



HAL
open science

Étude de la validité de l'évaluation interne conduite en cours préparatoire, en français et en mathématiques

Christophe Blanc

► **To cite this version:**

Christophe Blanc. Étude de la validité de l'évaluation interne conduite en cours préparatoire, en français et en mathématiques. Education. Université Sorbonne Paris Cité, 2017. Français. NNT : 2017USPCB255 . tel-01908815

HAL Id: tel-01908815

<https://theses.hal.science/tel-01908815>

Submitted on 30 Oct 2018

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.

Université Paris Descartes

École doctorale 180 : « Cultures, individus, sociétés »

Laboratoire EDA – EA 4071 Éducation, Discours, Apprentissages

Étude de la validité de l'évaluation interne conduite en Cours Préparatoire, en Français et en Mathématiques.

Par Christophe BLANC

Thèse de doctorat de Sciences de l'Éducation

Dirigée par Sylvette MAURY et Marc VANTOUROUT

Présentée et soutenue publiquement le 24 mai 2017

Devant un jury composé de :

Mme Brigitte GRUGEON-ALLYS	Professeure des Universités – Université Paris Est Créteil	Rapporteure
Mme Sylvette MAURY	Professeure des Universités émérite - Université Paris Descartes	Directrice
Mme Lucie MOTTIER LOPEZ	Professeure des Universités - Université de Genève	Examinatrice
M Yves REUTER	Professeur des Universités émérite – Université Ch. de Gaule, Lille 3	Rapporteur
M Éric RODITI	Professeur des Universités - Université Paris Descartes	Examinateur
M Marc VANTOUROUT	Maître de conférences - Université Paris Descartes	Co-directeur

Remerciements

Au terme de la rédaction de la thèse, je souhaite d'abord remercier mes deux directeurs Sylvette Maury et Marc Vantourout pour la confiance qu'ils m'ont accordée. Leurs encouragements, et leurs relectures méticuleuses, tout au long de ces années ont été constants et porteurs. Je les remercie de m'avoir incité à m'engager dans ce travail de recherche et de m'avoir appris le métier de chercheur avec patience et rigueur. L'attention qu'ils portent dans leurs travaux à la prise en compte de l'activité (des évaluateurs, des évalués), à l'articulation de points de vues cognitifs et didactiques, à la dimension didactique de l'évaluation ont orienté mon questionnement de la validité des évaluations.

J'exprime tous mes remerciements à Brigitte Grugeon-Allys et à Yves Reuter pour avoir accepté d'être les rapporteurs de ma thèse. Leurs remarques constructives, les échanges que nous avons eu lors de colloques de l'ADMEE, pour l'une, et aux comités de rédaction des Cahiers Pédagogiques, pour l'autre, et les questions qu'ils ont soulevées m'ont permis de nourrir mes réflexions et de revisiter mon travail.

Je remercie chaleureusement Lucie Mottier Lopez d'avoir accepté de faire partie du jury : ses encouragements de la première heure et son intérêt pour différents aspects de mon travail ont été motivants et ont permis de tendre des passerelles entre des approches différentes mais non exclusives de l'évaluation comme le symposium de l'ICAP (2014) nous a permis de les présenter et de les discuter.

Je suis heureux qu'Éric Roditi, directeur du laboratoire EDA, ait accepté d'être membre du jury : sa connaissance des pratiques enseignantes et son engagement dans la formation des enseignants du premier degré en mathématiques permettent de placer mon travail dans la perspective de contribuer à la formation des enseignants.

Je remercie également Sabine Rodriguez-Smith, « compère » de thèse au laboratoire EDA, et Rémi Goasdoué. Leur questionnement aiguisé et leurs encouragements amicaux ont été d'un réel soutien.

Je tiens à remercier chaleureusement ma famille et mes amis pour leur soutien inconditionnel et leur esprit de cordée. Ils savent l'ampleur de l'effort et de l'investissement qu'a demandé un tel travail réalisé à côté d'un emploi à plein temps.

J'adresse une pensée particulière à mes anciens élèves. Ils ont toujours nourri ma réflexion et mon engagement.

Table des matières

Partie I- L'évaluation interne : introduction et question de recherche	7
Chapitre 1. Finalité(s) de l'évaluation	9
Chapitre 2. Le référent et le référé de l'évaluation.....	13
2.1. Référent, référentiel, référentialisation	13
2.2. Référé	22
2.3. Référentiel et référé pour le chercheur	23
Chapitre 3. La validité : une qualité plurielle de l'évaluation	24
3.1. Validité de contenu des tests.....	25
3.2. Validité psycho-didactique des épreuves.....	28
3.3. Validité globale et locale	29
Chapitre 4. Synthèse et formulation de la problématique	30
Chapitre 5. Présentation du corpus de données recueillies.....	32
Partie II – Étude de la validité curriculaire de l'évaluation.....	36
Chapitre 1. Les progressions annexées aux programmes, référentiel des enseignants	38
1.1. Programmes et progressions, un lien ambiguë.....	38
1.2. Les items « lecture » des progressions.....	39
1.3. Les items du sous-domaine des nombres et du calcul des progressions.....	41
Chapitre 2. Première approche de la validité curriculaire	43
2.1. Les livrets utilisés par les enseignants.....	43
2.2. Le taux de remplissage des livrets.....	52

2.3. Discussion	54
Chapitre 3. Analyse de la validité curriculaire	56
3.1. Couverture des domaines des programmes réalisée par l'évaluation	56
3.2. Couverture du sous-domaine « lecture » du domaine du Français	63
3.3. Couverture du sous-domaine des nombres et du calcul du domaine des Mathématiques.....	65
Synthèse de la partie II	67
Partie III – Étude de la validité psycho-didactique des épreuves.....	69
Chapitre 1. Cadre conceptuel d'analyse de la tâche	69
1.1. La tâche, première approche.....	69
1.2. A quelle tâche répond l'activité ?	70
1.3. Adaptation du modèle pour notre étude.....	72
1.4. Vers un cadre méthodologique pour analyser les épreuves d'évaluation.....	74
Chapitre 2. Méthodologie d'analyse psycho-didactique d'une épreuve d'évaluation	75
2.1. Méthodologie.....	75
2.2. Application à deux épreuves de référence.....	76
2.3. Discussion :.....	81
Chapitre 3. Le dessin : un simple « habillage » de la tâche ?.....	82
3.1. Le dessin dans des tâches d'évaluation de compétences complexes.....	83
3.2. Le dessin dans des tâches visant à évaluer des habiletés « de base » en lecture..	92
3.3. Le dessin dans des tâches de dénombrement.....	100
3.4. Synthèse du chapitre 3.....	110
Chapitre 4. Les questions écrites de compréhension en lecture	111
4.1. L'item relatif à la compréhension en lecture.....	112
4.2. Les questionnaires pour évaluer la compréhension en lecture :.....	117
4.3. Synthèse du chapitre 4.....	129
Chapitre 5. Le système de réponse	131
5.1. Rédiger une (phrase) réponse complète.....	132

5.2. Trouver la bonne réponse parmi plusieurs propositions de réponse	136
5.3. Trouver la bonne réponse et rédiger la réponse :	150
5.4. Dessiner en guise de réponse	153
5.5. Lire à haute voix	161
5.6. Synthèse du chapitre 5	170
Chapitre 6. Le choix des tâches d'évaluation	171
6.1. Tâches empruntées au web	171
6.2. Tâches empruntées à des manuels	178
Synthèse de la partie III	183
Partie IV – Étude de la validité globale et locale des tests et des épreuves.....	185
Chapitre 1. Analyse des tests et des épreuves d'évaluation des compétences relevant des nombres et du calcul.....	186
1.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences du domaine des nombres et du calcul en CP.	186
1.2. Analyse des tests et des épreuves	196
1.3. Synthèse	223
Chapitre 2. Analyse des tests et des épreuves d'évaluation des compétences de résolution de problème (additif).....	226
2.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences de résolution de problèmes additifs en CP	226
2.2. Analyse des tests et des épreuves	230
2.3. Discussion autour du « problème simple à une opération ».....	234
2.4. Synthèse	250
Chapitre 3. Analyse des tests et des épreuves d'évaluation des compétences en lecture	251
3.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences de lecture en CP.....	252
3.2. Analyse de la dimension phonographique des tests et des épreuves du 1er trimestre	276
3.3. Analyse de la dimension lecture des tests et des épreuves du troisième trimestre	291

3.4. Analyse de la dimension compréhension des tests et des épreuves du troisième trimestre	295
Synthèse de la partie IV	313
Conclusion.....	314
1. La construction de référentiels : une étape clé	315
2. Différentes formes de validité de l'évaluation	317
3. Limites et ouvertures.....	321
Sommaire détaillé.....	323
Table des figures.....	330
Table des tableaux	333
Bibliographie	335
Résumé	354
Abstract.....	356

Partie I- L'évaluation interne : introduction et question de recherche

L'institution scolaire est productrice d'un grand nombre d'évaluations dont les objectifs variés ont été soulignés par Cardinet (1989, 1990) dans les années 1980. L'amélioration des décisions relatives aux apprentissages des élèves, et l'établissement d'une communication avec l'enfant et ses parents au sujet de la progression de l'enfant constituent deux d'entre ces objectifs. La certification de compétences attendues à certains paliers de la scolarité, et l'information des décideurs en vue de l'amélioration de la qualité de l'enseignement en général complètent ce panel d'objectifs de l'évaluation dans le champ scolaire. Pratique incontournable, l'évaluation recouvre ainsi un grand nombre de réalités et de finalités diverses. Le constat est fait d'une augmentation globale de cette pratique à l'heure actuelle : les évaluations se multiplient dans tous les segments du système éducatif conduisant Gonthier-Maurin à dénoncer sévèrement « une pression évaluative institutionnelle [...] sans qu'il ne soit proposé [aux enseignants] ni accompagnement, ni cadre d'action cohérent, ni hiérarchisation des priorités » (2012). Cette inflation évaluative actuelle, qualifiée de « délire évaluatif » par certains chercheurs (Hadji, 2012), ne doit pourtant rien au hasard : elle est fortement dépendante d'enjeux éducatifs au sens le plus large qui s'inscrivent dans l'histoire de notre système éducatif (Guimard, 2010 ; Pons, 2010), et qui se trouvent réactualisés par les comparaisons internationales, l'apparition des compétences dans le champ des apprentissages scolaires et l'irruption d'une demande institutionnelle récente d'évaluation « positive » et « bienveillante » en particulier en Maternelle.

Parmi les nombreuses évaluations réalisées dans le champ des apprentissages scolaires, nous nous intéressons dans notre travail à la forme la plus quotidienne de celles-ci. Il s'agit de l'évaluation des apprentissages scolaires telle qu'elle est réalisée « au quotidien de la classe » par les enseignants. Ces évaluations sont qualifiées d'internes en ce sens qu'elles relèvent exclusivement de la responsabilité de chaque enseignant dans sa classe, et qu'elles sont destinées à un usage interne à la classe ou à l'école.

Le champ des évaluations internes est un sujet assez mal connu comme le relève le Conseil national d'évaluation du système scolaire (CNESCO) :

« Bien qu'au cœur historiquement de l'institution scolaire, et analysées, à travers différents champs scientifiques (les sciences de l'éducation, la psychologie...), par des chercheurs s'intéressant à la pédagogie (en particulier aux effets des évaluations sur les apprentissages des élèves), [ces évaluations] ont suscité peu d'états des lieux des pratiques des enseignants dans leurs classes » (2014, p.7).

Certaines études, telles celle de E. Fijalkow (2003) portant sur l'enseignement de la lecture/écriture en CP, se sont intéressées, parmi d'autres aspects, à l'évaluation réalisée par les enseignants, mais les données collectées sont des réponses à des questionnaires soumis aux enseignants et relèvent donc de pratiques déclarées. Ainsi que le rapportent également O. Rey et Feyfant (2014) dans un dossier de veille de l'Institut Français de l'éducation consacré à l'évaluation, nous disposons donc d'assez peu de données sur ce type d'évaluation, pourtant au cœur de l'activité quotidienne des enseignants et de l'activité des élèves, et à la lumière de laquelle certains chercheurs invitent à penser l'échec scolaire. Celui-ci peut être ainsi considéré selon Crahay (1996, p. 294) « comme une réalité dépendante des pratiques d'évaluation des enseignants ».

A côté de ces évaluations internes, se sont développées depuis les années 1990 un grand nombre d'évaluations dont les modalités de passation et les protocoles de correction sont standardisés. Ces évaluations, par opposition aux précédentes, sont qualifiées d'externes : l'élaboration est confiée à des acteurs extérieurs à la classe, et les résultats ne sont pas directement destinés aux enseignants. L'enquête PISA en est un exemple typique. Il vise la comparaison des systèmes éducatifs nationaux. Ces évaluations externes sont elles-mêmes de différents types, ce que Trosseille et Rocher (2015) soulignent, selon que l'on procède à une évaluation de masse ou à une évaluation sur échantillon (comme c'est le cas pour les Cycles d'Évaluations Disciplinaires Réalisées sur Échantillon CEDRE) en fin de CM2 par la Direction de l'Évaluation, de la Prospective et de la Performance (DEPP). Les résultats de ces évaluations externes sont largement médiatisés et constituent des indicateurs pour les décideurs des politiques éducatives. Ce type d'évaluation a fait surgir un vaste questionnement sur l'évaluation. Des questions se posent particulièrement quant à la stabilité des diagnostics que l'évaluation permet d'établir, dans le cas où l'évaluation poursuit une visée diagnostique (Loye, 2012). D'autres questions se posent quant aux qualités métriques des items lorsque l'on vise une distribution des élèves selon la réussite aux items du test, et une caractérisation des compétences des groupes d'élèves en fonction de la difficulté des items du test (cf. Pini, 2012). Différentes théories (théorie classique des scores, théorie de la généralisabilité, ou théorie de réponse à l'item) ont été élaborées pour déterminer les items d'un test qui sont informatifs et qui discriminent bien les sujets. D'autres questions émergent encore quant au contenu de ces évaluations : Grapin (2015) soulève ainsi celle de la « validité » des évaluations externes et de la qualité des liens entre les questions posées et les ressources identifiées comme nécessaires à une bonne réponse. L'évaluation externe, donnant lieu à des recherches variées, est ainsi un champ bien documenté contrairement à celui de l'évaluation interne. Certaines questions se posent pourtant dans les mêmes termes. Ainsi en est-il de la validité, qui constitue une question centrale de notre recherche.

Dans ce travail, nous nous intéressons à la partie explicite de l'évaluation interne réalisée par des enseignants, i.e. celle à laquelle on accède en prenant en considération les épreuves soumises par l'enseignant aux élèves à des fins d'évaluation, et qui servent de supports aux jugements portés dans les « livrets scolaires ». Elle peut être qualifiée d'« instituée » dans la mesure où elle correspond à « un acte délibéré et socialement organisé aboutissant à la production d'un jugement de valeur » comme l'entend Barbier (1985, pp. 34-35). Cette définition de départ comporte l'émission d'un jugement évaluatif. Une première question se pose à propos de l'usage qu'il sera fait de ce jugement, lequel, souligne Hadji (1992, p.25-26), dépend du type de décision susceptible d'être prise en fonction des résultats obtenus.

Chapitre 1. Finalité(s) de l'évaluation

Diverses fonctions sont classiquement attribuées à l'évaluation et sont parfois combinées. Elles correspondent à trois modes d'articulation entre l'évaluation et l'action pédagogique.

A la suite d'Hadji (1989, p. 57), on entend par fonction « le rôle caractéristique d'un élément ou d'un objet dans l'ensemble où il est intégré ». En ce sens, pour cet auteur, la fonction de l'évaluation correspond au rôle joué par cette activité dans l'ensemble des activités de l'enseignant. S'appuyant sur les travaux de De Landsheere (1976, pp. 60-63), Hadji (idem) propose de repérer des fonctions de l'évaluation à partir des objets possibles de l'évaluation scolaire :

« L'évaluation de l'élève peut avoir une fonction pronostique si elle permet de le guider et de l'orienter dans ses choix scolaires par une prédiction des futures performances. Elle peut avoir une fonction diagnostique lorsqu'elle sert à situer l'élève dans son processus d'apprentissage et à diagnostiquer ses éventuelles difficultés en regard des savoirs et savoir-faire qui devraient être acquis. Elle vise une régulation de l'enseignement/apprentissage. Elle peut enfin viser l'inventaire ou le bilan des connaissances et des acquisitions, mesurer les apprentissages réalisés ».

Trois objectifs sont dès lors à distinguer qui entraînent chacun, selon Hadji (idem) la mise en œuvre d'une stratégie d'évaluation particulière :

« Si l'objectif est d'orienter, alors l'évaluation porte sur les aspects considérés comme des prérequis pour de futures acquisitions.

Si l'objectif est de réguler (guider constamment le processus d'apprentissage), alors l'évaluation doit permettre à l'évaluateur d'obtenir des informations précises sur les stratégies utilisées par les élèves dans telle ou telle situation et les difficultés manifestées.

Si l'objectif est de certifier, alors l'évaluation porte sur des composants globaux et socialement significatifs ».

Pour désigner les pratiques qui s'organisent autour de ces trois grandes fonctions (orienter, réguler, certifier), on parle « d'évaluation diagnostique ou pronostique, ou prédictive ; d'évaluation formative ; et d'évaluation sommative » résume Hadji (1989, p. 58)¹. Dans la littérature, il apparaît que les distinctions entre ces fonctions ne sont pas toujours tranchées.

Ainsi :

- Au sujet de l'évaluation pronostique. Grapin (2015, p.15) citant Scallon (1988) souligne le fait que le terme diagnostic peut renvoyer à plusieurs types de diagnostics. Dans ce qui précède, il apparaît que l'évaluation diagnostique citée par Hadji (idem) renvoie à une fonction pronostique de l'évaluation, ce qui n'est pas de même nature que le diagnostic que l'on établit lorsque l'on vise la compréhension de la situation de l'élève. Le diagnostic peut porter sur :

- le repérage du niveau de connaissances d'un élève (dans le cas d'une évaluation située en amont),
- le repérage de causes de difficultés persistantes (quelle que soit la position de l'évaluation dans le cycle d'apprentissage)

¹ On peut situer ces travaux dans le prolongement des travaux pionniers d'Allal (1979).

- ou bien encore, comme dans le cas de l'évaluation développée par Grugeon-Allys dans le cadre de Pépite, la recherche des cohérences de fonctionnement qui permettent de repérer des besoins d'apprentissages des élèves².

Le diagnostic ainsi réalisé ne poursuit plus l'objectif d'orientation et de pronostic relevé par Hadji concernant l'évaluation diagnostique. Ces auteurs sont donc conduits à considérer l'évaluation diagnostique comme poursuivant le but d'une régulation de l'enseignement/apprentissage comme Cardinet (1989) le suggérait déjà, ne différenciant pas l'évaluation diagnostique de l'évaluation formative.

- Dans la littérature de l'évaluation en éducation, l'idée d'une évaluation « au service des apprentissages » est inmanquablement associée à l'évaluation formative proposée par Scriven en 1967 laquelle vise à s'assurer que les moyens de formation correspondent aux caractéristiques des élèves. Elle attribue « moins d'importance aux décideurs et se préoccupe davantage de l'utilisateur » soulignent Amigues et Zerbato-Poudou (1996, p.146) pour expliquer ce qui, selon eux, a fait le succès de cette approche.

Brun (1979, p.174) appréhende l'évaluation formative en mathématiques, comme fondée sur « l'analyse du comment l'élève organise une connaissance au fur et à mesure qu'il progresse dans sa construction ». Pour cet auteur, cette forme d'évaluation requiert une exigence de compréhension de l'activité propre du sujet. L'analyse des procédures des élèves « devient alors la question centrale d'une compréhension de la démarche d'acquisition des notions, et par là-même d'une évaluation formative en relation avec une théorie de la formation des connaissances » (opus cité, p.179).

Elle exige parallèlement une analyse des tâches didactiques « en dégagant les critères d'adaptation de problèmes qui mobilisent l'activité (mathématiques) des élèves et la font évoluer » (idem.). Cette articulation évaluation/didactique est renforcée par Bain (1987), rapporté par Perrenoud (1991), qui, dans le cadre d'une évaluation intégrée à l'enseignement du français, plaide en faveur « d'une entrée par la didactique, autrement dit, pour une problématique de l'évaluation formative construite à partir des contenus et des structures spécifiques du savoir, aussi bien que des mécanismes d'apprentissages correspondants ».

En 1993, Allal, Bain & Perrenoud (cf. pp. 17-23) s'accordent pour reconnaître des convergences incontournables entre l'évaluation formative et la didactique dans ses dimensions épistémologiques (définition des objets d'enseignement, organisation des savoirs), psychologiques (élucidation des conditions d'appropriation des savoirs par les élèves, obstacles, représentations, stratégies d'apprentissages etc.) et praxéologiques (réflexion sur les interventions des enseignants, organisation des situations didactiques etc.) Les recherches conduites en didactique du français sur l'évaluation des écrits (Groupe EVA, INRP, 1991) constituent une manifestation de ces convergences.

Nunziati proposa dans les années 1970 une approche originale de l'évaluation qualifiée d'évaluation formatrice par Scallon en 1982 pour la différencier de l'évaluation formative dont elle est conceptuellement l'héritière. Selon Nunziati (1990), « les études en docimologie d'une part, et les premiers travaux sur le concept d'évaluation formative d'autre part, fournissaient des éléments théoriques sur lesquels travailler, mais le dispositif d'évaluation formative restait à inventer et à tester ». Pour ces chercheurs, l'évaluation, formatrice, est

² Cela pourrait faire écho à la proposition de Barbier (1987) qui estime que le terme « identification » serait plus adéquat pour désigner l'évaluation diagnostique / pronostique et insiste sur le fait que cette identification devrait intégrer les réussites, lesquelles constitueraient alors autant de points d'appui pour la formation.

considérée comme une démarche didactique, dans laquelle « la régulation est essentiellement le fait de l'élève » (1990, p.52). Il semble que l'on peut voir dans cette approche de l'évaluation, les prémices d'un renouveau évoqué par Scallon (2004, pp.12-16) dans le contexte nord-américain, en réaction à la pratique des tests standardisés. Le terme « assessment » permettrait de distinguer l'évaluation de la mesure et du testing. L'auteur cite Linn et Gronlund (2000, p.31) pour qui, selon sa propre traduction, « l'assessment est un terme général qui désigne l'étendue complète des procédés utilisés pour obtenir des informations au sujet de l'apprentissage de chaque étudiant (observations, jugements de performances ou de projets, tests papier-crayon) et la formation de jugements de valeur touchant la progression » (2004, p.15). Pour souligner la nuance de sens entre les deux termes, Scallon utilise l'analogie selon laquelle « l'évaluation est à l'assessment ce que la recherche d'un objet précis est à l'enquête » (2004, p. 26). Le terme « student assessment-as-learning », employé par le collègue Alverno qui fait figure de pionnier en la matière (Loacker, 1995), traduit l'idée de donner une place de choix à l'évaluation au cœur même des situations d'apprentissage. Cette visée formatrice intègre les « feed-back » apportés aux élèves ce qui correspond à la notion de régulation interactive développée par Allal (1979) dans le cadre de l'évaluation formative.

Mais, bien que, selon Vantourout & Goasdoué (2014), « elle soit fortement affirmée dans les discours prescriptifs et institutionnels », un bilan globalement négatif au sujet de la mise en œuvre d'une évaluation formative peut être dressé. Il est également dressé par les inspecteurs généraux de l'Education nationale (Bardi & Mégard, 2009, p.131). Pour expliquer ce bilan négatif, Vantourout & Goasdoué avancent plusieurs arguments. Ils soulignent entre autres le rapport complexe entretenu par la didactique des mathématiques, en particulier, avec l'évaluation, même si à l'inverse, les didacticiens du français se sont intéressés assez tôt à la question de l'évaluation (cf. les travaux du groupe EVA INRP sur les écrits à l'école primaire évoqués plus haut cf. Mas et coll. 1991).

- Perrenoud (2001) s'interroge sur le fait de savoir si l'évaluation formative et l'évaluation sommative (i.e. lorsque l'évaluation intervient en fin de cycle d'apprentissage) engagent l'adoption de postures contradictoires ou complémentaires. Il suggère de dépasser cette opposition estimant que « le conflit n'est pas entre le formatif et le sommatif mais entre logique de formation et logique d'exclusion ou de sélection ».

Ainsi, comme l'indique Hadji (1989, p. 61), les distinctions entre les différents types d'évaluation ne sont pas aussi tranchées que la typologie présentée pourrait le laisser penser. Il est finalement peu aisé de distinguer clairement les trois fonctions précédemment décrites au sein de l'évaluation conduite par l'enseignant dans sa classe, d'autant que, selon Bardi et Mégard (2009), « les fonctions diagnostique et formative sont demeurées bien peu pratiquées dans les classes » françaises. Hadji (1992, p.27) estime que, par-delà les fonctions de l'évaluation classiquement distinguées que nous venons de préciser, « on peut finalement considérer que l'évaluation, quelle que soit l'intention dominante qui oriente l'évaluateur vers son objet, peut contribuer à fonder deux grandes séries de décisions : des décisions d'ordre pédagogique (évaluation prédictive ou formative), des décisions d'ordre social (évaluation sommative certificative). Dans le premier cas, Hadji (1992, p. 27) souligne que « l'évaluation sert à formuler des jugements de valeur utiles à l'activité de l'enseignant, et qui puissent lui permettre d'organiser celle-ci de façon plus efficace, parce que plus rationnelle ». Dans le deuxième cas, l'auteur indique (idem) que l'évaluation vise de pouvoir « formuler des jugements de valeur utiles à la société en tant qu'elle a besoin de connaître la

valeur scolaire des sujets, pour « orienter » ceux-ci vers tel ou tel type d'activité socialement et économiquement nécessaire ».

Le regard sur l'évaluation peut ainsi se placer au niveau de ce qui permet à l'enseignant de fonder un jugement évaluatif et une décision d'ordre pédagogique. La qualité de la décision étant dépendante de la qualité de l'évaluation, c'est donc à la qualité de l'évaluation que nous nous intéressons. Nous suivons pleinement les inspecteurs généraux lorsqu'ils estiment que « la qualité de l'évaluation [des acquis des élèves] est tout aussi déterminante que l'usage que le maître en fait » (IGEN-IGAENR, 2005, p.68). La qualité didactique de l'évaluation (caractéristiques des épreuves, couverture des domaines) constitue ainsi une question centrale de notre travail, au-delà de la fonction que l'on attribue à l'évaluation. Ce point de vue nous conduira à questionner la validité didactique de l'évaluation considérée par Bain et Schneuwly (1993, p. 71) comme « la pertinence et l'utilité du contrôle évaluatif pour la régulation de l'enseignement et de l'apprentissage dans le cadre d'une séquence didactique ».

L'évaluation régulière que nous examinons dans ce travail est réalisée par des enseignants en classe de Cours Préparatoire (CP). Dans le cadre des programmes de 2008 en vigueur au moment où nous recueillons les données exploitées dans notre recherche, le CP se situe au milieu du cycle 2 de l'école primaire (lequel comprend avant le CP la Grande Section de maternelle et, après le CP, le CE1). Seul le CE1, situé en fin de ce cycle, est pleinement concerné par la question de la validation de compétences telles que définies par le palier 1 du Socle commun de compétences et de connaissances. L'évaluation n'a donc pas, en CP, de finalité sommative certificative relativement à ces compétences : nous ne sommes pas en fin de cycle d'apprentissage, il n'est pas attendu que les élèves valident un certain nombre de compétences du Socle. Quel que soit le moment où elle intervient dans la séquence d'enseignement/apprentissage, l'évaluation à laquelle nous nous intéressons dans ce travail engendre, ou contribue à fonder, des décisions qui sont avant tout des décisions d'ordre pédagogique. Celles-ci font l'objet d'une communication périodique aux familles conformément au Code de l'Education³ :

« L'enseignant de la classe est responsable de l'évaluation régulière des acquis de l'élève. Les représentants légaux sont tenus périodiquement informés des résultats et de la situation scolaire de leur enfant. Dès que des difficultés apparaissent, un dialogue renforcé est engagé avec eux ».

L'enseignant est donc tenu d'évaluer régulièrement les acquis des élèves, de communiquer les résultats de l'évaluation et les décisions qu'il prend ou envisage concernant les élèves en fonction de ces résultats. Le livret scolaire régulièrement remis aux familles⁴ joue un rôle important à ce niveau. Dans le cadre des programmes de 2008, il n'existe pas de livret scolaire unique dont l'architecture soit commune à tous les élèves (ce qui est désormais le cas depuis la rentrée 2016). Il échoit donc à chaque enseignant d'élaborer son propre livret, son propre référentiel. Nous interrogeons à présent cette notion de référentiel propre à chaque enseignant. Comment, et en référence à quoi, le référentiel de l'enseignant pour évaluer est-il conçu ou choisi ?

³ Extrait de l'article D321-6

⁴ Assimilable aux anciens bulletins trimestriels, il prend part au Livret Personnel de Compétences (LPC). Le LPC comprend en plus des résultats périodiques, l'évaluation des compétences du Socle aux Paliers 1 (fin CE1) et 2 (fin CM2), ainsi qu'un certain nombre d'attestations (Sécurité routière ...).

Chapitre 2. Le référent et le référé de l'évaluation

Selon Barbier (1985, p.64), l'acte d'évaluation est assimilable à un « procès de transformation de représentations ». Evaluer est « une lecture orientée de la réalité » (Hadji, 1992, p. 29) qui consiste à « lire » une représentation de la réalité (le « référé » de l'évaluation) au moyen d'une grille de lecture adéquate (le « référent » de l'évaluation) construite pour cela. Hadji précise que le référent est construit en fonction d'attentes concernant l'objet évalué (1992, p.32), et que le référé est également un construit, constitué de l'ensemble « des indicateurs prélevés dans la réalité témoignant de la « présence » des critères souhaités et recherchés » (1992, p.33)

S'intéresser à l'évaluation conduit à examiner successivement ces deux construits (le référé et le référent), et à les mettre en relation suivant ainsi la définition donnée par Lesnes (1984, p.132) et reprise par Barbier (1985), Hadji (1989, p.25) et Figari (1989, p.25) :

« 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.) ».

Selon De Ketele (2012, p.206), la confrontation d'un référé à un référent constitue un élément unificateur des différentes conceptualisations de l'évaluation, au-delà de la diversité des paradigmes et de leurs usages. Nous situons notre travail et notre réflexion dans ce cadre général.

Figari (2006, p. 103) prévient toutefois que « le « référent », avec ses dérivés (référentiel, référentialisation, etc.), constitue peut-être la notion qui a introduit le plus de changements et développé le plus de malentendus et de dérives ». Nous nous intéressons donc particulièrement, dans ce qui suit, à préciser l'usage que nous ferons du terme référent auquel nous préfererons, pour la suite, le terme de référentiel.

2.1. Référent, référentiel, référentialisation

Le référent constitue la « grille de lecture » particulière des référés indique Hadji (1992, p. 31). « Lorsqu'il est explicité », précise cet auteur, « il peut être défini comme l'ensemble des critères au nom desquels on va se prononcer » (idem). Nous prendrons à notre compte cette nécessité d'expliciter le référent utilisé par l'évaluateur en la rapportant au domaine de l'évaluation interne des apprentissages scolaires auquel nous nous intéressons.

Le référent constitue donc la grille de lecture permettant à l'enseignant de produire du sens et à attribuer une valeur au produit de l'activité de l'élève. La définition précédemment présentée souligne que la mise en relation du (ou des) référé(s) avec le référent est implicite ou explicite. Mottier Lopez (2015) suggère ainsi de distinguer « ce qui relève du référentiel formel de l'évaluation (objectifs des plans d'études, transposés et reformulés par les enseignant.e.s, les critères d'évaluation définis) et des référents souvent informels associés aux situations, aux contrats, pratiques, normes privilégiées dans les micro-cultures de classe, en tant qu'arrière-fond interprétatif recelant des significations et des attendus souvent implicites mais qui contribuent également aux processus évaluatifs de construction de sens et de valeur. »

Cette conception permet de détailler la définition précitée en intégrant une dimension temporelle. Une partie du référent est explicite et peut être désignée sous le terme de référentiel. Cette partie du référent préexiste à l'évaluation, c'est à partir de celle-ci que l'évaluation est réalisée. Son caractère formel peut toutefois être discuté comme nous le verrons ci-dessous. Une autre partie est plus informelle. Elle est contemporaine au temps de la passation de l'évaluation. Elle ne correspond pas à ce que nous étudions dans ce travail, nous l'évoquerons simplement pour montrer en quoi elle diffère.

a. En amont de la passation de l'évaluation par les élèves : le référentiel

Selon Barbier (1985, p. 73), le référent apparaît « dans son statut, comme lié à des objectifs ». Pour qu'il y ait évaluation, il faut qu'une référence à des objectifs intervienne. Il souligne en outre que « dans sa fonction, le référent joue un rôle instrumental dans la production d'un jugement de valeur » (idem). Le référent fonctionne ici comme un référentiel (Figari, 2006, p.103) qui permet de rapporter les résultats de l'évaluation au point de vue construit par l'évaluateur.

Dans notre travail, les programmes scolaires constituent le référentiel formel le plus évident. Ils constituent le système d'attente de l'institution scolaire. Ils orientent évidemment l'activité évaluative de l'enseignant qui répond à une demande d'évaluer des objectifs d'apprentissage désormais désignés comme des compétences. Mais, d'autres référentiels formels côtoient les programmes. Dans le cas de l'évaluation interne que nous étudions, les manuels constituent par exemple des aides à la mise en œuvre de l'enseignement susceptibles de constituer en eux-mêmes d'autres référentiels formels : dans la mesure où l'enseignant utilise un manuel particulier, il est possible qu'il soit tenté d'utiliser les épreuves que celui-ci propose pour évaluer lorsqu'il en contient. Il faudrait s'assurer que le référentiel formel du concepteur du manuel est strictement en conformité avec le référentiel formel des programmes. Plus encore, dans le cas des programmes de 2008, des progressions sont annexées aux programmes. Présentées comme de simples repères, ce que nous analyserons en détail dans la partie II de la thèse, elles constituent en fait une liste d'objectifs plus opérationnels que le texte des programmes. Elles sont donc susceptibles de constituer le véritable référentiel formel de l'enseignant pour évaluer, le « référentiel alpha », en quelque sorte, à partir duquel chaque enseignant élabore son propre référentiel formel.

Pour évaluer les apprentissages de ses élèves, l'enseignant dispose donc non pas, comme on le constate, d'un référentiel formel, mais d'une pluralité de référentiels formels qui, s'ils sont tous rattachés aux programmes, constituent autant de référentiels plus ou moins opérationnels, plus ou moins bien articulés. Ils occasionnent un travail de reconstruction, de « référentialisation », de la part de l'enseignant.

Figari (1994) et Figari & Achouche (2001) ont évoqué cette notion de référentialisation pour désigner « l'ensemble des procédures consistant à modéliser l'évaluation et à élaborer un protocole qui établit un lien logique entre référents, critères et indicateurs de l'évaluation » (Figari, 2006, p. 104). Selon ces auteurs, toute évaluation est dépendante de sa référentialisation laquelle peut être considérée « comme une notion active, définissant la partie de l'activité évaluative qui consiste à invoquer des référents justifiant le choix des critères et des indicateurs de l'évaluation ». L'explicitation de la référentialisation permet ainsi selon Figari (2006, p. 106) « d'interpréter les référentiels » et, particulièrement, la valeur des interprétations qu'ils conditionnent.

Au cours du processus de référentialisation, les connaissances de l'enseignant (sur l'objet évalué mais aussi sur les élèves) entrent en jeu. Le processus de référentialisation de l'évaluation entrepris par l'enseignant se matérialise par deux éléments : le livret scolaire qui

constitue la part la plus formelle du référent de l'enseignant pour évaluer les acquis de ses élèves (il est conçu à partir du référentiel « alpha » que sont les programmes), et les épreuves qui s'y rattachent. Un modèle de l'activité de l'élève au cours de la réalisation de la tâche, ou tout du moins une représentation des connaissances que l'élève mobilisera, sous-tend le choix de l'épreuve et des critères retenus pour en apprécier la réalisation.

La partie « amont » du référent de l'enseignant pour évaluer (i.e. située en amont de la passation des épreuves d'évaluation) résulte donc d'un processus de référentialisation élaboré à partir de référentiels (cf. Figari, 2006, p.107). Ce processus, articulé au référentiel formel que constituent les programmes, fait intervenir les connaissances personnelles de l'enseignant relativement aux objets évalués et aux élèves.

Nous qualifierons cette part « amont » du référent, disjointe de la production des référés, de référentiel.

