HAL open science

Concevoir et mettre en oeuvre des évaluations au service des apprentissages numériques des élèves au cycle 2

Nadine Grapin, Eric Mounier

To cite this version:

Nadine Grapin, Eric Mounier. Concevoir et mettre en oeuvre des évaluations au service des apprentissages numériques des élèves au cycle 2. Nouvelles perspectives en didactique : géométrie, évaluation des apprentissages mathématiques, II, La Pensée Sauvage, pp.391-410, 2019, 978 2 85919 335 5. hal-02912917

HAL Id: hal-02912917

https://hal.science/hal-02912917v1

Submitted on 23 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONCEVOIR ET METTRE EN ŒUVRE DES EVALUATIONS AU SERVICE DES APPRENTISSAGES NUMERIQUES DES ELEVES AU CYCLE 2

Nadine Grapin, Eric Mounier

RESUME
En nous appuyant sur la mise en œuvre d'un dispositif d'évaluation des connaissances des élèves en fin de CP, nous abordons des questions relatives à la conception d'une évaluation et les considérons dans un cadre didactique. Initialement définies en théorie des tests, les conditions de pertinence, validité et fiabilité sont ici réinterprétées et nous montrons la façon dont nous les avons considérées pour concevoir notre dispositif d'évaluation. En lien avec les conditions de validité et l'interprétation des réponses des élèves, nous illustrons ensuite sur une tâche de dénombrement l'intérêt de considérer un cadre d'analyse didactique pour interpréter des productions d'élèves. Enfin, la présentation du dispositif d'évaluation dans sa globalité, à travers les tâches proposées et les conditions de passation, permet de soulever divers points de vigilance quant à sa mise en œuvre, tels que la passation des consignes, l'influence des conditions passations sur l'activité de l'élève, la formation des évaluateurs au codage des productions, etc. Nous proposons le nouveau label « tâches d'évaluation robustes » pour prendre en compte ces différentes dimensions.

Mots clefs : évaluation, validité, numération, dénombrement

ABSTRACT
Based on the implementation of a mathematical assessment's device at the end of grade 1, several questions about the test and his design are discussed in a didactical framework. The conditions of validity, reliability and relevance, initially defined in the context of testing, are reinterpreted and we explain how we have considered them to design our assessment device. First, in keeping with the conditions of validity and the interpretation of student's answers, we illustrate the interest of considering a didactical framework for interpreting the answers given on a counting task. After, the presentation of the assessment device in his globality, with tasks and assessment course, gives rise to several points of vigilance about his implementation, like the handover of guidelines, the impact of the conditions of course on the student's activity, the training of evaluators for coding answers, etc. Finally, we suggest the new label "robust assessment task " for taking into account these different dimensions.

Key words : assessment, validity, numeration, counting

INTRODUCTION

Les recherches qui ont été présentées dans ces Travaux Dirigés (TD) émanent principalement des actions menées dans le cadre du LEA[1] EvalNumC2[2] et sont dans la continuité de nos travaux respectifs. Depuis sa thèse, Mounier (2010) s'intéresse aux différentes interprétations de la numération parlée et étudie la façon dont se construisent, dans l'enseignement et dans l'apprentissage, les liens entre la numération écrite chiffrée et la numération parlée. Ses travaux l'ont conduit, entre autres, à définir un cadre pour explorer ces liens et les étudier à travers les procédures mises en œuvre dans des tâches de dénombrement (Mounier 2017). Les recherches de Grapin (2015) sont orientées sur l'évaluation des connaissances des élèves et plus spécifiquement sur les conditions de validité d'une évaluation externe standardisée. Dans ce cadre, elle a développé une méthodologie d'analyse du contenu d'un test reposant sur la complémentarité d'approches psychométriques et didactiques et l'a exploité pour étudier le contenu sur le domaine numérique (nombres entiers) d'une évaluation nationale standardisée

1 Les LEA (Lieux d'Etude Associés) sont « définis dans le programme scientifique de l'IFÉ (Institut français de l'éducation) comme des lieux à enjeux d'éducation, rassemblant un questionnement des acteurs, l'implication d'une équipe de recherche, le soutien du pilotage de l'établissement, et la construction conjointe d'un projet dans la durée. » http://ife.ens-lyon.fr/lea

2 Evaluation (Eval) des connaissances numériques (Num) des élèves de Cycle 2 (C2) à Montreuil (Seine-Saint-Denis). Chercheurs : Aline Blanchouin, Nadine Grapin, Eric Mounier et Nathalie Sayac. Correspondants IFE : Elodie Foussard, Zola Zemrani

passée en fin d'école. L'étude des apprentissages des élèves au cycle 2 sur le nombre apparaît donc comme un objet à la croisée de leurs recherches respectives.

En menant nos travaux dans le cadre du LEA EvalNumC2, nous poursuivons trois objectifs :

- concevoir des outils d'évaluation valides pour les chercheurs et pour les enseignants,
- dresser un état des lieux des connaissances des élèves de cycle 2, élèves âgés de 6 à 9 ans, et plus spécifiquement de fin de CP (Cours Préparatoire, 1re année d'enseignement obligatoire, élèves âgés de 6-7 ans),
- et étudier les pratiques des enseignants à partir de traces de leur activité en classe (vidéo, cahier de bord) et des productions de leurs élèves aux évaluations.

Ce TD permet d'aborder ces différents axes tout en situant l'évaluation tour à tour comme un objet au service de l'institution (par le biais des évaluations externes), des chercheurs, de l'enseignant et de l'élève. Durant la première séance de ce TD, qui en compte trois, nous nous intéressons au contenu d'une évaluation externe afin d'introduire des questions relatives à sa « qualité ». Dans le champ de l'évaluation, une évaluation se doit d'être pertinente, valide et fiable (De Ketele & Gerard, 2005). Ces critères peuvent-ils être définis dans un cadre didactique ? De quelle façon ? Dans quel but ?

Dans la deuxième séance, nous montrons l'intérêt d'un cadre didactique pour interpréter les réponses des élèves à une tâche spécifique sur le nombre, le dénombrement de collections non manipulables. Nous travaillons à partir de productions écrites d'élèves mais aussi de vidéos afin de pouvoir qualifier de façon plus précise les procédures. De premiers résultats émanant de notre recherche en cours sont présentés et illustrent l'intérêt de ce cadre.

La dernière séance est consacrée à la méthodologie mise en œuvre dans le cadre du LEA pour étudier les connaissances des élèves ; nous montrons aux participants les différentes évaluations que nous avons proposées aux élèves (tâches données collectivement avec ardoise et papier-crayon ainsi qu'individuellement avec matériel et oralement avec le chercheur) et la façon dont nous pouvons mettre en relation les exercices afin de rechercher des cohérences dans les réponses des élèves et de déterminer la validité (ou non) de certaines des questions posées. Les échanges entre les participants au TD ont été enregistrés et certains de leurs propos sont repris au fil de notre texte.

ETUDE DE LA VALIDITE D'EVALUATIONS EXTERNES : UN POINT DE VUE DIDACTIQUE

Nous avons choisi d'introduire les questions relatives à la « qualité » d'une évaluation à partir d'une évaluation externe, construite par une circonscription pour des élèves de fin de CP (Annexe 1). Les participants ont été invités dans un premier temps à déterminer dans quelle mesure l'évaluation conçue dans ce cadre permettait de renseigner les connaissances des élèves en fin de CP.

1. Constats initiaux sur la « qualité » d'une évaluation

Les premières remarques des participants ont porté sur la conception même de cette évaluation et sur son utilisation : les enseignants ont-ils été obligés de la mettre en place dans leur classe ? Par qui a-t-elle été conçue ? Dans quel but ? Comme le décrivent Laveault & Grégoire (2014), l'enjeu de l'évaluation et de ses objectifs déterminent en premier son contenu et par conséquent sa passation.

Deux pistes ont été proposées par les participants pour étudier le contenu de l'évaluation :

- s'appuyer sur les contenus des programmes de CP, ce qui rejoint l'enjeu éventuellement institutionnel que peut revêtir cette évaluation,

- s'intéresser aux aspects du nombre abordés ou non dans ce test, avec une entrée didactique pour les définir.

Ces deux propositions correspondent à deux logiques de conception des évaluations tenant compte de leurs enjeux, mais aussi du référent choisi. Ainsi, l'évaluation étudiée ici étant conçue dans un cadre institutionnel, par des enseignants ou des conseillers pédagogiques, elle vise à évaluer les connaissances des élèves relativement aux objectifs des programmes ; il en est de même à une échelle nationale pour les bilans CEDRE (Cycle des évaluations disciplinaires réalisées sur échantillon) conçus par la DEPP (Direction de l'évaluation, de la prospective et de la performance) en fin d'école élémentaire ou en fin de collège (Dalibard & Pastor 2015, Dalibard & Arzoumanian 2015). En revanche, pour déterminer les connaissances numériques des élèves selon les différents aspects du nombre, il est nécessaire de définir au préalable un modèle de référence (Laveault & Grégoire 2014, p. 165) et de concevoir le contenu du test à partir de ce dernier.

