18*(2), 139-190.

ROBERT, A. et ROGALSKI, 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.

SAYAC, N. et GRAPIN, N. (2013). *Facteurs de compétence et de complexité en mathématique: un outil au service de la formation des enseignants.* Actes du 25e colloque de l'ADMEE-Europe, Fribourg.

SAYAC, N. et GRAPIN, N. (accepté). Évaluer par QCM en fin d'école: stratégies et degrés de certitude. *Annales de didactique et de sciences cognitives.*

VAN DEN HEUVEL-PANHUIZEN, M. (2003). The didactical use of models in realistic mathematics education. An example from a longitudinal trajectory on percentage. *Educational Studies in Mathematics, 54,* 9-35.

VERGNAUD, G. (1990). La théorie des champs conceptuels. *Recherches en didactique des mathématiques, 10*(2/3), 133-170.

Annexe 1

Classification des types de problèmes selon Vergnaud (1990), à partir de Fénichel et Pfaff (2005)

		Nombre d'items
STRUCTURES ADDITIVES		**17**
TRANSFORMATION DE MESURES	Recherche de l'état final	2
	Recherche de l'état initial	5
	Recherche de la transformation	2
COMPOSITION DE MESURES	Recherche d'une des deux mesures composantes	2
	Recherche de la mesure composée	5
COMPARAISON DE MESURES	Recherche d'une mesure dans le sens de la comparaison	
	Recherche d'une mesure dans le sens contraire de la comparaison	1
	Recherche de la comparaison	
COMPOSITION DE TRANSFORMATIONS	Recherche de la transformation totale	
	Recherche d'une des deux transformations	
STRUCTURES MULTIPLICATIVES : problèmes d'isomorphisme de mesure		**22**
PROBLÈMES MULTIPLICATIFS (l'énoncé fait intervenir l'unité)		
Problèmes de recherche d'une quatrième proportionnelle	Recherche du tout	1
	Recherche de la valeur unitaire	3
	Recherche du nombre d'unités	1
Problème de rapport entre deux valeurs de grandeurs de même nature ou problème de comparaison	Recherche de la valeur finale	2
	Recherche de la valeur initiale	
	Recherche du coefficient multiplicateur	
PROBLÈMES DE PROPORTIONNALITÉ (l'énoncé ne fait pas intervenir l'unité)		
Problème de recherche de quatrième proportionnelle	Problème de type classique	12
	Problème de type pourcentage	
	Problème de comparaison	3
PROBLÈMES MIXTES		**7**

Annexe 2

Application des facteurs de complexité et niveaux de compétence sur différents items

Exemple extrait de Lescure et Pastor (2012); dans l'évaluation CEDRE 2008, les questions données en exemple ci-dessous ont été posées sous forme de QCM.

Pour faire une salade de fruits, on a utilisé la recette suivante : 800 g de fruits pour 160 g de sucre.		Facteur de complexité 1	Facteur de complexité 2	Niveau de compétence
Item 1	Combien faut-il de sucre pour 400 g de fruits?	1	1	1
Item 2	Combien faut-il de sucre pour 1 600 g de fruits?	1	2	1
Item 3	Combien faut-il de fruits pour 40 g de sucre?	1	2	1
Item 4	Combien faut-il de fruits pour 120 g de sucre?	1	3	3

Pour tous les items, le facteur 1 (relatif à l'énoncé) est le même : l'énoncé est simple et correspond à une situation que l'on peut considérer comme proche de l'élève.

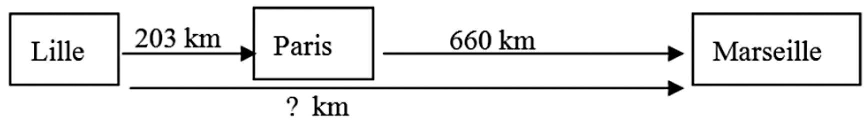
La différence de complexité entre les items se joue au niveau des relations entre les nombres en jeu; ainsi, le rapport entre 800 et 400 est reconnu plus facilement par les élèves que celui entre 40 et 160 ou encore celui entre 120 et 160. Ce qui explique la différence de points attribués au facteur de complexité 2, correspondant au savoir mathématique en jeu.