Le référentiel doit donc être compris comme étant le fruit d'un processus de référentialisation propre à chaque enseignant, que nous n'étudions pas en lui-même, mais dont nous étudions les effets produits au niveau du choix des épreuves et de leur adéquation à évaluer ce que l'enseignant veut évaluer.

b. Pendant et après la passation : le référent

D'autres travaux s'intéressant à l'activité évaluative des enseignants ont permis de considérer que le référent de l'évaluateur était susceptible de varier au cours de la correction des épreuves sous l'influence de différents facteurs (cf. Noizet et Caverni, 1978, Abernot dans Bonniol & Vial, 1997, pp. 59-60 ; et, plus récemment Vantourout, 2004, 2006, 2010 et Goasdoué et Vantourout, 2017). Cela conduit à considérer un autre aspect du référent, dynamique et contemporain de la production des référés.

Barbier estime (1985, p.74) que les processus d'élaboration des référents entraînent « une grande instabilité des actes d'évaluation marquée par la variabilité des jugements produits » : des variations interpersonnelles sont susceptibles d'intervenir de façon non contrôlée, faisant évoquer à certains auteurs l'idée d'une évaluation subjective (cf. De Ketele, 1993, pour une présentation d'un « paradigme de l'intuition pragmatique » en évaluation), conduite sans explication d'objectifs ni de critères. Le processus de constitution du référent comprend pour Barbier (1985, pp. 74-78) « un processus d'identification des objectifs poursuivis indépendamment de l'acte d'évaluation et pris en compte par lui » (idem) (que nous avons dénommé le référentiel) et « un processus de spécification de normes ou de critères à partir de ces objectifs », dépendant de l'acte d'évaluation (idem). Au cours de la correction, au moment de l'analyse des réponses des élèves et de la production du jugement évaluatif, un certain nombre d'effets sont susceptibles de se manifester (cf. par exemple les différents effets influençant la notation dont l'effet d'ordre examinés par Bonniol, 1981; Caverni, 1975 ou encore Gilly, 1980) confirmant que le référent élaboré par l'enseignant n'est pas une caractéristique totalement stable et prédéterminée mais une entité dynamique préformée qui conditionne l'évaluation tout autant qu'elle dépend d'elle.

Dans notre travail, nous nous intéressons aux référentiels construits par les enseignants. Nous analysons l'effet des choix effectués, au terme du processus de référentialisation, qui les conduisent à concevoir un livret scolaire et des épreuves permettant d'évaluer les compétences qui composent le livret. Nous nous intéressons ainsi au produit du processus de référentialisation.

2.1.1. Les programmes : un référentiel de compétences

Les programmes de 2008 sont explicitement conçus en référence à une approche par les compétences. Cela constitue un élément à prendre en compte pour comprendre les référentiels des enseignants. Disons simplement que dans les programmes de 1985 pour l'école primaire, apprendre à lire était un seul objectif, décliné dans les programmes de 2008 pour le CP en une dizaine de compétences.

Pour B. Rey et ses collaborateurs (2006, p.26), indépendamment d'un ancrage disciplinaire particulier, une « authentique compétence » serait « la capacité à répondre à des situations complexes et inédites par une combinaison nouvelle de procédures connues ; et non pas seulement à répondre par une procédure stéréotypée à un signal préétabli ». Dans ce type d'approche, dite approche par les compétences, la qualité de l'évaluation serait assurée si l'évaluation permettait bien d'évaluer des compétences. C'est une préoccupation dont se font écho De Ketele et Gérard (2005). Ces auteurs défendent « avec force, le principe que la pertinence [la pertinence est pour ces mêmes auteurs (opus cité) le caractère plus ou moins approprié de l'épreuve, selon qu'elle s'inscrit dans la ligne des objectifs visés] est première par rapport à la validité et à la fiabilité ». Pour ces derniers, évaluer dans le cadre d'une approche par les compétences consiste à, ou du moins aurait pour finalité, évaluer de véritables compétences au moyen de situations dites complexes.

Du côté des enseignants, cette nouvelle approche par les compétences « nourrit la confusion » (O. Rey, 2012). Endrizzi et O. Rey (2008) soulignent que, « dans le vocabulaire courant de l'évaluation, le terme de compétence peut aussi bien servir à qualifier une connaissance ou un savoir-faire qu'à désigner une approche complexe et intégrée qui s'oppose à la décomposition en objectifs détaillés et fragmentés ». Cela conduit certains auteurs dont Amigues (2004) à dénoncer l'existence d'un « effet de réalité » à leur endroit : « dès lors que l'on peut les nommer, les mesurer, les compétences existent » (idem).

La définition de Gillet (1991) reprise par Allal (2012, p. 79) diffère de celle proposée par B. Rey et coll. (2006) et permet de déplacer le débat sur le fait de savoir si on évalue des compétences à un débat sur la nature des situations qui permettent de le faire. Pour Gillet (1991), une compétence est « un système de connaissances, conceptuelles et procédurales, organisées en schémas opératoires et qui permettent, à l'intérieur d'une famille de situations, l'identification d'une tâche-problème et sa résolution par une action efficace » (idem). Évaluer une compétence signifie alors, selon Allal, « évaluer tout à la fois la qualité du système de ressource disponible et la faculté que le sujet compétent le mobilise à bon escient dans une situation « problème » relevant d'une famille de situations » (idem).

Cette définition ne mentionne pas l'idée de complexité et surtout d'inédit contenue dans la définition de la Communauté Francophone de Belgique précitée, reprise par B. Rey et coll. (2006) et critiquée par Crahay (2006) qui estime qu'une telle définition de la compétence est problématique car elle renvoie à « une norme de complexité inédite ». Crahay considère que l'approche par compétences s'attaque à un vrai problème – celui de la mobilisation des connaissances en situation de problème – « mais propose une solution bancale. Rejetant l'entrée par le disciplinaire, cette approche se confronte à des questions épineuses sinon impossibles : parmi celles-ci, on trouve la notion de familles de situations. En conclusion, l'auteur suggère d'oublier la notion de compétence pour repenser celle de l'apprentissage » (idem).

Cette préoccupation est partagée par Perrenoud (2011, p. 11). Ouvrant une réflexion sur ce à quoi l'école prétend préparer les élèves, il questionne le fait de savoir si l'on peut développer des compétences sans interroger les connaissances, voire, comme le font Jonnaert, Ettayebi & Defise (2009), si l'on peut développer des compétences sans écrire des programmes privilégiant des entrées par familles de situations. Or, les programmes de 2008 ne sont pas rédigés d'une telle manière. Une telle détermination de familles de situations est à ce jour inexistante pour tous les champs disciplinaires, ce qui constitue un problème central des approches par les compétences : Chenu, Crahay et Lafontaine (2014) critiquant la faiblesse de la notion de famille de situations, indiquent que « la possibilité de regrouper des situations en familles sur une base objective semble bien discutable ». B. Rey (2011, p.43) rejoint ici Crahay : « il n'existe pas de situation objective qui serait, dans les faits, délimitée par l'espace et le temps et qui comporterait des caractéristiques déterminées ».

S'il l'on est conduit à s'interroger sur ce qui constitue véritablement un enseignement de compétences, on peut concevoir dans le même temps qu'évaluer des compétences pose donc un certain nombre de problèmes non encore résolus. O. Rey (2012) souligne ainsi que « la généralisation des livrets de compétence induit des questions concrètes qui font souvent de l'évaluation le vrai point d'entrée des compétences dans les classes ». L'une des questions concrètes que se posent les enseignants est celle de la détermination de la compétence et de la différence qu'il convient de faire avec les objectifs : « on ne voit pas tellement la différence » dit l'un des enseignants au cours d'un entretien post-collecte. Pour Carette (2009), « les objectifs dans une approche par les objectifs et les compétences dans une approche par les compétences sont les points d'entrée pour les pratiques pédagogiques. Ils sont également, d'une certaine manière, les points de sortie, car au terme des apprentissages, on se prononcera sur la maîtrise des apprentissages des élèves au regard de ces référents ». La différence provient, pour cet auteur, du savoir-mobiliser qui constitue un objectif central d'une pédagogie par les compétences.

Il est nécessaire d'apporter quelques précisions sur les termes « connaissances », « savoirs », « compétences », « capacités » et « habiletés » fréquemment confondus par les enseignants, et sur les liens qui les unissent. Leplat et de Montmolin (2001, p.43) indiquent que « ces termes sont souvent expliqués l'un par l'autre et interchangeable ; ils s'éclairent aussi mutuellement ». Nous rappelons ainsi brièvement que l'on entend généralement par « savoir » le corps de connaissances historiquement élaboré, stabilisé et validé socialement, et diffusé par diverses institutions dont particulièrement l'institution scolaire qui transmet des savoirs classiques (au sens de ce « qu'on enseigne dans les classes », selon le dictionnaire Le Robert). L'école transmet des savoirs qui ne sont ni des savoirs savants ni des savoirs simplifiés, mais des savoirs scolaires spécifiquement reconstruits et présentés comme tels pour être transmis. La théorie de la transposition didactique de Chevallard (1985) rend compte de ce processus de transition, de fabrication et de transmission de contenus scolaires. On ne peut ainsi confondre savoir et connaissance dans la mesure où chaque individu construit des connaissances en interagissant avec une situation ou un objet de savoir. Compétence et connaissance sont intimement liées de sorte que la compétence semble se situer plutôt du côté du rapport au savoir que du savoir lui-même. Vergnaud⁵ estime d'ailleurs que « la plus grande partie de nos connaissances se situent dans nos compétences, souvent de manière implicite, voire inconsciente ». Cela renvoie à la faculté de se servir du savoir pour penser et agir, inhérente au savoir même mais discutée lorsqu'il

⁵ Repéré à http://pedagogpsy.eu/competences_vergnaud.html

s'agit des savoirs scolaires qui ne sont « ni théoriques ni pratiques » (cf. Perrenoud, 2011, p.63)

Perrenoud (2011, p. 66) propose de distinguer les habiletés des compétences dans une formule condensée : « les habiletés sont des ressources au service de compétences plus globales, donc chacune vise la maîtrise de toutes les composantes des situations appartenant à la même famille ». L'auteur rapproche ainsi l'habileté (que l'on peut assimiler au terme capacité) du savoir-faire et précise que « l'important est de ne pas confondre une habileté avec la compétence qui la mobilise, même lorsque cette habileté est une ressource majeure, même lorsqu'elle donne son nom à la tâche, à la situation ou à la compétence elle-même » (id. p. 68). Dans une perspective d'unification conceptuelle de ces différents concepts au concept de compétences, B. Rey et coll. (2006, p.26) estiment que les habiletés constituent les compétences élémentaires (ces auteurs distinguent trois degrés de compétence : « compétence élémentaire », « compétence élémentaire avec cadrage », « compétence complexe »). Cela contribue à répondre à la critique de la complexité érigée en norme formulée par Crahay (2006) (cf. Carette et B. Rey répondent (2010, p. 89)). Dans le cadre de notre étude, on peut penser que les enseignants évaluent davantage des habiletés, au sens large adopté par Perrenoud, que des compétences.

Vergnaud (idem) préfère distinguer deux formes de la connaissance : une forme opératoire et une forme prédicative. « La première est celle qui permet d'agir en situation. Elle ne s'oppose pas aux connaissances académiques classiquement transmises par l'école et l'université, mais il existe un décalage parfois impressionnant entre ce qu'une personne peut faire en situation, et ce qu'elle est capable d'en dire. La forme opératoire de la connaissance est en général plus riche, plus subtile, que « la forme prédicative », celle qui énonce les propriétés et les relations des objets de pensée ». La distinction établie entre forme opératoire et prédicative de la connaissance renvoie directement aux situations dans lesquelles la connaissance est mobilisée. Cette préoccupation se retrouve dans l'énoncé de la distinction qu'établit Perrenoud (2011) entre capacité et compétence, lequel propose la convention suivante, illustrant bien la difficulté qu'il y a à cerner précisément la compétence dans le cadre de l'apprentissage scolaire :

- *« lorsqu'on désigne des opérations qui ne se réfèrent pas à une famille identifiée de situations, on parlera de capacités, désignant alors des potentialités relativement indépendantes des contextes, sachant que leur mise en œuvre et leur efficacité peuvent dépendre de leur contextualisation et de leur orchestration avec d'autres ressources ; une capacité, dans cette acception, ne prétend pas gérer la globalité d'une situation, elle n'est qu'un ingrédient, une ressource parmi d'autres ;*
- *lorsqu'on désigne des actions prenant en compte la globalité d'une situation appartenant à une famille identifiée, ce qui amène le sujet à mobiliser de multiples ressources, dont des capacités, on parlera de compétence. »*

En résumé, que l'on s'intéresse à des connaissances ou des compétences, la question centrale est celle des situations dans lesquelles celles-ci sont mobilisées, et des tâches pour évaluer. « Que ce soit des compétences ou des connaissances, on ne peut pas définir les apprentissages évalués indépendamment des tâches et plus généralement des situations et contextes, qui, pour ces derniers, contribuent également à définir les attentes évaluatives » souligne ainsi Mottier Lopez (2015) qui rejoint Vergnaud (2001) pour qui « la compétence conduit à nous intéresser à l'activité elle-même, et pas seulement à son résultat. [...] ». Une

préoccupation au cœur des approches psycho-didactiques (cf. Vantourout et Goasdoué, 2014).

Pour Carette et Rey (2010, p. 73), « la compétence ne se définit pas par les processus psychologiques ou physiologiques qui doivent être activés dans l'exercice de la compétence, mais bien par la tâche ou la gamme de tâche qu'elle permet d'accomplir ».

Le lien entre la compétence et la tâche constitue un point central de notre travail. Nous reviendrons sur les notions de prescrit et de réalisé dans la partie III mais, pour éclairer le débat sur la compétence nous suivons Leplat et de Montmolin (2001, p. 45), qui utilisant ces concepts que Leplat a contribué à définir dans le champ de la psychologie ergonomique, souligne que la correspondance entre tâche et compétence n'est pas triviale. Comme le font ces auteurs, on peut « distinguer la tâche prescrite et la tâche effective », « la compétence supposée correspondre à la tâche prescrite et la compétence réelle correspondant à la tâche effective » et, « l'activité supposée par les prescriptions et l'activité effective » (idem)

Pour ces auteurs, « la traduction de la tâche en terme de compétence revêt parfois un caractère artificiel » (idem). Cet éclairage permet de dire que la compétence ne se définit pas seulement par la tâche. Leplat précise d'ailleurs que « de mêmes exigences de la tâche peuvent être satisfaites avec des compétences différentes », ce qui est « d'autant plus vrai que la tâche est complexe ». (Idem).

Pour analyser la compétence précisent Leplat et de Montmolin (2001, p.56), il faut disposer d'un modèle qui définit ses composants et leur organisation. Les auteurs renvoient au modèle de Rogalski (1995) et dégagent des éléments essentiels de la compétence. Les connaissances concernant le monde du travail mises en jeu dans l'activité constituent l'un de ces éléments. Rapporté à notre domaine d'étude, cela correspondrait aux connaissances enseignées. Les connaissances concernant le mode d'exécution des tâches constituent un autre élément : « connaissances pragmatiques, connaissances contextualisés, procédures ou heuristiques propres à traiter les problèmes susceptibles de se présenter » (idem). L'expérience constitue un autre élément important elle permet d'exploiter le raisonnement analogique etc. La compétence ne peut ainsi se résumer à la tâche.

Les données collectées ne nous permettent pas d'analyser l'activité évaluative de l'enseignant ni l'activité de l'élève en situation de réponse. Nous nous intéressons en revanche à l'analyse des tâches, en relation avec le référentiel construit par l'enseignant. Particulièrement, nous examinons la ou les tâches induites par les épreuves d'évaluation en émettant des hypothèses sur leur potentiel à entraîner une activité de réponse correspondant à la compétence visée.

Vergnaud (1991) estime que l'on peut distinguer deux grandes classes de situations suivant les connaissances accessibles à l'élève. Il distingue :

- « des classes de situations pour lesquelles le sujet dispose dans son répertoire, à un moment donné de son développement et sous certaines circonstances, des compétences nécessaires au traitement immédiat de la situation ;
- des classes de situations pour lesquelles le sujet ne dispose pas de toutes les compétences nécessaires, ce qui l'oblige à un temps de réflexion et d'exploration, à des hésitations, à des tentatives avortées, et le conduit éventuellement à la réussite, éventuellement à l'échec. »

La distinction opérée par cet auteur pose clairement que l'appartenance d'une situation à l'une des deux classes n'est pas une caractéristique propre à la situation elle-même mais

qu'elle est déterminée par la disponibilité des compétences permettant à l'élève de traiter la situation. La même situation peut ainsi appartenir à l'une ou l'autre classe suivant les élèves.

Cette approche centrée sur l'élève diffère de celle de Rey et de ses collaborateurs (2006, p.26) qui distinguent des degrés de compétence (« compétence élémentaire », « compétence élémentaire avec cadrage », « compétence complexe »), mais qui les opérationnalisent en fournissant des situations découpées en fonction de niveaux de complexité de la tâche préalablement déterminés. Les épreuves d'évaluation comprennent systématiquement trois phases pour ces auteurs (2006, p.46) : une tâche complexe, une tâche complexe découpée en tâches élémentaires, et enfin une série de tâches simples décontextualisés. Cette approche pose plusieurs questions non résolues. Tout d'abord celle du caractère aidant des décompositions que Dierendonck et Fagnant (2014) estiment « variable suivant la tâche complexe considérée », puis, celui du « danger », non seulement en termes de fiabilité de la mesure mais aussi sur le plan pédagogique, d'évaluer la compétence (..) des élèves uniquement au départ de quelques tâches complexes de troisième degré » (idem).

Cette réflexion conduit à mettre en avant la nécessité de construire des cadres de référence d'ordre didactique pour analyser l'évaluation, et d'ordre psychologique et didactique pour analyser l'activité de l'élève au cours de la réalisation de la tâche. Pour Vergnaud (2001) en effet :

« La complexité n'est pas que dans le faire, elle est aussi dans le dire. L'énonciation des objets et de leurs propriétés est essentielle dans les processus de conceptualisation. Parmi les difficultés rencontrées par les élèves dans l'apprentissage des mathématiques, on peut mettre presque à égalité d'une part la complexité des classes de problèmes à résoudre et des opérations de pensée nécessaires pour les traiter, et d'autre part la complexité de certains énoncés et de certains symbolismes mathématiques. Que l'expérience soit incontournable dans la formation de la compétence n'est pas une découverte. Ce qui est plus nouveau c'est l'analyse cognitive de l'activité, c'est-à-dire l'identification des formes stables d'organisation de l'activité (les schèmes), face à des situations d'une certaine variété; ainsi que l'analyse des relations entre elles de différentes situations, qu'on peut ainsi classer. Ce sont ces différentes classes de situations et le réseau des concepts qui permettent de les traiter que j'appelle champ conceptuel. Si les schèmes ne sont pas des stéréotypes, mais sont au contraire flexibles, c'est justement parce qu'ils reposent sur des invariants opératoires (conceptuels et propositionnels) qui tirent leur sens et leur pertinence d'une variété de situations. Les enseignants sont des médiateurs. Leur premier acte de médiation est le choix des situations à proposer aux élèves. Bref, il lui faut s'intéresser aux différentes composantes du schème qu'il souhaite voir émerger chez les élèves ».

L'approche par les compétences à laquelle se réfèrent les programmes et qui constitue le cadre de référence de l'enseignant pour évaluer implique, si l'on se réfère à Vergnaud, une dimension « situation » et une dimension « activité des élèves au cours de l'activité de réponse » impensables hors de cadres d'analyse psychologiques et/ou didactiques. En effet, au-delà de la complexité, qui renvoie aux ressources à mobiliser, il est impératif souligne Scallon « de savoir quelles ressources l'élève devra mobiliser pour mener à bien la tâche, quels savoir et/ou savoir-faire il devra puiser de lui-même dans son répertoire de

ressources les plus diverses tout en capitalisant sur ses attitudes et ses valeurs » (2004, p.156).

Dans la théorie des situations didactiques en mathématiques (TSDM) de Brousseau (1998), la situation occupe une place centrale : selon cet auteur, « les conditions d'une des utilisations particulières d'une connaissance mathématique sont considérées comme formant un système appelé « situation » ». Elle occupe aussi une place importante dans la théorie des champs conceptuels de Vergnaud à laquelle nous nous référerons pour catégoriser les problèmes utilisés dans les épreuves d'évaluation. Cette approche qui, selon Brousseau lui-même, est tout à fait « compatible avec la TSDM » (2007, p.54), décrit très bien sur le plan cognitivo-didactique, les champs conceptuels des relations additives et celui des relations multiplicatives. Ces cadres théoriques, pensés en didactique des mathématiques, n'offrent pas d'équivalent en didactique du français. On ne trouve pas, par exemple, de description de classes et de familles de situations en lecture.

La demande d'évaluer des compétences renvoie donc à la détermination de l'objet à évaluer, à l'analyse des tâches pour évaluer, à l'analyse de l'activité de l'élève au cours de l'effectuation de la tâche. L'analyse des tâches peut être ainsi conduite selon plusieurs directions comme le propose Vergnaud dans la citation précédente de 2001. L'analyse didactique de la tâche constitue l'une des directions que l'analyse peut prendre, l'analyse de la complexité de la tâche en constitue une autre. Cela constituera une dimension importante et exploratoire de notre travail. Nous détaillerons dans chacune des parties III et IV les référentiels que nous utiliserons pour permettre d'analyser les évaluations collectées.

2.1.2. Les livrets scolaires : des produits de processus de référentialisation

Nous avons précédemment souligné que l'évaluation orientait l'enseignant vers une activité de communication du jugement évaluatif au moyen d'un livret scolaire. Les livrets scolaires occupent une place tout à fait singulière et finalement peu prise en compte dans les travaux relatifs à l'évaluation. On peut pourtant les étudier en ce qu'ils sont un produit du processus de référentialisation de l'évaluation de l'enseignant : ce sera l'objet d'une partie de notre travail (Partie II). Sans entrer dans les détails, il est important de considérer, comme le fait un syndicat d'enseignants, que « le modèle de livret scolaire n'est pas imposé. Il est souvent le fruit d'un travail de cycle ou d'école à partir des programmes »⁶. La question du livret est donc une question importante pour les enseignants.

Dans le cadre des programmes de 2008, une circulaire⁷ précise la mise en œuvre de ce livret : il est conçu pour être « un instrument de liaison entre les maîtres ainsi qu'entre l'école et les parents ». Le livret scolaire est ainsi adressé à plusieurs destinataires en même temps : les professionnels d'une part, les familles d'autre part. Cela constitue une caractéristique majeure de cet outil : selon le destinataire privilégié par l'enseignant qui remplit le livret (ses collègues, les parents voire l'élève lui-même) et suivant le niveau présumé de compréhension du récepteur du livret, l'énonciation du jugement évaluatif est susceptible de varier dans son expression. Il est concevable que certains livrets listent l'ensemble des objectifs à atteindre et positionnent les réussites ou échecs de l'élève en fonction de ceux-ci quand d'autres les regroupent pour une communication plus fluide et plus synthétique. Le livret étant un outil de communication, il faut considérer que dans un certain nombre de cas (c'est un cas fréquent dans les environnements sociaux éloignés des codes

⁶ Repéré à <http://95.snuipp.fr/neovo/spip.php?article87>

⁷ Circulaire n° 2008-155 du 24-11-2008

de l'école), les enseignants reçoivent individuellement les parents pour le commenter, voire pour commenter directement le jugement évaluatif prononcé suite à l'analyse des résultats obtenus par les élèves aux épreuves. Ce cas de figure permet de considérer que le livret scolaire peut ne pas être considéré par l'enseignant comme le meilleur outil de communication avec les parents : la communication orale étant plus accessible et certainement plus efficace. Dans ce cas également, le fait de remplir le livret scolaire apparaît comme vidé d'une partie de sa substance et peut donc être considéré par les enseignants comme une simple contrainte institutionnelle. Dans la mesure où la circulaire précitée rappelle l'obligation de produire un livret scolaire, nous nous intéressons à la constitution des livrets en ce qu'ils sont des construits d'enseignants qui organisent les résultats de l'évaluation d'une manière communicable en fonction d'un certain nombre d'objectifs déterminés par les programmes et d'un certain nombre d'objectifs de communication propres à chaque enseignant. Ils résultent du processus de référentialisation de l'évaluation évoqué plus haut.

Pour résumé, le référentiel de l'enseignant pour évaluer est constitué d'éléments tangibles provenant de l'institution scolaire (les programmes en tout premier rang), lesquels ont des implications didactiques sur l'évaluation des connaissances acquises par les élèves réalisées au moyen de situations d'évaluation. Le référent est constitué d'autres éléments moins accessibles que nous ne pouvons décrire au moyen des données collectées et qui nécessiteraient des cadres d'analyse de l'activité. Le livret scolaire, manifestation explicite du processus de référentialisation de l'enseignant pour évaluer, constitue pour chaque enseignant le cadre opérationnel, le référentiel formel, de l'évaluation. Il indique ce en fonction de quoi l'enseignant évalue et, ce en fonction de quoi il organise les résultats de l'évaluation qui sont communiqués aux parents. Les épreuves d'évaluation s'inscrivent dans ce cadre : leur usage répond à la nécessité de correspondre à un point précis (une compétence) du référentiel.

2.2. Référé

Barbier (1985, p.65) définit le référé comme « appartenant à l'ordre des représentations sur les faits ». Le référé de l'évaluation n'est pas la réalité : c'est une représentation finalisée de l'objet d'évaluation. Il correspond à « ce à partir de quoi un jugement de valeur est porté ». Pour l'enseignant, le référé de l'évaluation est élaboré à partir des réponses des élèves à une épreuve.

Comme Vantourout (2004) et Vantourout & Goasdoué (2014), Mottier Lopez (2015) estime que « l'apprentissage de l'élève ne peut s'appréhender – et donc s'évaluer – que dans la saisie de l'activité de l'élève en interaction avec la tâche, elle-même saisie dans des situations et réseaux de significations plus larges ».

Cela souligne la nécessité de prendre en compte l'activité de l'élève, ce qui sous-entend la nécessité de considérer ce qu'il fait au moment où il le fait, et non pas les seules traces de son activité. Nous développerons la distinction entre tâche et activité dans la partie III de la thèse : la prise en compte de l'activité de l'élève est au cœur des approches psychodidactiques de l'évaluation. Nous indiquons simplement à ce stade que nous nous intéresserons à l'analyse de traces de l'activité d'élèves, que nous rapprocherons de l'activité supposée visée par la tâche (ou les tâches) d'évaluation proposée(s) par l'enseignant.

2.3. Référentiel et référé pour le chercheur

Notre travail s'intéresse aux livrets scolaires utilisés par les enseignants pour communiquer les résultats de l'évaluation. Ces livrets constituent une trace du processus de référentialisation entrepris par chaque enseignant à partir du référentiel formel que constituent les programmes. Ils deviennent à leur tour le référentiel formel de l'enseignant pour évaluer. Ils sont une partie du référé que le chercheur élabore pour étudier l'évaluation. Notre travail s'intéresse également aux épreuves d'évaluation. Elles sont des traductions des compétences du référentiel élaboré par l'enseignant en termes de tâches prescrites aux élèves. Elles constituent une autre partie du référé que le chercheur construit pour l'analyse. Rapporté à notre étude, le référé est ainsi constitué des épreuves et des livrets collectés.

Plusieurs types de référentiels peuvent être considérés et construits pour analyser les référés. Ils dépendent de la qualité du lien référé/référent que l'on étudie et du point de vue adopté sur l'évaluation :

Une façon de questionner l'évaluation se situe du côté de la conformité de l'évaluation au contrat passé par l'institution scolaire avec les familles. Ce contrat peut se traduire par « l'école doit évaluer les acquis des élèves conformément aux programmes ». Pour répondre, il est nécessaire d'analyser le lien de conformité entre le référentiel « alpha » (les progressions) et le référentiel formel élaboré par l'enseignant à l'issue du processus de référentialisation, mais aussi le lien de conformité entre les épreuves et les items du livret de chaque enseignant.

Un point de vue didactique peut être adopté. Il conduit à s'intéresser à la qualité didactique des tests et des épreuves, indépendamment du fait de savoir si l'évaluation couvre les programmes. Pour répondre aux questions concernant la qualité des épreuves et la qualité des tests, il est nécessaire de disposer d'un référentiel d'ordre didactique pour analyser les épreuves, ou de plusieurs en fonction des domaines disciplinaires étudiés.

Ces interrogations renvoient à la question centrale de notre recherche : les épreuves évaluent-elles bien ce pourquoi elles sont employées ?

La validité des épreuves, i.e. leur capacité à évaluer ce pourquoi elles sont employées, constitue l'objet central du lien référé(référentiel) auquel nous nous intéressons dans ce travail. Plusieurs types de validité peuvent être considérés.

Chapitre 3. La validité : une qualité plurielle de l'évaluation

Comme le souligne Loyer (2012), la validité et la fiabilité (ou fidélité) d'un test psychométrique constituent des exigences de qualité spécifiques à une évaluation « objective ». La fidélité d'un test concerne la capacité à mesurer un construit de manière cohérente. La validité a trait à la capacité du test à évaluer ce pour quoi il a été utilisé, le construit mesuré.

Dans le domaine des apprentissages scolaires, certains auteurs, tels De Ketele et Gérard (2005) font état d'un troisième critère, de nature pédagogique, la pertinence. Il correspond, dans le cadre des approches par les compétences, à vérifier que l'on évalue bien des compétences. Pour ces auteurs (idem), trois conditions sont ainsi nécessaires pour une évaluation de qualité dans le domaine:

«- 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 & coll., 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 et 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. »

Nous nous intéressons dans notre travail à la notion de validité : les tests, les épreuves permettent-elles bien d'évaluer ce pourquoi ils sont utilisés ?

Les questions de validité sont classiques lorsque l'on s'intéresse à un test et un certain nombre de types de preuves de validité (du test) peuvent être apportées. Elles ont été caractérisées par Laveault et Grégoire (2014, p. 165) et renvoient au contenu du test (relevant d'une formalisation par des experts), aux processus de réponse (adéquation des caractéristiques visées par le test et celles effectivement mises en œuvre par les répondants), à la structure du test (adéquation entre les items et les composants du test), aux relations avec d'autres variables (externes au test) et enfin, aux conséquences du testing.

L'évaluation que nous étudions peut être grossièrement rapprochée d'un test si l'on considère que toutes les épreuves utilisées par un enseignant pour évaluer les connaissances acquises de ses élèves à un trimestre donné forment un ensemble. Chaque ensemble d'épreuves permet à l'enseignant d'appréhender des caractéristiques pédagogiques d'ordre essentiellement cognitif. Les tests que nous étudions dans notre travail sont des tests forts éloignés de ce que l'on désigne usuellement par ce terme en psychométrie : ils ne permettent pas d'exploitation des scores selon une méthodologie statistique. Les enseignants ne s'intéressent pas non plus à l'étude des qualités métriques

ou techniques des items lors de la phase d'élaboration de leur test. Enfin, les tests ne font pas l'objet de pré-tests permettant d'affiner la qualité des items.

Ces précisions étant apportées, on peut toutefois s'intéresser à la validité de chaque ensemble d'épreuves dès lors assimilé à un test, et rechercher des preuves de validité (ou de non-validité) particulièrement en ce qui concerne le contenu du test, à l'exclusion, dans notre travail, de la recherche d'autres types de preuves. Les tests étudiés ne donnent pas lieu à des calculs de corrélation entre les différents items du test. De ce fait, il est inutile de rechercher des preuves de validité liées à la structure interne du test tout comme il est inutile de rechercher des preuves de validité basées sur les relations avec d'autres variables. Nous n'étudions pas dans le cadre de ce travail les conséquences du testing (sur les pratiques enseignantes par exemple). Enfin, pour s'assurer que l'élève mobilise effectivement les processus attendus pour produire la réponse, il serait nécessaire de se doter d'une méthodologie qui permette d'observer le processus de réponse de chacun des élèves. Cette méthodologie, dont la mise en œuvre est assez lourde, n'a pas été adoptée pour collecter nos données.

3.1. Validité de contenu des tests

Demeuse & Henry (2004, p. 164) indiquent que la validité de contenu, « comme son nom l'indique, repose sur la nature du contenu du test par rapport à l'objet à mesurer. Dans ce type de validité, le contenu du test doit donc être en rapport direct avec ce que le test est supposé mesurer. »

Deux points de vue sur le contenu peuvent être adoptés. Ils dépendent du référentiel choisi et ce que l'on entend par « ce que le test est supposé évaluer ». Le premier point de vue se réfère aux programmes. Le contenu du test est-il en adéquation avec le contenu des programmes, i.e. avec ce que les enseignants doivent avoir enseigné (et, supposément, avec ce que les élèves doivent avoir acquis) ? Le second point de vue se réfère aux didactiques. Le contenu du test couvre-t-il ce que les didactiques concernées considèreraient comme étant les principaux composants du domaine disciplinaire?

3.1.1. Validité curriculaire

Comme nous l'avons dit, pour l'enseignant, l'objet à mesurer est déterminé en référence aux programmes qui constituent le référentiel de l'évaluation à partir duquel chaque enseignant élabore, au terme d'un processus de référentialisation, son propre référentiel pour évaluer : le livret scolaire. A supposer que l'évaluation doive couvrir les programmes, s'intéresser à la validité de contenu du test invite à étudier si celui-ci couvre bien les programmes. Cette validité est qualifiée de validité curriculaire par De Landsheere (1988).

Précisons l'existence d'une différence de terminologie entre Scallon et De Landsheere au niveau du terme validité curriculaire. Pour Scallon (2004/2009, p.241), la recherche de validité curriculaire renvoie au besoin « d'établir que le contenu d'un examen, ou d'un procédé de vérification, correspond bien à ce qui a été enseigné ». De Landsheere (1988) qualifie plutôt cette validité de validité pédagogique i.e. portant sur les contenus réellement enseignés (cela fait écho à la distinction effectuée entre le curriculum officiel et le curriculum réellement implanté dans les classes cf. Crahay, 1996, p. 39). Dans notre travail nous utiliserons le terme de validité curriculaire au sens de De Landsheere.

3.1.2. Validité de couverture (didactique) des domaines

Le test a pour objectif d'évaluer la maîtrise de savoirs enseignés. La validité des épreuves d'évaluation fondée sur les contenus ou sur les objectifs se caractérise, selon De Ketele et Dufays (2003), par le « caractère représentatif de l'échantillon d'items par rapport à l'univers de référence ». Comme le souligne De Landsheere au sujet de la validité de contenu, il faut que chaque item du test corresponde au domaine évalué, et surtout que « l'ensemble des items couvre toutes les composantes⁸ du domaine » (1988, p.81).

Apporter des preuves de validité de contenu des tests nécessite de vérifier que l'ensemble des principaux composants des domaines (lesquels sont représentatifs du domaine) est bien évalué.

Grugeon-Allys et ses collaborateurs se sont intéressés à cette question au sujet de l'évaluation externe en mathématiques. Ils ont développé un outil diagnostique dans le champ conceptuel de l'algèbre élémentaire en fin de scolarité obligatoire. En référence à la théorie anthropologique du didactique, « chaque tâche est caractérisée par : un type de tâche, les registres sémiotiques et la complexité des objets algébriques mis en jeu, le niveau d'intervention des organisations mathématiques dans la tâche, des techniques attendues relativement aux éléments technologiques et théoriques visés » (Grugeon-Allys et coll., 2012). La validité à laquelle s'intéressent ces chercheurs est donc une validité de contenu, pensée *a priori* à partir d'une analyse épistémologique et didactique d'un champ de savoir.

Cette approche « épistémo-didactique » de la validité de contenu de l'évaluation doit être aménagée pour pouvoir être utilisée dans le domaine de la lecture en Français. Dans ce dernier champ, nous ne disposons pas d'une épistémologie mais d'une conjonction de plusieurs épistémologies relatives aux champs concernés par la lecture (la linguistique, la psychologie cognitive, la narratologie...).

Deux notions sont à prendre en compte pour approcher la possibilité de déterminer les domaines concernant la lecture et les nombres et le calcul auxquels nous nous intéressons dans notre travail : la transposition didactique de savoirs savants d'une part, et la notion de pratiques sociales de référence d'autre part. En effet, si l'existence de savoirs savants (i.e. reconnus socialement comme tels) ne fait aucun doute dans certaines disciplines scolaires et contribue à la légitimité de celles-ci (en mathématiques par exemple), ils font défaut dans d'autres disciplines et/ou ne suffisent pas à la décrire entièrement (c'est le cas des disciplines linguistiques ou artistiques, éducation sportive ...). Ces deux notions se rejoignent et constituent selon Perrenoud (1998) les « deux sources de la transposition didactique : d'une part des savoirs, savants ou experts, d'autre part des pratiques sociales ».