Le contenu de l'évaluation et l'interprétation qui sera faite des réponses des élèves dépend donc du référent choisi par l'évaluateur, comme l'explique Blanc (2017, p. 13), citant Lesnes (1984, p.132) :

> « Evaluer c'est mettre en relation, de façon explicite ou implicite, un référé (ce qui est constaté ou appréhendé de façon immédiate, ce qui fait l'objet d'une investigation systématique ou d'une mesure) avec un référent (ce qui joue le rôle de norme, ce qui doit être, ce qui est le modèle, l'objectif poursuivi, etc.) »

Le référent peut ainsi être déterminé par les programmes ou dans un cadre didactique, par le champ conceptuel du nombre (Vergnaud 1990, Mounier 2010) ou par une organisation mathématique de référence du domaine des nombres entiers (Bosch & Gascon 2005, Grapin 2015) ; quelle que soit la définition du référent, si l'on souhaite faire un état des lieux des connaissances des élèves, l'ensemble des tâches du test se doit de recouvrir le domaine tel qu'il est défini par le référent (Rey & Feyfant 2014, Grugeon-Allys & Grapin 2015) ou sinon, d'en être un échantillon représentatif (Chevallard 2007). Par exemple, dans l'évaluation proposée, un des participants a souligné qu'il manquait une tâche de comparaison de nombres, alors qu'inversement, dans l'exercice 2, des tâches similaires, qui nécessitaient une même procédure étaient proposées à plusieurs reprises.

L'équilibre des tâches proposées en termes de complexité est un autre axe de réflexion pour concevoir le contenu d'un test. Afin de déterminer la disponibilité des connaissances des élèves, l'évaluateur peut proposer des tâches plus ou moins simples. L'équilibre à trouver entra tâches simples et tâches plus complexes est fortement lié aux enjeux du test et, en didactique des mathématiques, plusieurs outils permettent de définir *a priori* la complexité d'une tâche, comme par exemple : les niveaux de mise en fonctionnement des connaissances et les adaptations de connaissances que nécessite une tâche pour être résolue (Robert 2005), les niveaux de convocation d'un type de tâche (Castela 2008) ou encore les facteurs de complexité et de compétences (Sayac & Grapin 2015).

Si ces premiers critères sur le choix des tâches sont liés à un référent et à l'objectif de l'évaluation, ils ne peuvent être dissociés d'une analyse *a priori* de chacune des tâches. Menée usuellement dans le cas de situations d'apprentissage, il s'agit ici de penser cette analyse avec l'objectif d'évaluation assigné à la tâche et à l'interprétation potentielle de la réponse, ou plus globalement, de la production de l'élève.

Ainsi, un des participants s'est interrogé pour chacun des exercices, s'il était « en mesure d'interpréter la réponse de l'élève à partir d'une erreur classique ». Il illustre son propos avec la question 2 de l'exercice 3 dans lequel il est demandé de ranger du plus petit au plus grand 72 – 53 – 35 – 27 : « si je les classe selon les chiffres des unités, je vais obtenir le même résultat que si je les avais classés par ordre décroissant, et si j'ai ça, je ne sais pas si l'élève a mal interprété croissant-décroissant ou s'il a raisonné sur le chiffre des unités ». Une telle analyse montre ainsi l'importance du choix des valeurs de variables didactiques, rendant

invalides certaines procédures et permettant une meilleure interprétation des réponses des élèves. De façon plus générale, il s'agit d'interroger ce qu'une tâche permet ou non de révéler des connaissances de l'élève à partir des traces recueillies sur sa production. Par exemple, dans l'exemple précédent, il peut être difficile d'interpréter une réponse fausse ou de déterminer la procédure employée par l'élève pour répondre. Ces questions relatives à l'interprétation des réponses de l'élève ne peuvent être traitées indépendamment de la conception du test ; nous les abordons à nouveau lors de la 2e séance du TD.

La compréhension de la tâche a été ensuite interrogée, tout d'abord en lien avec la formulation de certains énoncés. Certains énoncés peuvent ne pas être explicites pour les élèves : par exemple, le terme d'« étiquette » dans l'exercice 6 est-il judicieusement choisi pour que l'élève comprenne ce qu'il doit faire ? Est-ce que c'est un terme qui est employé régulièrement avec les élèves ? Ensuite, pour comprendre la tâche, il est nécessaire que l'élève ait connaissance de la consigne et de l'énoncé de la tâche : si la lecture de l'énoncé est à la charge de l'élève, les élèves présentant des difficultés en lecture risquent de ne pas réussir l'exercice, non pas parce qu'ils n'ont pas les connaissances mathématiques mais parce qu'ils n'ont pas compris ce qu'ils avaient à faire. Dans l'évaluation proposée, les conditions de passation précisaient à l'enseignant ce qu'il devait lire et expliciter aux élèves pour chacun des exercices (annexe 1).

Nous avons conclu ce premier temps d'échanges avec une liste non exhaustive d'indicateurs didactiques qui figurent dans l'analyse *a priori* des tâches et qui permettent de contrôler la variété des tâches proposées afin d'assurer la couverture du domaine, sans redondance inutile. Parmi ces indicateurs, nous avons évoqué lors de cette première séance : le type de représentation du nombre en jeu dans la tâche (écriture chiffrée, écriture en lettres, décompositions en unités de numération, collections organisées ou non, etc.), la taille des nombres, l'organisation des collections, l'ancienneté de l'objet de savoir, le format de la question, le niveau de langue utilisé. Nous retrouvons en partie les types de critères didactiques exploités par Grugeon & Grapin (2015) pour l'étude didactique du contenu d'une évaluation ; les procédures de résolution des tâches n'ont pas été développées lors de cette séance, mais le seront lors des suivantes.

2. *Pertinence, validité et fiabilité comme critères de qualité d'une évaluation*

Dans les ouvrages généraux sur l'évaluation, et notamment en théorie des tests, une évaluation, pour être de qualité, se doit d'être pertinente, valide et fiable (Laveault & Grégoire 2014). Nous avons choisi de travailler à partir d'un extrait de l'article de De Ketele & Gerard (2005) dans lequel ces trois critères (pertinence, validité, fiabilité) sont définis de la façon suivante :

> « La pertinence est le caractère plus ou moins approprié de l'épreuve, selon qu'elle s'inscrit dans la ligne des objectifs visés (DE KETELE et alii, 1989). C'est *son degré de « compatibilité » avec les autres éléments du système auquel elle appartient* (RAYNAL & RIEUNIER, 1997, 2003).
> La validité est le degré d'adéquation entre ce que l'on déclare faire (évaluer telle ou telle dimension) et ce que l'on fait réellement, *entre ce que l'outil mesure et ce qu'il prétend mesurer* (LAVEAULT & GRÉGOIRE, 1997, 2002).
> La fiabilité est le degré de confiance que l'on peut accorder aux résultats observés : seront-ils les mêmes si on recueille l'information à un autre moment, avec un autre outil, par une autre personne, etc. ? Elle nous renseigne sur le degré de relation qui existe entre la note obtenue et la note vraie (CARDINET & TOURNEUR, 1985 ; LAVEAULT & GRÉGOIRE, 1997, 2002). Il ne faut cependant pas perdre de vue que la note vraie est une abstraction, un point de convergence souhaité indépendant des évaluateurs et des circonstances. » (De Ketele & Gerard 2005)

Si ces dimensions donnent des « conditions nécessaires » (ibid.) garantissant la qualité d'une évaluation, elles ne sont pas suffisantes si elles sont prises isolément. Comme nous le constatons à la lecture de l'extrait, ces conditions sont plutôt discutées (dans le cas d'une

approche par compétence) à partir de méthodes de validation spécifiques à la théorie des tests : recours à la mesure dans la théorie classique des scores ou à des méthodes de validation *a priori* (par des juges) ou *a posteriori* (avec des indicateurs statistiques). Nous pensons cependant qu'il est possible de les interpréter dans un cadre didactique et d'en fournir des indicateurs autres que statistiques. Les échanges que nous avons eus avec les participants lors de cette première séance ont contribué à faire avancer la réflexion sur ce point ; les paragraphes ci-après en rendent compte et constituent le point de départ de notre réflexion, et non son aboutissement.

Définir les objectifs assignés à l'évaluation (sans mesure, ni interprétation des résultats préalable) est la condition première pour pouvoir étudier la pertinence. Ces objectifs diffèrent bien souvent selon l'évaluateur. Par exemple :

- l'institution, dans le cas d'évaluations externes, cherche à déterminer si les connaissances des élèves sont celles attendues par les programmes,
- l'enseignant souhaite plutôt évaluer les connaissances acquises par ses élèves au regard des exigences propres de son enseignement,
- le chercheur enfin vise à déterminer les connaissances acquises par les élèves en fonction de la question de recherche posée.

Par conséquent, une épreuve peut se révéler pertinente pour un type d'évaluateur, mais pas pour un autre.

Parce qu'elle s'intéresse en particulier au contenu du test et à l'interprétation des résultats, la validité est un critère davantage étudié en didactique des mathématiques. Grugeon-Allys & Grapin (2015) ont ainsi défini une liste d'indicateurs et une méthodologie permettant d'analyser le contenu d'une évaluation externe. S'appuyant sur l'analyse *a priori* des tâches, elle détermine des preuves de validité du côté du savoir (approche épistémo-didactique) et du côté de l'élève (approche psycho-didactique) à un niveau local (tâche par tâche) et global, sur l'ensemble du test.