Enfin, pour les trois premiers items, nous considérons qu'il s'agit d'une application directe de connaissance (niveau de compétence 1), mais que la tâche est moins usuelle pour le dernier (niveau de compétence 3) : les procédures à utiliser pour répondre sont moins disponibles à ce niveau (le rapport entre 120 et 160 n'est pas entier, ni fractionnaire simple; l'élève peut utiliser la question précédente ou encore le retour à l'unité).

Annexe 3

Analyse et score de réussite à trois problèmes additifs extraits de CEDRE 2008, posés sous forme ouverte (o) et sous forme de QCM (qcm).

Problème 1 :

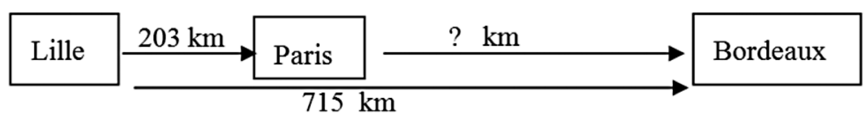


Quelle est la distance entre Lille et Marseille ?

□ 457 km □ 473 km □ 663 km □ 863 km

Scores de réussite	
Problème 1 o	Problème 1 qcm
Réussite: 85,39 %	Réponse 1 : 2,78 %
Autres réponses : 7,52 %	Réponse 2 : 0,82 %
Pas de réponse : 7,09 %	Réponse 3 : 3,75 %
	Réponse 4 : 86,67 %
	Pas de réponse : 5,83 %

Problème 2 :

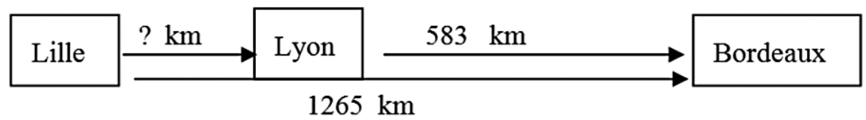


Quelle est la distance entre Paris et Bordeaux ?

□ 512 km □ 518 km □ 912 km □ 918 km

Scores de réussite	
Problème 2 o	Problème 2 qcm
Réussite: 70,79 %	**Réponse 1 : 68,62 %**
Autres réponses : 11,51 %	Réponse 2 : 10,07 %
Pas de réponse : 17,7 %	Réponse 3 : 3,56 %
	Réponse 4 : 10,32 %
	Pas de réponse : 7,42 %

Problème 2 :



Quelle est la distance entre Lille et Lyon ?

□ 682 km □ 1082 km □ 1265 km □ 1848 km

Scores de réussite	
Problème 3 o	Problème 3 qcm
Réussite: 62,8 %	**Réponse 1 : 72,7 %**
Autres réponses : 17,38 %	Réponse 2 : 7,22 %
Pas de réponse : 19,83 %	Réponse 3 : 3,85 %
	Réponse 4 : 8,25 %
	Pas de réponse : 7,86 %

Annexe 4

Problème 1 (QROC)

Lors des soldes, dans un grand magasin, on a vendu 2 768 vêtements en trois jours.
Le premier jour, 634 vêtements ont été vendus. Le deuxième jour, 1 012 vêtements ont été vendus.
Combien de vêtements ont été vendus le troisième jour?

Problème 1 (QCM)

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?
□ 1 868 km
□ 1 434 km
□ 1 756 km
□ 1 011 km

Problème 2 (QROC)
Pour faire des gâteaux, un pâtissier mélange 300 g de sucre avec 400 g de farine.
Combien faut-il mettre de sucre pour 1 kg de farine?

Problème 2 (QCM)
Pour faire une bonne confiture, il faut mettre 300 g de sucre pour 400 g de fruits.
Combien faut-il mettre de sucre pour 1 kg de fruits?
▫ 600 g
▫ 700 g
▫ 750 g
▫ 800 g

Annexe 5

S1 : L'élève effectue la tâche demandée mentalement ou explicitement, puis il 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 une solution parmi les choix possibles.
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.