Dans la théorie anthropologique du didactique de Chevallard (1985), le didactique est considéré comme un domaine de réalité constitutif d'un univers culturel. Ce domaine de réalité est contemporain de la production d'un savoir de ce domaine de réalité lequel est, pour cet auteur, empreint d'épistémologie : « une institution didactique relativement à un savoir X reconnaît une ou plusieurs formations épistémologiques de référence, définissant pour elle des savoirs X référentiels. Parmi ces savoirs, le savoir de type savant Xs occupe une position dominante dans le système des références » (Chevallard, 1988). La transposition didactique résulte, toujours selon Chevallard, du fait qu'un savoir

⁸ Nous utilisons le terme masculin dans notre travail.

(épistémologiquement fondé) est transposé dans une institution en vue d'y être étudié. C'est, dit-il, « le passage d'un contenu de savoir précis à une version didactique de cet objet de savoir » (1991, p.38).

Ce concept, importé en didactique (des mathématiques) par cet auteur (1985, 1991) qui l'emprunte au sociologue Michel Verret (1975)⁹, est, selon Garcia-Debanc (2008), « de nature à permettre de penser la possibilité d'une science didactique ».

L'objet d'étude des didactiques consiste, selon ce point de vue, en une analyse des écarts qui permettent de rendre compte des passages d'un contenu de savoir précis à une version didactique de cet objet de savoir.

Ce cadre didactique renvoie à la possibilité de déterminer des savoirs précis. Or, ce que souligne Mercier (2008), « certains objets d'enseignement n'ont pas de statut scientifique mais répondent seulement à un enjeu social ». S'il n'existe pas de possibilité d'étudier l'écart précédemment évoqué entre un contenu de savoir précis et une version didactique de ce savoir en ce qui concerne la lecture, on peut probablement analyser la version didactique d'une pratique de référence (un savoir-lire) laquelle est éventuellement nourrie de savoirs savants (lesquels renvoient aussi à des pratiques sociales comme Chervel (1988) l'indique pour la grammaire. Pour le savoir-lire, un savoir provenant de la narratologie (« le schéma quinaire ») peut être convoqué lorsqu'il s'agit de lire un récit).

La notion de pratique sociale de référence a été définie par Jean-Louis Martinand dans les années 1980 à propos de la technologie et de l'informatique, mais « elle convient aussi aux disciplines linguistiques ou artistiques, à l'éducation physique et aux formations professionnelles » précise Perrenoud (1998). Martinand (1986) explique que les pratiques sociales (i.e. les activités réelles d'un groupe social identifié) peuvent servir de référence pour la conception ou l'analyse d'activités scolaires. Cette notion permet de faire un pas de côté par rapport à la transposition didactique (de savoirs savants). La lecture enseignée à l'école, que nous étudierons dans cette thèse, renvoie à plusieurs disciplines de recherche constituées (linguistiques, psychologie(s), histoire...) et séparées, disposant chacune d'épistémologies, et renvoyant éventuellement à plusieurs pratiques sociales de référence.

Si, dans le domaine des nombres et du calcul une analyse épistémologique et didactique peut être conduite au sujet des tests tels que nous les définissons, cela n'est pas transposable au domaine de la lecture laquelle n'est pas décrite par une discipline de référence unique disposant elle-même d'une épistémologie mais par plusieurs comme nous l'avons déjà dit.

De ce fait, chercher à analyser la validité de contenu de l'évaluation nécessite au préalable de déterminer des cadres de référence épistémologiques et/ou didactiques permettant de le faire.

⁹ Selon Perrenoud (1998) « Verret (1975) cherchait, en sociologue, à désigner un phénomène qui dépasse l'école et les disciplines d'enseignement. Il s'intéressait à la façon dont toute action humaine qui vise la transmission de savoirs est amenée à les apprêter, à les mettre en forme pour les rendre "enseignables" et susceptibles d'être appris. La transposition didactique passe, selon Verret, par des transformations plus radicales. Il en décrit cinq :

-La désynchronisation du savoir, autrement dit sa structuration en champs et domaines distincts. Les savoirs savants sont déjà organisés en disciplines, mais on ne trouve pas l'équivalent pour les autres savoirs humains.

-La dépersonnalisation du savoir, qui le détache des individus et des groupes qui le produisent ou s'en servent.

-Une programmation, qui tient au fait qu'un savoir étendu ne peut être assimilé en une fois et passe par un chemin de formation balisé.

-Une publicité du savoir, qui trouve son achèvement dans les référentiels et les programmes qui permettent à chacun de saisir sur quoi porte l'intention d'instruire (Hameline, 1971).

-Un contrôle des acquisitions ».

Nous nous sommes intéressé à la validité de contenu du test en indiquant qu'elle pouvait être abordée du point de vue épistémologique et/ou didactique, et du point de vue curriculaire. Cette préoccupation centrale dans toute approche d'un test laisse toutefois de côté un aspect que nous estimons primordial car, préviennent Bastien et Bastien-Tonniazzo (2004, p.21), « les tests d'évaluation sont parfois trompeurs ». Ces auteurs ont en effet montré l'importance de prendre en compte les composantes des tâches et les conditions de réalisation de celles-ci, parce que la réussite à une tâche ne dépend pas seulement du contenu conceptuel supposé de la tâche et donc du savoir acquis par l'élève.

3.2. Validité psycho-didactique des épreuves

Nous avons abordé jusqu'ici la question de la validité sans nous intéresser aux tâches elles-mêmes. C'est pourtant un questionnement que certains didacticiens ont pris en charge à la naissance même des didactiques, sans le formuler en termes de validité, et que nous estimons fondamental pour conduire une analyse de la validité de l'évaluation. Maury a ainsi montré, en didactique des mathématiques, que « de nombreux facteurs autres que ceux directement liés à la structure mathématique du problème sont susceptibles d'influencer les comportements des élèves » (Maury, 1984) et, que « de légères modifications de la présentation d'un problème sont susceptibles d'avoir une influence sur les conduites des élèves » (Maury, 1985).

Nous consacrons un chapitre à déterminer un cadre conceptuel d'analyse de la tâche (cf. partie III, chapitres 1 et 2) mais nous indiquons brièvement ici l'importance de distinguer la tâche de l'activité et, la tâche de la situation.

Selon Leplat et Hoc (1983, p.50) « la tâche indique ce qui est à faire, l'activité, ce qui se fait ». Cette clarification conceptuelle est pour nous essentielle : l'analyse d'épreuves d'évaluation nécessite une double analyse. Celle-ci correspond d'une part à l'analyse de la situation dans laquelle se situe la tâche (laquelle se fait au moyen de cadres didactiques et/ou épistémologiques, elle est indépendante du sujet que l'on qualifie alors de sujet épistémique), et d'autre part à l'analyse de l'activité des élèves lorsqu'ils effectuent la tâche (cela implique ce que fait le sujet et s'analyse au moyen de cadres psychologiques). A défaut de pouvoir observer l'activité des élèves, on peut s'appuyer sur les traces de leur activité pour tenter de comprendre comment certaines caractéristiques de la tâche sont susceptibles d'orienter l'activité.

Il semble pleinement justifié, comme l'affirment notamment Vergnaud (2001) et Maury (2001), de considérer que des entrées psychologiques (au sens de la psychologie cognitive) et didactiques soient pertinentes pour analyser les objets d'enseignements et les tâches d'apprentissage. Les approches psycho-didactiques (des apprentissages, des évaluations) qualifient des approches qui prennent pour point de départ les connaissances à évaluer et les contenus impliqués dans les évaluations. Selon Vantourout et Goasdoué (2014), l'analyse de la validité dans le cadre des approches psycho-didactiques de l'évaluation, porte non pas sur des tests mais sur les tâches qui les composent. Elle porte à la fois sur les contenus et les connaissances mobilisés dans les tâches, les caractéristiques des tâches et les processus de réponse des évalués.

Pour souligner l'intérêt de procéder à de telles analyses, les auteurs prennent l'exemple d'une tâche issue de notre corpus (cf. III- 3-2 épreuve n°9). Cette tâche demande aux élèves d'entourer dans une liste de mots « la lettre qui fait [i] dans les mots ». Elle peut être traitée par un élève sans aucun recours aux valeurs prises par la lettre i dans le contexte d'un mot : en effet, dans les mots proposés, chaque fois que le mot contient la lettre i, le son [i] s'entend. De ce fait, bien que l'épreuve porte sur une composante essentielle de la lecture

(la maîtrise des correspondances entre les phonèmes et les graphèmes), la réussite de la tâche ne permet pas d'affirmer que cette connaissance a été mobilisée. Ceci corrobore des résultats décrits par Bastien et Bastien-Tonniazzo (2004) ainsi résumés par Richard (2004, p.11) :

« La réussite à une tâche ne dépend pas seulement du contenu conceptuel de la tâche et donc du savoir acquis par l'élève, mais également des composantes de la tâche ou des conditions de réalisation de celle-ci, qui sont choisies par les maîtres mais sont jugées non pertinentes par ceux-ci : leur importance n'est donc pas prise en compte par ces derniers dans les choix qu'ils font pour définir les conditions de travail de l'élève ».

Il apparaît ainsi nécessaire de s'intéresser à la façon dont chaque épreuve d'évaluation est proposée aux élèves de façon à établir si la tâche qu'elle contient permet bien aux élèves de mobiliser ce que l'on souhaite qu'ils mobilisent, de façon à éviter des réponses fausses fondées sur des connaissances justes (idem p. 36-47) et/ou des réponses justes fondées sur des connaissances fausses (idem p.48-54).

Rechercher à établir la validité psycho-didactique d'une épreuve revient à estimer si elle permet bien d'engager (de par les contenus mis en jeu, de par leur mise en contexte etc.) une activité de l'élève correspondante à ce que l'enseignant cherche à observer (ce qui est formulé par l'item).

3.3. Validité globale et locale

Grugeon-Allys (2015), Grapin (2015) et Grugeon-Allys et Grapin¹⁰ (2015, 2016) se sont intéressés à l'analyse de tests d'évaluation (externe) d'un double point de vue : global et local.

Pour ces auteurs, l'analyse de validité se joue à un premier niveau, local, avec une analyse *a priori* de chacun des items. Leur approche épistémo-didactique, située dans le cadre de la Théorie Anthropologique du Didactique (Chevallard, 1985), s'attache à « préciser le type de tâches auquel l'item¹¹ correspond, les techniques de résolution possibles, les éléments technologico-théoriques sous-jacents » (Grugeon-Allys et Grapin, 2016).

L'analyse de validité se joue également à un deuxième niveau, global, portant sur l'ensemble des items de façon à « étudier le recouvrement du domaine par les tâches proposées dans l'évaluation, la représentativité des types de tâches et la variété de leurs niveaux de convocation ; cela permet ainsi de repérer d'éventuels manques ou redondances dans les types de tâches et d'observer s'il existe un certain équilibre selon la complexité, les registres de représentation en jeu, etc. » (Grugeon-Allys et Grapin, 2016).

Nous nous inspirons du principe d'une approche centrée sur les épreuves et sur les tests utilisée par ces auteurs pour conduire nos analyses. Rapporté au cadre de l'évaluation interne à laquelle nous nous intéressons dans notre travail, l'analyse locale correspond à l'analyse psycho-didactique des épreuves tandis que l'analyse globale renvoie à l'analyse de la couverture de domaine réalisée par chaque test utilisé.

¹⁰ https://evaluationmaths.sciencesconf.org/conference/evaluationmaths/pages/Axe2_Recherche_Grugeon_Grapin.pdf

¹¹ Dans ce cadre, l'item est une question à laquelle sont associées plusieurs propositions parmi lesquelles le sujet évalué doit faire un choix. Dans notre travail, item renvoie à la formulation d'une compétence dans les livrets ou les programmes.

Chapitre 4. Synthèse et formulation de la problématique

Indépendamment des fonctions que l'on peut attribuer à l'évaluation, nous nous intéressons dans ce travail à la qualité de l'évaluation instituée conduite « au quotidien de la classe » par des enseignants de CP, dans deux sous-domaines à fort enjeu socio-scolaire, à savoir : le sous-domaine de la lecture en français, et le sous-domaine des nombres et du calcul en mathématiques. La qualité de l'évaluation que nous étudions dans notre travail renvoie à la validité dont nous avons mis en exergue trois dimensions :

1. une dimension curriculaire permettant d'apprécier comment l'évaluation remplit le « cahier des charges » établi par l'institution scolaire ;
2. une dimension psycho-didactique, permettant d'apprécier en quoi chacune des épreuves est susceptible - ou non - de remplir les objectifs qui lui sont assignés.
3. une dimension didactique permettant de juger comment l'ensemble des épreuves assure la couverture conceptuelle des champs disciplinaires concernés laquelle ne peut être appréciée sans que l'on soit assuré de la validité de chaque épreuve du point de vue psycho-didactique.

La confrontation d'un référé à un référent constitue un élément unificateur des différentes conceptualisations de l'évaluation. La construction de référés et de référents constitue ainsi l'ossature de notre étude. L'opération de confrontation entre les données recueillies (les livrets scolaires, et les ensembles d'épreuves que nous assimilons à des tests) et les référentiels construits prend différentes directions selon la qualité (le type de validité) recherchée et structure notre problématique qui comprend trois axes :

Axe n°1 : l'étude de la validité curriculaire de l'évaluation constitue le premier axe de notre problématique. Il s'agit de construire une méthodologie d'analyse permettant de confronter l'évaluation mise en œuvre par les enseignants au référentiel formel que représentent les (progressions annexées aux) programmes. Cette méthodologie vise à permettre de répondre à la question de savoir si l'évaluation est valide du point de vue curriculaire.

Au-delà des résultats obtenus, nous nous intéressons surtout à **l'effet des choix et reconfigurations effectués sur l'évaluation** réalisée parce que les enseignants établissent leur référentiel pour évaluer à l'issue d'un processus de référentialisation propre à chacun. Se dégage-t-il une hiérarchie entre les compétences évaluées ? Certaines compétences sont-elles délaissées ou d'autres au contraire davantage évaluées ? La question centrale est donc celle de l'impact de la prescription d'évaluer adressée aux enseignants par l'institution scolaire sur l'évaluation mise en œuvre au quotidien de la classe.

Axe n°2 : l'axe central de notre travail concerne **l'étude de la validité psycho-didactique des épreuves**. Pour conduire une analyse de cette nature, il est nécessaire d'élaborer une méthodologie d'analyse permettant d'étudier le potentiel de chaque épreuve à engager les élèves dans l'activité de réponse souhaitée par l'évaluateur, à engager l'élève à mobiliser les compétences évaluées. Cela nécessite d'explicitier des référentiels d'ordre didactique en fonction des domaines étudiés.

Cela nécessite également de prendre en considération des composants de l'épreuve qui auraient pu échapper à l'analyse proprement conceptuelle de la tâche, et qui pourraient cependant s'avérer des déterminants majeurs du comportement de réponse des élèves.

Axe n°3 : l'explicitation de référentiels épistémologiques et/ou didactiques permet de déterminer les principaux composants de chaque domaine. Cela constitue le travail préalable essentiel du troisième axe de notre problématique visant à élargir l'analyse de la validité aux tests. Que peut-on considérer comme étant les principaux composants des domaines étudiés ?

Suite à la détermination de ces composants, une analyse globale (des tests) et locale (des épreuves) pourra être entreprise. **Dans quelle mesure le test recouvre le (ou les) domaine(s) évalués d'un point de vue didactique ? Dans quelle mesure chaque épreuve engage une activité de l'élève correspondant à ce qui est attendu ?**

L'enjeu de notre étude est de montrer que la validité des épreuves est une dimension centrale de la validité de l'évaluation. Notre travail s'inscrit dans une perspective de formation des enseignants à l'évaluation, à ce jour fort discrète bien qu'elle ait un impact très important sur les parcours scolaires des élèves.

Nous avons souligné dans cette première partie qu'indépendamment des buts assignés à l'évaluation, l'évaluation est fondamentalement une opération de comparaison de référés à un référent. Nous avons précisé que nous ne nous intéressons qu'au référentiel formel de l'enseignant (en amont des passations) ce qui place de côté la partie du référent contemporaine de l'évaluation. L'explicitation du référentiel de l'évaluation au moyen duquel les référés sont analysés constitue un enjeu central pour étudier l'évaluation. Nous avons également souligné que la validité de l'évaluation était une qualité pouvant s'observer de plusieurs manières.

Dans la **partie II**, nous nous intéressons aux produits des processus de référentialisation conduits par les enseignants et aux effets des choix effectués sur l'évaluation réalisée. Nous les analysons sous l'angle de la couverture des programmes dans les domaines de la lecture, des nombres et du calcul.

Après avoir décrit la méthodologie que nous utiliserons pour étudier la validité psychodidactique des épreuves dans les chapitres 1 et 2 de la **partie III**, nous présenterons nos analyses selon différents composants de la tâche susceptibles d'avoir une influence sur le processus de réponse de l'élève confronté à l'épreuve (chapitres 3 à 5). Nous nous interrogerons également sur la validité attribuée par défaut à des épreuves accessibles sur le web ou dans des manuels (chapitre 6).

L'analyse conduite dans la **partie IV** prend deux directions : une analyse globale des tests, de l'ordre de la couverture didactique des composants du domaine, au moyen de référentiels préalablement explicités combinée à une analyse locale, psychodidactique, des épreuves .

Chapitre 5. Présentation du corpus de données recueillies

5.1. Population

Les 16 enseignants auprès de qui nous avons collecté nos données sont tous des enseignants expérimentés¹². Ils exercent dans 13 écoles différentes implantées à Paris ou proche banlieue. Leurs écoles sont réparties sur 6 circonscriptions scolaires, ce qui constitue une façon de minimiser l'impact que pourrait avoir une action de circonscription vis-à-vis de l'évaluation (format de livrets suggérés, utilisation d'épreuves de circonscription par exemple). Ces enseignants exercent dans des écoles que l'on qualifie parfois de « ni-ni » pour indiquer grossièrement qu'elles ne sont « ni » situées dans des zones d'éducation prioritaires, « ni » dans des zones particulièrement privilégiées. L'un des enseignants occupe des fonctions de professeur des écoles-maître formateur, il est également co-auteur d'une récente méthodologie d'apprentissage de la lecture. Quatre autres enseignants passent l'examen de certification aux fonctions de professeur maître-formateur. Plus d'un quart de ces enseignants sont donc formateurs, ou en passe de le devenir, et susceptibles de prendre part à ce que nous qualifierons plus loin la prescription secondaire.

Dans la thèse, chaque enseignant est représenté par une lettre : de la lettre A à la lettre P. Le tableau ci-dessous indique la circonscription et l'école d'exercice des enseignants.

	Circonscription	École
Enseignant B	1	1
Enseignant C	1	2
Enseignant G	2	3
Enseignant H	2	4
Enseignant I	2	4
Enseignant J	2	5
Enseignant K	2	5
Enseignant L	2	6
Enseignant M	2	6
Enseignant N	2	7
Enseignant O	2	8
Enseignant P	3	9
Enseignant A	4	10
Enseignant D	5	11
Enseignant E	6	12
Enseignant F	6	13

Tableau 1 : Répartition des enseignants formant la population étudiée, par circonscription et par école

¹² I.e. une expérience d'enseignement supérieure à 5 ans, et/ou, une expérience d'enseignement dans cette classe supérieure à 3 ans

5.2. Collecte des données

La collecte des données s'est déroulée au cours des années scolaires 2010-2011 et 2011-2012, période dans lesquelles les programmes de 2008 s'appliquent.

Deux types de données ont été collectés auprès de chaque enseignant :

- Toutes les copies d'un même élève, pour chacune des épreuves d'évaluation ayant servi, au cours du trimestre, à l'enseignant à renseigner le livret scolaire de l'élève. Dans de rares cas, lorsqu'il n'a pas été possible de collecter les copies, nous avons collecté les épreuves vierges.
- Un exemplaire du livret scolaire de l'élève, dont les copies ont été collectées, transmis aux familles.

Les documents ont été collectés à deux moments de l'année de façon à pouvoir analyser l'évolution de l'évaluation entre le début et la fin du CP :

- à la fin du premier trimestre,
- à la fin du troisième trimestre.

Des entretiens post-collecte, semi-directifs ont été entrepris auprès des enseignants visant à préciser le référencement des épreuves.

5.3. Présentation du corpus de données recueillies

Une certaine confusion règne au niveau terminologique dans les divers documents analysés, particulièrement en ce qui concerne les termes « item » et « programme ».

Nous argumenterons nos choix en partie II (chapitre 2) mais indiquons que, dans la thèse :

- l'item (du livret) correspond en réalité à une ligne du livret de l'enseignant (en général une phrase),
- le programme (au singulier) renvoie aux progressions annexées aux programmes de 2008 concernant le CP. Le terme programme (de CP) est toutefois impropre puisque les programmes de 2008 sont des programmes de cycle,
- l'item du programme correspond à une phrase (délimitée par un alinéa) figurant dans les progressions annexées aux programmes,
- l'épreuve désigne une (ou plusieurs) tâche(s) utilisée(s) dans une situation d'évaluation explicite. Les enseignants lui préfèrent le terme « exercice ».

1 184 épreuves d'évaluation et 16 livrets ont été collectés.

Nous avons catégorisé les épreuves recueillies en fonction des 61 items du programme (lequel comporte 40 items en français, 21 items en mathématiques) dans les volumes d'annexe A et B. Cette catégorisation a été effectuée sur la base des déclarations des enseignants, confirmées par les entretiens post-collecte. Il se produit toutefois quelques cas irréguliers faisant l'objet d'une catégorie « Autres ». Cela correspond aux épreuves qui évaluent plusieurs items à la fois. Dans ce cas, pour une exploitation quantitative, nous n'avons pas comptabilisé l'épreuve comme évaluant chaque item une fois, mais l'avons comptabilisée une seule fois dans cette seule rubrique. En revanche, pour une exploitation qualitative nous l'avons placée dans le volume d'annexe sous chacun des items qu'elle participe à évaluer. A ceci, il faut ajouter deux épreuves de mathématiques du premier trimestre qui sont destinées à évaluer un item non référencé dans les programmes. Dans ce cas, chaque épreuve est comptabilisée dans la rubrique Autre.

L'ensemble des données recueillies fait l'objet de trois volumes d'annexes structurés comme suit :

Annexe A :

-Une présentation chronologique de l'évolution des programmes depuis la loi d'orientation de 1985.

-Les livrets utilisés par les 16 enseignants qui constituent la population étudiée,

-Les 430 épreuves d'évaluation collectées aux premiers et troisièmes trimestres relatives aux sous-domaines « Oral ; Ecriture ; Vocabulaire ; Grammaire ; Orthographe » du domaine disciplinaire du Français des programmes, et « Géométrie ; Grandeurs et mesure ; Organisation et gestion des données » du domaine disciplinaire des Mathématiques. Ces épreuves ne seront pas analysées en détail dans la thèse.

Annexe B :

-Les 754 autres épreuves des domaines disciplinaires « Français » et « Mathématiques » du programme collectées, relatives aux sous-domaines « Lecture » (186 épreuves pour le premier trimestre et 107 pour le troisième trimestre) et « Nombre et calcul » (210 épreuves pour le premier trimestre et 251 pour le troisième trimestre).

Les épreuves sont rangées par enseignant et par item du programme auxquels elles sont déclarées être rattachées par l'enseignant.

Les items sont numérotés comme dans le livret SCEREN que nous présenterons plus loin. Le groupement en sous-domaines des programmes est respecté (Langage Oral, Lecture, Ecriture, Vocabulaire, Grammaire, Orthographe pour le Français et Nombres et Calcul, Géométrie, Grandeurs et Mesures, Organisation et gestion de données en Mathématiques).

Le corpus de données est également trié par trimestre.

Pour une référence simple et rapide nous avons opté pour le codage suivant : F pour désigner que l'épreuve dont il est question renvoie au domaine du Français, et M pour les Mathématiques ; T1 pour indiquer que l'épreuve en question est une épreuve utilisée au premier trimestre, T3 pour le troisième trimestre.

Les épreuves ayant été collectées sur une période étalée, chaque feuille d'épreuves a été numérotée. Puis, pour chaque feuille, les épreuves qu'elle contient ont été numérotées. Par exemple 125-3 indique qu'il s'agit de la 125^e feuille d'évaluation collectée et que l'épreuve dont il est question est en 3^e position sur la page.

Exemple de codage complet :

L'épreuve « FT3 ; 63-3 ; enseignant E » indique qu'il s'agit de la 3^e épreuve située sur la feuille n°63 de l'ensemble des épreuves collectées au troisième trimestre en Français. Cette épreuve est utilisée par l'enseignant E.

Pour une lecture fluide nous avons évité autant que possible de mentionner ce type de codage. Nous avons préféré renvoyer directement à une page du volume d'annexe en faisant simplement mention de l'enseignant, du domaine disciplinaire et du trimestre.

Annexe C :

Les mêmes épreuves que celles figurant déjà dans les Annexes précédentes sont reprises et réorganisées en fonction des composants essentiels des domaines étudiés que la mise à jour de référentiels d'ordre didactique aura permis de réaliser, indépendamment des affectations (aux items du livret et/ou du programme) indiquées par les enseignants. C'est ce qui justifie un volume d'annexe particulier. La catégorisation réalisée explique que nous obtenions un volume d'épreuve légèrement différent (par rapport à la catégorisation utilisée pour organiser l'Annexe B) : 183 épreuves d'évaluation de la lecture et 219 épreuves d'évaluation du nombre (problèmes inclus) pour le 1^{er} trimestre ; et 101 de lecture et 203 épreuves d'évaluation du nombre pour le 3^e trimestre.

Partie II – Étude de la validité curriculaire de l'évaluation

Le livret scolaire de l'élève utilisé par chaque enseignant constitue l'aboutissement d'un processus de référentialisation dont nous nous intéressons à analyser les effets dans cette partie. En particulier, nous nous intéressons à l'effet que produit la détermination, par chaque enseignant, de ce qu'il y a à évaluer parmi ce qu'il est demandé d'enseigner, sur l'évaluation réellement mise en œuvre par rapport à un curriculum prescrit (nous n'avons pas collecté de données permettant de décrire le curriculum réalisé). Cela renvoie à la validité curriculaire de l'évaluation au sens de De Landsheere (cf. I. 3.1.1). L'analyse de la validité curriculaire de l'évaluation fournit un indice de l'activité évaluative de l'enseignant en amont des passations des épreuves par les élèves.

Nous avons brièvement indiqué en introduction (cf. I.2.1) que le processus de référentialisation de l'évaluation mis en œuvre par chaque enseignant prend appui sur une conjonction de référentiels découlant tous d'un référentiel « alpha » (les progressions). Il est possible de les considérer comme des « organisateurs du travail » (Amigues, 2009). Parmi les organisateurs du travail de l'enseignant, Goigoux (2007) distingue deux types de prescriptions. Une prescription « primaire » est portée par « l'organisation hiérarchique du travail » émanant directement de l'institution scolaire pour aider le professeur à « concevoir, à organiser et à réaliser son travail »¹³. Une prescription « secondaire » émane « des instituts de formation professionnelle » et est « élaborée et diffusée par des formateurs qui non seulement reformulent, interprètent ou concrétisent les injonctions hiérarchiques mais développent un ensemble de recommandations autonomes » (idem).

Pour conduire nos analyses, deux organisateurs du travail issus de la prescription primaire nous intéressent particulièrement en raison de leur caractère opérationnel pour mettre en œuvre l'évaluation. Il s'agit des progressions annexées aux programmes de 2008 et du livret scolaire publié par le SCEREN CRDP de l'académie d'Orléans-Tours.

L'institution est contrainte de prolonger localement la commande faite aux enseignants afin de la rendre opératoire. Ceci la conduit à produire de la prescription pour préciser le contenu du livret scolaire. Par exemple, l'inspection de l'Education nationale de Tarentaise souligne¹⁴ que « les documents des évaluations périodiques en usage dans l'école (bulletins périodiques ou trimestriels) sont en prise directe avec les éléments des programmes et les progressions en français et mathématiques parus au B.O. hors-série n° 3 du 19 juin 2008 et

¹³ Elle regroupe selon cet auteur « les programmes d'enseignement et autres instructions officielles, les lois et règlements de la fonction publique d'État, l'évaluation du travail enseignant réalisée par les inspecteurs de l'Éducation nationale, l'évaluation des acquisitions des élèves, etc. » (idem)

¹⁴ Repéré à http://www.ac-grenoble.fr/savoie/pedagogie/docs_pedas/evaluation_2009/index.php?num_lie=866

permettent un suivi des progrès des élèves ». Cet exemple permet de mettre en évidence qu'une liaison directe entre les programmes et le livret scolaire est implicitement pensée, quand bien même le modèle de livret n'est pas imposé (depuis la rentrée 2016 et la mise en place du Livret Scolaire Unique, ce n'est plus le cas).

A un autre niveau, académique ou national, des relais institutionnels, comme le SCEREN CRDP de l'académie d'Orléans-Tours, ont publié un modèle de livret articulé aux programmes destiné à être employé tel quel par les enseignants.

Comment ces diverses prescriptions, ces divers organisateurs du travail, s'articulent-ils au cours du processus de référentialisation mené par chaque enseignant ?

Chapitre 1. Les progressions annexées aux programmes, référentiel des enseignants

1.1. Programmes et progressions, un lien ambigu

Avant d'analyser les livrets scolaires et leur articulation avec les programmes, nous revenons sur un élément brièvement mentionné dans l'introduction (cf. I.2.1). Les programmes de 2008, qui sont comme tous les programmes depuis 1989 des programmes de cycles, présentent la particularité d'adjoindre en annexe une liste de repères pour organiser la progressivité des apprentissages par niveau d'enseignement, en Français et en Mathématiques : des « progressions ». Les programmes¹⁵ de l'école primaire sont donc présentés comme une entité bicéphale comprenant deux parties distinctes « mais indissociables : les programmes proprement dits et les progressions annuelles, qui vont, en français, de la petite section de maternelle au CM2 et, en mathématiques, du CP au CM2 » (B.O. 2008, p.11).

Rédigés sous la forme d'un texte (B.O. 2008, pp.17-19), les programmes sont constitués :

- d'une introduction générale qui indique notamment que « les enseignements en français et en mathématiques font l'objet d'une progression par année scolaire, jointe au présent programme »,
- d'une partie allouée au domaine du Français elle-même subdivisée en une introduction et autant de paragraphes que ce domaine comporte de sous-domaines (5) : Langage oral, Lecture et écriture, Vocabulaire, Grammaire, Orthographe.
- d'une partie allouée au domaine des Mathématiques elle-même subdivisée en une introduction et autant de paragraphes que ce domaine compte de sous-domaines (4) : Nombres et calcul, Géométrie, Grandeurs et mesures, Organisation et gestion des données
- d'une dernière partie correspondant aux autres domaines d'enseignement : Education physique et sportive, Langue vivante, Découverte du monde, Pratiques artistiques et histoire des arts, Instruction civique et morale.

Rédigées sous la forme de tableaux (B.O. 2008, pp.31-32), des progressions annuelles listent par niveau et pour deux domaines d'enseignement (le Français et les Mathématiques), un certain nombre de « repères aux équipes pédagogiques pour organiser la progressivité des apprentissages ». Ces repères sont des compléments ou des reformulations du texte des programmes. Elles sont annexées aux programmes.

De vives critiques ont été émises quant à la rédaction de ces programmes. Goigoux (2008) dénonce notamment le décalage que nous pointons entre ce que le texte affirme définir pour chaque domaine d'enseignement en termes de connaissances et de compétences à atteindre en fin de cycle, et la progression annuelle proposée. Les documents de planification annuelle annexés, « probablement rédigés par d'autres

¹⁵ Afin de ne pas alourdir la lecture de la thèse, l'analyse chronologique des textes officiels de 1989 à 2013 est renvoyée au volume d'Annexe A (pp. 2-23). Elle nous a été utile pour comprendre et mesurer la « pression évaluative croissante » et « l'inflation de la prescription » auxquelles sont soumis les enseignants et que dénoncent les parlementaires (Gonthier-Maurin, 2012). Elle nous a permis d'explicitier quelques éléments jalonnant l'installation d'une culture de l'évaluation chez les enseignants, influencée par l'évaluation standardisée (cf. par exemple les évaluations CE2-6ième).

conseillers, prétendent n'être là que pour donner des points de repères aux équipes pédagogiques : « ils ne constituent pas des référentiels pour une évaluation normative des élèves à la fin de chaque section ». Comprenne qui pourra » conclu, dubitatif, cet auteur (idem).

Il est de ce fait hautement probable que les items des progressions constituent le cadre de référence (le référentiel « alpha ») à partir duquel l'enseignant construit le livret scolaire : ce sont les éléments les plus opérationnels des programmes pour évaluer.

Nous présentons, dans ce qui suit, ce référentiel et particulièrement les items sur lesquels porteront principalement nos analyses dans les parties suivantes (les items du sous-domaine du Français intitulé « Lecture », et ceux du sous-domaine des Mathématiques intitulé « Nombres et calcul »).

1.2. Les items « Lecture » des progressions

Une série de 11 connaissances et compétences (des items) sont présentées pour le CP. Nous les reproduisons ci-dessous :

-
- Connaître le nom des lettres et l'ordre alphabétique.
 - Distinguer entre la lettre et le son qu'elle transcrit ; connaître les correspondances entre les lettres et les sons dans les graphies simples (ex. *f* ; *o*) et complexes (ex. *ph* ; *au*, *eau*).
 - Savoir qu'une syllabe est composée d'une ou plusieurs graphies, qu'un mot est composé d'une ou plusieurs syllabes ; être capable de repérer ces éléments (graphies, syllabes) dans un mot.
 - Connaître les correspondances entre minuscules et majuscules d'imprimerie, minuscules et majuscules cursives.
 - Lire aisément les mots étudiés.
 - Déchiffrer des mots réguliers inconnus.
 - Lire aisément les mots les plus fréquemment rencontrés (dits mots-outils).
 - Lire à haute voix un texte court dont les mots ont été étudiés, en articulant correctement et en respectant la ponctuation.
-
- Connaître et utiliser le vocabulaire spécifique de la lecture d'un texte : le livre, la couverture, la page, la ligne ; l'auteur, le titre ; le texte, la phrase, le mot ; le début, la fin, le personnage, l'histoire.
 - Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens.
 - Ecouter lire des œuvres intégrales, notamment de littérature jeunesse.
-

Tableau 2 : Connaissances et compétences en Lecture (Progressions pour le CP, MEN BO, 2008, p. 31)

Nous avons reproduit ce tableau en introduisant un saut de paragraphe. De ce fait, deux blocs d'items apparaissent :

1. le premier bloc est une série de connaissances et compétences portant des aspects relatifs à l'identification des mots,
2. le second bloc est constitué d'une série de connaissances et compétences portant davantage sur des aspects relatifs à la compréhension.

Le découpage effectué (en deux blocs) permet de constater que le document est plus disert sur ce qui fait maintenant consensus et n'est plus remis en cause. A savoir, outre le développement des compétences langagières, « la conscience alphabétique et la conscience phonologique, la capacité d'identifier les graphèmes, les phonèmes et de les

mettre en correspondance, l'automatisation du traitement du code de reconnaissance/déchiffrage des mots » (cf. PIREF, 2003, p.3).

Ce découpage permet également de considérer que la part de la compréhension est moindre, une critique formulée par Bianco (2014¹⁶) qui déplore, qu'au niveau de l'ensemble des programmes, « la place de la compréhension en lecture est inexistante ».

Les progressions sont relativement évasives, renvoyant la compréhension d'un texte lu (par l'élève ? le maître ? par l'ensemble des élèves ?) aux seules détections du thème (ou sujet) du texte lu (« dire de qui ou de quoi parle le texte lu ») et aux seules capacités à répondre correctement à des questions portant sur le texte, ou reformulations. Cèbe, Goigoux et Thomazet (2004) ont pourtant souligné la relative passivité des élèves dits « faibles lecteurs » qui attendent de disposer du questionnaire pour s'intéresser véritablement au texte et qui se contentent d'un repérage thématique (« de quoi parle le texte » ?) au détriment d'une construction problématique (« que veut dire le texte ? Quel est le problème ? »).

On retrouve dans le sous-domaine « langage oral » un item formulé en des termes proches de l'item précédent portant sur la compréhension d'un texte en lecture : « manifester sa compréhension d'un récit ou d'un texte documentaire lu par un tiers en répondant à des questions le concernant ; reformuler le contenu d'un paragraphe ou d'un texte, identifier les personnages principaux d'un récit ». Cela laisse sous-entendre que la compréhension d'un texte lu seul n'est pas foncièrement de nature différente de celle d'un texte entendu lire (nous détaillerons la référence faite au modèle de la « simple vue » en lecture dans les parties suivantes, cf. Gough et Tunmer (1986). Tout cela justifierait à lui seul un développement, une déclinaison sur la base de propositions telles que, par exemple, celles formulées par Fayol (2003) en vue de la conférence de consensus de 2003 sur l'enseignement de la lecture à l'école primaire. L'auteur citait notamment la nécessité de travailler les processus inférentiels et les relations anaphoriques. Ces deux dimensions ne sont pas explicitement présentes dans les items des progressions.