La fiabilité soulève quant à elle la question de la reproductibilité de la tâche d'évaluation. La formulation de la définition proposée par l'article se réfère aux épreuves externes et à une mesure donnée (note). Nous l'avons interprétée différemment, sans référence à la mesure, de la façon suivante : toutes choses étant égales par ailleurs, obtient-on les mêmes réponses des élèves sur une tâche « identique » ? Plusieurs éléments permettent de contrôler le fait que la tâche cognitive soit effectivement identique dans l'activité cognitive qu'elle demande *a priori* :

- des critères épistémo-didactiques, du côté du savoir en jeu,
- des critères psycho-didactiques (Vantourout & Goasdoué, 2014), afférents à l'activité de réponse de l'élève,
- les conditions de passation : une évaluation conduite par un enseignant dans sa classe n'induit pas les mêmes règles implicites du contrat didactique que lorsqu'elle est menée par un chercheur inconnu des élèves,
- le choix du support de l'évaluation (papier-crayon, oral, informatique),
- etc.

Au-delà des critères psychométriques définis pour la qualité des évaluations externes, nous avons ainsi pointé que des conditions didactiques sur le contenu de l'évaluation se devaient d'être respectées tout en tenant compte de différentes contraintes (de passation notamment, nous y reviendrons). Ce constat, établi depuis bien longtemps dans les travaux sur l'évaluation, se formule ainsi dans un cadre didactique. Ce sont les apports de ce champ de recherche qui permettront de définir des modèles théoriques, des indicateurs et des contraintes en lien avec le savoir mathématique évalué, l'élève, l'enseignant et l'institution. Nous illustrons de manière concrète ces apports dans les deux séances de TD suivantes.

ETUDE DE CAS : UNE TACHE EXTRAITE D'UNE EVALUATION MENEE PAR UN CHERCHEUR

Dans la 2e séance nous souhaitons montrer le lien entre les tâches proposées, l'objectif de l'évaluation et le cadre didactique pour interpréter les résultats. Pour l'aborder, nous nous sommes centrés sur une seule tâche, donnée dans le cadre de nos évaluations CP du LEA EvalNumC2, celle emblématique du dénombrement d'une collection non organisée et non manipulable (voir Annexe 2). L'ensemble des tâches données lors de cette évaluation de fin de CP est étudié dans la 3e séance du TD.

Les participants ont disposé de la feuille comportant la collection de 52 carrés fournie aux élèves et de la consigne « Ecris le nombre de ronds qu'il y a sur la feuille », mais aussi de quelques productions d'élèves (Annexe 2). Ils ont eu à déterminer les objectifs d'évaluation qui pourraient être assignés à cette tâche et, au regard de ces objectifs, à interpréter les productions fournies (Tableau 1).

	Résultat	Traces sur la copie
Elève 1	52	Chaque carré est marqué
Elève 2	52	50 carrés sont marqués + groupements par 10
Elève 3	52	Groupements par 10
Elève 4	52	Des carrés sont marqués + groupements par 10
Elève 5	51	Tous les carrés sauf un sont numérotés
Elève 6	55	Tous les carrés sont marqués ; une marque distinctive selon des groupements (de taille variable)
Elève 7	91	Tous les carrés sauf un sont marqués ; groupements de 5 ou 6 ou 10

Tableau 1. *– Description des productions des élèves concernant la tâche analysée*

Les participants ont indiqué que l'évaluation pourrait avoir pour objectif de déterminer si l'élève est capable de dénombrer une collection et quelle procédure il utilise. Deux procédures semblent en effet à l'œuvre concernant les sept élèves pris en exemple : le comptage un à un et un comptage avec des dizaines. Le comptage un par un demande une connaissance de la comptine numérique, une synchronisation entre chaque objet pris en compte et cette comptine mais aussi le fait de savoir que le dernier nombre prononcé correspond au cardinal de la collection (Briand 1999). Des connaissances liées à l'énumération (Briand 1999, Margolinas et Wozniak 2016) sont en jeu pour la réussite de la tâche proposée : tous les éléments doivent être passés en revue une et une seule fois. Un oubli ne signifie pas que l'élève ne sait pas qu'un nombre désigne une quantité. Ainsi ne prendre en compte que la réussite ou l'échec n'est pas suffisant pour évaluer les connaissances que pourraient témoigner les élèves concernant le dénombrement.

En fin de CP, une autre procédure est disponible, en lien avec la numération décimale, celle avec des dizaines. Plusieurs questions se posent cependant. Il peut y avoir un effet de contrat didactique qui incite l'élève à faire des dizaines alors que dans un autre contexte il n'en ferait pas. Un des résultats obtenus lors de recherches précédentes (Mounier 2016) a montré que des tâches similaires pouvaient être traitées différemment selon que l'élève se trouve seul avec le chercheur ou bien en classe avec l'enseignant. L'hypothèse avancée est celle de l'effet d'un contrat didactique « dizaine » établi en classe, mais aussi le fait que pour

un élève bien rôdé au comptage un à un (procédure largement travaillée en amont du CP), la tâche peut se résoudre efficacement sans faire de dizaines. Ainsi ce n'est pas parce que l'élève ne montre pas certaines connaissances dans une tâche donnée qu'il ne les a pas. On retrouve ici des questions afférentes aux conditions de passation et à l'analyse didactique de la tâche qui ont été abordées dans la séance 1 du TD. Elles seront reprises dans la séance 3 qui aborde l'importance de la diversité des tâches proposées dans une évaluation et celle des différentes modalités de passation. Mais ici une autre question émerge. De quelles connaissances l'emploi de la dizaine témoigne-t-il ? La question se pose aussi pour l'emploi de groupements non forcément effectués de manière maximale, et sans qu'ils soient tous de dix ou tous équipotents, comme on le voit chez l'élève 7 (Tableau 1 et Annexe 2).

Cette tâche de dénombrement des 52 carrés s'inscrivait en fait dans une évaluation plus large dont l'objectif était celui de mieux connaître les connaissances des élèves sur la numération décimale. Que peuvent alors nous dire les productions des élèves ? Cette question demande un apport théorique sur le lien entre procédure et numération. La thèse sur la numération au CP (Mounier 2010) a montré que des élèves pouvaient solliciter la dizaine dans les tâches de dénombrement sans pour autant comprendre le lien avec l'écriture chiffrée : c'est le cas par exemple lorsque la procédure mobilisée est le comptage oral dix, vingt, trente, quarante, cinquante, cinquante-et-un, cinquante-deux, puis l'écriture de « 52 » à partir du nom (oral) du nombre « cinquante-deux ». Un cadre d'analyse plus fin des procédures a alors été élaboré (Mounier 2017) pour relier les connaissances en jeu dans une tâche de dénombrement avec les propriétés de la numération orale et celles de la numération écrite chiffrée (le lien procédure/connaissance est celui de la théorie des champs conceptuels de Vergnaud, 1990). Il nécessite alors d'observer les procédures de manière plus fine, notamment si l'élève passe nécessairement par le nom du nombre pour obtenir l'écriture chiffrée (comme la procédure évoquée si avant) ou non (dans ce cas il peut compter les dizaines, un, deux, trois, quatre, cinq et les unités restantes, un, deux, et accoler les deux chiffres dans le bon ordre « 52 »).

Cet exemple montre le rôle des apports proprement didactiques nécessaires lorsqu'il s'agit d'évaluer des connaissances sur le nombre et pas uniquement de rendre compte de la diversité des procédures.

Au-delà des objectifs différents selon les chercheurs, l'enseignant en classe et l'institution, trois éléments ont finalement émergé concernant l'analyse didactique nécessaire dans une évaluation :

- l'intérêt d'aller au-delà des résultats (vrai/faux) et de considérer les procédures. Est-ce que les productions recueillies suffisent ? Que doit-on observer ? Quels moyens pour recueillir ces productions ?
- la nécessité de donner un ensemble de tâches dans une évaluation : lesquelles ?
- l'adoption d'un cadre théorique pour interpréter les procédures utilisées en termes de connaissances afférente à une notion mathématique.

Concernant les moyens pour recueillir non seulement les productions mais aussi les procédures, se posent des questions pratiques liées au temps nécessaire et au nombre de chercheurs à mobiliser pour réaliser les évaluations, d'autant que le deuxième point met l'accent sur la diversité des tâches à proposer. Ces aspects sont développés dans la troisième séance du TD mais signalons ici que ces questions se posent de manière différente dans le cadre d'évaluation en classe pour l'enseignant ou d'évaluations institutionnelles. Les enseignants peuvent prendre des informations sur les procédures à différents moments, par exemple en observant l'activité des élèves durant la réalisation de tâches d'apprentissage ou de réinvestissement des connaissances. Dans le cas des évaluations externes, l'utilisation de cahiers tournants permet de proposer un nombre de tâches important, chaque élève ne passant pas l'ensemble des tâches proposées, mais seulement un échantillon. Le modèle de réponse à

l'item permet ensuite d'attribuer une probabilité de réussite à chaque tâche pour tous les élèves.

Finalement, il apparait que les objectifs d'évaluation peuvent plus volontiers se décliner en termes de procédures (savoir-faire) ou de savoirs (dans le sens d'une référence à un texte de savoir, comme les tables de multiplication ou les propriétés du carré), qu'en termes de connaissances sur une notion comme nous l'avons montré dans l'exemple sur la numération (les numérations devrions-nous dire). L'interprétation des procédures demande en effet de mobiliser des outils didactiques spécifiques. La diversité des éléments à recueillir dépend donc non seulement des objectifs et donc d'une référence (voir la séance 1), mais aussi du cadre d'analyse que l'on adopte et des résultats de recherches déjà obtenus. Ces trois aspects, objectif/référence, résultats de recherches antérieures sur les connaissances à évaluer et cadre d'analyse ne peuvent être dissociés.

DISPOSITIF D'EVALUATION EN FIN DE CP : ELEMENTS METHODOLOGIQUES

Le dispositif d'évaluation que nous avons présenté dans cette troisième séance est celui mis en œuvre dans le cadre du LEA EvalNumC2 fin juin 2017. L'objectif de ce dispositif est de déterminer les connaissances sur le nombre des élèves de fin de CP. Au-delà de dresser un état des lieux de ces connaissances, nous souhaitons utiliser les résultats à deux fins : déterminer un ensemble de tâches d'évaluation pertinentes, valides et fiables et interroger les pratiques des enseignants au filtre des productions de leurs élèves.

En nous référant au cadre théorique développé par Mounier (2017) et grâce aux tâches choisies, il sera ainsi possible de décrire des profils d'élèves. Le terme de profil se réfère à celui utilisé par Grugeon (1997) et Chenevotot-Quentin, Grugeon & Delozanne (2009) décrivant le rapport personnel d'un élève à un domaine mathématique donné et étant établi à partir de cohérences de fonctionnement repérées sur l'ensemble d'un test. Les géographies des classes (Chenevotot-Quentin & al., 2009) selon ces profils pourront être comparées.

1. *Dispositif d'évaluation mis en œuvre*

Les tâches d'évaluation ont été conçues pour que l'ensemble du test réponde aux critères de fiabilité, de pertinence et de validité tels que nous les avons conçus dans un cadre didactique. Nous décrivons dans ce paragraphe le dispositif d'évaluation dans sa globalité avec les modalités de passation des différentes tâches.

Sur le domaine des nombres entiers, nous avons choisi comme référence le champ conceptuel relatif aux nombres (Vergnaud 1990), repris par Mounier (2010), et avons exploité l'organisation mathématique de référence définie sur ce même domaine mathématique par Grapin (2015) dans le cadre de la Théorie Anthropologique du Didactique (Chevallard 1999). En définissant notre référence selon ces deux approches didactiques, il nous parait possible de définir les apprentissages d'un élève, en tant que sujet cognitif, au regard des différentes institutions (les programmes et la classe). Chacune des tâches est donc définie comme appartenant à une classe de problèmes donnée et comme représentant un type de tâche donnée, selon l'approche choisie. La couverture du domaine est assurée au regard du référent ; le tableau présenté en annexe 3 explicite la répartition des tâches selon qu'elles relèvent de la résolution de problèmes additifs, des numérations ou du calcul (Grapin 2015). Les procédures de résolution attendues (ou les techniques et technologies, si on se place dans le cadre de la TAD) sont définies selon les programmes en vigueur.

Tâches relevant de la résolution de problèmes

En ce qui concerne la résolution de problèmes arithmétiques, nous avons retenu douze problèmes additifs, répartis en problèmes de transformation d'état (avec recherche de l'état initial, final et de la transformation) et de composition de mesures (avec recherche d'une des parties ou du tout). Tous les problèmes ont un même contexte, à savoir, une boîte et des jetons (que l'on ajoute, retire ou que l'on assemble) et l'énoncé est proposé sous forme écrite, accompagné systématiquement d'un schéma comme le montre l'exemple ci-dessous (Figure 1) :

Problème n°1

On cherche le nombre de jetons dans la boîte à la fin.

Au début, il y a 5 jetons dans la boîte.
Je mets encore 3 jetons dans la boite.

Quel est le nombre de jetons dans la boîte à la fin ?
Écris ta réponse sur les pointillés.

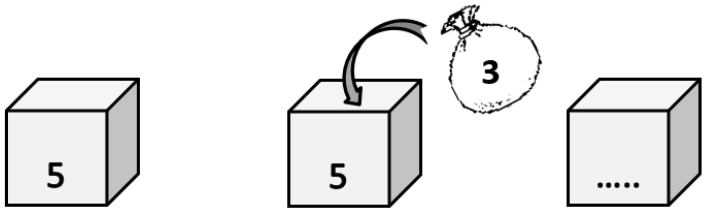


***Figure 1**. – Exemple d'énoncé de problème additif proposé dans l'évaluation*

Lors de la passation, chaque énoncé est lu par le chercheur et l'action schématisée dans l'énoncé est mimé avec des gestes et avec des objets réels (une boîte et des jetons). Nous avons proposé deux exercices similaires pour chaque classe de problèmes, l'un avec des nombres inférieurs à 10 et l'autre avec des nombres supérieurs à 10. Ce choix nous permet ainsi, en fonction de la réussite mais aussi des traces écrites laissées par les élèves, de distinguer les élèves qui comptent de ceux qui calculent, mais aussi de voir ceux pour qui la taille des nombres ferait obstacle à la résolution du problème.

Tâches relevant des numérations et du calcul

Les tâches de numération s'inscrivent dans différentes classes de problèmes du champ conceptuel. Elles mettent en jeu diverses représentations du nombre : écriture chiffrée, nom du nombre à l'oral, écriture en unités de numération (c'est-à-dire une formulation en termes de dizaines et unités), représentation d'une collection d'objets organisée ou non, d'une collection de doigts de la main ou d'une constellation de dé. En complément de tâches de conversion d'une représentation à l'autre, ont été proposées des tâches visant à évaluer la maitrise de la comparaison (de collections, de nombres en écriture chiffrée) et du dénombrement de collections organisées ou non en dizaines – unités. Les nombres en jeu dans les exercices sont de tailles différentes mais restent inférieurs à 100, et certaines tâches, notamment de dénombrement, sont contextualisées.

En ce qui concerne le calcul, nous proposons des additions posées de deux ou trois nombres (avec et sans retenue) et des calculs à effectuer mentalement. Ces derniers portent sur la connaissance de répertoires additifs, de compléments à 10 et d'ajout d'un nombre entier de dizaines (ajouter 20, ajouter 40…).

Passation de l'évaluation

Le déroulement de l'évaluation dans sa globalité est réparti sur quatre demi-journées. Des passations collectives en classe, sur ardoise et sur papier-crayon sont menées par un chercheur pendant trois séances de 45 minutes environ chacune (Tableau 2).

	Demi-journée 1	Demi-journée 2	Demi-journée 3	Demi-journée 4
	Passation collective en classe (45 minutes)			Passation individuelle avec le chercheur (15 min par élève)
1er temps	Calcul mental sur ardoise			Dictée et écriture de nombres
2e temps	Evaluation sur papier avec consignes lues oralement par le chercheur. Trois cahiers contenant dans l'ordre : - 4 problèmes additifs ; - des tâches de numération et de calcul.			Tâches de dénombrement contextualisées

Tableau 2. *– Dispositif d'évaluation*

Concernant les passations menées individuellement par un chercheur, les procédures mises en jeu par l'élève sont ainsi directement observées et peuvent faire l'objet d'explicitations selon les demandes du chercheur. Les modalités de passation de chacune des tâches sont fortement liées à la validité du contenu. Par exemple, il est important que certaines tâches soient réalisées en un temps limité pour le calcul mental, mais aussi pour certaines des tâches de dénombrement afin de rendre inefficaces les procédures par comptage un à un. Les modalités de passation ainsi que les procédures (ou les techniques) attendues sont donc ainsi définies pour chacune des tâches selon son objectif d'évaluation, ce qui permet de contrôler la pertinence de la tâche d'évaluation choisie.

2. *Echanges autour du dispositif*

Les participants ont eu en main les trois cahiers de l'évaluation passée collectivement avec les réponses d'élèves (les trois cahiers d'un même élève). Sans occulter l'ensemble des tâches proposées, nous nous sommes centrés sur certaines, celles qui peuvent renseigner les connaissances concernant le système de numération écrit chiffré. Il s'agit de regarder comment mettre en relation les réponses d'un même élève à différentes tâches afin de repérer des cohérences de fonctionnement. L'objectif est ici de se mettre à la place du chercheur qui va devoir élaborer l'évaluation, la faire passer et l'interpréter. Nous rendons compte des remarques et interventions diverses et de la façon dont nous avons répondu à ces questions lors de la préparation puis de la passation du test.

Les participants sont tout d'abord revenus sur ***la prise d'information à partir des réponses afin de les interpréter*** en prenant en compte :

- certaines erreurs, attendues et « révélatrices » ; par exemple certaines tâches mettent en jeu des groupements, qui ne sont visiblement pas de dix, mais que les élèves peuvent considérer comme tels,
- des distracteurs, comme la proposition « 4 unités et 2 dizaines » pour l'écriture chiffrée « 42 »,
- certaines traces dans les productions (lesquelles retenir ?),
- plusieurs types de procédures menant à une réponse exacte.

Il a été souligné à nouveau certaines limites à des évaluations dont on ne voit que les résultats ou productions mais dont les procédures ne peuvent être déterminées. Ces

procédures sont néanmoins nécessaires pour investiguer certaines connaissances, comme celles sur la numération écrite chiffrée : compter de dix en dix (dix, vingt, trente, etc.) ne témoigne pas des mêmes connaissances que le comptage des dizaines (un, deux, trois dizaines, etc.).