Goigoux et Cèbe estiment ainsi que ces programmes « survalorisent (en CP) la maîtrise du code au détriment de la compréhension, renvoyée au CE1 faute de distinguer les lectures autonomes des élèves de celles réalisées par ou avec l'aide du maître. ».

Probablement par tradition scolaire, le travail sur le sens et la morphologie est renvoyé à des champs tendant à se constituer de fait comme externes à l'apprentissage de la lecture. Vocabulaire, Grammaire, Orthographe constituent trois sous-domaines du Français distincts, à part égale (en volume) avec Lecture et Écriture. Ceci semble de nature à entretenir une confusion quant à la place du travail morphologique qui aurait un effet positif sur les performances en lecture selon Gombert (2002, p.9) ou encore Goigoux qui précise que (2000/2002, p. 70) « Que l'on traite de morphèmes lexicaux ayant une fonction sémantique ou de morphèmes grammaticaux, ayant une fonction syntaxique, la focalisation de l'élève sur les aspects morphologiques serait de nature à faciliter l'acquisition de l'orthographe, grammaticale ou lexicale »

Enfin, le texte des programmes liait lecture et écriture. La présentation conjointe dans un même sous-domaine dans le texte du programme laissait augurer une relation forte marquée

¹⁹http://cache.media.education.gouv.fr/file/Le_systeme_educatif/70/7/Bianco_Maryse_-_MCF_-_CSP_Contribution_362707.pdf

entre décodage (automatisation des procédures d'identification des mots écrits) et encodage. Cette liaison n'est plus apparente dans les progressions : l'écriture devient un sous-domaine à part entière, distinct de celui de la lecture.

1.3. Les items du sous-domaine des nombres et du calcul des progressions

10 items décrivent ce sous domaine. Nous les reproduisons ci-dessous :

-
- Connaître (savoir écrire et nommer) les nombres entiers naturels inférieurs à 100.
 - Produire et reconnaître les décompositions additives des nombres inférieurs à 20 ("table d'addition").
 - Comparer, ranger, encadrer ces nombres.
 - Écrire une suite de nombres dans l'ordre croissant ou décroissant.
 - Connaître les doubles des nombres inférieurs à 10 et les moitiés des nombres pairs inférieurs à 20.
 - Connaître la table de multiplication par 2.
 - Calculer mentalement des sommes et des différences.
 - Calculer en ligne des sommes, des différences, des opérations à trous.
 - Connaître et utiliser les techniques opératoires de l'addition et commencer à utiliser celles de la soustraction (sur les nombres inférieurs à 100).
-
- Résoudre des problèmes simples à une opération.
-

Tableau 3 : Connaissances et compétences relatives aux nombres et au calcul (Progressions pour le CP, MEN BO, 2008, p. 33)

Comme pour la lecture, nous avons reproduit ce tableau en introduisant un saut de paragraphe. De ce fait deux blocs d'items apparaissent :

1. le premier est une série d'items portant sur des aspects relatifs à ce qui est reporté en premier la connaissance des nombre entiers naturels,
2. le second est constitué d'un seul item portant sur la résolution de problèmes numériques.

Le découpage effectué permet de constater que, tant en lecture que dans le sous-domaine des nombres et du calcul, les compétences qui s'appuieraient sur la résolution de tâches complexes (lire un texte, résoudre un problème) sont sous-représentées.

Dans le premier bloc d'item aucun objectif particulier, explicite, en matière de dénombrement (recours spontané au dénombrement, maîtrise du dénombrement, constitution d'une collection de cardinal donné) n'est assigné aux enseignants comme le texte du programme le mentionne pourtant (B.O. 2008, p.13) en indiquant que les élèves «dénombrer¹⁷ des collections, connaissent la suite des nombres, comparent et rangent ». Le calcul est fortement représenté traduisant un objectif prioritaire des programmes parfois énoncé de façon synthétique par la formule « Lire, écrire, compter ».

On ne peut pas véritablement parler d'un second bloc puisqu'un seul objectif le définit. Alors que les programmes de 2002 détaillaient abondamment le type de problèmes à proposer aux élèves, les progressions réduisent ce sujet à une formule concise : le problème est dit « simple » et ne comporte qu'une seule opération.

¹⁷ C'est la seule occurrence de ce mot dans le texte des programmes. Nous n'avons trouvé aucune occurrence de ce mot dans les progressions.

Notons que le terme « problème » est aussi utilisé dans le sous-domaine « grandeurs et mesures ». Les programmes précisent que « les élèves commencent à résoudre des problèmes portant sur des longueurs, des masses, des durées ou des prix ». Le problème est ici rattaché à la vie courante.

-
- Repérer des événements de la journée en utilisant les heures et les demi-heures.
 - Comparer et classer des objets selon leur longueur et leur masse.
 - Utiliser la règle graduée pour tracer des segments, comparer des longueurs.
 - Connaître et utiliser l'euro.
 - Résoudre des problèmes de vie courante.
-

Tableau 4 : Connaissances et compétences en Grandeurs et mesures (Progressions pour le CP, MEN BO, 2008, p. 33)

Nous soulignons ici une évolution marquée depuis les programmes 2002 (BO 2002 HS du 14 février) qui faisaient des problèmes un sous-domaine à part entière (distinguant ceux résolus par une procédure personnelle de ceux résolus par une procédure experte). Les problèmes se rapportaient prioritairement au domaine numérique et une typologie sous-jacente (faisant référence à celle de Vergnaud que nous détaillerons plus loin dans la partie IV, chapitre 2) permettait de les hiérarchiser. Ces précisions ont disparu dans les programmes de 2008 dans lesquels l'accent est placé sur les automatismes ainsi qu'il est expressément dit dans la préface du document d'accompagnement des programmes de 2008 en mathématiques et paru en 2012 intitulé « Le nombre au cycle 2 » (MEN, 2012, p.5).

Enfin, il n'est pas question, dans les progressions, de « problèmes de groupements et de partage permettant d'approcher une première approche de la division » (ibid.) alors que le texte des programmes le mentionne explicitement.

Les progressions sont donc une traduction parfois imparfaite du texte des programmes. Elles sont susceptibles de laisser de côté certains aspects pourtant essentiels ou de minimiser leur importance (le dénombrement en mathématiques par exemple).

Chapitre 2. Première approche de la validité curriculaire

Après avoir souligné le rôle de référentiel tenu par les progressions annexées aux programmes, nous nous intéressons à présent aux livrets conçus ou utilisés par les enseignants. Nous cherchons tout d'abord à analyser les choix de formats de livret réalisés par les enseignants du point de vue de leur articulation au référentiel que constituent les items des progressions.

Suite à cela, nous examinerons le taux de remplissage de ces livrets : il constitue une première estimation de la validité curriculaire de l'évaluation réalisée. En effet, si le livret est articulé aux items des progressions et que tous les items qu'il contient sont évalués en fin d'année, alors on peut en déduire que l'évaluation couvre le programme (alors réduit aux progressions).

2.1. Les livrets utilisés par les enseignants

Dans la mesure où il concorde avec le référentiel que constituent les items des progressions, le livret publié par le SCEREN est adopté par plusieurs enseignants de notre population.

2.1.1. Le livret SCEREN

Dans le but d'outiller la demande faite aux enseignants d'établir la communication aux familles, et à l'institution, au moyen d'un livret scolaire, un tel document a été publié en 2010 par le SCEREN – CRDP de l'Académie d'Orléans-Tours. Préfacé par Philippe Claus, inspecteur général, doyen du groupe de l'enseignement primaire, il se situe du côté de la prescription primaire. Il tente d'établir une liaison entre les programmes et l'évaluation, s'insérant dans une dynamique de guidage de l'action. Nous analysons les choix faits dans ce livret en les comparant au référentiel « alpha » précité.

Ce livret est un véritable livret scolaire de l'élève au sens où le définit l'institution. Il comporte en effet les deux volets réglementaires :

- d'une part il contient les évaluations des compétences du Socle aux paliers 1 (fin CE1) et 2 (fin CM2) et les différentes attestations requises,
- d'autre part il adopte les contours d'un livret traditionnel tri- ou quadri- annuel relevant de la responsabilité de chaque enseignant.

Il a été largement adopté en raison du fait notamment de sa facture, jugée pratique (il va de la Petite Section de maternelle au CM2), de son prix de revient (plus économique que le tirage papier chaque trimestre d'un livret local). L'archivage qu'il permet d'une année à l'autre des acquisitions des élèves, et l'harmonisation du format de livret au niveau de la même école constituent d'autres atouts souvent plébiscités par les enseignants.

Trois parties retiennent particulièrement notre attention : le mode d'emploi, la préface, les pages consacrées au CP.

2.1.1a Le mode d'emploi du livret

Le mode d'emploi reproduit ci-dessous affirme que le livret est articulé aux connaissances et compétences qui constituent les objectifs des programmes.

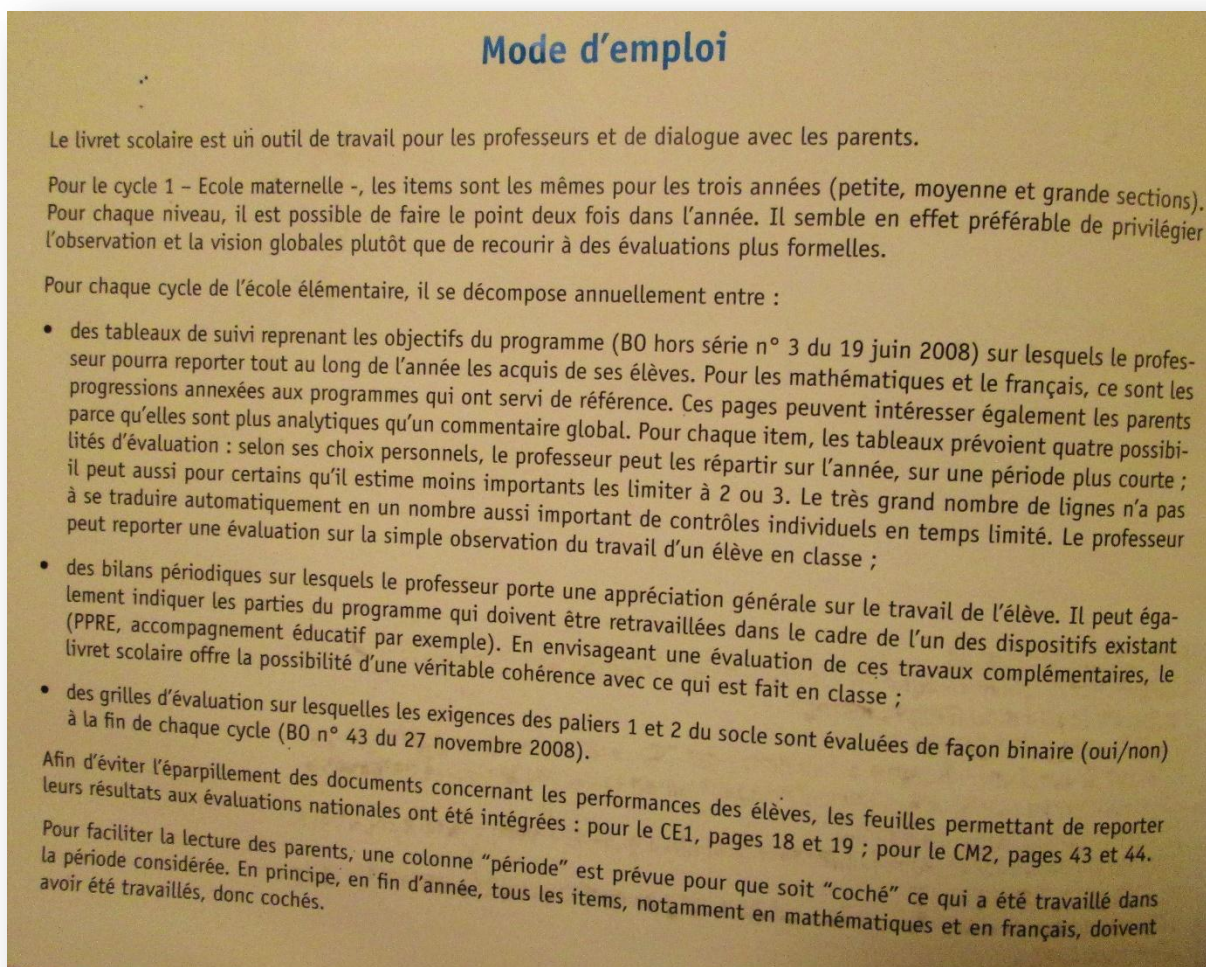


Figure 1 : Mode d'emploi du livret SCEREN

Le terme utilisé dans le livret est pourtant celui d'« item » ce qui nécessite une clarification. Les auteurs du livret désignent du terme « item » chaque alinéa du livret correspondant à une connaissance ou compétence des progressions. Il n'est pas précisé dans ce document ce qui relève d'une connaissance, d'une attitude ou d'une compétence : l'item est un terme générique. L'item est une ligne du tableau. Par exemple : « connaître le nom des lettres et l'ordre alphabétique ».

Suite à une préconisation faite au niveau de la temporalité de l'évaluation, les auteurs du livret précisent que « le très grand nombre de lignes n'a pas à se traduire automatiquement en un nombre aussi important de contrôles individuels en temps limité. Le professeur peut reporter une évaluation sur la simple observation du travail d'un élève en classe ». Il est important de souligner que cette proposition d'opérationnalisation de l'évaluation en vue de renseigner un livret est inédite. Les auteurs du livret ont manifestement conscience de la grande quantité d'items contenus dans les progressions (40 en français, 21 en mathématiques), et indiquent toutefois que cela n'implique pas de tout évaluer par des contrôles individuels en temps limité. Ils précisent qu'« en principe, en fin d'année, tous les

items, notamment en mathématiques et en français, doivent avoir été travaillés, donc cochés ». Il est sous-entendu que l'évaluateur doit tout évaluer mais pas nécessairement de façon individuelle en temps limité : une « simple observation » suffit. Un fort risque existe alors d'une évaluation subjective et implicite. D'aucuns ont discuté de la place de la subjectivité dans l'acte d'évaluer pour la « blanchir » (Weiss, 1986) parce que (Weiss, 1992, p. 28) « les enseignants pourraient bien être ceux qui apprécient le mieux et le plus valablement les performances des élèves (parce que) ce sont eux qui précisément connaissent le mieux les élèves, mais aussi les exigences scolaires formelles, les attentes sociales dominantes, souvent implicites, mais aussi le fonctionnement complexe du système éducatif et les ingrédients intellectuels, comportementaux, et sociaux qui entrent dans la composition du succès scolaire ». L'approche technicienne de l'évaluation instrumentée, « sous des apparences d'objectivité ne la garantit toutefois pas » précise encore cet auteur (idem p.28).

Mais, s'il est admis qu'évaluer comporte une part de subjectivité, il semble qu'il existe un risque à affirmer qu'évaluer peut se faire par « simple observation », tant que l'on n'aura pas défini précisément ce qu'elle est. On ne sait par exemple pas ce que doit observer l'évaluateur, ni dans quelle situation pour l'élève (exercice, transfert) il doit le faire. Les grilles de références pour l'évaluation et la validation des compétences du socle commun au palier 1 précisent bien (2011, p. 2) que « Les temps d'évaluation sont distincts des temps d'apprentissage et des temps d'entraînement nécessaires à l'appropriation des connaissances et à la maîtrise des compétences par les élèves ». On conviendra que l'idée même de simple observation, pour économique et pragmatique qu'elle soit, pose de nouveaux problèmes autant qu'elle en résout.

2.1.1b La préface du livret

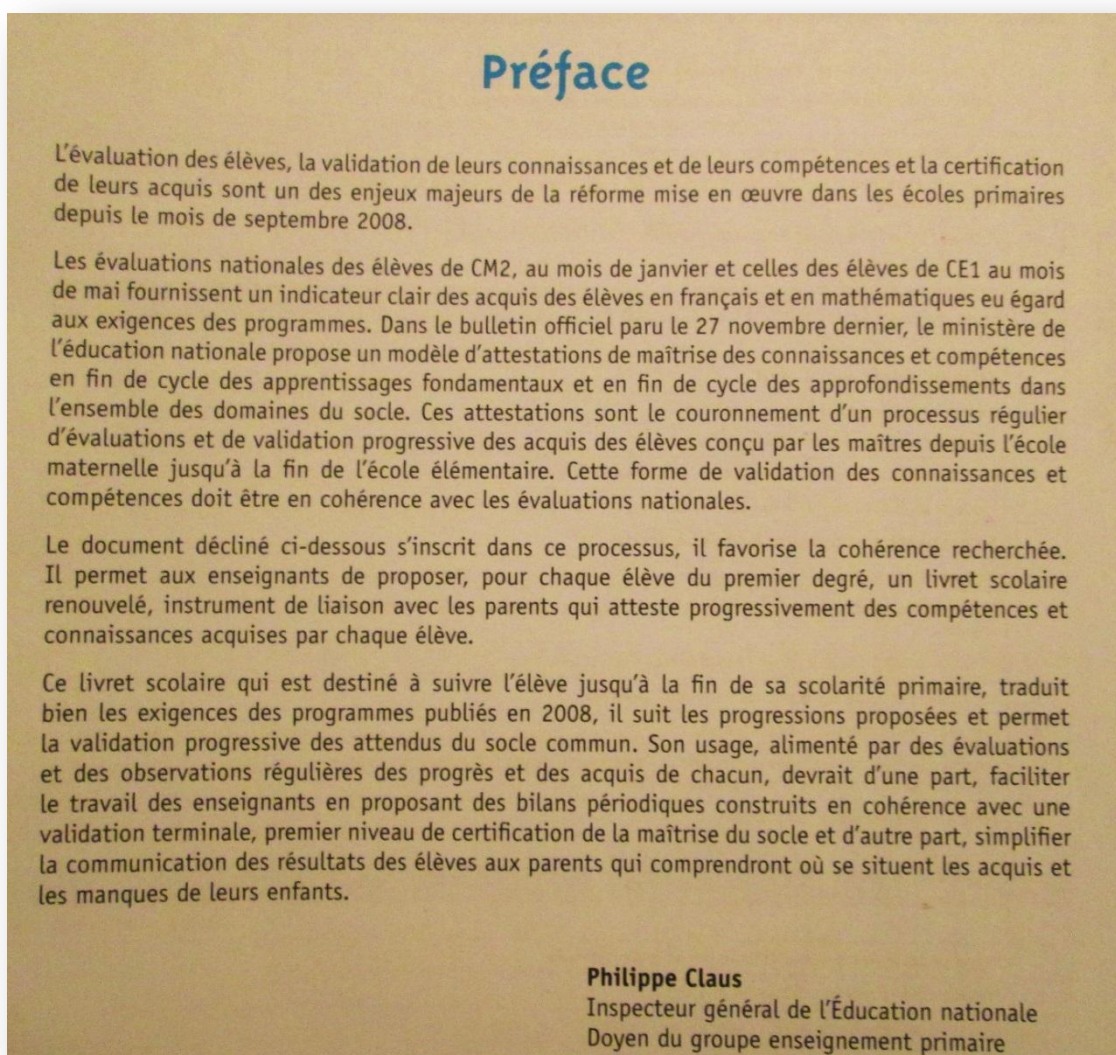


Figure 2 : Préface du livret SCEREN

« L'évaluation des élèves, la validation de leurs connaissances et de leurs compétences et la certification de leur acquis sont un des enjeux majeurs de la réforme mise en œuvre dans les écoles primaires depuis le mois de septembre 2008 » annonce l'inspecteur général en introduction du livret. Une précision de taille est apportée quant à la nature de l'évaluation : quand le mode d'emploi suggère de reporter une évaluation sur la base d'une « simple observation », cette préface précise que l'évaluation prend sa place dans un processus progressif de validation et de certification particulièrement accentué en fin de CE1 et en fin de CM2 par la validation des compétences des paliers 1 et 2. L'évaluation régulière participe ainsi à la validation des compétences.

2.1.1c Les pages CP du livret¹⁸

Le livret présente pour chaque niveau de classe une liste d'items. Cette liste est calquée sur le programme d'enseignement, à deux aménagements près : la reformulation et la suppression d'items des progressions. Pour faciliter la comparaison nous avons reproduit les items concernant la lecture. En gras, la partie de l'item des progressions retenue dans le livret du SCEREN. En italique, les items fusionnés (n°12 et A) ou retirés (B).

N° d'item	L'item dans le livret SCEREN	L'item dans les progressions
Item 8	Connaître le nom des lettres et l'ordre alphabétique.	Connaître le nom des lettres et l'ordre alphabétique.
Item 9	Connaître les correspondances entre les lettres et les sons	Distinguer entre la lettre et le son qu'elle transcrit ; connaître les correspondances entre les lettres et les sons dans les graphies simples (ex. f ; o) et complexes (ex. ph ; au, eau).
Item 10	Repérer graphies et syllabes dans un mot.	Savoir qu'une syllabe est composée d'une ou plusieurs graphies, qu'un mot est composé d'une ou plusieurs syllabes ; être capable de repérer ces éléments (graphies, syllabes) dans un mot.
Item 11	Connaître les correspondances entre minuscules et majuscules d'imprimerie, minuscules et majuscules cursives.	Connaître les correspondances entre minuscules et majuscules d'imprimerie, minuscules et majuscules cursives.
Item 12	<i>Lire aisément les mots étudiés, les mots les plus fréquemment rencontrés (dits mots outils).</i>	Lire aisément les mots étudiés.
Item 13	Déchiffrer des mots réguliers inconnus.	Déchiffrer des mots réguliers inconnus.
A- Non numéroté		Lire aisément les mots les plus fréquemment rencontrés (dits mots outils).
Item 14	Lire à haute voix un texte court dont les mots ont été étudiés	Lire à haute voix un texte court dont les mots ont été étudiés , en articulant correctement et en respectant la ponctuation.
Item 15	Utiliser le vocabulaire spécifique de la lecture : le livre, la couverture, la page, la ligne ; l'auteur, le titre	Connaître et utiliser le vocabulaire spécifique de la lecture d'un texte : le livre, la couverture, la page, la ligne ; l'auteur, le titre ; le texte, la phrase, le mot ; le début, la fin, le personnage, l'histoire.
Item 16	Dire de qui ou de quoi parle le texte lu	Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens.
B-Non numéroté	-	Écouter lire des œuvres intégrales, notamment de littérature de jeunesse.

Tableau 5 : Correspondances des items du livret SCEREN aux items des progressions (MEN, BO 2008 p. 31)

18 Elles sont reproduites en Annexe A, pp.29-31

La reproduction in extenso des items des programmes occupe une place très importante peu compatible avec des contraintes de fabrication. De ce fait, le premier aménagement réalisé est celui d'une reformulation des items du programme par réduction du texte de manière à les réduire à l'essentiel au risque de délaissier une partie du contenu de l'item. L'item n°16 du livret est par exemple réduit à la première de ses trois composantes.

Pour ce qui concerne le CP, l'intégralité des 21 items de Mathématiques du programme est reproduite dans le livret. Cela n'est pas le cas en Français : seuls 36 des 40 items du programme sont reproduits (si l'on considère au vu du tableau précédent que l'item 12 du livret se subdivise en deux items dans le programme).

Quatre items du programme ne sont pas reproduits dans le livret :

- « Manifester sa compréhension d'un récit ou d'un texte documentaire lu par un tiers en répondant à des questions le concernant : reformuler le contenu d'un paragraphe ou d'un texte, identifier les personnages principaux d'un récit » (sous-domaine Langage oral).
- « Ecouter lire des œuvres intégrales, notamment de littérature jeunesse » (sous-domaine Lecture).
- « Commencer à utiliser de manière autonome les marques du genre et du nombre (pluriel du nom, féminin de l'adjectif, terminaisons –nt des verbes) » et « Commencer à utiliser correctement la majuscule (début de phrase, noms propres de personnes) » (sous-domaine Orthographe).

On peut questionner ces choix éditoriaux du point de vue des conséquences (sélectionner les items que l'on peut ne pas faire figurer dans le livret) en regard de la finalité affirmée « tous les items devraient être cochés ».

Bien entendu on pourra toujours réaffirmer que seuls les programmes font force de loi, mais si les enseignants utilisent ces documents présentés par un inspecteur général de l'éducation nationale, ces derniers risquent de s'y substituer.

2.1.1d Discussion

Le livret SCEREN est constitué de traductions-reformulations des items des progressions annexées aux programmes, bien que ces progressions soient présentées comme des « repères pour organiser la progressivité des apprentissages ».

Par le biais du livret, les « repères » deviennent des normes pour l'évaluation et se substituent aux programmes. Le livret SCEREN joue un rôle extrêmement important : il est une validation institutionnelle de l'usage des items des progressions comme étant le référentiel formel pour évaluer.

De surcroît, une logique de validation (de chacun des items) fonde cet outil ce qui renvoie à un autre type de décision que des seules décisions d'ordre pédagogiques soulignées en introduction (cf. Hadji I.1). Elles sont orientées vers la certification.

Une qualité de ce livret est de synthétiser les attentes institutionnelles et de proposer un début d'opérationnalisation en matière d'évaluation. Ce début d'opérationnalisation ne repose que sur la liste des différents items qui peuvent, ou non, être validés. On peut concevoir que, dans le cas particulier de l'évaluation, la prescription conduite à un guidage de l'évaluation par les items du livret. On pourrait ainsi qualifier l'approche de l'évaluation induite par les progressions d'approche « par les items ». Parce que la dimension « situations » est absente, dans l'outil livret, ce guidage par les items du livret risque de se faire au détriment des situations. Or, celles-ci sont fondamentales dans une approche par

compétences dans laquelle le système éducatif s'est inscrit depuis 2005. Cela est susceptible de véhiculer une conception de la compétence dans laquelle la dimension situation est absente.

2.1.2. Autres livrets

Le livret SCEREN n'est toutefois qu'une offre, et nous avons souligné le fait que les enseignants n'ont pas de contrainte imposée en ce qui concerne le format du livret qu'ils utilisent. L'analyse que nous venons d'en faire nous a simplement permis de montrer le rôle important tenu par les items des progressions qui constituent la partie des programmes la plus opérationnelle en vue de conduire l'évaluation.

Si les enseignants n'utilisent pas nécessairement cet outil, ils peuvent utiliser la même démarche et utiliser les progressions comme référentiel pour construire leur livret comme l'institution les y encourage parfois (cf. citation de l'inspection de l'éducation nationale de Tarentaise reproduite en introduction de cette partie). Dans ce cas le livret ressemble très fortement au livret SCEREN - en ce qui concerne tout du moins la section relative à l'évaluation en CP.

Nous pouvons isoler un troisième cas. Il correspond au cas où les enseignants se réfèrent également au référentiel des progressions mais particularisent le livret en fonction de ce qu'ils ont enseigné tout au long de la période considérée.

Les entretiens post-collecte, menés auprès des enseignants, nous ont permis de leur demander comment les items du livret étaient choisis. Les réponses fournies indiquent, pour tous, une référence aux items des progressions. Cela conforte l'hypothèse initialement émise selon laquelle les items de progressions constituent le référentiel formel des enseignants pour évaluer.

Lorsque les enseignants se réfèrent directement aux progressions, les livrets sont établis pour l'année entière. Les items du livret représentent l'ensemble du programme de l'année. Certains items sont susceptibles de ne pas être évalués à chaque trimestre selon qu'ils sont considérés comme relevant du début d'apprentissage (« connaître le nom des lettres ») ou, au contraire, qu'ils sont considérés comme relevant de la fin d'apprentissage (« dire de qui ou de quoi parle le texte lu »). Au contraire, lorsque les enseignants listent les items correspondant aux enseignements dispensés et évalués sur la période considérée dans le livret, il est prévisible que tous les items soient renseignés.

Le tableau ci-dessous présente le choix de format de livret fait par chacun des 16 enseignants qui forme la population étudiée. Il présente également la circonscription et l'école d'exercice de chaque enseignant.

	Circonscription	École n°	Format de livret selon le trimestre (T) considéré
Enseignant H	2	4	Livret T1 et livret T3 différent
Enseignant I	2	4	Livret T1 et livret T3 différent
Enseignant J	2	5	Livret T1 et livret T3 différent
Enseignant K	2	5	Livret T1 et livret T3 différent
Enseignant M	2	6	Livret T1 et livret T3 différent
Enseignant O	2	8	Livret T1 et livret T3 différent
Enseignant A	4	10	Livret édité pour l'année (SCEREN)
Enseignant D	5	11	Livret édité pour l'année (SCEREN)
Enseignant F	6	12	Livret édité pour l'année (SCEREN)
Enseignant B	1	1	Livret édité pour l'année
Enseignant C	1	2	Livret édité pour l'année
Enseignant E	6	13	Livret édité pour l'année
Enseignant L	2	6	Livret édité pour l'année
Enseignant G	2	3	Livret édité pour l'année
Enseignant N	2	7	Livret édité pour l'année
Enseignant P	3	9	Livret édité pour l'année

Tableau 6 : Formats de livrets utilisés par les enseignants

Cette présentation permet tout d'abord de souligner un fait important : comme les enseignants déterminent (seul ou en équipe) le livret qu'ils utilisent, il y a presque autant de livrets différents qu'il y a d'enseignants. Cette diversité pose un certain nombre de problèmes de lisibilité pour les destinataires : entre deux enfants d'une même fratrie, entre deux enfants dans deux classes du même niveau, entre les différents enseignants qu'un élève aurait eu, les livrets et les échelles de notation peuvent différer. Sans entrer dans le détail, nous pouvons indiquer que l'on retrouve dans les données collectées un résultat décrit par Amigues et Zerbato-Poudou (1996, p. 130) au sujet du discours (évaluatif) des enseignants lequel « s'appuie sur un très large registre : des appréciations qualitatives non ordonnées (annotations de copies, appréciations sur les livrets scolaires) ou ordonnées (TB, B, AB...) à des appréciations chiffrées (notation), etc. »

Le tableau n°6 permet de remarquer que les formats de livret choisis sont multiples.

Trois sous-groupes peuvent être distingués :

1. Le premier sous-groupe est composé des 6 enseignants H ; I ; J ; K ; M ; O. Ces enseignants optent pour un livret présentant uniquement les items évalués au cours de la période considérée. Ils exercent tous dans la même circonscription (circonscription n°2). Toutefois, 3 autres enseignants exerçant dans la même circonscription optent pour un autre choix, de sorte que l'on ne peut attribuer le format de livret utilisé à un quelconque projet de circonscription. Le choix du livret apparaît de ce fait relever d'une décision personnelle plus ou moins influencée par un choix d'équipe ou une incitation de circonscription.
2. Le deuxième sous-groupe est constitué des 3 enseignants A ; D ; F. Ils optent pour le livret du SCEREN. Ils exercent dans 3 circonscriptions différentes soit dans la

moitié des circonscriptions dans lesquelles les enseignants de la population étudiée exercent. Cela permet de mettre en lumière la diffusion de ce livret.

3. Le troisième sous-groupe est constitué des 7 enseignants B ; C ; E ; L ; G ; N ; P. Les livrets qu'ils utilisent ressemblent au livret SCEREN. Les items qu'ils contiennent sont très proches des items des progressions.

Les livrets utilisés présentent, dans la plupart des cas, l'ensemble des items des progressions dès le 1er trimestre, y compris lorsque ces livrets sont conçus par les enseignants eux-mêmes (sous-groupe 3).

Il semble prévisible que les livrets de ce type seront peu renseignés au premier trimestre et qu'ils devraient l'être davantage au troisième trimestre, le programme ayant été traité en entier en fin d'année.

Les livrets utilisés par les enseignants traduisent en somme deux conceptions de l'évaluation différentes : l'une consiste à faire état de ce qui a été enseigné sur la période considérée et seulement cela (ce qui pourrait conduire à l'examen de la validité pédagogique de l'évaluation cf. I.3.1.1), l'autre consiste à faire état de ce qui a été enseigné parmi tous les items du programme (ce qui permet d'examiner la validité curriculaire de l'évaluation).

2.2. Le taux de remplissage des livrets

Nous cherchons à observer le volume d'items contenu dans chaque livret et le taux de remplissage de ce dernier. Dans le cas où les enseignants utilisent un livret identique pour l'année (i.e. les 10 enseignants formant les sous-groupes 2 et 3 précédemment définis), on s'attend à ce que tous les items qu'il contient soient évalués tout au long de l'année ce qui donne une première indication de la validité curriculaire de l'évaluation.

Le tableau suivant détaille enseignant par enseignant (groupés selon les trois sous-groupes précédemment définis), domaine disciplinaire par domaine disciplinaire (M pour Mathématiques, F pour Français) et trimestre par trimestre (T1 pour 1^{er} trimestre, T3 pour 3^e trimestre), les valeurs prises par 2 variables (L et R) et le 2 ratio (R/L) calculé à partir de ces variables.

Les deux variables choisies représentent respectivement : le nombre total d'items (L) contenus dans les livrets, et le nombre total d'items (R) réellement renseignés. A partir de ces variables on obtient le taux R/L qui exprime le pourcentage d'items (R) réellement renseignés par rapport au nombre total d'items (L) composant le livret. Ce taux correspond au taux de remplissage des livrets.

	FT1			MT1			FT3			MT3		
	L	R	R/L	L	R	R/L	L	R	R/L	L	R	R/L
Sous-groupe 1 (livret différent T1/T3)												
Enseignant H	19	19	100%	9	9	100%	19	19	100%	10	10	100%
Enseignant I	18	18	100%	8	8	100%	18	18	100%	11	11	100%
Enseignant J	26	26	100%	12	12	100%	24	24	100%	18	18	100%
Enseignant K	23	23	100%	9	9	100%	28	26	93%	16	16	100%
Enseignant M	26	26	100%	12	12	100%	23	23	100%	15	15	100%
Enseignant O	10	10	100%	11	11	100%	9	9	100%	13	13	100%
Sous-groupe 2 (livret SCEREN identique T1/T3)												
Enseignant A	37	29	78%	21	10	48%	37	33	89%	21	16	76%
Enseignant D	36	16	44%	21	9	43%	36	16	44%	21	13	62%
Enseignant F	36	25	69%	21	13	62%	36	35	97%	21	20	95%
Sous-groupe 3 (livret identique T1/T3)												
Enseignant B	26	22	85%	19	10	53%	26	25	96%	19	17	89%
Enseignant C	15	11	73%	18	11	61%	15	15	100%	18	18	100%
Enseignant E	27	15	56%	17	4	24%	27	17	63%	17	14	82%
Enseignant L	41	24	59%	27	10	37%	41	41	100%	27	25	93%
Enseignant G	21	15	71%	18	8	44%	21	19	90%	18	15	83%
Enseignant N	15	7	47%	23	10	43%	15	10	67%	23	20	87%
Enseignant P	34	17	50%	32	8	25%	34	18	53%	32	14	44%

Tableau 7 : Distribution des variables L ; R et du taux R/L par enseignant (en ligne), par domaine disciplinaire et trimestre (en colonne).

Les enseignants du sous-groupe 1 (i.e. qui établissent un livret correspondant à ce qui a été évalué sur la période considérée) disposent de façon tout à fait prévisible d'une couverture de livret de 100% à une exception près (enseignant K) pour lequel la couverture est de toute façon très élevée (93%). Leurs livrets contiennent, dans une large mesure, moins d'items que tous les autres types de livrets : par exemple le livret du troisième trimestre de l'enseignant O contient quatre fois moins d'items de Français que le livret SCEREN. On peut

émettre l'hypothèse que dans ce cas les élèves sont soumis à moins d'épreuves d'évaluation.

Les enseignants du sous-groupe 2, comme l'enseignant A dont nous reproduisons ci-dessous un extrait du livret, utilisent le livret SCEREN lequel, comme nous l'avons souligné, contient beaucoup d'items.

Pour autant le taux de remplissage du livret prend des valeurs disparates, alors même que, dans ce sous-groupe, le lien entre les progressions et l'évaluation est le plus fort. Si le taux R/L avoisine 95% pour l'enseignant F en fin d'année dans chacun des deux domaines (Français et Mathématiques), il ne dépasse pas 62% en fin d'année pour l'enseignant D, ce qui signifie que plus d'un tiers des items du livret ne sont pas renseignés en fin d'année. En Français, pour cet enseignant, moins d'un item sur deux est renseigné dans ce livret quel que soit le trimestre. Précisons que, dans tous les cas, les items renseignés au premier trimestre le sont aussi au troisième trimestre comme l'extrait de livret reproduit ci-dessous permet de le constater.