Concernant l'interprétation des données recueillies, il a été à nouveau souligné la nécessité de prendre en compte l'ensemble des tâches. Nous avons déjà indiqué dans le paragraphe précédent la manière dont nous avons procédé pour assurer ***le recouvrement du domaine*** relatif aux nombres. Précisons ici que nous avons ***recouper/doubler certaines tâches*** données collectivement par des tâches individuelles dans lesquelles nous avons observé les procédures. Ceci nous permet en particulier de voir si un effet de contrat est intervenu mais aussi d'interpréter certaines des réponses ou productions des évaluations collectives. Notre but est de pouvoir faire une sélection de tâches pour élaborer une évaluation qui peut se proposer de manière collective, le recueil des réponses et des productions nous permettant de dresser un état de lieux des connaissances des élèves. L'interprétation des résultats en termes de connaissances des élèves s'effectue en tenant compte d'une cohérence d'ensemble, qui peut s'effectuer grâce au recouvrement du domaine.

Un autre sujet de discussion a concerné ***les contraintes matérielles*** liées aux passations collectives. Le fait de vouloir faire un état des lieux des connaissances des élèves à la fin d'une année scolaire demande de concentrer les tests sur une durée réduite, en fin d'année (fin juin) et ceci pour un nombre important de classes. Le fonctionnement des classes à cette période de l'année, en particulier des classes de jeunes enfants, peut largement impacter les possibilités de faire les tests : fêtes de l'école et fêtes religieuses, assiduité voire absence de certains élèves en fin d'année. ***La nécessaire mobilisation de nombreux évaluateurs***, non tous nécessairement directement impliqués dans la recherche menée, ***pose en outre la question de leur formation***. Un protocole est indispensable pour assurer une uniformité des conditions de passation afin que tous les élèves de toutes les classes aient effectivement répondu aux mêmes tâches. Il s'agit parfois de résister à donner des procédures qui peuvent être demandées par les élèves mais il s'agit aussi de leur permettre de se concentrer. Les évaluateurs doivent avoir par ailleurs les connaissances leur permettant de reconnaître en situation les procédures utilisées afin de les coder immédiatement si cela est nécessaire. Filmer le moment de l'évaluation permet à la fois de s'assurer de l'uniformité des conditions de passation et de la pertinence des codages effectués. Cependant, s'il concerne beaucoup d'élèves, ce dispositif peut s'avérer assez couteux en temps d'analyse lors de passations individuelles, alors qu'il permet difficilement de recueillir les procédures dans le cas de passations collectives : on peut en effet imaginer mettre une caméra au fond de la classe pour obtenir des informations sur les conditions et modalités de passation réelles, mais moins facilement filmer simultanément chaque élève de la classe. Il semble alors pertinent de proposer en amont une véritable formation à l'évaluation, en utilisant par exemple des vidéos avec des passations des tests proposés. Cette formation nous semble devoir aborder non seulement les modalités de passation en situation et les codages - respecter l'esprit plutôt que la lettre, donc s'adapter - mais aussi le nécessaire climat de confiance à instaurer avec les enseignants qui « prêtent » leur classe. Cet aspect « relation humaine », que nous ne développons pas ici, nous semble néanmoins essentiel pour que les chercheurs puissent réaliser des évaluations dans des classes dites ordinaires et pour que les professeurs en puissent en recueillir des enseignements.

En lien avec ce qui précède, nous avons évoqué les ***différents moyens de pouvoir indiquer une tâche*** (ici à de jeunes élèves) sans leur indiquer la procédure. La question est plus sensible lors de passations collectives. Nous avons fait une liste non exhaustive de ces moyens : ce qui est indiqué sur le papier, les gestes, les actions, le discours « méta », le matériel, les représentations utilisées (suffisamment « évocatrices »), l'utilisation du temps

pour bloquer les procédures (il est possible de faire tourner la page comme marqueur de fin de l'activité). Si ces moyens ne sont pas forcément très différents de ceux utilisés usuellement, il y a cependant un contrat différent de celui utilisé pour l'apprentissage avec l'enseignant, un contrat en évaluation (Sayac, à paraître) auquel l'élève est plus ou moins sensible ou sensibilisé selon la classe. Prendre conscience de cette rupture de contrat nous semble important pour mener une évaluation puis interpréter les résultats.

Dans la continuité de la discussion sur les conditions de passation, nous avons abordé la question des conditions matérielles. Nous faisons l'hypothèse de l'intention de l'élève à produire une réponse raisonnée, hypothèse qui nous est indispensable pour interpréter les résultats des évaluations en termes de connaissances mathématiques. Un « contrôle » de la cohérence des réponses produites peut se faire *a posteriori* via les apports didactiques sur les notions en jeu. Par exemple on peut s'attendre à ce que les élèves qui réussissent à telle tâche réussissent aussi à telle autre. Si tel n'est pas le cas, il doit être possible de pouvoir en chercher un éclairage d'ordre cognitif. Il s'agit donc d'installer les conditions de production des réponses afin que l'élève puisse effectivement produire cette réponse « raisonnée ». Ceci pose la question de la prise en compte de la diversité des classes confrontées à une même évaluation, formatée pour une classe lambda. ***Certains facteurs matériels peuvent directement influencer les procédures et réponses des élèves***, notamment quand il s'agit d'une passation collective. Certains de ces facteurs relèvent des conditions de passation, d'autres des tâches proposées. Ainsi, la disposition des élèves peut permettre ou non le copiage. La présence et le bruit, voire la rapidité des autres influe aussi sur l'activité, notamment chez de jeunes élèves. Il se peut aussi que certains des contextes des tâches d'évaluation soient plus familiers que d'autres, que tel type de matériel ait été fréquenté ou non. Certaines de ces contingences sont plus sensibles en fonction de l'âge des élèves. Dans l'évaluation que nous avons proposée en juin 2017, nous avons essayé de limiter les disparités entre passations. Les élèves sont isolés le plus possible afin qu'ils puissent réaliser la tâche seul, le matériel utilisé est neutre, c'est-à-dire qu'il n'est pas lié à un contexte précis (par exemple des cubes ou des ronds dessinés, et non pas des billes ou des figurines ou des confettis), les problèmes sont décontextualisés (des boites et des cubes). Notons que ce dernier point interroge cependant la nature des tâches à proposer et des modalités de passation lorsqu'il s'agit d'évaluer la mobilisation des connaissances en contexte.

L'importance du ***statut des connaissances évaluées*** a été relevée c'est-à-dire le fait qu'elles soient anciennes ou récentes. Il se peut même qu'elles n'aient pas été encore objet d'enseignement au moment où les élèves ont été évalués. Il est possible d'avoir une idée de ce statut des connaissances car il existe un ordre dans les apprentissages qui respecte l'avancée didactique, la programmation étant en outre à la fois indiquée dans les programmes et les manuels scolaires. Par exemple, donner du sens à l'algorithme de l'addition posée au CP nécessite auparavant de donner du sens aux écritures chiffrées, ce qui rejette d'ailleurs souvent l'enseignement de l'algorithme en fin d'année (Mounier & Priolet 2016). Il est ainsi possible de prévoir qu'au moment de l'évaluation certaines tâches, comme les additions posées en colonnes au CP, ne puissent être traitées comme l'institution l'attend. Notons que, dans le cas des évaluations faites en juin 2017, nous avons pu connaitre précisément la programmation adoptée par le professeur pour 6 des 8 classes dont les élèves ont été évalués.

CONCLUSION

Nous avons voulu montrer lors de ce TD les multiples questions auxquelles nous avons été confrontés en tant que chercheurs en didactique en mathématiques pour concevoir et mettre en œuvre une évaluation des connaissances des élèves, à un instant donné de leur scolarité, et les éléments théoriques que nous avons convoqués pour y répondre. En cherchant à réinterpréter

les notions de pertinence, validité et fiabilité et en montrant l'intérêt de choisir un cadre didactique pour l'interprétation des réponses des élèves, nous souhaitons que les spécificités liées à la mise en œuvre d'évaluations dans les classes soient aussi considérées par les chercheurs, notamment quand ils exploitent les évaluations dans leurs propres recherches.

Une évaluation ne peut se concevoir indépendamment de ses enjeux : ainsi, celle que proposera un chercheur en lien avec sa problématique ne poursuit les mêmes objectifs que celle de l'enseignant pour sa classe ou qu'une évaluation externe visant à piloter le système éducatif. Les conditions de validité, pertinence et fiabilité telles que nous les avons décrites se doivent donc d'être considérées différemment pour une évaluation interne conçue par l'enseignant pour ses élèves en tenant compte à la fois du contrat didactique en évaluation existant dans la classe et du jugement professionnel en évaluation de l'enseignant (Sayac, à paraître). Pour les évaluations externes menées par l'institution (évaluations nationales et internationales) les acquis des élèves sont déterminés à partir de la réussite ou de l'échec à certaines tâches, sans prise en compte systématique des procédures ni des connaissances en jeu sous-jacentes dans la résolution. Pourtant, par exemple, les réussites et les échecs aux opérations posées en colonnes pourraient être aussi reliés à des réponses données dans d'autres tâches, comme celles mettant en jeu le principe décimal de la numération écrite chiffrée. Ce sont de tels recoupements qui permettent d'évaluer la maîtrise de savoirs mathématiques sur un domaine donné mais aussi de repérer des manques expliquant les erreurs des élèves.

Pour inférer les connaissances des élèves nous avons besoin de connaître les procédures qu'ils utilisent dans les tâches qu'on leur propose (Vergnaud 1990). Pour les interpréter, nous avons montré alors la nécessité d'apports théoriques sur les notions mathématiques en jeu. C'est une approche proprement didactique. Si pour le chercheur, il y a un gain qualitatif important à considérer les procédures des élèves plutôt que leurs seules réponses, les passations individuelles avec un observateur semblent alors plus appropriées. Pour autant, lorsqu'il s'agit d'évaluer un grand nombre d'élèves, le coût de ces passations est important, en temps et en analyse, donc en moyens humains. Un de nos objectifs est de déterminer ce qui pourrait être révélateur des connaissances des élèves lors de passations collectives sans observation directe des procédures. C'est pourquoi nous avons doublé certaines tâches, fait des recoupements, afin de sélectionner celles qui permettent d'inférer les procédures à travers les réponses ou productions. En s'assurant en outre de la couverture du domaine relatif aux nombres et en prenant en compte le déroulement des évaluations, nous cherchons ainsi à déterminer des tâches qui résistent à une certaine variabilité des conditions de passation et qui permettent d'inférer les connaissances des élèves. Nous proposons de qualifier ces tâches de ***tâches d'évaluation robustes***. Il s'agit de faire un parallèle avec les notions de pertinence, validité et fiabilité (De Ketele & Gerard, 2005) mais aussi avec les tâches robustes pour l'enseignement en classe ordinaire, telles qu'elles sont définies par Robert (2007) :

> « Une tâche robuste donne lieu à des activités possibles, voire a minima, peu différentes des activités analysées a priori, quelles que soient les interventions de l'enseignant […]. La robustesse correspond à un potentiel de « non variabilité » des activités attachées à un énoncé. »

Dans notre approche, l'étude de la robustesse de tâches d'évaluation conjugue ainsi la didactique des mathématiques et les recherches en évaluation.

Enfin, si nous avons pu montrer certains résultats globaux lors de la deuxième séance du TD, le traitement des données recueillies est encore en cours afin de pouvoir dresser des profils d'élèves à partir des tâches proposées. Au-delà de l'état des lieux des connaissances des élèves en fin de CP que nous souhaitons dresser et des réflexions méthodologiques sur l'évaluation pour le chercheur en didactique des mathématiques que nous avons menées, nous pensons aussi exploiter les résultats de cette recherche en formation d'enseignants dans le cadre du LEA. Nous espérons par ailleurs vivement que nos réflexions et les outils que nous avons élaborés participent plus globalement à la conception d'évaluations dans le cadre de

dispositifs de recherche ou d'évaluations externes, notamment celles prévues en France en fin de CP.

REFERENCES BIBLIOGRAPHIQUES

BLANC, C. (2017) Étude de la validité de l'évaluation interne conduite en Cours Préparatoire, en Français et en Mathématiques. Thèse de doctorat. Université Paris Descartes. Paris.

BOSCH, M., GASCON, J. (2005) La praxéologie comme unité d'analyse des processus didactiques. In Mercier A. & Margolinas C. (Eds) *Balises pour la didactique des mathématiques* (p. 197-122). Grenoble : La Pensée Sauvage.

BRIAND J. (1999) Contribution à la réorganisation des savoirs pré-numériques et numériques. Etude et réalisation d'une situation d'enseignement de l'énumération dans le domaine pré-numérique. *Recherches en Didactique des Mathématiques, 19,* 41-75.

CASTELA, C. (2008) Travailler avec, travailler sur la notion de praxéologie mathématique pour décrire les besoins d'apprentissage ignorés par les institutions d'apprentissage. *Recherches en Didactique des Mathématiques*, 28(2), 135 – 182.

CHEVALLARD Y. (1997) L'analyse des pratiques enseignantes en théorie anthropologique du didactique. *Recherches en didactique des mathématiques* 19(2) 221 – 266.

CHEVALLARD Y. (2007) Une épreuve expérimentale de mathématiques. http://yves.chevallard.free.fr/spip/spip/IMG/pdf/Une_epreuve_experimentale_de_mathematiques.pdf (consulté le 20/11/2017).

CHENEVOTOT-QUENTIN F., GRUGEON B., DELOZANNE E. (2011) Vers un diagnostic cognitif dynamique en algèbre élémentaire. In Kuzniak A., Sokhna M. (Eds.) Actes du colloque Espace Mathématique Francophone EMF2009, Enseignement des mathématiques et développement : enjeux de société et de formation (pp. 827−842), Dakar.

DALIBARD, E., ARZOUMANIAN, P. (2015). CEDRE 2014 - Mathématiques en fin de collège : une augmentation importante du pourcentage d'élèves de faible niveau. *Note d'information*, 19, MEN - DEPP.

DALIBARD, E., PASTOR, J.-M. (2015). CEDRE 2014 - Mathématiques en fin d'école primaire : les élèves qui arrivent au collège ont des niveaux très hétérogènes. *Note d'information*, 18, MEN - DEPP.

DE KETELE, J-M., GERARD, F-M (2005). La validation des épreuves d'évaluation selon l'approche par les compétences, *Mesure et Évaluation en Éducation*, 28-3, 1-26.

GRAPIN N. (2015). Étude de la validité de dispositifs d'évaluation et conception d'un modèle d'analyse multidimensionnelle des connaissances des élèves de fin d'école. Thèse de doctorat. Université Paris Diderot. Paris.

Grugeon, B. (1997). Conception et exploitation d'une structure d'analyse multidimensionnelle en algèbre élémentaire. *Recherches en didactique des mathématiques*, 17(2), 167 - 210.

GRUGEON-ALLYS B., GRAPIN N. (2015). Validité d'une évaluation externe : complémentarité entre une approche didactique et psychométrique In *Actes du séminaire national de didactique des mathématiques*. (Eds A.C. Mathé, E. Mounier).

LAVEAULT D., GREGOIRE J. (2014) *Introduction aux théories des tests en sciences humaines.* Bruxelles : De Boeck.

MARGOLINAS, C. & WOZNIAK, F. (2016). Situations d'énumération et exploration des collections. *Recherches en Didactique des Mathématiques, 35*(2), 183-220.

MOUNIER E. (2010) *Une analyse de l'enseignement de la numération au CP. Vers de nouvelles pistes* (Thèse de Doctorat). Université Paris.Diderot, Paris.

MOUNIER E. (2016) Évaluer les connaissances mathématiques dans des tâches de dénombrement chez les élèves du Cours Préparatoire, élèves âgés de 6-7 ans : un nouvel outil d'analyse, des premiers résultats. In *Actes du colloque de l'Association pour le Développement des Méthodologies d'Evaluation en Education ADMEE Europe*. ADMEE : Lisbonne

MOUNIER E, PRIOLET MY. (2016) La programmation des techniques opératoires dans les manuels scolaires de l'école élémentaire. Le cas de l'addition et de la soustraction, *Grand N*, 98, 5-26.

MOUNIER E. (2017) Nouveaux outils d'analyse des procédures de dénombrement pour explorer leur lien avec la numération écrite chiffrée et la numération parlée. *Recherches en didactique des mathématiques*, 36(3), 347-396.

REY O., FEYFANT A. (2014) Évaluer pour (mieux) faire apprendre. Dossier de veille de l'Ifé, 94.

ROBERT A. (2007) Stabilité des pratiques des enseignants de mathématique : une hypothèse, des inférences en formation. *Recherches en Didactique des Mathématiques, des mathématiques* 27/3 271-312.

SAYAC, N. (à paraître) Approche didactique de l'évaluation et de ses pratiques en mathématiques : enjeux d'apprentissages et de formation. Habilitation à diriger des recherches. Université Paris Diderot, Paris.

SAYAC, N., GRAPIN, N. (2015) Évaluation externe et didactique et didactique des mathématiques : un regard croisé nécessaire et constructif. *Recherche en didactique des mathématiques*, 35(1), 101-126.

VERGNAUD, G. (1990). La théorie des champs conceptuels. *Recherche en didactique des mathématiques*, 10-2.3, 133 – 170.

ANNEXE 1

Exercices extraits de l'évaluation fin CP proposée en 2016 par la DSDEN de la Haute Vienne (http://pedagogie.ac-limoges.fr/ia87/spip.php?article1175).

Ont été fournis aux participants les énoncés des exercices tels qu'ils figurent dans le cahier de l'élève et les consignes de passation et de correction données à l'enseignant ; nous n'avons pas retenu les exercices portant sur la résolution de problèmes additifs ou multiplicatifs. Nous les avons réorganisés dans cette annexe en réduisant la taille des énoncés et en faisant figurer les consignes et le codage à la suite de l'énoncé de l'exercice.

Exercice NC1 : *L'élève connaît-il les désignations écrites et chiffrées des nombres jusqu'à 100 ?*

Exercice NC1 :

Écris les nombres en lettres

40 .....

72 ..... 1 ☐

90 .....

Écris les nombres en chiffres

quatre-vingt-douze .....

trente-sept ..... 2 ☐

soixante-quatorze .....

Dire : « Sur la feuille nous avons écrit des nombres en chiffres, tu dois écrire ces nombres en lettres sur la ligne juste à côté. », puis « Sur la feuille, nous avons écrit des nombres en lettres, tu dois les écrire en chiffres sur la ligne juste à côté. »

Correction, codage :

Item 1 réussi si 2 nombres en lettres sur 3 sont correctement orthographiés.