Code	N°	FRANÇAIS	T1	T2	T3
		Langage oral			
1		S'exprimer de façon correcte.	A	A	A
2		Rapporter un événement, une information très simple.	B	A	A
3		Raconter une histoire connue en s'appuyant sur des illustrations.	B	A	A
4		Décrire des images.	B	A	A
5		Reformuler des consignes.	A	A	A
6		Prendre part à des échanges en sachant écouter les autres ; poser des questions.	A	A	A
7		Réciter des comptines ou de courts poèmes (une dizaine).	A	A	A
		Lecture			
8		Connaître le nom des lettres et l'ordre alphabétique.	A	A	A
9		Connaître les correspondances entre les lettres et les sons.	A	A	A
10		Repérer graphies et syllabes dans un mot.	A	A	A
11		Connaître les correspondances entre minuscules et majuscules d'imprimerie, minuscules et majuscules cursives.	A	A	A
12		Lire aisément les mots étudiés, les mots les plus fréquemment rencontrés (mots-outils).	A	A	A
13		Déchiffrer des mots réguliers inconnus.	A	A	B
14		Lire à haute voix un texte court dont les mots ont été étudiés.	A	A	B ⁺
15		Utiliser le vocabulaire spécifique de la lecture : livre, couverture, auteur, titre...	B	A	A
16		Dire de quoi et de qui parle le texte lu.	B	A	B ⁺
		Écriture			
17		Copier un texte très court dans une écriture cursive lisible.	A	B ⁺	A
18		Écrire sous la dictée des syllabes, des mots et des courtes phrases dont les graphies ont été étudiées.	A	A	A
19		Écrire de manière autonome des mots simples.	B	B ⁺	B
20		Concevoir et écrire collectivement avec l'aide du professeur une phrase simple cohérente, puis plusieurs.	B	A	B
21		Comparer sa production écrite à un modèle, rectifier ses erreurs.	B	A	A
22		Produire un travail écrit soigné.	A	A	A
		Vocabulaire			
23		Utiliser des mots précis pour s'exprimer.	B	A	A
24		Commencer à classer les mots par catégories (personnes, animaux, choses, fruits...).			A
25		Trouver des noms correspondant à une catégorie donnée.			A
26		Trouver des mots de sens opposé (adjectif qualificatif, verbe).			
27		Ranger des mots par ordre alphabétique.			A

Figure 3 : Extrait du livret utilisé par l'enseignant A (Annexe A, p.32)

Ainsi dans le cas de l'enseignant D, le taux R/L obtenu en français conduit-il à constater que plus de la moitié des items du programme n'a pas été évalué. L'analyse du taux R/L dans les livrets des enseignants du sous-groupe 2 permet de considérer que la couverture du programme n'est pas homogène pour les utilisateurs de ces livrets et qu'elle n'est jamais égale à 100% quel que soit le trimestre considéré.

Les enseignants du sous-groupe 3 (i.e. ceux qui utilisent le même livret entre le 1^{er} et le 3^e trimestre), utilisent des livrets situés dans un entre-deux : ils contiennent plutôt moins d'items que les progressions (et donc que le livret SCEREN qu'utilisent les enseignants du sous-groupe 2) et plutôt davantage d'items que les livrets utilisés par les enseignants du sous-groupe 1. Nous constatons que l'enseignant P ne renseigne pas la moitié des items du livret qu'il a conçu. Pourtant, comme dans l'exemple reproduit ci-dessus, tous les items du 1^{er} trimestre sont renseignés au 3^e trimestre. Près d'un item du livret sur deux n'est pas renseigné dans ce livret (alors que celui-ci présente davantage d'items que le programme n'en contient). Cela conduit à constater que, comme pour l'enseignant D, environ la moitié des items n'a pas été évaluée.

2.3. Discussion

Dans la plupart des cas, les enseignants conçoivent un livret scolaire fidèlement articulé aux progressions dont ils reproduisent une large sélection, voire l'ensemble, des items comme c'est le cas pour les livrets utilisés par les enseignants des sous-groupes 2 et 3. Cela confirme le statut de référentiel formel, à partir duquel les enseignants élaborent leur propre référentiel pour évaluer, que l'on peut accorder aux progressions. Dans ces cas, nous constatons pourtant que la couverture des livrets n'est (à de rares exceptions près) pas égale à 100%. Puisque les enseignants évaluent au troisième trimestre des items également évalués au premier trimestre, cela signifie que certains items ne sont jamais évalués. Le guidage de l'évaluation par les items des progressions constitue, dans ces cas, une contrainte trop importante à respecter : il y a trop d'items à évaluer. La validité curriculaire de l'évaluation est ainsi, en première approche, une qualité que l'on ne peut attribuer à l'évaluation mise en œuvre par ces enseignants.

A l'inverse, les enseignants qui conçoivent leurs livrets en fonction de ce qui a été évalué sur la période, dégagés de cette contrainte, présentent un taux de couverture du livret de 100% (à une exception près). Cela ne permet toutefois pas d'apprécier, à l'issue de cette simple analyse, la couverture des programmes.

Nous avons questionné les enseignants lors d'entretiens post-collecte semi-directifs, à propos de l'utilité qu'ils attribuaient au livret et de la manière dont ils l'avaient conçu. Les réponses apportées permettent d'éclairer le rapport au livret qu'entretiennent les enseignants.

Nous avons successivement posé les questions suivantes : « A quoi sert le livret ? Comment est-il conçu ? »

L'enseignant F appartient à une équipe qui a choisi d'utiliser le livret SCEREN. Il souligne le poids de la prescription : « on nous demande trop, c'est délirant de remplir toutes les cases on a autre chose à faire, c'est pour cette raison que cette année on ne donnera les livrets qu'en janvier et en juin. On l'a annoncé comme ça au conseil d'école ». Dans cette école (l'école n°12), aucun élève n'a donc eu de livret au deuxième trimestre : i.e. les parents d'élèves n'ont pas eu accès à une information synthétique, émanant de l'enseignant, des progrès (ou difficultés) de leurs enfants entre janvier et juin ce qui est parfaitement non conforme à la commande institutionnelle que nous rappelions en tout début de cette partie. Ces propos permettent de considérer que l'enseignant en question utilise un outil pourtant prôné par l'institution mais qu'il perçoit comme une contrainte totalement étrangère à sa pratique : il lui semble devoir remplir toutes les cases dit-il, ce qu'il fait puisque le taux R/L de remplissage de son livret est supérieur ou égal à 95% en Français et en Mathématiques au troisième trimestre. L'enseignant I, qui pourtant conçoit un livret en fonction de ce qu'il a

enseigné sur la période considérée, indique lui aussi répondre avant tout à une obligation : « c'est parce qu'on doit évaluer qu'on le fait. On n'en a pas besoin au quotidien. Ça conforte le fait que je ne me suis pas plantée sur l'enfant, c'est pour les parents ». De la même manière, les enseignants M et L soulignent leur difficulté à remplir les livrets bien que le premier ait choisi le même format de livret que l'enseignant I (il appartient au premier sous-groupe) : « il y a trop d'items, c'est difficile de remplir tout ça, on le fait parce qu'il faut ». Le guidage de l'évaluation par les items des progressions se révèle ici contre-productif, voire totalement contre-productif dans le cas de l'enseignant F.

Ces exemples permettent de révéler une autre fonction de l'évaluation non apparente dans la réflexion sur les fonctions de l'évaluation (cf. I.1) mais non moins importante à prendre en compte : il s'agit d'une fonction purement institutionnelle liée au respect formel des obligations de service et déconnectée d'une « vraie » évaluation, plus intuitive et pragmatique implicitement évoquée plus haut par l'enseignant I. Evaluer comprend alors une dimension formelle totalement inutile à l'évaluation des apprentissages scolaires des élèves et à la communication des résultats (qui peut passer par d'autres canaux bien plus efficaces) qui correspond au fait de remplir un livret (une expression très en vogue dans les écoles désigne les livrets comme des « champs de croix » (au sens où l'enseignant coche des cases correspondantes à des degrés d'acquisitions pour chaque item)).

C'est sans doute ce qui conduit l'enseignant N à adopter une position radicalement opposée. Son livret est l'un de ceux qui présente le moins d'items (cf. tableau 7) : il contient environ deux fois moins d'items en Français que celui du SCEREN, il contient toutefois un peu plus d'item que celui du SCEREN en Mathématiques. « Je ne respecte pas les IO plus que ça. C'est clair comme ça pour les parents et pour nous » indique-t-il pour justifier le fait que son livret tient sur une seule page. On voit poindre ici la prise en compte de la dimension de communication aux parents d'élèves. Certes le livret sert à rendre compte des acquis des élèves en fonction de ce qui a été enseigné et évalué, mais il doit être recevable pour des parents qui peuvent se sentir démunis devant le nombre d'items que le livret contient.

Les livrets ne sont pas seulement le choix de l'enseignant. Ils reflètent aussi parfois des incitations faites par les inspecteurs. L'enseignant P adopte ainsi une position inverse de celle de l'enseignant N : il ajoute aux items des programmes de nouveaux items, pour mieux répondre à une commande institutionnelle : « on a eu plusieurs conférences pédagogiques imposées sur ce sujet, fallait qu'on le fasse ». Ce faisant le taux R/L de couverture du livret n'avoisine que les 50% en fin d'année en Français et en Mathématiques.

En conclusion, les livrets apparaissent symptomatiques d'une prescription primaire très imprécise voire même paradoxale en matière d'évaluation : en imposant les items des progressions comme le référentiel formel de l'évaluation via le livret SCEREN, l'institution transforme des « repères » en normes (nous avons même souligné que pour des contraintes de place probablement, la formulation des items est reprise et que certains items disparaissaient en Français notamment). La prescription ne lève pas les obstacles que rencontrent les enseignants au moment d'opérationnaliser la commande d'évaluation (que convient-il d'évaluer ? dans quel format rendre compte des résultats ?) et aboutit à conduire certains enseignants à la considérer (tout du moins dans le discours) comme une contrainte formelle extérieure à leur pratique.

Chapitre 3. Analyse de la validité curriculaire

3.1. Couverture des domaines des programmes réalisée par l'évaluation

Nous ne prenons pas en compte le sous-domaine « Langage oral » du Français. Nous n'avons en effet collecté que de rares épreuves relatives à ce sous-domaine (1 épreuve au premier trimestre, 2 épreuves au troisième trimestre sur l'ensemble du corpus). Ceci laisse à penser qu'évaluer les compétences orales se fait de façon exclusive au fil de la conduite de la classe, sur la base de critères qui ne nous ont pas été explicités au moment de la collecte.

3.1.1. Au premier trimestre

T1	Lecture	Écriture	Vocabulaire	Grammaire	Orthographe	Autre ¹⁹ F	Nombres et calcul	Géométrie	Grandeurs et mesures	OGD	Autre M	Total d'épreuves	
Enseignant A	11	3	0	0	0	2	15	5	8	0	0	44	
Enseignant B	7	3	0	2	2	5	13	2	1	1	0	36	
Enseignant C	13	2	0	0	2	0	11	0	0	0	0	28	
Enseignant D	7	2	0	0	1	1	11	3	1	0	0	26	
Enseignant E	16	3	0	0	0	3	13	3	0	0	0	38	
Enseignant F	11	0	1	1	0	3	16	4	2	1	0	39	
Enseignant G	10	5	0	0	1	0	26	1	4	0	0	47	
Enseignant H	22	4	0	0	0	0	14	1	0	0	0	41	
Enseignant I	22	4	0	0	0	0	19	1	0	0	0	46	
Enseignant J	14	3	0	1	2	2	16	4	2	0	0	44	
Enseignant K	16	0	0	0	0	0	13	1	0	0	2	32	
Enseignant L	9	1	1	3	5	0	10	1	0	1	0	31	
Enseignant M	8	6	0	0	0	0	7	1	0	0	0	22	
Enseignant N	6	2	0	1	0	1	8	2	0	0	0	20	
Enseignant O	4	2	0	0	1	1	9	4	1	0	0	22	
Enseignant P	10	2	0	0	2	4	9	2	0	0	0	29	
Total d'épreuves	186	42	2	8	16	22	210	35	19	3	2	545	
% d'épreuves	34%	8%	0%	1%	3%	4%	39%	6%	3%	1%	0%		
	51%						49%						

Tableau 8 : Nombre d'épreuves d'évaluation par sous-domaine du Français et des Mathématiques et par enseignant pour le 1^{er} trimestre

¹⁹ La catégorie Autre sert à rendre compte d'épreuves évaluant plusieurs items (dans plusieurs sous-domaines ou dans le même sous-domaine). Par exemple une épreuve d'orthographe pour évaluer l'item « écrire sans erreur sous la dictée, des mots » (écriture) et l'item « écrire sans erreur des mots appris » (orthographe).

D'une façon globale, les sous-domaines de la lecture et des nombres et du calcul concentrent la plus grande part des épreuves d'évaluation au premier trimestre : les épreuves d'évaluation liées à ces sous-domaines représentent respectivement 34% et 39 % du nombre total d'épreuves. Ces deux sous-domaines concentrent ainsi, environ, les $\frac{3}{4}$ des épreuves d'évaluation au premier trimestre (73%).

A l'inverse, les items des trois sous-domaines « vocabulaire, grammaire, organisation et gestion des données » (OGD dans le tableau) sont évalués de façon très sporadique, et les items du programme des sous-domaines « orthographe » et « grandeurs et mesures » sont peu évalués.

L'ensemble des épreuves permettant d'évaluer les items du programme correspondant à ces sous-domaines représente moins de 10% du total d'épreuves.

La rubrique Autre indique qu'un volume non négligeable d'épreuves (4% du nombre total d'épreuves) sert à évaluer plusieurs items en même temps. 10 enseignants ont au moins recours à une épreuve pour évaluer plusieurs items à la fois, mais plutôt en français.

On constate également qu'il y a à peu près autant d'épreuves destinées à évaluer les items de Français que d'épreuves destinées à évaluer les items de Mathématiques. Si l'on rapporte cela au nombre d'items des progressions (nous avons indiqué que ceux-ci contiennent environ deux fois plus d'items en Français qu'en Mathématiques), il apparaît qu'un réajustement du déséquilibre du nombre d'items consacrés au Français et aux Mathématiques dans les progressions semble s'opérer au niveau de l'évaluation.

On constate que les compétences rédactionnelles et/ou liées à l'écriture (graphie) sont (4 fois) moins évaluées que les compétences en lecture. Ce résultat peut être rapproché de l'analyse développée par l'ONL (2011, pp. 22-23) au sujet de l'apprentissage de la lecture et de l'écriture proposée dans les manuels de lecture : « les manuels récents traitent peu l'écriture. Il semble qu'on y considère l'apprentissage du lire-écrire d'une manière chronologique du lire vers l'écrire : on apprendrait à lire puis on produirait de l'écrit, comme si le transfert de la réception d'écrit vers sa production était naturel. On peut convenir que les manuels de notre sélection apprennent surtout à lire et bien moins à écrire. (...) Finalement, en production d'écrits, les propositions restent en deçà des attentes, et il n'y a pas de changements notables ces dernières années ». Cela est symptomatique d'un déséquilibre entre les sous-domaines au niveau de l'évaluation et laisserait sous-entendre que certains sous-domaines ne s'évaluent qu'en fin d'année.

On constate également qu'un nombre non négligeable d'épreuves permettent aux enseignants d'évaluer plusieurs items à la fois. Certains items étant proches, il ne semble pas incongru de les évaluer ensemble comme c'est le cas, semble-t-il, d'épreuves²⁰ qui évaluent en même temps les items n°18 d'écriture (« écrire sans erreur sous la dictée des syllabes, des mots et de courtes phrases dont les graphies ont été étudiées ») et n°34 d'orthographe (« écrire sans erreur des mots appris »). Ce résultat peut aussi s'interpréter comme une adaptation des items aux épreuves. Par exemple, une même épreuve²¹ sert à évaluer les items de lecture 9 (maîtrise du code) et 16 (compréhension) au premier trimestre.

²⁰ Epreuves 19-2 p. 154 ; 45-2 et 45-3 p. 156 ; 56-2 p. 157 et 70-1, pp. 246 Annexe A

²¹ Epreuve 113-3 p. 22 Annexe B

Nous avons souligné que les sous-domaines de la lecture et des nombres et du calcul sont les deux sous-domaines les plus évalués au premier trimestre. Mais, les variations inter évaluateurs sont toutefois importantes et contrastent ce résultat : l'enseignant H propose presque 4 fois plus d'épreuves de lecture que l'enseignant N ; l'enseignant L propose plus d'épreuves pour évaluer les compétences d'orthographe que l'enseignant O n'en propose pour évaluer la lecture. Cette prédominance peut devenir exclusive à l'image de l'enseignant K qui, en français, n'évalue rien d'autre que la lecture au premier trimestre ou de l'enseignant C qui n'évalue rien d'autre que les aspects du Nombre et du calcul en mathématiques. Certains enseignants (comme l'enseignant L) se préoccupent au contraire très tôt dans l'année d'évaluer les acquis des élèves en vocabulaire, grammaire et orthographe, alors que plusieurs autres, comme l'enseignant K, ne proposent aucune épreuve dans ces sous-domaines. Ainsi, de fortes disparités inter-enseignant sont constatées au niveau du volume d'épreuves et du nombre de sous-domaines évalués.

3.1.2. Au troisième trimestre

T3	Lecture	Écriture	Vocabulaire	Grammaire	Orthographe	Autre F	Nombres et calcul	Géométrie	Grandeurs et mesures	OGD	Autre M	Total d'épreuves	
Enseignant A	14	1	3	0	0	0	20	9	8	0	0	55	
Enseignant B	9	2	0	3	1	3	12	9	4	1	0	44	
Enseignant C	3	4	0	0	2	0	16	6	2	0	0	33	
Enseignant D	8	3	1	1	1	1	15	0	4	0	0	34	
Enseignant E	7	1	0	9	3	0	28	3	6	1	1	59	
Enseignant F	12	1	4	9	0	3	28	5	4	2	2	70	
Enseignant G	1	0	2	6	0	0	9	6	4	1	0	29	
Enseignant H	0	1	2	3	2	2	8	3	1	0	0	22	
Enseignant I	5	1	0	3	4	2	8	2	1	0	0	26	
Enseignant J	10	2	0	4	0	0	7	2	5	0	0	30	
Enseignant K	0	0	0	0	0	0	24	3	7	0	0	34	
Enseignant L	16	1	3	9	3	3	23	5	6	2	0	71	
Enseignant M	4	0	0	2	1	1	13	4	0	0	0	25	
Enseignant N	6	6	0	3	0	0	14	5	3	0	0	37	
Enseignant O	2	3	0	0	1	1	14	2	1	0	0	24	
Enseignant P	10	1	3	1	3	6	12	4	1	5	0	46	
Total d'épreuves	107	27	18	53	21	22	251	68	57	12	3	639	
% d'épreuves	17%	4%	3%	8%	3%	3%	39%	11%	9%	2%	0%		
	39%						61%						

Tableau 9 : Nombre d'épreuves d'évaluation par sous-domaine du Français et des Mathématiques et par enseignant pour le 3^e trimestre

Les épreuves d'évaluations des items des sous-domaines de la lecture et des nombres et du calcul représentent, comme au premier trimestre, la plus large part des épreuves d'évaluation : respectivement 17% et 39 %, soit 56% du nombre total d'épreuves d'évaluation au troisième trimestre. Cette part cumulée est en baisse par rapport au premier trimestre, la part de l'évaluation de la lecture diminuant de moitié entre le 1^{er} et le 3^e trimestre. Elle peut s'expliquer par l'utilisation d'épreuves plus longues.

On remarque que les items du sous-domaine de la grammaire constituent les items les plus évalués en Français après la lecture. Ce sous-domaine supplante en termes de nombre d'épreuves le sous-domaine de l'écriture. Ce dernier comporte environ autant d'épreuves que les autres sous-domaines du Français.

Le nombre d'épreuves permettant d'évaluer plusieurs items au moyen d'une même épreuve reste stable.

On constate qu'en volume d'épreuves, la part des mathématiques augmente entre le premier et le troisième trimestre : elle passe de 49% à 61%. Le domaine des nombres et du calcul reste, tout au long de l'année, le sous-domaine disciplinaire le plus évalué. En fin de CP, l'évaluation des compétences liées aux nombres et au calcul supplante très nettement en volume l'évaluation des compétences liées à la lecture. Pourtant, les deux sous-domaines contiennent chacun des items qui nécessitent un certain nombre d'épreuves mobilisant des compétences complexes (lire un texte, résoudre un problème). Quand bien même les progressions détaillent davantage les items de Français, les enseignants rétablissent le « déséquilibre » formel constaté entre le nombre d'items de Mathématiques et de Français.

Un effet de bascule semble s'opérer entre la lecture et les sous-domaines « vocabulaire, grammaire, orthographe » dont la part (13.8%) s'approche en volume de la part de la lecture (16.7%) tandis que la part de l'écriture s'amenuise (elle passe de 7.7% à 4.2%). L'écriture n'est plus la deuxième préoccupation évaluative des enseignants en nombre d'épreuves. La grammaire la supplante (en moyenne). Cette « montée en puissance » de la grammaire se matérialise par un nombre d'épreuves d'évaluation parfois même supérieure à celles utilisées pour la lecture comme c'est le cas pour les enseignants E ; G et H. Un résultat à temporiser du fait de la structure même de ces épreuves qui peuvent n'être que très ponctuelles. Il est un fait notable que, dès la fin du CP, les aspects formels de la langue (ou sous-entendus comme tels par le prescripteur qui les a constitués comme des sous-domaines détachés de la lecture et de l'écriture) supplantent largement l'écriture et atteignent quasiment le volume d'épreuves consacrées à la lecture.

3.1.3. Discussion

L'analyse de la répartition des épreuves par sous-domaines du Français et des Mathématiques et par trimestre, nous a conduit à constater que certains sous-domaines des programmes font l'objet d'un plus grand nombre d'épreuves d'évaluation que d'autres. Les sous-domaines de la lecture et des nombres et du calcul entrent dans cette catégorie a contrario de l'écriture dont la part baisse en pourcentage d'épreuves entre le 1^{er} et le 3^e trimestre (bien que le nombre d'items d'écriture (6 items) soit proche de celui de la lecture (9 items) et des nombres et du calcul (10 items)).

Ainsi, la couverture des sous-domaines des programmes évalués n'est-elle pas homogène.

Considérant la couverture des sous-domaines réalisée par l'évaluation, la validité curriculaire de l'évaluation n'est pas assurée dans bien des cas : au troisième trimestre seuls les enseignants D, L et P proposent des épreuves qui permettent d'évaluer des items relatifs à chacun des sous-domaines disciplinaires du Français, et seuls les enseignants B, E, F, L et P proposent des épreuves qui permettent d'évaluer des items relatifs à chacun des sous-domaines disciplinaires des Mathématiques. Ainsi, seuls les enseignants L et P proposent des épreuves dans chacun des sous-domaines disciplinaires étudiés en Français et en Mathématiques.

Ce résultat fondé sur la remarque d'un déséquilibre entre les sous-domaines rappelle un résultat ancien décrit par Grisay en 1992. Dans le tableau ci-dessous, nous rapportons la distribution des épreuves de Français collectées au troisième trimestre²² dans notre étude et celle constatée par Grisay au cours d'une étude portant sur les questions de français recueillies dans les épreuves d'examen de 33 classes de Première primaire (équivalent du CP en Belgique) en 1992 :

	Lecture (Déchiffrage et Compréhension)	Ecriture	Voc.	Gramm. et Conjug.	Orth.
% d'épreuves se rapportant au sous-domaine dans notre travail au troisième trimestre N = 16 classes	43%	15%	7%	21%	13%
% d'épreuves se rapportant au sous-domaine relevé par Grisay (1992) N= 33 classes	27%	12%	18%	18%	24%

Tableau 10 : Comparaison des distributions des épreuves de Français dans les tests collectés pour notre étude et ceux étudiés par Grisay (1992, p. 139)

Une certaine similitude des distributions peut être constatée entre les deux études. Dans les deux cas, la part des « compétences formelles » (cf. Grisay, 1992, p.128) (elles renvoient aux sous-domaines « grammaire, conjugaison, orthographe ») dans l'évaluation des compétences de Français est déjà importante en fin de CP et dépasse la part de l'évaluation des compétences en écriture.

La principale différence provient de l'évaluation des compétences en lecture. Nous émettons l'hypothèse que le guidage de l'évaluation par les items des progressions est de nature à l'expliquer. Le découpage du savoir-lire en une dizaine d'items dans les progressions est en effet susceptible d'occasionner l'utilisation de plusieurs épreuves pour évaluer ces items là même où, en 1992, on ne trouvait peut-être qu'une description assez générale du savoir-lire. Pour comparaison, les programmes français de 1985²³ ne mentionnent qu'un seul objectif concernant la lecture en CP : « apprentissage de la lecture ».

²² 9% d'épreuves évaluent plusieurs items principalement des sous-domaines « écriture » et « grammaire ». Nous les prenons en compte dans le tableau en répartissant équitablement cette valeur dans les deux sous-domaines.

²³ Repéré à

http://www.formapex.com/telechargementpublic/textesofficiels/1985_1.pdf?616d13afc6835dd26137b409becc9f87=6a80c47a47996b3f37db8d67a7989539 (page 7)

3.2. Couverture du sous-domaine de la lecture du domaine du Français

3.2.1. Au premier trimestre

Le tableau reproduit ci-dessous indique le nombre d'épreuves que chaque enseignant déclare utiliser pour évaluer chacun des items des progressions du sous-domaine de la lecture (numérotés de 8 à 16 comme dans le livret SCEREN pour plus de commodité, cf. II.2.1.1c). Lorsque les items du livret de l'enseignant différaient des items des progressions, nous lui avons demandé auquel de ces items il référait l'épreuve en question. Les résultats renseignent ainsi sur la validité curriculaire de l'évaluation conduite en lecture au premier trimestre.

Enseignant / N° item	8	9	10	11	12	13	14	15	16
Enseignant A	0	2	3	1	1	2	0	0	2
Enseignant B	1	1	1	0	2	1	1	0	0
Enseignant C	0	9	0	0	0	0	1	1	2
Enseignant D	1	1	1	1	1	1	1	0	1
Enseignant E	2	4	3	2	2	0	1	0	4
Enseignant F	0	1	2	1	1	1	1	1	2
Enseignant G	0	1	2	0	0	2	0	0	5
Enseignant H	0	2	0	2	5	1	1	0	9
Enseignant I	0	3	0	1	5	0	0	0	13
Enseignant J	4	1	5	0	3	1	1	0	4
Enseignant K	2	1	1	1	2	2	1	0	2
Enseignant L	0	1	2	1	4	1	0	0	1
Enseignant M	0	2	3	0	1	0	2	0	1
Enseignant N	0	2	0	0	0	0	0	0	1
Enseignant O	0	0	0	0	0	0	3	0	1
Enseignant P	0	3	0	1	2	1	0	0	3
Totaux	10	34	23	11	29	13	13	2	51

Tableau 11 : Nombre d'épreuves d'évaluation du 1^{er} trimestre pour chaque item de Lecture, par enseignant.

L'item n°15 porte sur des connaissances très localisées relative au vocabulaire du livre qui ne donnent pas nécessairement lieu à une épreuve. Si l'on ne tient pas compte de cet item, seuls les enseignants D et K évaluent l'ensemble des items de lecture des progressions au premier trimestre. Ils utilisent un livret dont le format est identique au premier et au troisième trimestre. L'enseignant D qui utilise le livret SCEREN utilise le protocole 1 item - 1 épreuve.

On pourrait concevoir qu'évaluer l'item n°16 « dire de qui ou de quoi parle le texte lu... » soit difficile à évaluer au premier trimestre parce que les élèves ne sont pas encore suffisamment experts en lecture mais les faits montrent que les enseignants s'intéressent à l'évaluation de « la compréhension de l'écrit » dès le premier trimestre quitte à réduire l'empan du texte à celui de la phrase.

Il est assez surprenant de constater que de nombreux enseignants n'évaluent pas les items n°8 ; n°9 ; n°10 et n°11 (disons en gros que ces items renvoient tous à la maîtrise des relations graphophonologiques et à des connaissances « de base » de reconnaissance des lettres et des graphies, cf. Partie IV, chapitre 3) : ce sont pourtant typiquement des items

renvoyant à des compétences de début d'apprentissage de la lecture. Seuls les enseignants D, E et K proposent des épreuves portant sur l'ensemble de ces aspects.

Seuls les enseignants B, D, F, H, J et K évaluent chacun des items n°12, n°13 et n°14 (disons en gros que ces items sont relatifs à l'identification de mots écrit et au recodage phonologique) au moyen d'au moins une épreuve.

En prenant les items des progressions pour référentiel dont on retire l'item n°15, la validité curriculaire de l'évaluation réalisée au premier trimestre en lecture n'est constatée que dans deux cas (12,5% des tests). Si l'on tient compte de cet item, alors la validité curriculaire est nulle, l'item n'étant évalué que dans de très rares cas.

3.2.2. Au troisième trimestre

Le tableau reproduit ci-dessous indique le nombre d'épreuves que chaque enseignant déclare utiliser pour évaluer chacun des items des progressions (numérotés de 8 à 16 comme dans le livret SCEREN pour plus de commodité, cf. II.2.1.1c) au troisième trimestre.

Enseignant / N° item	8	9	10	11	12	13	14	15	16
Enseignant A	0	2	1	0	1	3	2	0	5
Enseignant B	1	1	0	0	2	1	2	0	2
Enseignant C	0	0	0	0	0	0	0	0	3
Enseignant D	1	1	1	1	1	1	1	0	1
Enseignant E	1	0	1	0	0	1	1	0	3
Enseignant F	0	1	3	1	2	1	1	1	2
Enseignant G	0	0	0	0	0	0	0	0	1
Enseignant H	0	0	0	0	0	0	1 ²⁴	0	1
Enseignant I	0	0	0	0	0	0	0	0	5
Enseignant J	0	2	0	0	2	0	0	0	6
Enseignant K	0	0	0	0	0	0	0	0	0
Enseignant L	3	3	1	1	3	1	2	0	2
Enseignant M	0	0	0	0	1	1	0	0	2
Enseignant N	0	3	0	0	0	0	1	0	2
Enseignant O	0	0	0	0	0	0	1	0	1
Enseignant P	0	3	0	1	0	0	1	0	5
Totaux	6	16	7	4	12	9	12	1	40

Tableau 12 : Nombre d'épreuves d'évaluation du 3^e trimestre pour chaque item de lecture, par enseignant

Les résultats obtenus permettent de considérer qu'aucun enseignant n'évalue l'ensemble des items de lecture au moyen d'au moins une épreuve. Si l'on fait abstraction de l'item 15 (pour des raisons développées plus haut) seuls les tests proposés par les enseignants D et L sont alors valides du point de vue curriculaire.

Notons le cas très particulier de l'enseignant K qui ne propose aucune épreuve mais qui transmet tout de même un livret renseigné aux parents : l'enseignant indique avoir été débordé par des projets de classe et ne pas avoir eu le temps de procéder à une évaluation instituée. Pourtant il estime être à même de prononcer un jugement évaluatif sur la base de ses observations quotidiennes (une modalité d'évaluation prônée par le livret SCEREN dont on mesure ici l'aspect problématique). L'enseignant D utilise le même protocole au premier

²⁴ La même épreuve évalue les items n°14 et n°16

et au troisième trimestre pour évaluer « 1 item - 1 épreuve » : nous analysons cela comme la matérialisation du rôle de guidage pris par les items des progressions.

La validité curriculaire de certains tests d'évaluation apparaît très faible (<50% des items évalués) pour les tests utilisés par un peu plus de la moitié des enseignants de la population étudiée²⁵. Cette sous-population d'enseignants semble considérer que les épreuves plus complexes de compréhension d'un texte (l'item 16) évaluent aussi par nature les items portant sur ce que nous désignerons dans les parties III et IV comme des « habiletés de base ». De ce fait, ces habiletés ne sont plus évaluées en fin de CP ce qui peut être problématique pour réguler l'enseignement de la lecture auprès des plus faibles lecteurs.

3.3. Couverture du sous-domaine des nombres et du calcul du domaine des Mathématiques

3.3.1. Au premier trimestre

Le tableau reproduit ci-dessous indique le nombre d'épreuves que chaque enseignant déclare utiliser pour évaluer chacun des items des progressions du sous-domaine « Nombre et calcul » (numérotés de 1 à 10 cf. II.1.3) au premier trimestre.

Enseignant / N° item	1	2	3	4	5	6	7	8	9	10
Enseignant A	1	4	2	1	0	0	2	4	0	1
Enseignant B	3	0	2	0	1	0	1	1	0	5
Enseignant C	8	0	1	0	0	0	0	2	0	0
Enseignant D	2	0	2	1	0	0	1	1	0	4
Enseignant E	4	0	2	2	0	0	5	0	0	0
Enseignant F	2	1	4	1	1	0	2	2	0	3
Enseignant G	7	4	1	0	0	0	8	6	0	0
Enseignant H	6	0	4	0	0	0	0	4	0	0
Enseignant I	8	0	3	0	0	0	1	5	0	2
Enseignant J	10	1	3	0	0	0	1	1	0	0
Enseignant K	7	0	4	0	0	0	2	0	0	0
Enseignant L	4	0	2	1	0	0	3	0	0	0
Enseignant M	2	0	2	1	0	0	0	2	0	0
Enseignant N	3	0	3	0	0	0	1	1	0	0
Enseignant O	2	0	3	0	0	0	1	1	0	2
Enseignant P	5	1	2	0	0	0	1	0	0	0
Totaux	74	11	40	7	2	0	29	30	0	17

Tableau 13 : Nombre d'épreuves d'évaluation du 1^{er} trimestre pour chaque item du sous-domaine des nombres et du calcul, par enseignant

Les résultats obtenus permettent de considérer que la validité curriculaire de l'évaluation réalisée au premier trimestre dans ce sous-domaine n'est constatée dans aucun des tests étudiés. Cela s'explique pour partie parce que certaines compétences ne sont enseignées qu'en fin d'année de CP comme par exemple celles formulées par les items n°9 « connaître et utiliser les techniques opératoires de l'addition ... », et n°6 « connaître la table de multiplication par 2 » qui ne donnent lieu à aucune épreuve. Seul l'enseignant F évalue tous les autres items au moyen d'au moins une épreuve.

²⁵ Enseignants C, G, H, I, J, K, M, N, O et P

Certains items font l'objet d'un grand nombre d'épreuves d'évaluation à l'image des items n°1 (« connaître (savoir écrire et nommer) les nombres entiers naturels inférieurs à 100 ») et n°3 (« comparer, ranger, encadre ces nombres »). Ce résultat plaide en faveur d'une centration importante et unanime des enseignants sur la connaissance du nombre en début de CP. Une telle unanimité n'est jamais constatée en ce qui concerne le sous-domaine de la lecture.

3.3.2. Au troisième trimestre

Le tableau reproduit ci-dessous indique le nombre d'épreuves que chaque enseignant déclare utiliser pour évaluer chacun des items des progressions du sous-domaine « Nombre et calcul » (numérotés de 1 à 10 cf. II.1.3) au troisième trimestre.

Enseignant / N° item	1	2	3	4	5	6	7	8	9	10
Enseignant A	6	0	4	1	0	0	2	3	1	3
Enseignant B	0	1	4	0	1	0	1	1	1	3
Enseignant C	4	3	4	0	0	0	0	2	1	2
Enseignant D	2	1	2	2	1	0	1	1	1	4
Enseignant E	4	2	4	4	3	2	3	4	1	1
Enseignant F	9	1	4	1	2	1	2	3	2	3
Enseignant G	0	0	1	1	0	0	0	0	2	5
Enseignant H	1	0	2	0	0	0	2	0	1	2
Enseignant I	2	0	4	0	0	0	0	0	2	0
Enseignant J	0	3	0	0	0	0	0	2	2	0
Enseignant K	9	0	7	0	0	0	1	4	3	0
Enseignant L	6	0	5	2	2	1	1	2	2	2
Enseignant M	3	1	3	1	2	0	0	1	1	1
Enseignant N	4	1	2	5	0	0	0	0	2	0
Enseignant O	4	0	2	0	1	0	1	2	1	3
Enseignant P	3	2	3	1	0	0	0	2	1	0
Totaux	57	15	51	18	12	4	14	27	24	29

Tableau 14 : Nombre d'épreuves d'évaluation du 3^e trimestre pour chaque item du sous-domaine des nombres et du calcul, par enseignant

Les résultats permettent de considérer que seuls les tests proposés par les enseignants E et F sont valides du point de vue curriculaire. Comparé au sous-domaine de la lecture, l'item renvoyant à des compétences complexes (item n°10 « résoudre un problème simple à une opération ») est moins évalué : 5 enseignants²⁶ ne proposent aucune épreuve de résolution de problème au troisième trimestre. Parmi eux, 3 ne proposaient pas non plus d'épreuves d'évaluation portant sur cet item au premier trimestre : les élèves des enseignants J ; K et P ne sont pas explicitement évalués sur cette compétence en début et en fin d'année.