Item 2 réussi si 2 nombres sur 3 sont correctement chiffrés.

Exercice NC2 : *L'élève connait il les décompositions additives des nombres inférieurs à 20 ?*

Exercice NC2 :

Remplis chaque case comme dans l'exemple :

18	c'est 6 + 12		5 + 5	c'est 10	
16	c'est	3 ☐	4 + 6	c'est	4 ☐
17	c'est		8 + 12	c'est	
11	c'est		6 + 14	c'est	

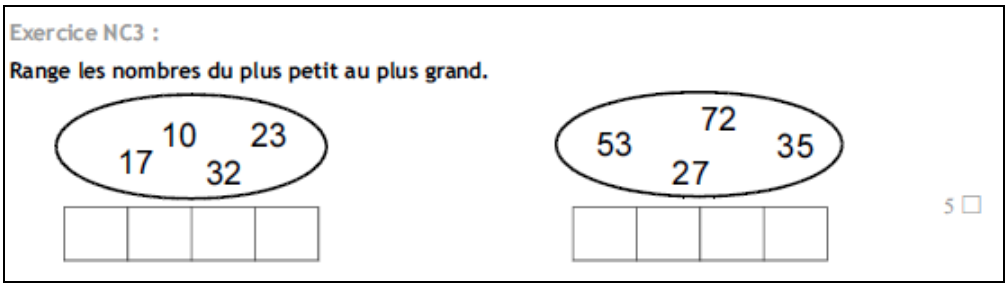
Dire : « Dans la première case, tu vois un exemple de ce qu'il faut faire : 18 c'est 6+12. Il y a plusieurs réponses possibles (demander à la classe d'autres exemples de réponse). Maintenant trouve une réponse pour chacun des nombres en dessous. »

Correction, codage : Item 3 réussi si 2 réponses sur 3 sont justes.

Dire : « Dans la deuxième case tu vois une addition : 5 +5 c'est 10. Donne les autres résultats. »

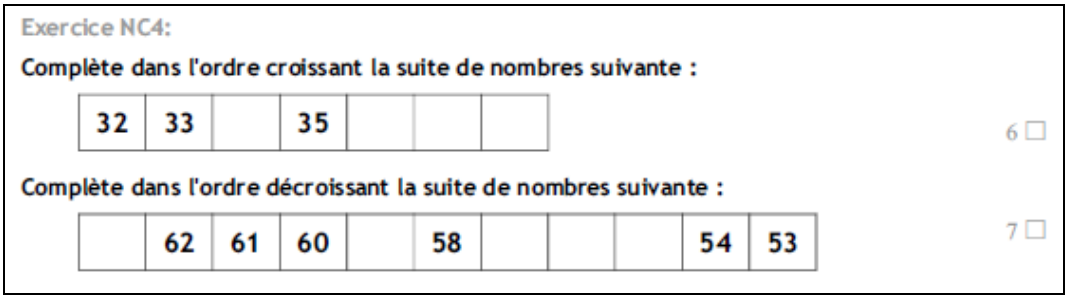
Correction, codage : Item 4 réussi si toutes les réponses sont justes.

Exercice NC3 : L'élève est-il capable de ranger des nombres < 100 du plus petit au plus grand ?



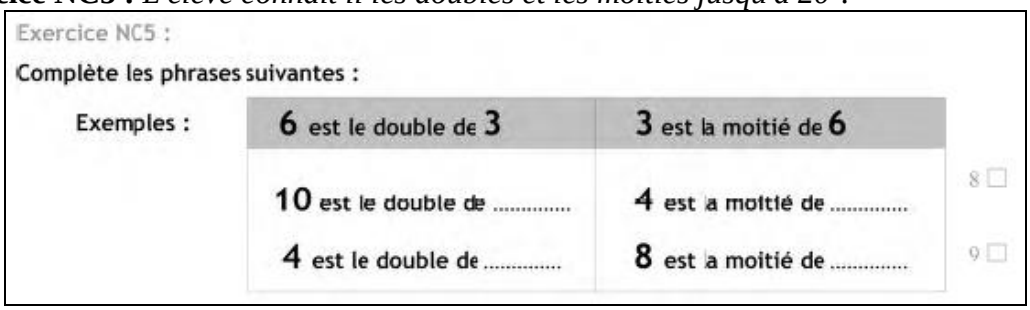
Dire : « Tu vois 2 sacs de nombres sur ta feuille, il faut sortir les nombres du sac et les écrire dans les bandes qui sont au-dessous, en les rangeant du plus petit au plus grand. »
Correction, codage : Item 5 réussi si les 2 bandes numériques sont bien renseignées.

Exercice NC4 : *L'élève est-il capable d'écrire une suite de nombres dans l'ordre croissant ou décroissant ?*



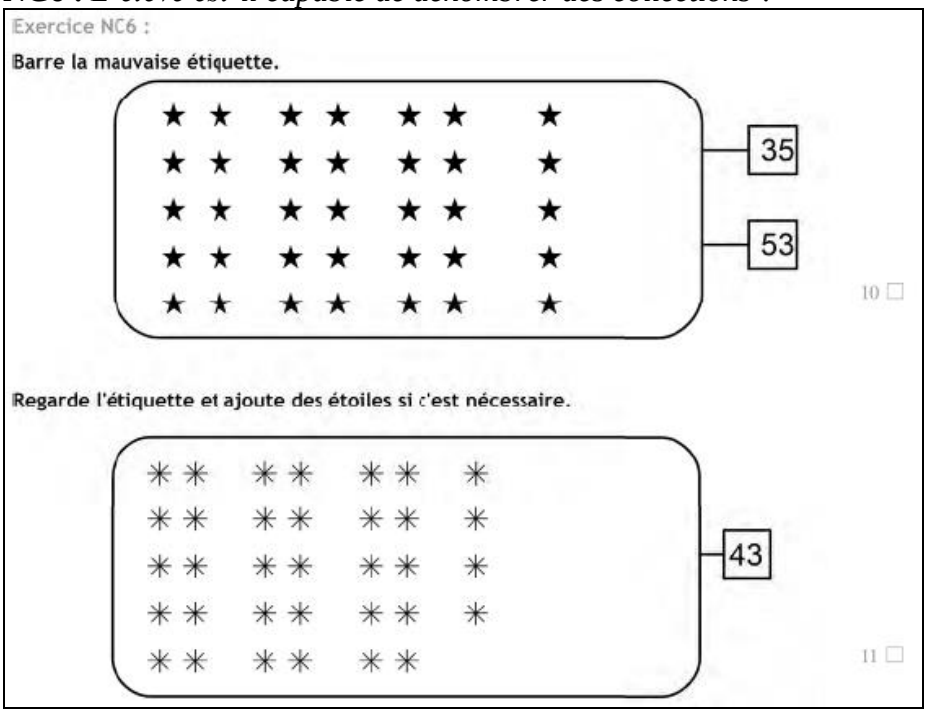
Dire : « Dans la première bande, tu vois des nombres dans les cases : 32, 33, tu dois poursuivre dans l'ordre croissant cette suite de nombres. »
Correction, codage : Item 6 réussi si les 4 nombres manquants sont justes.
Dire : « Sur la deuxième bande, il y a une suite de nombres dans l'ordre décroissant. Écris les nombres manquants. »
Correction, codage : Item 7 réussi si 4 nombres sur 5 sont justes.

Exercice NC5 : *L'élève connait il les doubles et les moitiés jusqu'à 20 ?*



Dire : « Dans la première colonne du tableau, tu vois un exemple : 6 est le double de 3. Tu dois compléter les deux lignes suivantes par le nombre qui convient. »
Correction, codage : Item 8 réussi si 5 et 2 sont notés.
Dire : « Dans la deuxième colonne, tu vois un exemple : 3 est la moitié de 6, tu dois compléter les deux lignes suivantes par le nombre qui convient. »
Correction, codage : Item 9 réussi si 8 et 16 sont notés.

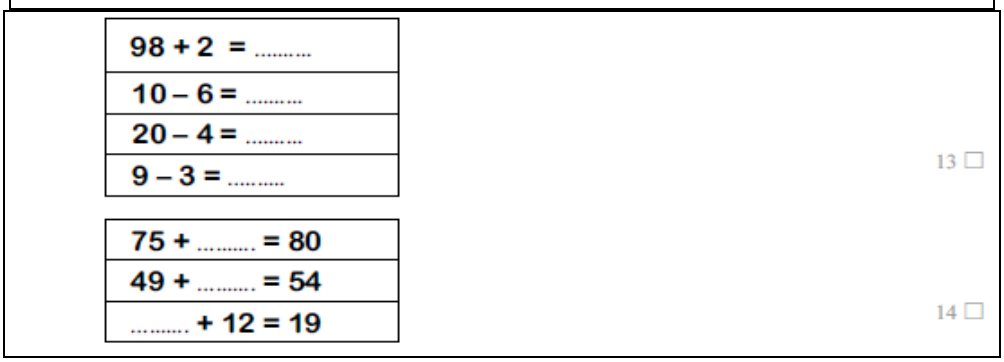
Exercice NC6 : *L'élève est-il capable de dénombrer des collections ?*



Dire : « Sur ta feuille, nous avons dessiné une collection d'étoiles mais 2 étiquettes nombres ont été écrites. Barre la mauvaise étiquette. », puis « Pour la 2ème collection la bonne étiquette est écrite. Vérifie que le nombre d'étoiles correspond bien à cette étiquette. Si nécessaire, complète le dessin en fonction de l'étiquette ».