²⁶ Enseignants I, J, K, N et P

Synthèse de la partie II

Dans cette partie, nous nous sommes posé la question de la validité curriculaire de l'évaluation conduite par les 16 enseignants de la population étudiée, en Français et en Mathématiques, et plus particulièrement dans les sous-domaines qui concentrent le plus d'épreuves : la lecture, les nombres et le calcul.

Pour répondre, nous avons été conduit à expliciter le référentiel pour évaluer de ces enseignants. Nous avons indiqué que les items des progressions prennent le statut de référentiel alpha pour tous. Ces enseignants reconfigurent ce référentiel en suivant un processus de référentialisation propre à chacun. Cela les conduit à élaborer, ou choisir, une grande variété de livrets. Ces livrets ont en commun de décliner les sous-domaines considérés (par exemple la lecture) en plusieurs items, lesquels peuvent être considérés comme une formulation des compétences visées.

L'analyse conduite est fondée sur les items (des progressions) que les enseignants déclarent évaluer au moyen des épreuves de leur test. Cette dimension sera approfondie dans les parties suivantes mais, certains constats peuvent déjà être dressés.

Nous avons constaté que certains sous-domaines des domaines disciplinaires étudiés (Français et Mathématiques) étaient sous-représentés, voire absents de l'évaluation qui, de ce fait, n'est pas, dans de très nombreux cas, valide du point de vue curriculaire : l'évaluation ne couvre pas, dans une large part des cas, les contenus figurant dans les programmes. Quel que soit le trimestre considéré aucun test ne couvre l'ensemble des deux sous-domaines étudiés.

Nous nous sommes ensuite intéressé à la validité curriculaire des tests utilisés pour évaluer les items des sous-domaines de la lecture, des nombres et du calcul. Pour cela nous nous sommes plus particulièrement intéressé aux 10 enseignants qui élaborent un livret fidèle aux progressions (ils présentent tous les items des progressions pour chacun des trois trimestres de l'année, cf. II.2). Nous avons constaté que :

- Seuls deux tests (D et L²⁷) couvrent l'ensemble les items de lecture au troisième trimestre²⁸.
- Plus précisément, l'évaluation de la compréhension de l'écrit (texte ou phrase) est une préoccupation des enseignants dès le début de l'apprentissage comme en témoigne le volume d'épreuves qui se rapportent à l'item n°16. Les items relevant des habiletés « de base » en lecture (connaissances d'ordre graphophonologiques, connaissance des lettres et des graphies de lettres, items n°8 à 11) ne constituent en revanche pas, au premier trimestre, une base partagée de l'évaluation de la lecture (3 enseignants sur 16 proposent au moins une épreuve relativement à chacun de ces items).
- Seuls deux tests (E et F) couvrent l'ensemble des items du sous-domaine « Nombre et calcul » au troisième trimestre.

²⁷ Pour simplifier la lecture, nous désignons le test utilisé par un enseignant et l'enseignant par la même lettre.

²⁸ Exception faite de l'item n°15 pour des raisons développées plus haut (II.3.2.1)

- Plus précisément, l'item n°10 relatif à la résolution de problème n'est pas évalué par tous les enseignants aux deux trimestres considérés. Ceci constitue un résultat à rapprocher du précédent en matière de compréhension en lecture : les compétences les plus expertes (ou complexes) sont davantage évaluées en lecture qu'en mathématiques. Contrairement à la lecture également, deux items sont unanimement évalués au premier trimestre au moyen d'une épreuve au moins (les items n°1 et n°3 relatifs à la connaissance des nombres). L'évaluation de ces connaissances semble constituer une base partagée de l'évaluation des connaissances en mathématiques.

Pour résumer ces résultats, nous pouvons indiquer que les différents organisateurs du travail de l'enseignant en matière d'évaluation (programmes, progressions, livret SCEREN ...) s'accordent mal. Ils conduisent les enseignants à déterminer des référentiels conséquents en termes de nombre d'items, proches de la liste d'items fournie dans les progressions annexées aux programmes.

Ces référentiels résultent tous de la mise en œuvre de processus individuels de référentialisation conduits à partir des items des progressions. Cela a pour effet l'adoption par les enseignants d'une liste d'items non hiérarchisés, et dont la couverture n'est pas assurée au sens strict par les épreuves d'évaluation dans la plupart des cas, aux trimestres et dans les domaines considérés. La validité curriculaire de l'évaluation ne caractérise ainsi pas l'évaluation interne instituée telle qu'elle est mise en œuvre par les enseignants qui forment la population étudiée.

Si les progressions constituent de fait le référentiel formel alpha pour évaluer des enseignants, la qualité de la reconfiguration de ce référentiel est remise en question par l'analyse de l'évaluation réalisée. Nous avons de surcroît souligné que le statut de référentiel des progressions est remis en cause dans le discours par certains enseignants. Nous avons en effet relevé que certains enseignants utilisent le référentiel des progressions en réponse à ce qu'ils perçoivent davantage comme une commande formelle (l'évaluation revêt dans ce cas une nouvelle fonction de « répondre à une commande institutionnelle ») qu'une nécessité pédagogique renvoyant à la dimension de régulation de l'enseignement.

L'analyse conduite dans ce chapitre permet de mettre en avant une partie importante de l'activité évaluative des enseignants qui consiste à sélectionner à l'intérieur du référentiel (celui des progressions mais aussi le leur) ce qu'il faut évaluer en fonction de critères ou de besoins personnels que nos données ne nous permettent pas d'explorer mais qu'un travail complémentaire pourrait permettre d'explorer.

Nous poursuivons notre analyse de la validité de l'évaluation en nous intéressant à présent finement aux tâches engagées par les épreuves. Elles sont le support de l'activité de l'élève. Une analyse fine des tâches permet de faire des hypothèses sur l'activité de l'élève au cours du traitement de la tâche, elle permet une « descente vers le cognitif » chère à Vergnaud (2001, p.47). Pour lier cette analyse à l'analyse des référentiels pour évaluer utilisés par les enseignants conduite dans cette partie, nous examinons l'adéquation de l'épreuve à évaluer l'item auquel l'enseignant la rattache. Pour cela, nous nous intéressons aux caractéristiques de l'épreuve en nous préoccupant de l'activité de réponse (supposée) de l'élève.

Partie III – Étude de la validité psychodidactique des épreuves

Dans cette partie, nous nous intéressons aux tâches utilisées par les enseignants pour évaluer, dans le but d'en étudier la validité. Nous adoptons pour cela un point de vue psychodidactique. Caractériser la validité des épreuves selon cette approche conduit, pour Vantourout et Goasdoué (2014), à considérer les tâches prises individuellement, et nécessite de s'intéresser « aux contenus ²⁹ impliqués dans les évaluations, mais aussi au fonctionnement cognitif des évalués, aux processus de réponse provoqués par la tâche, tâche dont les caractéristiques doivent être analysées ».

Chapitre 1. Cadre conceptuel d'analyse de la tâche

Le concept de tâche est un concept central en psychologie ergonomique à partir duquel on peut questionner et analyser l'activité d'un sujet. Il a été théorisé par Leplat et Hoc (1983) et de nombreux travaux portant sur l'activité des enseignants y ont recours (Rogalski 1999, 2000, 2003, 2007 et Robert 2000, 2001, 2002 en mathématiques, Goigoux, 2001 en français). Il permet d'examiner ce qu'un sujet fait par rapport à ce qu'il doit faire et aux conditions dans lesquelles il doit le faire³⁰.

1.1. La tâche, première approche

Brousseau (2004, p.2) souligne qu'une tâche est tout en même temps « une suite d'actions, une activité et/ou son résultat » et « n'est pas ce que le sujet accomplit effectivement ». La tâche se distingue ainsi de l'activité, ce qui est clairement formulé par Leplat et Hoc (1983, p. 50), « la tâche indique ce qui est à faire, l'activité, ce qui se fait »³¹. Ces auteurs distinguent ce qui relève de l'activité du sujet au cours de la réalisation d'une tâche, de la tâche que l'on peut en première approche assimiler à un travail à réaliser, « pour répondre à la demande et aux attentes de la personne qui l'a donné à faire et qui en sera généralement le destinataire et/ou l'évaluateur » indique Delcambre (2007, p.217). La tâche se distingue donc fondamentalement de l'activité (cf. Rogalski, 2003) même si elle entretient un rapport étroit

²⁹ Les référentiels didactiques correspondant aux sous domaines des nombres et du calcul en mathématiques de la lecture en français, sont respectivement explicités en pages 186 à 196 et 226 à 230, puis en pages 252 à 275. Ils ont été également utilisés pour conduire l'analyse de la VPD présentée dans cette partie.

³⁰ Dans une approche psychologique ergonomique, on peut entreprendre l'analyse de l'activité en croisant une analyse du point de vue de la tâche prescrite avec une analyse du point de vue de l'agent, i.e. à partir de ses caractéristiques propres, indépendamment des propriétés de la tâche. Cf. Leplat, 1997, p.14-15.

³¹ Leplat indique qu'une ambiguïté peut provenir du terme anglophone « task ». « Dans « task analysis » « task » correspond à notre acception du mot tâche. « task description » ou « performance analysis » correspond à notre analyse de l'activité » (2000, p. 19)

avec elle. L'activité ne peut pas être directement appréhendée : on ne peut qu'en inférer une partie à partir d'observables.

L'analyse psychologique du travail permet de considérer que si l'activité est effectivement guidée par la tâche, « elle n'est pas pour autant contenue en elle » (Clot, 1995, p. 210) : elle provient de l'interaction de celui à qui elle est prescrite, avec la tâche. Pour Leplat (2000, p.7), la tâche est « le produit d'une activité en même temps qu'une source de l'activité ». Cela permet de préciser les propos de Brousseau : au cours de la réalisation de la tâche, une redéfinition de la tâche est engagée par le sujet qui nécessite de préciser la notion de tâche afin de savoir à quelle tâche répond son activité.

De ce fait, des décalages sont susceptibles d'être constatés entre ce que le concepteur/prescripteur attend de l'agent et ce que l'agent réalise i.e. entre la tâche prescrite et la tâche réalisée.

Rapporté à notre étude, cela revient à rechercher à quelle tâche répond l'élève, et à savoir si la tâche prescrite permet à chaque élève d'engager une activité ciblée par l'enseignant permettant à ce dernier d'évaluer des compétences précises.

1.2. A quelle tâche répond l'activité ?

L'analyse psychologique du travail³² s'est intéressée à trois notions essentielles : l'agent, la tâche et l'activité. Analyser le travail correspond, dans cette approche, à analyser les relations que ces trois notions entretiennent entre elles : une entreprise difficile car, selon Leplat (1997, p.14), « l'étude de chaque terme ne peut être conduite qu'en référence aux autres qui agissent sur lui et sur lesquels il agit ». L'activité dépend de la tâche et des caractéristiques de l'agent, mais elle peut contribuer en retour selon cet auteur (idem), « à la définition de la tâche et à la transformation de l'agent ». Ce point de vue conduit à estimer que deux agents dont l'expertise diffère n'exécutent pas la même tâche : ils poursuivent le même but mais les conditions qu'ils prennent en compte ne sont pas identiques. La tâche effectivement réalisée correspond à celle que se fixe l'agent à partir de la tâche prescrite, suite à une redéfinition de celle-ci, et de ses compétences.

Selon cette approche, la tâche prescrite n'est donc pas nécessairement celle à laquelle répond l'agent. Citant les travaux pionniers d'Ombredane et Faverge (1955) ; Haackman (1969) et Leplat et Hoc (1983), Leplat (1997, p.24) rappelle ce résultat fondamental des recherches portant sur l'analyse du travail : « la tâche effectivement réalisée par l'agent (la tâche effective) ne coïncide pas toujours avec la tâche prescrite ». Cette perspective conduit à analyser l'activité en termes de tâche, « non pas de tâche prescrite, mais de tâches que se fixe ou que réalise l'agent » (idem, p.16).

Leplat propose un modèle d'analyse psychologique de l'activité centrée sur la tâche dans lequel il montre comment, à partir de la tâche prescrite, se constitue la tâche à laquelle répond l'activité (cf. 1997, p.17). La tâche y est définie comme « un but à atteindre dans des conditions déterminées » (idem).

Il distingue tout d'abord la tâche à réaliser de la tâche prescrite. La tâche à réaliser est marquée du sceau des attentes de celui qui gère ou conçoit la tâche prescrite. La tâche prescrite en est une formulation à destination de celui qui doit l'exécuter. Leplat (2000, p.17) souligne que la transposition de la tâche à réaliser à la tâche prescrite constitue elle-même un problème car celui qui définit la tâche prescrite attend une certaine activité, un certain

³² La psychologie du travail considère le travail comme une activité

résultat. Or, son modèle (de l'activité de l'agent) peut être erroné en ce sens que « la tâche exécutée, même si elle suit fidèlement les instructions, n'est pas celle que le concepteur voulait faire réaliser » (idem). Le concepteur est amené à privilégier une transposition de la tâche à réaliser qu'il juge optimale, compte-tenu de la représentation qu'il se fait des moyens de cette exécution (1997 p.19), ce qui est lié au modèle qu'il se fait de l'agent. Au cours de la transposition, ou du passage, de la tâche à réaliser en tâche prescrite, le concepteur explicite plus ou moins la tâche prescrite (cf. figure 1 ci-dessous). Le degré d'explicitation dépend de la nature de la tâche indique Leplat (idem, p.21) : les tâches les plus simples peuvent donner lieu à des modes opératoires finement décrits, les tâches les plus complexes sont plus difficiles à procéduraliser et dans ce cas, on s'en remet à la compétence de l'agent pour découvrir les procédés ou procédures permettant d'atteindre le but.

S'intéressant à l'analyse de l'adéquation de la tâche prescrite à la tâche à réaliser, Leplat souligne (1997, p.22) que l'on peut concevoir une critique interne de la tâche prescrite indépendamment de l'activité : « elle consiste à examiner la cohérence et la faisabilité de l'exécution par l'application systématique de la prescription ». La tâche est-elle réalisable et ne contient-elle pas d'éléments qui conduisent à des impossibilités dans l'exécution ?

Cet auteur indique (idem) qu'une deuxième forme de critique consiste à « examiner dans quelle mesure la tâche prescrite est adéquate à celle qui est effectivement à réaliser, c'est-à-dire à évaluer la pertinence de la transposition ». Dans quelle mesure la tâche prescrite exécutée systématiquement est-elle compatible avec les moyens disponibles ?

La figure ci-dessous permet de représenter le passage de la tâche à réaliser à la tâche prescrite (pour l'agent) :

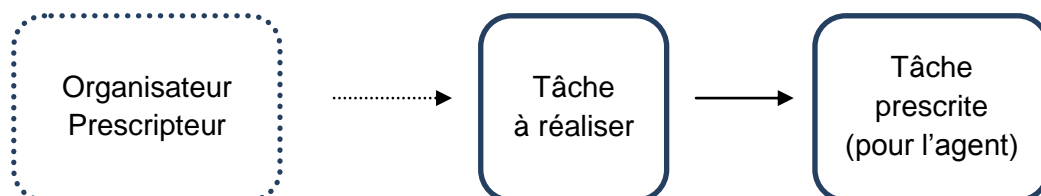


Figure 1 : de la tâche à réaliser à la tâche prescrite (d'après Leplat 1997, p.17 et 2000, p.17).

Du côté du concepteur,

- la tâche à réaliser correspond à une prescription émanant d'un organisateur ou système prescripteur, faite au concepteur.
- la tâche prescrite³³ est la tâche « fixée par celui qui en demande l'exécution » Leplat (2000, p.17).

L'analyse psychologique du travail s'intéresse à l'activité qui vise à l'exécution de la tâche prescrite. La connaissance de la tâche prescrite est donc une donnée essentielle pour se demander comment l'agent répond à la tâche, comment il la transforme éventuellement en fonction de ses caractéristiques et de ses propres finalités. Leplat souligne (1997, p.25) que l'activité est constituée « de l'exécution même de la tâche et des représentations qui l'accompagnent et la guident » : il y a interaction entre l'agent et la tâche prescrite. « L'agent n'est jamais un pur exécutant de la tâche prescrite » (idem, p.26) : il la redéfinit.

³³ Ou «tâche attendue » selon Rogalski, 2003. Nous privilégierons les termes choisis par Leplat pour faciliter la lecture.

Il y a toujours de l'implicite dans la tâche prescrite. La redéfinition de la tâche consiste en premier lieu, pour l'agent, à opérationnaliser celle-ci en fonction des conditions présentes. Leplat indique par ailleurs que la tâche s'inscrit dans l'histoire de l'agent (il peut viser d'autres buts que la réalisation de la tâche comme par exemple un but personnel) et que l'agent redéfinit la tâche prescrite par rapport à celle des autres : « la tâche redéfinie apparaît alors comme un compromis entre la tâche prescrite, des « rationalités » (par rapport à des objectifs matériels, par rapport à l'accomplissement de soi), et la qualité du vivre ensemble dans le travail » (idem, p.29). Nous ne développerons pas d'analyse de la tâche selon cette dimension dans notre travail.

Différentes caractéristiques de la tâche et de l'agent conduisent à une redéfinition de la tâche prescrite. Parmi celles-ci, Leplat cite la compétence de l'agent : « une source importante des décalages entre la tâche prescrite et la tâche redéfinie est liée à l'adaptation des compétences à la tâche prescrite » (1997, p.30).

Différents écarts existant entre la tâche prescrite et la tâche effectivement réalisée par l'agent peuvent être considérés, notamment (Leplat, 1997, p.32) :

- des écarts par réduction, lorsque tout ce qui figure dans la tâche n'est pas réalisé mais ce qui est réalisé fait partie de la tâche.
- des écarts par adjonction, lorsque l'agent a réalisé la tâche mais a ajouté d'autres choses
- des écarts par substitution, lorsque la nature et l'organisation des actions sont modifiées.

La figure ci-dessous permet de représenter le passage de la tâche prescrite pour l'agent à la tâche réalisée par l'agent :

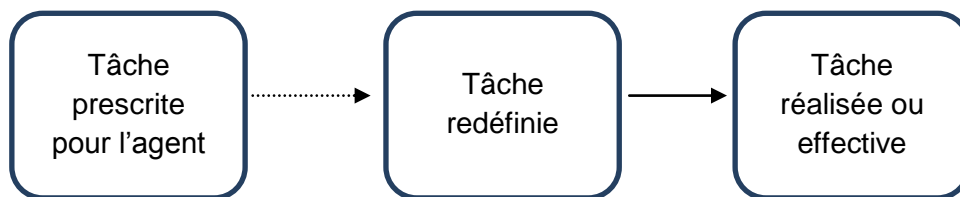


Figure 2 : de la tâche prescrite à la tâche réalisée (d'après Leplat 1997, p.17 et 2000, p.17).

Du côté de l'agent,

- La *tâche redéfinie* correspond à l'intention de l'agent, à ce qu'il propose de faire. Rogalski (2003) la qualifie comme « la représentation de la tâche que se donne le sujet »
- La *tâche effective* ou *tâche réalisée* (Leplat, 1997, p.17) est celle « dont l'activité serait l'exacte réalisation » (Leplat, 2000, p.18). Elle diffère de la tâche prescrite « par les moyens et les contraintes voire par l'interprétation des buts » précise Rogalski (idem).

1.3. Adaptation du modèle pour notre étude

Notre travail porte sur l'analyse d'épreuves d'évaluation des connaissances des élèves. Dans ce cadre, l'épreuve est une tâche (elle constitue un but à atteindre dans des conditions déterminées), l'enseignant en est le concepteur/prescripteur (même s'il l'emprunte à d'autres concepteurs) et les élèves sont les agents.

Les travaux de psychologie ergonomique que nous venons de présenter mettent en avant l'existence d'écarts entre les différents niveaux de la tâche.

Nous ne nous intéressons pas directement à l'analyse psychologique de l'activité des élèves ni des enseignants : en effet nous ne sommes pas en mesure d'analyser l'activité au moyen

des données recueillies qui ne sont que des traces d'activité et de la tâche réalisée. Nous nous servons toutefois de ce cadre pour qualifier et expliciter les écarts susceptibles de se produire entre la tâche à réaliser et la tâche prescrite et la tâche réalisée par les élèves. L'enjeu est d'essayer de mettre en lumière les écarts susceptibles d'être rencontrés entre ces différents niveaux de la tâche et, d'explicitier des éléments contenus dans la tâche prescrite susceptibles de les générer ou de les renforcer.

Deux niveaux de la tâche nous intéressent : celui de l'enseignant, celui de l'élève.

Au niveau de l'enseignant, la tâche à réaliser en ce qui concerne l'évaluation est multiple. Elle relève du contrat professionnel (Rogalski, 2007) : déterminer ce qu'il faut évaluer et, déterminer une épreuve qui permette de le faire (ce qui implique de déterminer *a priori* les comportements de réponse des élèves par rapport aux connaissances visées). Les livrets et les épreuves constituent deux observables de l'activité évaluative de l'enseignant :

- Le livret contient un certain nombre d'items qui sont les formulations des compétences évaluées. Il correspond à une double prescription du système prescripteur : évaluer et communiquer le résultat de l'évaluation.
- Les tâches prescrites aux élèves (les épreuves). Elles sont, pour l'enseignant, directement en lien avec le premier observable. Idéalement, l'activité de l'élève engendrée par chaque épreuve correspond à celle que l'enseignant souhaite mobiliser chez les élèves.

Au niveau de l'élève, l'activité se matérialise par une redéfinition de la tâche prescrite. L'élève élabore une représentation de la tâche, qui suivant Julo (1995, p. 11), est « le résultat d'une véritable activité mentale mettant en œuvre tout un ensemble de processus chargées de traiter les informations issues de cet environnement ». La copie que rend l'élève est l'observable qui résulte de la tâche réalisée. Elle est souvent l'observable à partir duquel l'activité de l'élève est inférée par l'enseignant notamment du point de vue de ses connaissances, de ses compétences.

Analyser une épreuve d'évaluation consistera donc, pour nous, dans ce qui suit, à analyser ces différents niveaux de la tâche.

1.4. Vers un cadre méthodologique pour analyser les épreuves d'évaluation

Considérant l'épreuve comme une tâche prescrite à l'élève, l'analyse de la tâche prend deux directions principales : celle du but et celle des conditions. Deux types de conditions de la tâche peuvent être distingués suivant leur lien (direct ou indirect) aux contenus mobilisés.

La première direction de l'analyse porte sur le but de la tâche et la transposition ou le passage de la tâche à réaliser de l'enseignant en tâche prescrite à l'élève. Cette analyse se place au niveau de la tâche qui concerne exclusivement l'enseignant. Quel référentiel utilise-t-il pour choisir les épreuves d'évaluation ?

La deuxième direction de l'analyse porte sur les conditions de la tâche. Elle se place au niveau de la tâche qui concerne l'élève :

- conditions liées aux aspects conceptuels sur lesquels repose la tâche prescrite. Quels connaissances, compétences sont mobilisées et évaluées d'un point de vue didactique ?
- conditions liées à d'autres composants de la tâche, susceptibles de générer un écart important entre la tâche prescrite à l'élève et la tâche effectivement réalisée qui ne seraient pas apparues au cours de l'analyse du contenu didactique de la tâche. L'analyse devra ainsi prendre en compte les réponses des élèves, ainsi qu'un certain nombre de phénomènes, ou de conditions, en ce qu'ils sont susceptibles de prendre part à la redéfinition de la tâche prescrite et susceptibles d'être défavorables aux élèves les plus sensibles à la formulation des énoncés (cf. Van Dooren, Verschaffel, Greer et coll., 2010, p.161). Ainsi, par exemple, « l'habillage de la tâche » au sujet duquel Goigoux met les enseignants en garde : « l'habillage excessif des tâches scolaires gêne le traitement des informations essentielles (l'attention est attirée par des détails non pertinents pour la résolution) » (1998b). Se pose précisément la question de savoir distinguer l'habillage de la tâche avec la tâche elle-même. Selon Jonnaert et Laveault (1994, p. 273), « l'habillage d'une situation est l'organisation de ces composantes en un énoncé intelligible pour le sujet, et susceptible de provoquer chez ce dernier une activité cognitive ». Pour ces auteurs l'habillage vise l'intelligibilité, terme qu'il faut également définir. Leplat (2004, p. 203) distingue plusieurs aspects de l'intelligibilité d'un document prescripteur : l'intelligibilité textuelle (le document veut-il bien dire ce que son concepteur veut lui faire dire ?), l'intelligibilité du message (est-il compréhensible ?) et enfin son intelligibilité fonctionnelle (conduit-il les sujets à comprendre ce qui est prescrit ?).

Dans le cadre des approches psycho-didactiques de l'évaluation, apprécier la validité d'une tâche prescrite par l'enseignant à l'élève consiste dès lors à apprécier :

- l'adéquation de la tâche prescrite à l'élève à celle qui est effectivement à réaliser par l'enseignant, (c'est-à-dire à évaluer la pertinence de la transposition de la tâche à réaliser de l'enseignant en tâche prescrite à l'élève)
- le potentiel de la tâche prescrite à engager l'élève dans une tâche effective qui corresponde bien à ce pourquoi elle est prescrite à l'élève.

Chapitre 2. Méthodologie d'analyse psycho-didactique d'une épreuve d'évaluation

2.1. Méthodologie

En nous appuyant sur le cadre précédemment décrit, nous avons élaboré une méthodologie d'analyse des épreuves d'évaluation et des tâches prescrites, dans le but de pouvoir apprécier leur validité psycho-didactique.

Trois niveaux d'analyse sont distingués :

1. l'item du livret,

Dans un premier temps, nous recherchons à déterminer l'item (ou les items) du livret que l'enseignant utilise et auquel il rattache l'épreuve. L'item peut être assimilé à une formulation de la compétence à évaluer. Il est l'énoncé le plus explicite de ce que l'enseignant souhaite évaluer au moyen de l'épreuve qu'il choisit ou conçoit. Il constitue le référentiel de l'évaluation pour l'enseignant. Il est le produit du passage de la tâche à réaliser fixée par l'institution (cf. figure 2).

2. l'épreuve

Nous analysons dans un second temps l'épreuve, une tâche prescrite à l'élève. Pour ce faire, nous entrecroisons deux types d'analyses :

- Le premier type d'analyse renvoie à une analyse dite « *a priori* », classique en didactique des mathématiques. Une telle analyse se préoccupe classiquement de procédures susceptibles d'être mises en œuvre par les élèves, de variables de commandes spécifiques (les variables didactiques), d'effets de ces variables sur les procédures. Elle est conduite sous le double regard de l'épistémologie et des données d'expériences antérieures. En ce sens, ce qui nous permet d'élargir au domaine du français ce type d'analyse, le terme d'analyse préalable pourrait être préféré à analyse *a priori*. Cette analyse permet d'observer si le savoir visé par l'enseignant et qu'il exprime par l'item est effectivement mobilisé.
- Le deuxième type d'analyse porte sur des éléments rarement pris en compte par l'analyse *a priori* parce que peu, voire pas du tout, ancrés dans les contenus, comme par exemple des éléments d'ordre contextuel (cadre de réponse suggéré ou imposé par la tâche, habillage de la tâche ...). Pour le cas très particulier du CP, nous ne perdrons ainsi jamais de vue que l'évaluation s'adresse à des élèves qui sont dans la première année de l'enseignement obligatoire de la lecture/écriture en particulier et qu'à ce titre, ils ne peuvent être considérés comme des lecteurs/scripteurs experts. L'analyse permet d'observer si la tâche est présentée de façon à ce que l'élève puisse mobiliser les connaissances évaluées, sans que d'autres connaissances intervenant comme un « écran » ne s'interposent.

Cette double analyse permet de faire des hypothèses sur la tâche redéfinie par l'élève à partir de la tâche qui lui est prescrite par l'enseignant.

3. la copie de l'élève

Dans un troisième temps, nous prenons en compte des éléments de réponse des élèves provenant des copies (ce sont des traces de l'activité de l'élève et de la tâche réalisée). Ils permettent d'affiner l'analyse psycho-didactique de la tâche prescrite. Ils fournissent

éventuellement de nouveaux indices pour pouvoir formuler de nouvelles hypothèses, infirmer ou confirmer les précédentes, sur la tâche réalisée par l'élève..

Du point de vue psycho-didactique, la tâche prescrite est valide si elle permet bien d'engager (de par les contenus mis en jeu, de par leur mise en contexte etc.) une activité de l'élève correspondante à ce que l'enseignant cherche à observer du point de vue des connaissances visées (ce qui est formulé par l'item).

L'analyse de réponses d'élèves peut participer à l'établissement de cette validité : elle peut notamment permettre de comprendre le recours à des procédures inadéquates lors de la redéfinition de la tâche par l'élève qui seraient imputables à la tâche elle-même (et non aux concepts qu'elle recouvre) comme par exemple la présentation des données de la tâche, le système de réponse etc.

2.2. Application à deux épreuves de référence

Nous illustrons cette méthodologie d'analyse pour les deux épreuves suivantes :

Épreuve n°1

a- l'item :

Comme indiqué en bas de l'épreuve reproduite dans la figure 4, l'enseignant P évalue l'item du livret « *lire une consigne, un énoncé et comprendre ce qui est attendu* ». Cet item ne renvoie à aucune compétence listée dans les programmes : il est ajouté par l'enseignant aux items du livret. L'enseignant indique lors d'un entretien post-collecte qu'il considère nécessaire cette déclinaison particulière de l'item « *Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens* ». Pour cet enseignant, comprendre une consigne est une déclinaison particulière de la compréhension en lecture qui justifie de positionner, dans le livret, un item spécifique. On peut y voir ici un effet de la prescription secondaire quant à l'importance à accorder au traitement des consignes (cf. Dossier « Attention aux consignes » n°483, *Cahiers Pédagogiques*). On peut aussi y voir l'influence d'une compétence « lire seul et comprendre un énoncé, une consigne simples » attendue en fin de cycle 2 (cf. Compétence 1 du Palier 1 du Socle commun de compétences et de connaissances – MEN 2008, p. 15). Quoiqu'il en soit, l'institution ne demande pas spécifiquement aux enseignants d'évaluer ceci de façon explicite en CP bien qu'il l'encourage d'une certaine façon en indiquant cet attendu en fin de CE1.

b- l'épreuve :

Pour évaluer l'item au troisième trimestre, l'enseignant utilise l'épreuve ci-après :

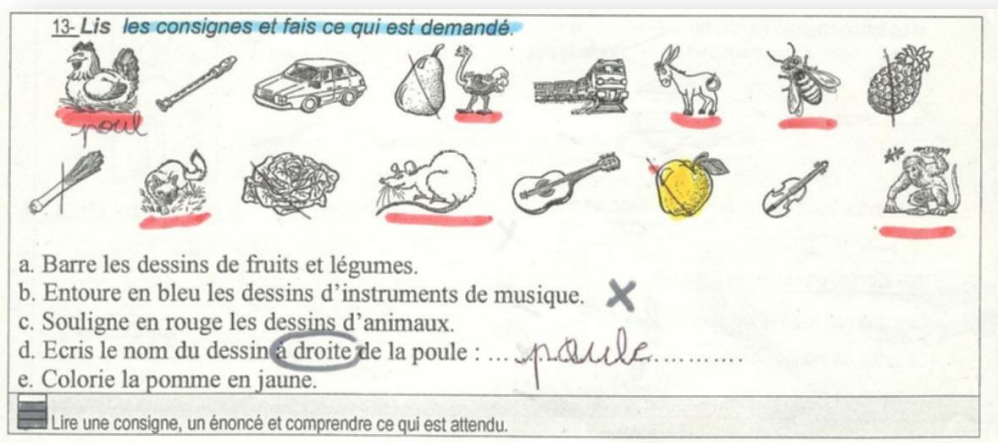


Figure 4 : Épreuve FT3 (Annexe B, p. 156) avec réponses d'élève

Pour réussir l'épreuve, l'élève doit lire la « méta-consigne » de l'épreuve : « lis les consignes et fais ce qui est demandé ». Celle-ci n'est pas lue par l'enseignant.

L'élève doit comprendre que la consigne de la tâche porte sur d'autres consignes (les consignes a, b, c, d et e) dont l'exécution indique, pour l'enseignant, la capacité de l'élève à « lire une consigne ». L'élève doit lire la série de consignes qui ne sont pas liées les unes aux autres, puis doit les exécuter.

Par exemple, après avoir lu la consigne a, « Barre les dessins de fruits et de légumes », l'élève doit barrer les dessins de la poire, de l'ananas, du poireau de la salade et de la pomme. Il est présumé par le concepteur de l'épreuve que les élèves accordent la valeur « poireau » au dessin situé en bas à gauche. Les verbes (mode impératif) sont placés en tête des phrases comme cela est souvent le cas dans les consignes, ce qui renvoie à un format de tâche usuel. La série de consignes engage l'élève à traiter simultanément plusieurs sous-ensembles (les fruits et les légumes), dans leur totalité, ce qui est formulé de façon implicite (il faut barrer (*tous*) les dessins d'instruments de musique, il faut entourer (*tous*) les dessins d'animaux). La tâche prescrite comprend à ce titre une tâche de catégorisation qui n'est pas envisagée pour elle-même (on pourrait ici questionner la fidélité de la représentation des dessins et la reconnaissance du sens que leur attribuent les élèves).

Réussir cette tâche nécessite d'avoir lu et réalisé la consigne de l'épreuve qui elle-même demandait de réaliser toutes les consignes.

Plusieurs décalages sont constatés :

- La tâche prescrite porte sur un système de consignes. La consigne « lis les consignes et fais ce qui est demandé » implique d'en réaliser d'autres, de sorte que c'est la réalisation des autres consignes qui signe la réalisation de la méta-consigne de la tâche.
- Les consignes peuvent par ailleurs être multiples. La consigne d nécessite de déterminer un mot qu'il faut ensuite écrire : ne pas l'avoir écrit peut signifier que l'élève l'ayant trouvé n'a pas su l'écrire, ou qu'il ne l'a pas trouvé. Il est attendu que

l'élève écrive le mot « flûte ». Ce mot peut poser des problèmes d'encodage à de nombreux élèves de CP³⁴.

c-la copie de l'élève:

Les réponses de l'élève permettent de constater que deux consignes ont été incorrectement réalisées : la consigne b (absence de réponse) et la consigne d.

-L'élève n'a pas exécuté la consigne b. A quoi attribuer l'absence de réponse ? On ne peut pas savoir s'il n'a pas su déchiffrer les mots qui compose la consigne, ou si l'effort de déchiffrement a été tel que, parvenu au terme de la phrase, l'élève ne s'est pas réengagé dans une relecture, ou bien encore s'il a simplement omis de la traiter en sautant de la consigne a à la consigne c sans considérer la consigne b.

Analysons la difficulté que le mot « instruments » représente en termes de déchiffrement (c'est la première partie de la consigne générale de l'épreuve « lis les consignes et fais ce qui est demandé »). C'est un mot très rarement rencontré par les élèves³⁵ et qui comporte de nombreuses difficultés du point de vue du recodage³⁶. L'incapacité à décoder ce mot est de ce fait prévisible pour un certain nombre d'élève : elle entraîne l'impossibilité de lire la consigne donc de la traiter.

-L'élève n'a pas correctement exécuté la consigne d. Nous pouvons toutefois analyser son échec. Deux possibilités sont envisageables pour « Écris le nom du dessin à droite de la poule ». Elles témoignent de deux interprétations possibles d'une partie de la consigne totalement laissée à la charge de l'élève.

On peut comprendre qu'il faut écrire à droite du mot « poule » sur les pointillés, le mot « flûte » (i.e. le nom du dessin situé à droite du dessin de la poule).

On peut comprendre aussi qu'il faut écrire à droite du dessin de la poule, le mot « poule » (i.e. le nom du dessin) ce que l'élève a probablement compris. N'ayant pas la place à droite du dessin, il a recopié le mot « poule » (en omettant le e final) en dessous. Puis, il a réécrit le mot sur les pointillés. Cela peut se comprendre comme un effet de contrat : habituellement les pointillés sont destinés à l'emplacement réservé à l'écriture d'une réponse.

Cette analyse permet de considérer que la tâche pose différents problèmes :

- l'absence de réponse à la consigne b ne peut être, de façon certaine, attribuée à une incapacité de l'élève à « lire une consigne, un énoncé et comprendre ce qui est demandé ». Différentes hypothèses sont plausibles et renvoient à la décision de ce qui est une réponse acceptable par l'évaluateur et aux inférences restant à la charge de l'évaluateur concernant l'activité de l'élève : peut-être que l'élève n'a pas pu lire (déchiffrer les mots de) la consigne, peut-être l'a-t-il parfaitement lue mais n'a-t-il pas

³⁴ Si l'on se réfère à l'Echelle d'acquisition en Orthographe LExicale (EOLE) de Pothier B. et P., le mot « flûte » est écrit selon la norme orthographique par 0% des élèves de CP, 37% des élèves de CE1, et près de 50% par les élèves de CE2, CM1 ou CM2 (2004, p.115). Il ne fait donc pas partie des 347 termes dont l'orthographe lexicale est acquise en CP (2002/2004, p.12).