Correction, codage : Item 10 réussi si 53 est barré. Item 11 réussi si l'élève a complété la 2ème collection avec les 9 étoiles qui manquent.

Exercice NC7 : *L'élève est-il capable de calculer en ligne des sommes et des différences, des opérations à trou ?*



Dire : « Observe et complète les opérations »

Correction, codage :

Item 12 réussi si 46, 62 et 100 sont notés.

Item 13 réussi si deux réponses sur trois sont données.

Item 14 réussi si deux réponses sur trois sont données.

Exercice NC8 : *L'élève connait il et utilise-t-il les techniques opératoires de l'addition et commence-t-il à utiliser celles de la soustraction (nombres inférieurs à 100) ?*

Exercice NC8 :

Pose et effectue les opérations suivantes :

29 + 15 =		27 + 32 =	

15 ☐

Dire : « Pose et effectue dans les cases les opérations ».
Correction, codage : Item 15 réussi si les deux opérations sont soigneusement posées et les résultats sont justes.

Exercice NC9 : *L'élève connaît-il la table de multiplication par deux ?*

Exercice NC9 :

Écris les résultats des opérations dans les cases suivantes :

16 ☐

Dire : « Dans la première case note le résultat de 2 fois 10, dans la deuxième case note le résultat de 2 fois 5, dans la troisième, note le résultat de 2 fois 8, dans la quatrième, note le résultat de 4 fois 2 et enfin dans la cinquième case, note le résultat de 3 fois 2. »
Correction, codage : Item 16 réussi si 4 réponses sur 5 sont justes.

Exercice NC10 : *L'élève est-il capable de calculer mentalement des sommes et des différences sur des nombres dont le résultat est inférieur à 40 ?*

Exercice NC10 :

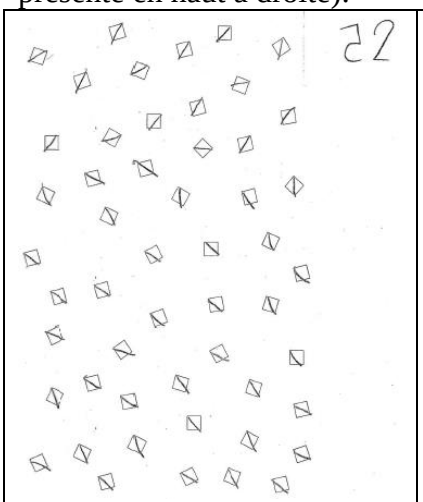
Écris les résultats des opérations dans les cases suivantes :

17 ☐

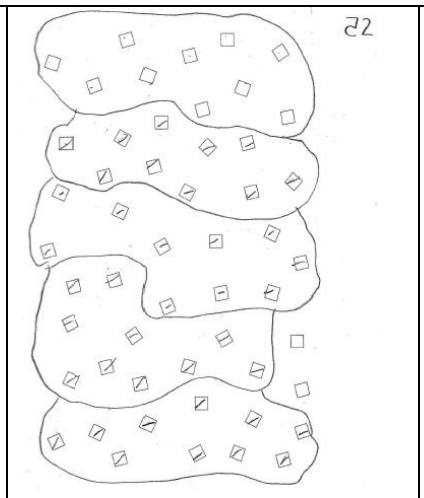
Dire : « Dans la première case tu indiques le résultat de 10 moins 3, dans la deuxième case tu notes le résultat de 20 moins 5, dans la troisième : 29 plus 10, dans la quatrième case tu notes le résultat de 35 plus 5 et enfin dans la dernière case tu notes le résultat de 19 plus 11. »
Correction, codage : Item 17 réussi si 4 réponses sur 5 sont justes.

ANNEXE 2 – TACHE DE DENOMBREMENT & PRODUCTIONS D'ELEVES

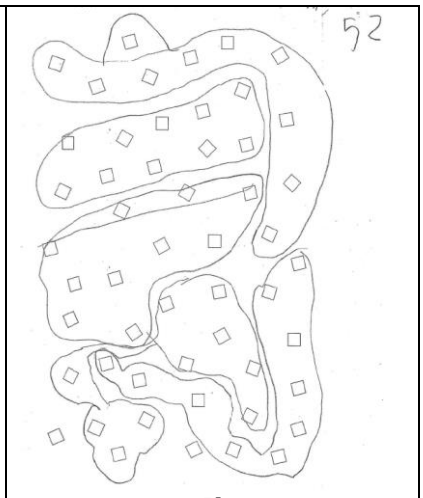
Tâche proposée : une collection de 52 carrés étant dessinée sur une feuille blanche, il est demandé à l'élève d'écrire le nombre de carrés sur un morceau de papier blanc (qui est ici présenté en haut à droite).



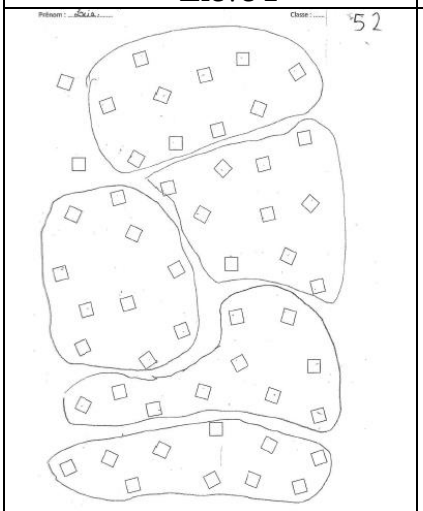
Elève 1



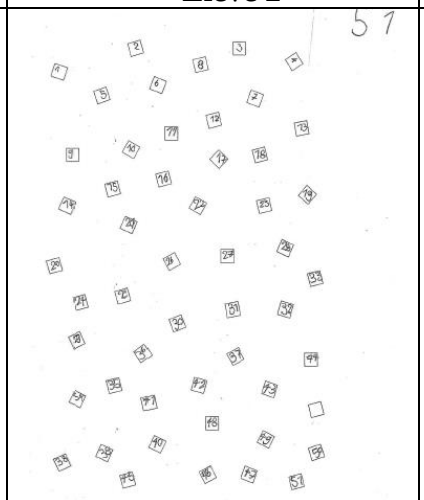
Elève 2



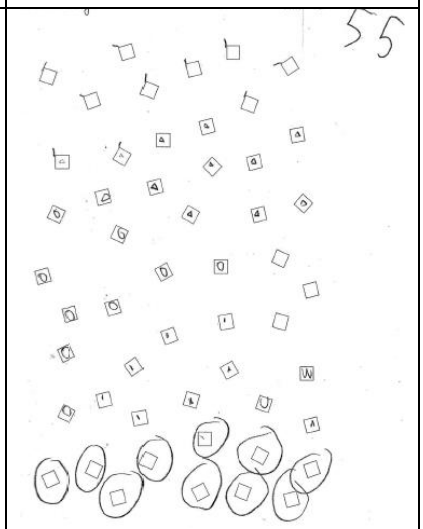
Elève 3



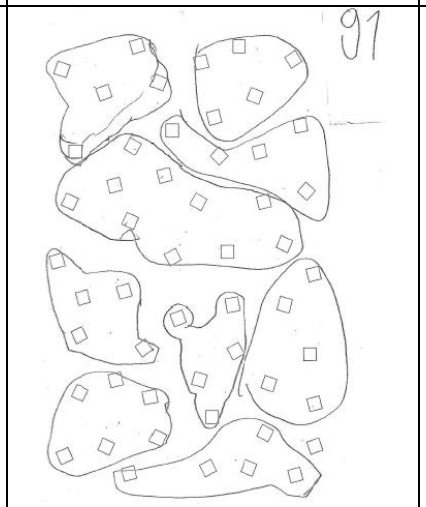
Elève 4



Elève 5



Elève 6



Elève 7

ANNEXE 3 – REPARTITION DES TACHES DANS LE DISPOSITIF D'EVALUATION

Résolution de problèmes arithmétiques

Classe de problème	Nb de tâches
Transformation d'état - Recherche de l'état final	2
Transformation d'état - Recherche de l'état initial	2
Transformation d'état - Recherche de la transformation	2
Composition de mesures - Recherche du tout	2
Composition de mesures - Recherche d'une des parties	2

Numération

Classe de problèmes	Nb de tâches
Dénombrement - Donner le cardinal	5
Dénombrement - Produire la collection	3
Comparer des nombres en EC	3
Comparer des collections	1
Passer du nom du nombre à son écriture chiffrée (le nom du nombre est donné oralement)	12
Passer de l'écriture chiffrée du nombre à son nom (le nom du nombre est donné oralement)	6
Connaître la comptine numérique	

Dans la classe de problèmes « dénombrement – donner le cardinal », des tâches mettant en jeu plusieurs étapes ont été proposé, notamment celle de donner le nombre d'éléments d'une collection après avoir ajouté une dizaine à une collection de cardinal connu.

Calcul

Classe de problèmes	Nb de tâches
Connaître les doubles	3
Ajouter 10	9
Décomposer 10 (a + …. = 10)	4
Décomposer 10 (a + b = *10)*	2
Connaître les répertoires (*autres que doubles ou décomposition de 10)*	7
Calcul d'additions posées	4