³⁵ L'accès à la compréhension dépend également du vocabulaire utilisé. La base Manulex (Lété, Sprenger-Charolles, et Colé, 1994), fondée sur la notion d'« exposition à l'écrit », permet d'estimer cette caractéristique du texte en procurant des mesures de fréquences de mots basées sur un corpus d'écrits adressés à l'enfant dans les manuels de lecture. L'indice SFI de ce mot dans la base Manulex est de 45,72 ce qui signifie que le mot survient tous les (environ) 750 000 mots dans les manuels d'apprentissage de la lecture : c'est un mot très rarement rencontré. L'exposition des élèves à ce mot est donc très faible.

³⁶ Du point de vue du recodage phonologique, il comporte des difficultés liées à un grand nombre de lettre (>7) et des syllabes qui représentent l'ensemble des complexités que l'on peut trouver : Consonne Consonne Voyelle graphique (-tru), Consonne Vnasale (-ment) et Voyelle nasale Consonne (ins-)

su la traiter (peu probable dans ce cas précis), peut-être l'a-t-il parfaitement lue mais l'a-t-il sautée (problème de mémoire de travail) ?

- il existe un écart entre le produit de l'activité de l'élève et la tâche prescrite imputable pour partie à la rédaction même de la consigne : c'est à l'élève de déterminer s'il doit relier « à droite de » à « écrire » ou à « le dessin ». C'est aussi à l'élève d'inférer une partie de la consigne (traiter tous les éléments d'un sous-ensemble par exemple).

d-discussion de la validité psycho-didactique de l'épreuve :

En conclusion de l'analyse, la validité psycho-didactique de cette épreuve n'est pas avérée.

Cette épreuve permet en outre de pointer une difficulté de codage des réussites. La détermination d'un seuil pour apprécier la réussite d'une série d'items est une préoccupation classique des concepteurs d'épreuves³⁷. Deux niveaux de seuils de réussite peuvent être distingués :

- l'un au niveau d'une composante de la tâche (à partir de combien de fruits et légumes barrés considère-t-on que la consigne est réussie ? le fait que la totalité des éléments à barrer est implicite, que faire si l'élève en oublie un ? à quoi attribuer la raison de l'oubli d'un élément) et,
- l'autre au niveau de la tâche entière (comment coter sur 4 (les rectangles indiqués en bas à gauche) une réalisation d'un ensemble contenant 5 demandes : sont-elles d'égale difficulté ? etc.)

Nous ne poursuivrons pas ce type d'investigation relative à la détermination des seuils de réussite dans notre travail mais soulignons le fait que la cotation des réponses représente un problème important pour l'enseignant.

Épreuve n°2

La tâche précédemment analysée n'est que peu ancrée conceptuellement dans l'un des deux domaines qui nous intéressent. Nous avons donc choisi de présenter une autre tâche qui l'est davantage, dans le domaine des mathématiques.

a- l'item :

Au premier trimestre, les enseignants H et I, qui ne formulent aucun item relatif à la résolution de problèmes, souhaitent évaluer pour le premier, l'item du livret « *Je sais calculer en lignes des sommes, des différences, des opérations à trous* » et, pour le second, l'item du livret « *Je sais calculer en ligne des sommes et des différences* ». Un décalage est susceptible d'être constaté entre ce que les évaluateurs souhaitent évaluer et ce qu'ils vont évaluer au moyen de la même tâche prescrite aux élèves reproduite ci-après.

b- l'épreuve :

Pour évaluer les items précédemment décrits, ces deux enseignants ont choisi d'utiliser la même épreuve reproduite ci-dessous, pourtant clairement identifiée (ce que l'on constate en haut à gauche de la figure suivante) comme un « Problème » :

³⁷Selon De Ketele et Gérard (2005), « 75% au moins de l'échelle des notes (*devrait être*) réservée aux critères minimaux et 25% au plus aux critères de perfectionnement (règle des 3/4) » et « le seuil de maîtrise minimal pour un critère minimal (*devrait être*) fixé à 2/3 : cette règle est basée sur le fait que, dans la majorité des cas, commettre une erreur sur trois (exemple : se tromper dans un calcul sur trois) n'est pas nécessairement le fait d'une non-maîtrise, mais peut être le fruit d'une distraction ou de tout autre facteur ; en prenant cette règle des 2/3, on essaie de minimiser les échecs abusifs ».



Figure 5 : Épreuve MT1 (Annexe B, p. 215 et 223) avec réponses d'élève

Réussir à résoudre ce problème nécessite de :

- Mathématiser la première phrase de l'énoncé et particulièrement le mot « chaque ». Ehrlich souligne que (pour des élèves de fin d'école élémentaire) la présence du terme « chaque » ou « chacun » entraîne un nombre anormalement élevé d'erreurs (1990, p. 70)³⁸. Le dessin soutient la compréhension du terme « chaque ». Il permet de considérer que « dans chaque enveloppe, il y a 4 images » se traduit par « dans la première enveloppe, il y a 4 images. Dans la deuxième enveloppe, il y a 4 images ». Le dessin est une donnée à part entière du problème. La figuration du nombre d'enveloppe dans le dessin évite de l'écrire au moyen d'une écriture chiffrée dans l'énoncé. De ce fait, l'énoncé ne contient qu'une seule donnée chiffrée. L'élève ne peut répondre au problème en additionnant les données numériques de l'énoncé (il pourrait éventuellement attribuer la valeur 1 à chacune des 4 parties du dessin d'une enveloppe et dénombrer 8 morceaux pour répondre : cela permet d'interroger le statut du dessin cf. III.3)

- Mathématiser la deuxième phrase : « Combien ... en tout ». Si l'on désigne par T le nombre total d'images à trouver, l'élève doit réaliser l'opération $T = e + e$, où e désigne le nombre d'image placé dans chaque enveloppe. Il doit donc opérer sur des données numériques.

Il reste encore à exprimer dans la numération chiffrée, dans une phrase « à trou », le résultat trouvé.

L'écart entre la tâche à réaliser et la tâche prescrite est aisément constaté : il ne s'agit pas d'une seule tâche de calcul en ligne mais bien d'une tâche de résolution de problème arithmétique.

c- la copie de l'élève:

La copie de l'élève que nous avons recueillie n'amène aucun élément nouveau.

d-discussion de la validité psycho-didactique de l'épreuve :

L'épreuve proposée n'est, du point de vue psycho-didactique, pas valide. Elle ne permet pas à l'évaluateur d'évaluer la capacité de l'élève à effectuer un calcul en ligne : la difficulté ne se situe pas, ou pas seulement, dans le seul fait d'additionner 4 et 4 pour trouver 8. Elle est bien plus que cela : elle est une tâche de résolution de problème qui, en outre, engage

³⁸ Nous avons fait état (C. Blanc, 2016) d'un débat mathématique entre élèves de Cm2 autour de l'utilisation des mots « chaque » et « autant » dans une question conçue par un élève portant sur un énoncé également conçu par lui : « Si chaque professeur devait avoir le même nombre de garçons et de filles. Combien y aura-t-il de garçons et de filles dans chaque classe ? ». Le débat conduit à éclaircir la demande « En fait tu demandes quoi : combien il y a d'élèves dans chaque classe ou combien il y a d'élèves si il y a autant de garçons et de filles dans chaque classe ? C'est pas pareil ! »

l'élève à traiter une situation dans laquelle les données sont présentées sous deux formes différentes mais complémentaires.

2.3. Discussion :

Si la deuxième épreuve avait été utilisée à des fins d'évaluation de l'item « Résoudre des problèmes simples à une opération », sa validité psycho-didactique serait à reconsidérer. Ceci tend à indiquer que la validité d'une épreuve d'évaluation n'est pas une qualité endogène, une tâche peut être didactiquement fondée, il faut encore considérer ce pour quoi elle est utilisée, surtout lorsqu'elle est employée à des fins d'évaluation, et la consigne qui permet à l'élève d'entrer dans la tâche.

Nous avons vu que l'analyse des traces de l'activité de l'élève permet d'affiner l'analyse de la tâche et de mettre en lumière des aspects qui ne sont pas apparents (la difficulté de traitement du mot « instrument » dans l'épreuve 2 par exemple).

Dans les chapitres suivants, nous nous attachons à analyser la validité psycho-didactique des épreuves d'évaluation que nous avons collectées suivant la méthodologie que nous venons de développer. Pour cela, nous les regroupons selon divers points communs. En examinant les différentes épreuves, plusieurs éléments d'ordre contextuel communs à un certain nombre d'épreuves nous sont en effet apparus. De nombreuses épreuves contiennent par exemple des dessins dont nous interrogeons la portée dans le chapitre 3. Il est fréquemment fait usage par les enseignants de questionnaires écrits à réponses écrites pour évaluer la compréhension en lecture. Nous examinons ce paramètre des épreuves d'évaluation dans le chapitre 4 et prolongeons cette réflexion autour du système de réponse imposé ou suggéré par l'épreuve dans le chapitre 5. Enfin, nous examinons dans le chapitre 6, l'adéquation d'épreuves conçues par d'autres concepteurs que les enseignants importent dans les tests qu'ils utilisent au risque de décalages que nous analysons.

Chapitre 3. Le dessin : un simple « habillage » de la tâche ?

C'est une évidence de dire que les élèves de CP se caractérisent plus que tous les autres dans l'enseignement obligatoire par le fait qu'ils ne sont pas des lecteurs experts. Ceci constitue une contrainte importante pour l'enseignant lorsqu'il définit les épreuves pour évaluer : l'élève doit pouvoir accéder à une bonne compréhension des informations de la tâche qu'il doit traiter. La tâche prescrite doit être intelligible pour l'élève i.e. dire correctement ce que l'enseignant veut lui faire dire, être compréhensible par l'élève (en termes d'identification des mots et de connaissance du vocabulaire employé), être présentée de façon que les élèves comprennent ce qu'ils doivent faire.

On peut ranger, de plus, du côté des contraintes de l'enseignant celle qui consiste à évaluer tous les élèves d'une classe (entre 25 et 30 élèves) sur une période réduite afin de renseigner un livret scolaire. Cette contrainte amène les enseignants à recourir massivement à des épreuves écrites, individuelles, simultanées (tous les élèves passent la même épreuve écrite en même temps) et à réponses écrites.

C'est ainsi qu'il est largement fait recours au dessin dont il est postulé sans doute qu'il concoure à rendre intelligible les tâches, mais aussi qu'il soit de nature à soutenir l'engagement dans la tâche et la réalisation autonome de celle-ci, voire à faciliter la tâche en elle-même. Dans le champ de la psychologie sociale, Monteil et Huguet (1991), puis Toczek (2009) (voir aussi Van Dooren et coll. 2010) ont montré que le contexte et plus particulièrement le mode de présentation de la tâche peuvent devenir de puissants régulateurs de performance scolaire. Le dessin pourrait constituer un mode de présentation de la tâche de nature à permettre une meilleure réussite même si Monteil et Huguet (2002, p.124) doutent que la composante ludique des apprentissages (l'habillage ludique d'une tâche au rang duquel on pourrait compter le dessin) soit systématiquement de nature à favoriser la performance.

Si l'on considère les différents modèles issus de la psychologie cognitive auxquels il est fait recours pour décrire la compréhension en lecture (modèles de Van Dijk et Kintsch (1978), Kintsch (1988) ou Van den Broek et coll. (1995 ; 1999)), ou la modélisation en mathématique (Nathan, Kintsch et Young (1992), Stillman (2012), Verschaffel, Greer et De Corte (2000), Verschaffel, De Corte et Crahay (2008), tous ont en commun la place accordée à la construction par l'élève d'un modèle de situation (qui doit ensuite être mathématisé dans le cas de la résolution de problèmes). En lecture, Van Dijk et Kintsch définissent un modèle de situation comme « une représentation cognitive des événements, des actions, des individus et de la situation générale évoquée par le texte » (1983, p. 11-12).

Nous formulons l'hypothèse que le dessin est susceptible d'intervenir au cours du processus d'élaboration du modèle de situation (donc au niveau de la redéfinition de la tâche), ce qui pourrait avoir pour conséquence de générer un écart entre la tâche prescrite et la tâche réalisée que ce soit dans des tâches complexes ou non.

3.1. Le dessin dans des tâches d'évaluation de compétences complexes

Du fait de la faible expertise en lecture des élèves de CP, une même tâche peut mobiliser des procédures « connues » et automatisées (cela correspond au premier degré de compétence du modèle de Rey et coll. 2006, p.14) pour certains élèves quand d'autres, au contraire, mobiliseront un ensemble de procédures « connues » non automatisées conférant à la tâche proposée le statut de situation complexe (ce qui correspond au deuxième voire au troisième degré de la compétence dans le modèle de Rey et coll. précité). Cet éclairage permet d'illustrer (cf. III-1.2) qu'une même tâche peut occasionner de fortes différences interindividuelles au moment de la redéfinition de la tâche par les élèves.

Postulons que les concepteurs aient recours au dessin dans des tâches supposées renvoyer à la mobilisation de procédures expertes de façon à les rendre plus intelligibles, et observons comment le dessin peut être susceptible d'agir au niveau de la redéfinition de la tâche dans quatre tâches différentes renvoyant à de telles compétences « complexes », de nature différente.

Épreuve n°3 : « huit »

a- l'item :

Au premier trimestre, l'enseignant G évalue conjointement les deux items du livret scolaire : « savoir écrire et nommer les nombres » et « produire les décompositions additives des nombres inférieurs à 20 ».

b- l'épreuve :

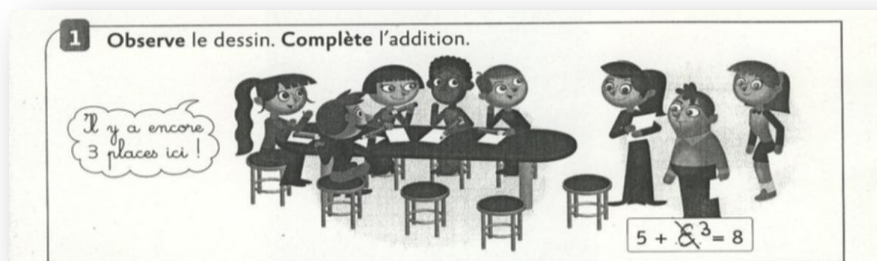


Figure 6 : Épreuve MT1 (Annexe B, p.163) avec réponses d'élève

Au vu des items du livret auxquels l'enseignant réfère l'épreuve, la tâche attendue est une tâche de décomposition additive, mobilisant des connaissances relatives à la désignation du nombre (lecture de « 5 », écriture de « 3 ») : l'élève devrait ainsi déterminer le complément à 8 de 5 (par calcul, ou par sur-comptage³⁹), tâche qui peut être réalisée sans recourir aux informations apportées par l'illustration.

Intéressons-nous au rôle potentiel joué par l'illustration dans la redéfinition de la tâche par un élève qui réaliserait cette consigne « à la lettre », en commençant par observer l'illustration.

³⁹ L'élève peut « sur-compter » à partir de 5 en énonçant la suite des mots-nombres jusqu'à obtenir 8 et en dénombrant parallèlement le nombre de mots-nombres prononcés au moyen d'un autre compteur (les doigts par exemple) : « six » (il lève un doigt), « sept » (il lève un deuxième doigt), « huit » (il lève un troisième doigt) soit trois mots nombres.

L'élève pourrait dénombrer les collections d'enfants représentés, « les cinq personnages situés à gauche » et « les trois personnages situés à droite », puis traduire le résultat de ces dénombrements dans la numération chiffrée. Mais il pourrait aussi bien dénombrer les sièges comme la remarque inscrite dans la bulle « il y a encore 3 places ici » pourrait l'inciter à le faire. Une difficulté pour un élève qui procéderait ainsi résiderait dans l'élaboration du modèle de situation associée à l'illustration: il peut y avoir en effet un doute sur la nature des informations pertinentes à retenir (les enfants ? les sièges ?), même si dans les deux cas la compétence à l'œuvre -celle du dénombrement d'objets- conduit au même résultat. Il se pourrait même que certains élèves, mis en difficulté dans l'élaboration du modèle de situation, se bornent à reporter dans l'addition à trou le seul chiffre présent dans l'illustration (le chiffre « 3 » figurant dans la bulle).

En admettant que cette difficulté soit surmontée, la réalisation de cette tâche, que l'élève s'attache aux données « enfants » ou « sièges » ne nécessiterait pas la production d'une décomposition additive : reporter « 3 » dans l'addition « à trou » suffit en effet pour réussir sans avoir eu à construire que l'addition de 5 et de 3 produisait 8 ou que 8 se décomposait en additionnant 5 et 3. Autrement dit la tâche redéfinie par l'élève serait une tâche de dénombrement.

c- la copie de l'élève:

La réponse que nous avons recueillie (fig.6) invite à penser que l'élève qui l'a produite sait traduire « trois » (résultat oral du comptage) en « 3 » (écriture chiffrée dans la numération écrite), sans que l'on puisse savoir s'il a effectué un dénombrement ou une addition à trou. On remarquera que, dans le cas de cet élève, la traduction opérée est encore hésitante (écriture symétrique du 3), ce qui est un élément de nature à contribuer à l'évaluation du premier item que l'enseignant veut évaluer au moyen de cette épreuve.

En définitive, la VPD de l'épreuve n'est pas avérée, la présence de l'illustration pouvant favoriser l'engagement de l'élève dans un dénombrement plutôt que dans une décomposition additive, sans que la qualité de la réponse produite (le report de « 3 ») permette de savoir si l'élève sait, ou a su, décomposer 8 en 5+3.

Épreuve n°4 : « les 10 canards »

a- l'item :

Au premier trimestre, l'enseignant l'évalue l'item : « résoudre des problèmes simples à une opération » au moyen de l'épreuve ci-dessous :

b- l'épreuve :

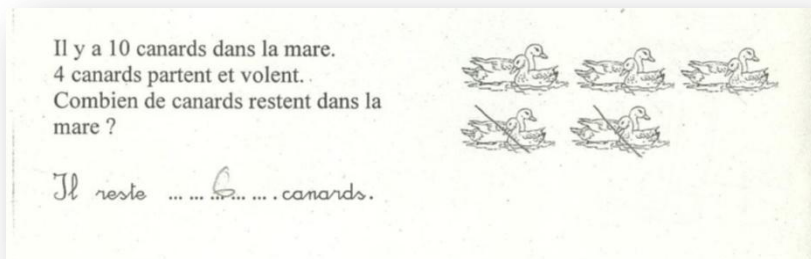


Figure 7 : Épreuve MT1 (Annexe B, p. 223) avec réponses d'élève

Au vu de l'item de référence, il est attendu que l'élève traite ce problème en mathématisant la situation décrite par l'énoncé en une opération (10-4) dont il doit ensuite calculer le résultat. D'après la classification de Vergnaud (1981), il s'agit d'un problème additif que l'on peut considérer comme assez simple car la mathématisation que doit effectuer l'élève est soutenue par le terme « partent » qui induit l'idée d'une « diminution »⁴⁰. L'élève doit ensuite écrire ce résultat dans une phrase « à trou » pour former la réponse. La réponse que produit l'élève ne se compose que d'une seule valeur numérique, l'énoncé prenant en charge la liaison de celle-ci (« 6 ») avec l'unité dans laquelle elle s'exprime (« canards »).

Intéressons-nous au rôle potentiel joué par l'illustration dans la redéfinition de la tâche. Sous la forme de 5 couples, 10 canards sont représentés ce qui traduit l'information contenue dans la première phrase de l'énoncé « il y a 10 canards ». Barrer 4 canards correspondrait au traitement de la deuxième phrase de l'énoncé « 4 canards partent et volent » (c'est ce qu'a fait l'élève dont nous donnons la réponse dans la figure ci-dessus). Pour répondre à la question « combien de canards restent dans la mare ? », l'élève n'a plus qu'à dénombrer les canards restant (« un, deux ... jusqu'à dire six ») puis à traduire le résultat du comptage dans la numération écrite « 6 ». La tâche ainsi redéfinie est une tâche de dénombrement.

Nous sommes donc amené à mettre en doute la validité de l'épreuve (au sens des approches psycho-didactiques de l'évaluation), la présence de l'illustration étant susceptible d'orienter l'élève vers une activité de dénombrement alors que l'enseignant cherche à évaluer sa compétence à résoudre un problème simple, c'est à dire à mathématiser une situation et à en effectuer le traitement opératoire à l'aide (dans le cas présent) d'une soustraction.

Épreuve n°5 : « le petit chaperon rouge »

a- l'item :

Au troisième trimestre, l'enseignant N souhaite évaluer l'item : « Je comprends un texte ». Cet item renvoie à l'item des programmes : « Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens ». Pour ce faire, il utilise une seule épreuve que nous reproduisons ci-dessous.

⁴⁰ Le terme « partent » agit comme un opérateur sémantique qui est ici de même signe (au sens d'Ehrlich, 1990), que l'opérateur mathématique, ce qui est la situation la plus favorable à la réussite.

b- l'épreuve :

6) Donner, après lecture, des renseignements sur un texte 8 / 10

Consigne : Observe la couverture du livre. Lis le texte. Coche les bonnes réponses.

**Le petit
Chaperon rouge**

D'après Jacob Grimm



Il était une fois une petite fille qu'on appelait le petit Chaperon rouge.

Un jour, en allant chez sa grand-mère, elle rencontra le loup.

Vif et rusé, l'animal réussit à dévorer la vieille femme, puis la petite fille.

Heureusement, un bûcheron arriva et tua le monstre. Il délivra les prisonnières.

Elles vécurent alors très heureuses.

A – Ce texte est :

une recette
 un conte
 une lettre

B – Son titre est :

La petite fille aux allumettes
 Le petit Chaperon rouge
 Le vilain petit canard

C – L'auteur de l'histoire s'appelle :

Jacob Grimm
 Hans Andersen
 Charles Perrault

D – C'est l'histoire :

de Cendrillon qui épouse le prince.
 du Chaperon rouge qui voit le loup
 de Blanche Neige et 7 nains.

E – Comment est le loup ?

maladroit
 rusé
 peureux

F – La petite fille rencontre le loup :

en allant chez sa tante.
 en allant chez le bûcheron.
 en allant chez sa grand-mère.

G – Qui tue l'animal ? *Fais une phrase pour répondre*

~~le loup~~

Figure 8 : Épreuve FT3 (Annexe B, pp. 152-153) avec réponses d'élève

Selon la consigne de la tâche, il est attendu que l'élève coche les bonnes réponses après avoir observé la couverture du livre et lu le texte. Pour ce faire, l'élève doit :

- comprendre l'image de la couverture de livre, ce qui implique une analyse particulière qui ne relève pas de la lecture,
- lire un texte,
- lire un questionnaire particulier (QCM) dans le sens où, par contrat, parmi les réponses proposées il faut localiser les bonnes réponses (i.e. elles sont contenues dans les propositions). Ceci mobilise une connaissance sur le fonctionnement des QCM non triviale.

La tâche de lecture intervient après celle d'observation du dessin. Observer le dessin engage très probablement l'élaboration d'une représentation. Or, lire engage également une

telle activité si l'on se réfère à N. Blanc (2010 ; voir aussi Fayol, 2003), pour qui lire « c'est construire une représentation mentale intégrée et cohérente de la situation décrite par ce texte - ce qu'on appelle souvent un modèle mental ». La question est de savoir comment les deux représentations cohabitent.

L'analyse du dessin permet à l'élève qui en disposerait⁴¹ de mobiliser des connaissances littéraires (cf. notion de réseau et de pensée en réseaux développée par Tauveron (2002), Devanne (1996, pp. 13-19 ; 1997)), pour établir qu'il s'agit du personnage de conte du petit chaperon rouge et le reconnaître : il n'a pas besoin de recourir à la construction d'un modèle mental de la situation décrite par le texte. Les réponses aux questions A, B et D sont immédiatement accessibles sans qu'il soit besoin de lire le texte : il suffit d'avoir reconnu le petit chaperon rouge sur le dessin (c'est le conte / intitulé le petit chaperon rouge/ il raconte l'histoire du petit chaperon rouge). Les réponses aux questions E, F et G peuvent découler de cette reconnaissance : elles sont accessibles sur la base de la connaissance de ce conte (le loup est rusé / la petite fille rencontre le loup en allant chez sa grand-mère / le bûcheron (en réalité le trappeur) tue le loup (seulement dans la version des frères Grimm, pas dans celle de Perrault)).

On note par ailleurs que la réponse à la question C ne nécessite pas non plus d'avoir lu le texte : ce dernier ne contient pas cette information. Ce type de questions est connu dans la littérature et décrit sous le nom de « passage-independent questions » (Keenan et Betjemann, 2006) et correspond au fait de pouvoir répondre à des questions indépendamment de la lecture du texte, sur la base de connaissances déjà disponibles⁴². Dans le cas du début d'apprentissage de la lecture qui nous occupe ici, et dont les enseignants souhaitent évaluer l'accès à l'autonomie et à la compréhension, on peut donc tout à fait concevoir d'avoir des élèves capables de répondre à presque toutes les questions de compréhension malgré de très nombreuses difficultés de déchiffrage : un cas décrit par Keenan et Betjemann (id. p. 365)⁴³.

L'élève qui ne disposerait pas de connaissances sur ce conte serait conduit à lire le texte, à parfaitement associer « le monstre » au loup, « les prisonnières » à « la vieille femme » et « la petite fille » = « le petit chaperon rouge ». La même tâche est donc plus ou moins complexe en fonction des connaissances disponibles des élèves ce que soulignait fort bien Vergnaud (cf. I. 2.1.a.2).

c- la copie de l'élève:

L'analyse des traces de l'activité de l'élève permet de considérer que l'élève produit une réponse erronée à la seule question ouverte. La réponse fournie interroge quant au modèle de situation que l'élève a élaboré en supposant que la réponse est fidèle à ce modèle. Il est difficile d'interpréter cette réponse sans demander à l'élève d'expliquer sa réponse. Plusieurs pistes sont possibles :

Il est possible que l'élève ne comprenne pas le lien entre le loup et son substitut lexical « l'animal » dans la question. Mais alors que penser alors de sa reconnaissance du substitut lexical dans le texte et de « le monstre », autre substitut également rencontré dans le texte.

⁴¹ Le conte du Petit Chaperon rouge fait partie de la liste de référence « La Littérature à l'école » pour l'école Maternelle. Il est donc susceptible de faire partie du bagage de connaissances littéraires des élèves de CP.

⁴² « Passage-independent questions can be answered by using information the reader already knows (e.g., What color was the banana ?) » (Keenan et Betjemann, 2006, p. 363)

⁴³ « We found that several children who were unable to decode many words of a passage were nonetheless able to answer nearly all the comprehension questions for that passage correctly » (p. 365)

Cela aurait sans doute pour conséquence que l'élève pense qu'un loup, un animal et un monstre peuplent l'histoire portée par le texte.

Il est possible que l'élève réponde en réalité à la question « qui meurt à la fin ? » correspondant à une question classique portant sur la situation finale d'un tel conte.

Il est encore possible que l'élève se méprenne dans l'identification du mot « tue » lu « est ». Ces différentes hypothèses illustrent bien la difficulté d'analyser les erreurs hors de la présence de l'élève.

La validité de la tâche est remise en cause : nous avons mis en évidence qu'au total, un élève pourrait répondre correctement à la grande majorité des questions du questionnaire sans avoir lu le texte dont il est sensé évaluer la compréhension. Les connaissances intertextuelles de l'élève mobilisées au cours de l'analyse du dessin en particulier peuvent totalement suffire et se substituer à la lecture. Le dessin est ici largement susceptible d'entrer en concurrence avec la tâche prescrite qui consiste à lire un texte et à faire état de sa compréhension. La conséquence de cela est importante si l'on se réfère au fait que l'enseignant évalue la compréhension en lecture de ses élèves au moyen de cette seule épreuve. Par ailleurs, l'élève ne doit pas véritablement « donner des renseignements sur un texte » comme annoncé dans la rédaction de l'item situé au-dessus de la consigne de la tâche et qui serait équivalent à l'item du livret « Je comprends un texte ». Il doit plutôt trouver dans un QCM les bonnes réponses aux questions posées, ce dont nous discuterons dans le chapitre 5.

Épreuve n°6 : « le chien Hector »

a- l'item :

Au troisième trimestre, l'enseignant L souhaite évaluer l'item « Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens ». Pour ce faire, il utilise deux épreuves dont celle que nous reproduisons ci-dessous et qui est extraite du protocole du ROLL⁴⁴.

La consigne de la tâche « écoute le texte suivant puis répond aux questions » permet de considérer la transposition effectuée par l'enseignant de la tâche à réaliser (« dire de qui ou de quoi parle le texte lu, répondre à des questions ») : il détermine une tâche de compréhension d'un texte lu par un tiers.

⁴⁴ Réseau des Observatoires Locaux de la Lecture. Mesnager J et Bentolila A. (1996). Cf. <http://www.roll-descartes.net/fr/accueil.php>. Le principe est que l'élève lise seul le texte puis, texte retourné, dise ce qu'il en a compris au cours d'un atelier dont le cadre est assez précis (Atelier de Questionnement de Texte).

b- l'épreuve :

Dire de qui ou de quoi parle le texte lu; trouver, dans le texte ou son illustration, la réponse à des questions concernant celui-ci; reformuler son sens.

A	AR	EA	NA
---	----	----	----

9) Ecoute le texte suivant lu par la maîtresse, puis réponds aux questions.

Hector

Notre vieux chien Hector est plutôt bizarre. Il n'aboie jamais !

Pourtant, une fois, on a entendu sa voix !

Maman, Hector et moi, nous marchions sur le trottoir.

Et puis j'ai voulu traverser la rue, au moment où une voiture arrivait ! Alors Hector a aboyé très fort, une fois, deux fois, trois fois.

Maman regardait un étalage. Surprise, elle s'est retournée. Elle m'a attrapé par le bras, juste devant la voiture. Il était temps !

Elle m'a grondé, elle a serré le chien très fort dans ses bras. Et même, elle a pleuré !

Depuis, on n'a plus jamais entendu la voix d'Hector.



Je m'appelle : la date :

1) Le chien Hector, qu'a-t-il de spécial ?
il n'aboie jamais

2) Où se passe l'histoire ?
dans la rue

3) La voiture va écraser l'enfant ! Mais il est sauvé ! Comment ?

Le conducteur s'arrête.

Le chien aboie et le garçon recule.

Le chien aboie et la maman attrape le garçon. /

Hector se casse la patte.

4) Après cet accident...

Hector n'aboie plus jamais. /

Hector aboie tout le temps.

Hector court après les voitures.

Hector mange les chats.

"Il naboie plus c'est plutôt bizarre."
 L'évaluateur trace une vague sous le mot
 « plus » et ajoute au-dessus le mot
 « jamais »

"dans la rue"

Figure 9 : Épreuve FT3 (Annexe B, p. 148) avec réponses d'élève.

Si l'on se réfère à la consigne de la tâche, il est attendu que l'élève écoute le texte lu par la maîtresse, puis qu'il réponde aux questions données après cette lecture. A strictement considérer cette consigne, il n'est pas requis que l'élève lise le texte⁴⁵.

Considérons le dessin : il représente la situation finale (lignes 11 et 12). On y voit une femme souriante, tenant dans ses bras un chien triomphant et un petit garçon à la mine déconfite. Les personnages se trouvent dans une zone délimitée sur le dessin par un double trait qui évoque un trottoir. L'arrière-plan représente un paysage urbain. Imaginons que l'élève ait eu le texte à lire seul (ce qui rappelons-le n'a pas été le cas). La réponse à la question 2 « Où se passe l'histoire ? » est orientée par l'analyse du dessin. Les réponses « l'histoire se passe en extérieur », « l'histoire se passe en ville », « l'histoire se passe dans la rue » sont toutes acceptables. Elles ne sont pas fondées sur la lecture du texte.

L'énoncé de la question 3 « la voiture va écraser l'enfant... » peut éventuellement justifier *a posteriori* la réponse à la question précédente.

La réponse à la question 3 est plus complexe à fournir. Si l'on observe les réponses proposées, aucune n'est conforme à l'illustration : la maman n'a pas attrapé le garçon mais le chien. De cela peut naître un conflit de représentations, un faible lecteur pouvant très bien s'arrêter au sens du dessin et à la représentation qu'il en a construit sans rechercher à en

⁴⁵ L'usage que l'enseignant a fait de ce texte ne correspond pas au cadre du ROLL qui nécessite une lecture individuelle du texte.

construire une autre qui le remette en cause. Une réponse erronée au questionnaire (visant à établir la compréhension du texte) pourrait ainsi provenir des représentations élaborées suite à l'analyse du dessin que l'élève ne mettrait pas en question.

c - la copie de l'élève :

L'analyse des réponses de l'élève permet de considérer que l'élève produit une réponse qui ne correspond pas à la situation initiale du texte (ce que la question interroge) mais à la situation finale (l'élève indique que le chien n'aboie plus, l'enseignant corrige en remplaçant « plus » par « jamais »). Cette réponse est probablement fondée sur la mémoire de la compréhension du texte, elle-même renforcée par le dessin qui porte sur la situation finale : le chien a aboyé mais n'aboie plus jamais depuis.

La tâche proposée n'apporte pas de preuve de validité au sens psycho-didactique. L'item de lecture « dire de qui ou de quoi parle le texte lu... » a été redéfini par l'enseignant en un item de compréhension orale d'un texte lu par un tiers. Autrement dit, la tâche évalue davantage la mémoire de la compréhension du texte que la compréhension du texte qui, à se référer à la consigne, n'a pas besoin d'être lu par l'élève. Une non-réponse peut donc être attribuée à un défaut de mémoire et non à un défaut de compréhension d'une lecture.

Par ailleurs, le dessin est susceptible d'entrer en concurrence avec la tâche prescrite (faire état de la compréhension d'un texte lu par un tiers) : il est susceptible d'engendrer une représentation particulière du texte correspondant davantage à la situation finale. Cette représentation pourrait bien ne pas être remise en cause pour répondre à une question portant sur la situation initiale ce que la copie de l'élève collectée permet de penser.

Discussion :

Ces quatre exemples permettent de souligner que, dans une épreuve d'évaluation d'un item complexe de mathématique ou dans une épreuve d'évaluation de la compréhension en lecture (texte lu seul ou entendu lire par un tiers), le dessin est un support d'information susceptible d'intervenir au niveau de la redéfinition de la tâche par l'élève. Il est à même de générer potentiellement un écart important entre la tâche à réaliser et la tâche prescrite à l'élève d'une part, et entre la tâche prescrite et la tâche réalisée d'autre part, le dessin pouvant interférer par exemple avec le modèle de situation que l'élève élabore.

Dans les deux épreuves de lecture (« Petit chaperon rouge » et « Le chien Hector »), et les deux épreuves de mathématiques (« $5+3=8$ » et « les 10 canards »), le dessin peut (ce n'est pas obligatoire) amener l'élève à substituer à la tâche prescrite une autre tâche (redéfinition de la tâche). Par exemple, l'élève peut substituer une tâche d'analyse d'image à une tâche de compréhension de texte, ou, substituer une tâche de dénombrement à une tâche de traitement d'une situation mettant en jeu des relations additives entre des données.

Pourtant, le dessin n'occupe pas la même fonction dans tous les cas :

- Pour les épreuves de français examinées, on pourrait se passer du dessin sans que cela change (*a priori*) la tâche prescrite. Le dessin occupe une fonction illustrative qui comporte comme principal écueil celui de focaliser l'attention de l'élève sur le sens qu'il faut lui attribuer plutôt que sur le sens qu'il faut attribuer au texte qui a permis à un illustrateur de réaliser le dessin. Le dessin de l'illustrateur joue de surcroît de codes qui peuvent être identifiés par le destinataire. Dans le cas de la tâche « Petit chaperon rouge », on peut aisément postuler qu'aucun élève de CP ne soit capable de dire qu'un chaperon est une ancienne coiffure enveloppant la tête et le cou,

descendant jusqu'aux épaules. Le texte ne mentionne pas le panier alors qu'il est pourtant représenté dans le dessin. Associer au dessin la valeur « petit chaperon rouge » nécessite de mobiliser des connaissances qui ne relèvent pas de la seule lecture. Dans le cas de la tâche « Le chien Hector », le dessin renvoie à ce que l'illustrateur considère lui-même comme important (l'idée principale) avec le risque qu'il entraîne l'élève sur cette focalisation excessivement placée. Dans les deux cas, le dessin oriente la redéfinition de la tâche de compréhension d'un texte (lu seul ou lu par un tiers) vers une tâche comportant une analyse de dessin susceptible de suffire pour répondre aux questions posées. Il n'y a pas d'adéquation de la tâche prescrite à la tâche à réaliser (évaluer la compréhension d'un texte lu de façon autonome).

- Pour les épreuves de mathématiques examinées, le dessin n'a pas qu'une valeur illustrative. Il comporte de l'information, numérique en particulier et oblige l'élève à opérer des changements de registres de représentations sémiotiques. Retirer le dessin dans l'épreuve « $5+3=8$ » obligerait l'élève à s'appuyer sur la seule addition à trou pour produire une décomposition additive qui passerait par l'écriture chiffrée de nombres (i.e. la tâche à réaliser) tandis que retirer le dessin dans l'épreuve « les 10 canards » obligerait l'élève à prendre en charge la représentation des données du problème et ne l'orienterait pas vers une procédure de dénombrement par défaut.

3.2. Le dessin dans des tâches visant à évaluer des habiletés « de base » en lecture

On retrouve fréquemment le terme « simple » ou « de base » accolé à certains items (« résoudre des problèmes simples » par exemple) ou à certaines compétences (habiletés de base en lecture par exemple). Il n'est pas certain qu'évaluer ce qui est simple soit également simple. Dans les tâches que nous analysons maintenant, nous avons choisi de présenter des tâches relevant des habiletés dites de base mobilisant toutes une forme de dessin. Trois d'entre elles (épreuves 7, 8 et 10 ci-après) proviennent du même test proposé par l'enseignant C : celui-ci est particulièrement soucieux dit-il de produire une évaluation « objective par rapport à un contrôle continu » des habiletés de base en début de CP. Pour ce faire, il utilise neuf épreuves dont celle que nous analysons ci-après. Il adapte les évaluations de la méthode de lecture Abracadalire qui ne lui convenaient pas et s'inspire d'un ouvrage de Borel-Maisonny (méthode de rééducation pour l'apprentissage de la lecture).

Épreuve n°7 : « silhouettes de mots »

a- l'item :

Au premier trimestre, l'enseignant C souhaite évaluer l'item du livret « Comprendre la combinatoire dans des graphies simples, dans des graphies complexes. Reconnaître les syllabes et les mots à l'écrit et à l'oral ». Cette rédaction relève du choix personnel de l'enseignant (elle ne correspond pas aux items des programmes). Elle renvoie à l'apprentissage des règles de correspondance graphophonologiques indispensables lorsque l'on doit déchiffrer un mot régulier jamais rencontré.

b- l'épreuve :

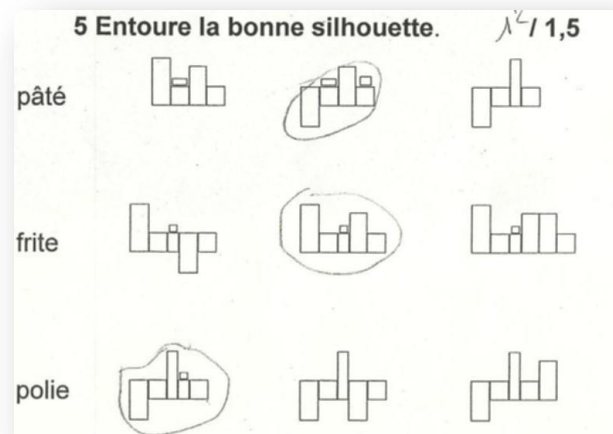


Figure 10 : Épreuve FT1 (Annexe B, p.13) avec réponses d'élève

Il est attendu que l'élève associe correctement le dessin d'un mot (sa « silhouette⁴⁶ ») avec une des trois silhouettes proposées en regard. Ces lettres sont représentées au moyen de formes géométriques les décrivant dans un lignage de type Seyès. Un carré (il est implicite que le côté inférieur du carré est posé sur la ligne) permet de représenter une lettre se situant dans le premier interligne. Un rectangle permet de représenter une lettre dépassant d'un interligne (il n'est pas fait de différence entre le « f » qui utilise deux interlignes au-dessus de la ligne d'écriture et le « t » qui n'en utilise qu'un seul). D'autres formes géométriques sont utilisées pour marquer les accents et les points.

L'élève ne doit considérer du mot que son enveloppe graphique, son dessin, ce qui renvoie à une évaluation d'habiletés visuelles. A aucun moment l'élève n'a besoin de mobiliser des connaissances de type graphophonologiques. Jacquier-Roux et Zorman, médecins de l'Education nationale, (2002, p.10) indiquent que « l'efficacité du système d'analyse visuelle est impliquée dans l'apprentissage de la lecture, et que l'entraînement visuel pratiqué dans le cadre (d'une étude décrite par les auteurs dans le texte) a un effet sur ce même apprentissage de la lecture ». Ils supposent même que « c'est principalement dans les classes sociales défavorisées que des enfants n'auraient pas acquis à l'entrée en CP l'habileté visuelle qui permet de saisir et traiter des graphèmes et que dans ce cas, l'entraînement visuel en CP serait un facteur de réduction des inégalités scolaires ». C'est sans doute ce qui fonde l'utilisation de ce type d'épreuves très particulières destinées à évaluer les habiletés visuelles des élèves. C'est un cas marginal dans notre corpus mais qui, il y a quelques années ne l'était pas⁴⁷.

Bastien et Bastien-Tonniazzo (2004, p.103-104) soulignent pourtant que dans un tel traitement logographique du mot, « ce n'est pas le mot qui est traité mais ce qui est autour », une lettre pouvant être utilisée comme indice saillant. Sans doute serait-il pertinent au regard

⁴⁶Par silhouette il faut entendre qu'il s'agit de la forme globale du mot constituée par les différentes formes graphiques de chacune des lettres composant le mot.

⁴⁷On se référera par exemple aux manuels d'apprentissage de la lecture « Croquepages » (Bordas, 1992- On trouve trace de ce type de tâche dans le « Cahier de lecture 1 » en pages 18 ; 19 ; 45) et « Ribambelle » (Hatier, 2004 – cf. « Cahier d'activités 1 » en page 29 ; 68)

de ces travaux de proposer aux élèves une épreuve nécessitant de discriminer perceptivement des mots différents partageant des lettres communes (Bastien, 1997, p.114). Il est aisé de considérer que l'enveloppe graphique des mots est une propriété inconstante des mots. Les formes pâté / PATE / Pâté / PÂTÉ encodent la même réalité sémantique et phonologique sans pour autant revêtir la même enveloppe graphique.

c- la copie de l'élève :

Prenons le cas du mot « polie » (troisième mot de l'épreuve). Dans la police utilisée, les lettres « l » et « i » ont la même hauteur. La silhouette de ce mot devrait donc présenter la représentation du « l » à égale hauteur de la représentation du « i » (point du i inclus) ce qui n'est le cas d'aucune des trois silhouettes. L'élève peut produire la bonne réponse par défaut (ce ne peut être ni la deuxième ni la troisième modalité de réponse).

La tâche prescrite à l'élève ne correspond pas à l'item. Cette tâche n'est pas une tâche mobilisant la reconnaissance de règles « qui permettent que des unités linguistiques abstraites puissent devenir productrices de sens » selon la définition que donne Charmeux au terme « combinatoire »⁴⁸.

Épreuve n°8 : « lettres symétriques p|q »

a- l'item :

Au premier trimestre, l'enseignant C souhaite évaluer l'item « Comprendre la combinatoire dans des graphies simples, dans des graphies complexes. Reconnaître les syllabes et les mots à l'écrit et à l'oral ». Pour ce faire, il utilise l'épreuve ci-dessous.

b- l'épreuve :

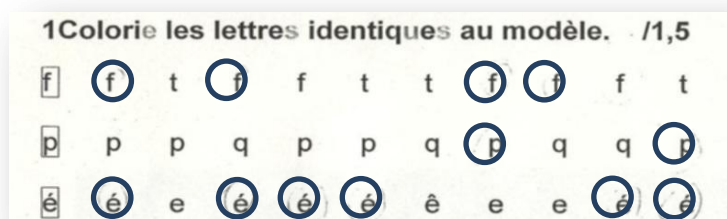


Figure 11 : Épreuve FT1 (Annexe B, p. 13) avec réponses d'élève

Cette épreuve ne permet pas d'évaluer la maîtrise de la combinatoire : elle ne fait aucunement appel au sens.

Analysons toutefois la demande faite à l'élève de colorier les lettres identiques à « p » (deuxième ligne). L'élève doit colorier les lettres identiques au modèle présenté en début de ligne. C'est une tâche de discrimination visuelle de lettres parmi des lettres visuellement proches (symétriques). Ce type d'épreuve est fondé sur la reconnaissance d'un modèle graphique orienté. Or, des difficultés dans la maîtrise de l'orientation des lettres symétriques (selon l'axe vertical p|q et b|d, ou selon une symétrie centrale pour les paires p|d et q|b) ont

⁴⁸Pour Charmeux, la combinatoire est « l'ensemble des règles, spécifiques à chaque langue, et par voie de conséquence, arbitraires mais organisées en système, qui permettent que des unités linguistiques abstraites puissent devenir productrices de sens par leur présence (ou leur absence) et leur ordre, dans des productions langagières, » (<http://www.charmeux.fr/combinatoire.html> consulté en avril 2015)

été mises en avant par Magnan (1995) chez des élèves de CP (plus de 30% d'erreurs chez les faibles lecteurs dans le cas de symétrie selon l'axe vertical ont été relevées). « Les confusions des lettres symétriques ne relèvent pas d'un trouble visuo-spatial mais sont déterminées par le mode même d'identification, lui-même résultant de la représentation que l'enfant a des mots » (2004, pp.103) soulignent Bastien et Bastien-Tonniazzo au sujet d'enfants scolarisés en grande section : tenir compte de l'orientation des lettres est « en contradiction totale avec le précurseur très primitif « ne pas tenir compte de l'orientation des lettres » » (idem p.104). Ces auteurs précisant que « même un bon lecteur peut être amené, lorsqu'il est fatigué par exemple, à faire des confusions de ces lettres symétriques » (idem p. 104).

c- la copie de l'élève :

La consigne demande d'entourer les lettres identiques au modèle. Il est implicite qu'il est attendu de l'élève qu'il les entoure toutes, ce que l'élève dont nous avons reproduit la copie n'a pas fait. La consigne contient un implicite de nature à désavantager les élèves qui ne l'aurait pas rendue explicite.

La validité psycho-didactique de cette épreuve peut être remise en cause : la tâche prescrite à l'élève (« colorier des lettres identiques aux modèles ») ne correspond pas à l'item (« comprendre la combinatoire ... »). L'item du livret que l'enseignant souhaite évaluer au moyen de l'épreuve discutée utilise le terme de « combinatoire » et semble renvoyer aux règles de correspondance graphophonologiques indispensables lorsque l'on doit déchiffrer un mot régulier jamais rencontré. La traduction ainsi opérée ne correspond pas à ce que l'on entend par le terme « combinatoire »⁴⁹, pas plus qu'elle ne correspond à la façon dont les items des programmes le traduisent. Les programmes évoquent bien la relation entre sons et lettres dans l'item « distinguer entre la lettre et le son qu'elle transcrit ; connaître les graphies simples et complexes », mais proposent un item portant spécifiquement sur les équivalences de graphies (« connaître les correspondances entre minuscules et majuscules d'imprimerie, minuscules et majuscules cursives ») qui semblerait se rapprocher de ce que l'enseignant évalue (ce qui n'est pas non plus le cas puisque toutes les lettres sont présentées dans une police scripte et en minuscule).

⁴⁹Cf. note précédente.

Épreuve n°9 : « i ou [i] ? »

a- l'item :

Au premier trimestre, l'enseignant E souhaite évaluer l'item « Connaître les correspondances entre les lettres et les sons ». Pour ce faire, il utilise quatre épreuves dont celle que nous reproduisons ci-dessous⁵⁰.

b- l'épreuve :

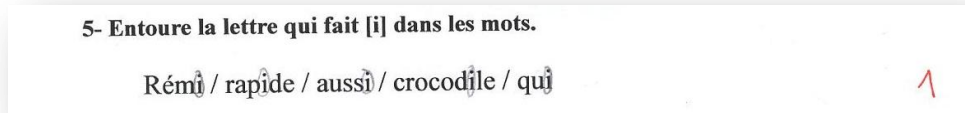


Figure 12 : Épreuve FT1 (Annexe B, p. 15) avec réponses d'élève

Il est attendu de l'élève qu'il « entoure la lettre qui fait [i] dans une liste de mots ».

La liste des mots proposés ne présente que des mots dans lesquels le graphème « i » code toujours le phonème [i]. Ainsi un élève qui n'aurait rien lu et aurait simplement entouré tous les « i » sans distinction aurait une réussite totale.

Ecalte et Magnan (2002, p.76) rapportent une étude de Magnan, Leonard et Aimar (1995) qui aboutit à l'hypothèse selon laquelle « les connaissances que les jeunes enfants possèdent sur le dessin sont transférées aux signes graphiques, les lettres, en cours d'apprentissage. Autrement dit, les lettres ont d'abord des propriétés de dessin avant d'acquiescer les propriétés spécifiques de l'écrit. »

Dans le cas ici présenté, on ne peut pas distinguer la lettre (le dessin d'une lettre) du graphème.

c- la copie de l'élève :

L'élève dont nous avons reproduit la copie est évalué positivement sans que l'on soit sûr qu'il sache établir une correspondance entre différentes graphies mobilisant la lettre « i » et la transcription du seul phonème [i].

La validité de cette tâche peut ainsi être discutée : la tâche prescrite ne correspond à l'item que dans le cas très particulier où le nom de la lettre correspond à un graphème (la lettre « i » code le phonème [i]) et qu'elle n'entre pas dans la composition d'un autre graphème (« in », « ai » par exemple). Cette tâche peut se transformer en une tâche de détection de dessin de lettre dans un ensemble de lettres, plutôt que d'établissement de correspondance graphophonologique.

⁵⁰ Cette tâche est prise en exemple dans le texte de Vantourout et Goasdoué (2014) pour illustrer l'intérêt d'une analyse psycho-didactique des tâches d'évaluation.

Épreuve n°10 : « double dessin »

a- l'item :

Au premier trimestre, l'enseignant C souhaite évaluer l'item « Comprendre la combinatoire dans des graphies simples, dans des graphies complexes. Reconnaître les syllabes et les mots à l'écrit et à l'oral. ». Il utilise l'épreuve reproduite ci-dessous :

b- l'épreuve :

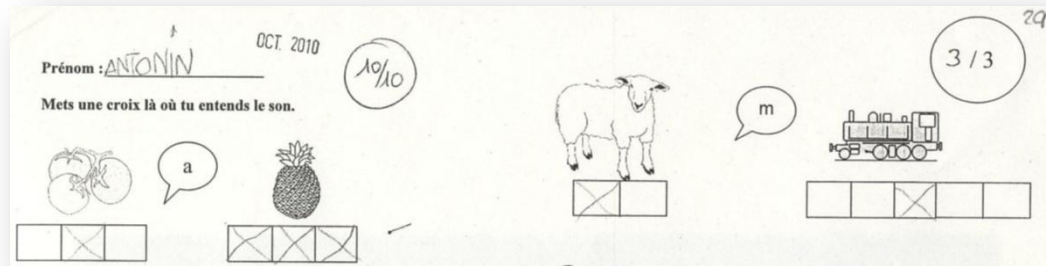


Figure 13 : Épreuve FT1 (Annexe B, p. 11) avec réponses d'élève

Il est attendu que l'élève mette une croix là où il entend le son (représenté par un archigraphème placé dans une bulle). Il est implicite et à la charge de l'élève, que la bulle contenant la lettre « m » concerne les mots « mouton » et « locomotive ».

L'usage du dessin dans la présentation des données se révèle problématique lorsqu'il occupe plusieurs fonctions au sein de la même tâche : représentation sémantique d'un mot et représentation d'un concept linguistique (celui de syllabe) comme c'est le cas de l'exemple que nous développons ici.

Depuis le début des années 2000, différentes recherches ont souligné la forte corrélation entre le niveau de conscience phonologique et l'acquisition de la lecture. Bien que les programmes ne mentionnent pas explicitement les habiletés phonologiques, les enseignants les évaluent dans le cadre de l'item du programme du Français « distinguer entre la lettre et le son qu'elle transcrit ; connaître les correspondances entre les lettres et les sons dans les graphies simples et complexes ». On distingue parmi ces tâches, des tâches de catégorisation de mots en fonction des rimes, les tâches de suppression de phonèmes, les tâches de substitution de phonèmes, les tâches de comptage de sons dans un mot, les tâches de détection d'unités communes (Ecale, Magnan et Bouchara, 2002 p.5).

Goigoux, Cèbe et Paour, en présentation d'un ouvrage destiné à développer les compétences phonologiques des élèves, croient « important de proposer également des activités sans support visuel qui, parfois, gêne plus les élèves qu'il ne les aide » (2004, p.26). Ils citent une épreuve d'identification d'attaques⁵¹ identiques dans « phoque, fleur, singe » dans laquelle les élèves échouent si on leur montre les images. « Ils opèrent un traitement sémantique et associent les deux animaux, alors qu'ils réussissent à associer phoque et fleur s'ils ne voient pas les images » (idem, voir la note 13 de cette référence). Le dessin ne faciliterait pas le traitement « non sémantique » du mot.

⁵¹L'attaque est la consonne ou le groupe consonantique initiales de la syllabe, la rime étant constituée de la voyelle et des phonèmes qui suivent. Dans le mot « cric » [kr] est l'attaque [ik] est la rime.

Le problème se pose quand, ajouté au dessin (dont le traitement sémantique est parfois difficilement évitable pour les élèves), se surajoute un autre dessin comme dans le cas discuté ici. On retrouve ainsi à plusieurs reprises le type de tâche suivant : un dessin représente un mot et, sous le mot, autant de carrés que le mot contient de syllabes orales sont placés côte à côte. Pour réussir la tâche, l'élève doit alors effectuer un grand nombre de sous-tâches :

- oraliser seul le nom du dessin ou écouter le nom du dessin dit par l'enseignant, ce, sans traiter la valeur sémantique du mot : [mutō]
- localiser dans la chaîne sonore, la syllabe dans laquelle se trouve le phonème cible, ce qui sous-entend d'avoir préalablement effectué la segmentation syllabique du mot : [mutō] se décompose en [mu-tō] et [m] s'entend dans [mu]
- associer chaque syllabe orale à un carré : [mu-tō] = □-□
- cocher le carré correspondant à la syllabe : □-□

Le dessin des carrés donne à l'élève le nombre de syllabes que le mot contient. L'élève doit adapter sa réponse (bonne ou mauvaise) au cadre fourni. La tâche présentée ci-dessus témoigne du problème de la représentation de la syllabe et de la stabilité accordée aux types de syllabes⁵² : parle-t-on de syllabe orale (il y en a une seule dans le mot « classe ») ou écrite (il y en a deux dans le mot « classe ») ? Dans l'exemple reproduit ci-dessus, pour réussir la tâche concernant le mot « locomotive » (4 syllabes orales mais 5 syllabes écrites), l'élève doit cocher la troisième case. Pour ce faire (outre qu'il peut attribuer la valeur « train » au dessin et ne cocher aucune case, ce qui a été discuté précédemment), il doit se donner oralement le mot « locomotive ». A ce stade, il est supposé que le mot qu'il oralise pour lui-même est identique au mot cible⁵³ ce dont l'enseignant n'a aucune preuve puisque cela ne donne lieu à aucune oralisation. Cela pourrait être d'ailleurs un élément intéressant d'évaluation de la production phonologique de l'élève. Puis, l'élève doit adopter l'une ou l'autre de deux stratégies conduisant l'une à la bonne réponse, l'autre à une réponse erronée : soit il repère que le phonème [m] se situe dans la troisième syllabe orale du mot et coche donc la troisième case, ce qui le conduit à manifester la bonne réponse ; soit il repère que le phonème [m] se situe dans l'avant dernière syllabe et coche alors la quatrième case, ce qui est une réponse erronée, le mot ayant été découpé en 5 syllabes (syllabes écrites). Le même type de remarque s'applique au mot « tomate » qui contient deux syllabes orales [to-mat] ou trois syllabes écrites TO/MA/TE.

Le recours au dessin dans les tâches de discrimination auditive n'apparaît pas toujours pertinent compte-tenu des obstacles soulignés ici : différence de valeurs (sémantique vs linguistique), différence de statut (le dessin pour représenter un mot, le dessin pour représenter un concept linguistique)

c- la copie de l'élève :

La copie de l'élève n'apporte aucun élément nouveau comme c'est généralement le cas lorsque les réponses sont justes.

⁵² La syllabe écrite est un découpage que l'on fait à l'intérieur d'un mot, comme si tout se prononçait. Exemple : lec/tu/re 3 syllabes écrites contre 2 syllabes (orales) lec/tur. « La segmentation en syllabes écrites est indispensable aussi plus tard dans l'apprentissage de la lecture et de l'écriture » selon TFL (consulté à l'adresse <http://www.uvp5.univ-paris5.fr/TFL/ac/AffQpeRep.asp?CleFiche=P8-2>)

⁵³ On suppose par exemple que l'élève se donne une image phonologique correcte du mot « train » et non une image phonologique incorrecte (ex : « crain ») liée à un défaut de prononciation, et à partir de laquelle il ne pourrait pas localiser le phonème cible. Avec cet exemple, il ne pourrait pas indiquer que [t] débute le mot. C'est une erreur fréquente sur ce type de syllabe.

Observant les élèves réaliser cette tâche (localiser le [m] de « mouton »), nous avons constaté que l'un d'eux ne se souciait pas du nombre de syllabes et ânonnait plusieurs fois le phonème [m] pour former la suite [mmm..]. Ce faisant il cochait alors la première case. La réponse fournie est correcte mais ne correspond pas à ce qui est attendu : il n'a pas segmenté le mot en deux syllabes. Il n'a donc pas pu localiser le phonème [m] dans la première des deux syllabes du mot. Il a en revanche oralisé plusieurs fois le phonème [m], conclu au fait que la lettre « m » codait bien ce phonème et coché la première case. Nous lui avons ensuite demandé combien le mot comptait de syllabes et avons obtenu une bonne réponse, argumentée : [mu] et [tō]. Réaliser la tâche mobilise une compétence de segmentation syllabique (conscience syllabique) qui est supposée acquise : il est possible que l'élève ne s'appuie pas sur cette dernière mais sur la seule attaque du mot.

La validité psycho-didactique de cette tâche est discutable. Donner le nom du dessin est à la charge de l'élève, ce qui constitue un premier obstacle éventuel. Assurer la segmentation syllabique du mot (en syllabes écrites et non en syllabes orales dans le cas discuté ici) et associer la syllabe oralisée par l'élève avec la représentation de la syllabe (carrés placés sous le mot) sont également à la charge de l'élève. Ils fonctionnent comme des pré requis. Pour reprendre les termes de la consigne, avant de « placer une croix là où (on) entend le son », l'élève doit avoir effectué une segmentation syllabique équivalente à celle utilisée pour présenter la tâche. Il faut enfin que l'élève soit capable de faire figurer la position du phonème cible dans le mot, ce qui n'est pas nécessairement synonyme de la capacité à repérer un phonème dans un mot.

Discussion :

Dans une épreuve (épreuve 7, « silhouettes de mots ») le dessin d'un carré est utilisé pour représenter une lettre minuscule scripte non accentuée quand dans une autre épreuve (épreuve 10, « double dessin ») ce même dessin est utilisé pour représenter une syllabe (écrite) d'un mot que l'élève doit se donner oralement à partir d'un dessin figuratif. Lorsque les lettres sont données comme dans l'épreuve 9, « i ou [i] ? », traiter la lettre comme un dessin permet de réussir l'épreuve, alors que dans l'épreuve 8 « lettres symétriques » on se préoccupe d'orientation de lettres minuscules scriptes. Trois de ces quatre tâches sont utilisées par le même évaluateur pour évaluer des connaissances dite « simples ».

A travers ces quatre exemples, on voit fonctionner un rapport au signe complexe que doit reconstruire l'élève alors même que J. Fijalkow (2000, p.113) souligne qu'« on est nullement assuré que l'enfant distingue clairement les deux types d'unités (unités graphiques et unités phoniques que l'adulte lui demande d'associer ... Lettres et phonèmes n'apparaissent pas comme des unités acquises en début d'apprentissage ».

Lorsque la lettre est présentée, comme dans l'épreuve 9, l'élève peut la traiter comme le simple dessin d'une forme, sur laquelle on attire par ailleurs l'attention de l'élève quant à l'orientation (cf. épreuve 8 « lettres symétriques »), indépendamment du signifié auquel il est rattaché.

Lorsque l'on présente à l'élève des carrés (dans les tâches « silhouette de mots » et « double dessin »), l'élève doit attribuer de lui-même un signifié (le carré est ici le support matériel de la lettre (1^{er} cas), ou le support matériel de la syllabe (2^{ième} cas)). C'est à lui de reconstruire ce lien (rappelons que ces deux épreuves sont utilisées par le même enseignant). Dans tous les cas, le dessin n'a pas la même valeur selon qu'il est présumé fonctionner comme un signifiant ou comme un signifié voire comme l'un et l'autre au sein de la même tâche (cf. épreuve 10 « double dessin »)

3.3. Le dessin dans des tâches de dénombrement

Dans 21 des 74 épreuves de mathématiques données au premier trimestre et destinées à évaluer l'item 1 (ou des items proches) « Connaître (savoir écrire et nommer) les nombres entiers naturels inférieurs à 100 », les élèves doivent dénombrer les éléments d'une collection représentée par un dessin ET exprimer le cardinal de l'ensemble sous forme chiffrée. C'est ce qui justifie que nous nous intéressions tout particulièrement à cet item et à la place du dessin dans les épreuves permettant de l'évaluer.

3.3-1. L'item relatif à la désignation des nombres en mathématiques.

Nous allons préalablement analyser une trace de la redéfinition de la tâche prescrite aux enseignants : celle qui préside, en amont, à la détermination de l'item du livret.

a- l'item des programmes.

Dans le texte très condensé des programmes de mathématiques du cycle 2 qui précède les listes d'items, il est précisé que « les élèves apprennent la numération décimale inférieure à 1 000. Ils dénombrent des collections, connaissent la suite des nombres, comparent et rangent » (B.O. 2008 p. 118).

Pour le CP, les programmes déclinent ensuite ce qui a trait à la désignation des nombres dans un item : « connaître (savoir écrire et nommer) les nombres entiers naturels inférieurs à 100 ». La connaissance des nombres renvoie ici à savoir écrire les nombres et à savoir les nommer.

Afin de guider les enseignants, le ministère a publié un document d'accompagnement des programmes⁵⁴ dans lequel deux difficultés sont explicitées. L'une d'elles concerne l'accès au symbolisme : « la mise en correspondance des quantités avec des systèmes de symboles, qu'il s'agisse de la suite orale des noms des nombres, des configurations de doigts, des abaques ou des chiffres arabes posent problème à tous les enfants » (2010, p.9). L'autre concerne le passage des transformations aux opérations : « le fait que les enfants perçoivent et comprennent très précocement et facilement les effets des transformations affectant la quantité laisse souvent penser à tort qu'ils maîtrisent ou au moins comprennent les opérations » (2010, p. 10). Dans ce document d'accompagnement, un chapitre entier est consacré à « Apprendre le nombre » lui-même décliné en trois sous chapitres « premières compétences pour accéder au dénombrement », « du comptage au calcul » « débiter la numération ». Il y est notamment fait mention des relations entre les différentes formes de représentations des nombres et les quantités (2010, p. 28). De cette rapide présentation, nous retiendrons que les documents d'accompagnement, même s'ils sont assez axés sur la question du calcul, font une place aux différentes désignations des nombres (orale, écrite ou collection équipotente) et qu'ils sont davantage articulés au dénombrement que ne le sont les progressions du programme qui ne l'évoquent plus.

b- l'item dans les livrets

Nous reproduisons dans le tableau ci-dessous les seize traductions de l'item collectées dans les livrets des enseignants.

⁵⁴Le nombre au cycle 2. DA 2010 MEN SCEREN CNDP

Formulations de l'item dans les livrets :

« Connaître (savoir écrire et nommer) les nombres entiers naturels inférieurs à 100 » ⁵⁵	Enseignant A ; D ; E ; F
Réciter, lire et écrire la suite des nombres	Enseignant B
Au premier trimestre « Je connais la suite des nombres jusqu'à 39 » « Je sais compléter la suite numérique (>39) » « Je sais associer la désignation orale et chiffrée d'un nombre » Au troisième trimestre « Je nomme et j'écris en chiffres les nombres de 0 à 100 »	Enseignant H
Au premier trimestre « Je connais la suite des nombres jusqu'à 29 » « Je sais compléter la suite numérique jusqu'à 29 » « Je sais associer la désignation orale et chiffrée d'un nombre » Au troisième trimestre « Je connais la suite des nombres jusqu'à 100 »	Enseignant I
« Dénombrer une collection <20, <50, <100 » « Dire la suite des nombres jusqu'à 20/50/100 » « Se repérer sur la frise numérique jusqu'à 20/50/100 »	Enseignant C
« Lire et écrire des nombres inférieurs à 100 » « Dénombrer et réaliser des collections »	Enseignant G
Au premier trimestre « Maîtriser la suite orale des nombres de 1 en 1 au moins jusqu'à 39 » « Maîtriser la suite écrite des nombres de 1 en 1 jusqu'à 39 » « Ecrire en chiffres les nombres dictés » « Dénombrer et réaliser des quantités inférieures à 39 » Au troisième trimestre « Maîtriser la suite orale des nombres de 1 en 1 au moins jusqu'à 100 » « Maîtriser la suite écrite des nombres de 1 en 1 » « Maîtriser la suite écrite des nombres de 10 en 10 » « Ecrire en chiffres les nombres dictés »	Enseignant J
Au premier trimestre « Savoir écrire et nommer les nombres entiers de 0 à 39 » « Maîtriser la suite écrite des nombres jusqu'à 39 » « Dénombrer le nombre d'objets dans une collection organisée » « Compléter une collection d'objets organisée » Au troisième trimestre « Maîtriser la suite orale des nombres jusqu'à 99 » « Maîtriser la suite écrite des nombres jusqu'à 99 » « Ecrire en chiffres et en lettres des nombres »	Enseignant K
« Connaître (savoir écrire et nommer) les nombres entiers naturels inférieurs à 100 » « Exprimer une quantité par un nombre » « Dénombrer une collection organisée par dizaines » (uniquement troisième trimestre) « Dénombrer une collection en utilisant des groupements par dix. »	Enseignant L
Au premier trimestre « Dénombrer une quantité » « Connaître la suite écrite des nombres de 1 à 40 » « Savoir écrire les nombres en chiffres » « Compléter une collection de 10 objets » Au troisième trimestre	Enseignant M

⁵⁵Ceci est l'intitulé exact de l'item des programmes. Il est repris à l'identique dans le livret du SCEREN.

« Ecrire en chiffres des nombres dictés » « Connaître la suite écrite des nombres de 1 à 100 » « Dénombrer des quantités jusqu'à 100 »	
« J'utilise des nombres ordinaux (1 ^{er} , 2 ^{ème} ...) » « Je dénombre une collection » « Je nomme et j'écris les nombres en chiffres jusqu'à 10 » « Je nomme et j'écris les nombres en chiffres jusqu'à 69 » « Je nomme et j'écris les nombres en chiffres jusqu'à 100 » « Je sais faire des groupements de dizaines et d'unités »	Enseignant N
Au premier trimestre « Ecrire les nombres en chiffre (entre 0 et 10) » « Dénombrer et coder des quantités » Au troisième trimestre « Lire les nombres (de 0 à 100) » « Ecrire les nombres en chiffres (de 0 à 100) » « Ecrire les nombres en lettres (de 0 à 10) » « Dénombrer et coder des quantités »	Enseignant O
« Dénombrer et coder des quantités » « Décomposer un nombre en centaines, dizaines, unités » « Produire des suites orales et écrites de nombres » « Ecrire des nombres en chiffres et en lettres »	Enseignant P

Tableau 15 : Formulations de l'item du programme « Connaître (savoir écrire et nommer) les nombres entiers naturels inférieurs à 100 » dans les livrets utilisés par les enseignants.

Analyse des formulations :

Nous pouvons établir le premier constat selon lequel tous les enseignants font figurer un item relatif à la désignation des nombres en mathématiques dans leurs livrets, au premier et au troisième trimestre.

Nous soulignerons deuxièmement que les rédactions choisies diffèrent suivant les enseignants ce qui traduit l'aspect générique de la formulation choisie dans les programmes:

- Cas 1 (majoritaire) : 11 enseignants⁵⁶ (sur 16) déclinent cet item en sous-composants dont le nombre est variable en quantité (de 2 à 6 pour l'enseignant N par exemple) et en qualité (les énoncés varient suivant les trimestres).
- Cas 2 (minoritaire) : 5 enseignants⁵⁷ (sur 16) formulent à l'identique cet item dans leurs livrets ou le reformulent en des termes proches.

Troisièmement, parmi les 11 enseignants qui ne reprennent pas tel quel l'item des programmes, on retrouve un souci marqué quant à la place à accorder au dénombrement de collection, voire à la possibilité d'organiser des groupements de 10. A part 2 d'entre eux⁵⁸, les 9 autres utilisent le terme « dénombrer » en le rapprochant du fait de déterminer le cardinal d'une collection. Cette dimension du nombre n'est pas explicite dans les programmes de CP : présent dans le texte du programme, elle est absente des progressions. Pour autant en fin de maternelle, il est attendu des élèves qu'ils sachent « comparer des quantités, résoudre des problèmes portant sur les quantités », « dénombrer une quantité en utilisant la suite orale des nombres connus », « associer le nom de nombres connus avec leur écriture chiffrée » -BO 2008, p.16)

La rédaction choisie par l'enseignant L résume assez bien le souci qu'une part des enseignants accorde au dénombrement. On peut faire l'hypothèse d'une empreinte encore

⁵⁶ Les enseignants C ; G ; H ; I ; J ; K ; L ; M ; N ; O et P

⁵⁷ Les enseignants A ; B ; D ; E et F.

⁵⁸ Les enseignants H et I

forte des programmes 2002 qui mentionnaient explicitement le dénombrement dans le paragraphe « 2.1 Désignations orales et écrites des nombres entiers naturels (inférieurs à 1000) : - dénombrer et réaliser des quantités en utilisant le comptage un à un ou des groupements et des échanges par dizaines et centaines.. »

Ce préambule nous invite à considérer que les enseignants ne semblent pas renoncer à un aspect du nombre qu'ils jugent majoritairement devant figurer dans les livrets : l'aspect cardinal (la taille d'une collection étant désigné par ce dernier).

3.3-2. Épreuves de dénombrement

Toutes les collections en jeu dans les tâches d'évaluation sont des collections figurées. Nous présentons dans ce qui suit l'analyse de six tâches impliquant des représentations différentes des collections. Trois représentations des collections sont utilisées : « stéréotypée » (figuration des doigts d'une main, constellation du dé, cartes à jouer), « pseudo-stéréotypée » (des doigts des deux mains pour une quantité inférieure à 5 par exemple, des groupements de stéréotypes) ou « sans organisation spécifique ». On peut étudier leur effet sur l'activité de l'élève.

Épreuve n°11 : « Groupes de 5 »

a- l'item :

Au premier trimestre, l'enseignant L souhaite évaluer les items « Dénombrer une collection organisée par dizaine (appui possible sur dix, vingt, trente) ». Cet item ne figurant pas dans les progressions, l'enseignant, au cours d'un entretien post-collecte, indique le référer à « connaître (savoir écrire et nommer) les nombres entiers naturels \leq à 100 ».

b- l'épreuve :

Pour ce faire, l'enseignant choisit d'utiliser l'épreuve reproduite ci-dessous :

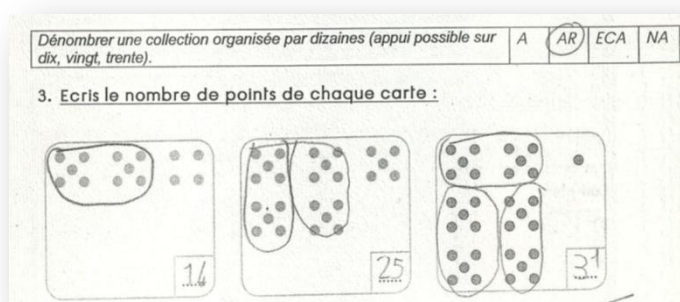


Figure 14 : Épreuve MT1 (Annexe B, p. 174) avec réponses d'élève

Il est attendu que l'élève dénombre les éléments des trois collections et qu'il transcrive le résultat obtenu dans la numération orale intériorisée (par ex. « trente-et-un ») puis dans la numération chiffrée (par ex. « 31 »). Les collections sont figurées au moyen de représentations pseudo-stéréotypées des nombres 4 et 5 : elles ne sont pas organisées par dizaines comme l'annonce la consigne de l'épreuve. La figuration des collections choisie sous-tend que l'élève combine des paires de « cinq » pour former lui-même des dizaines. Pour répondre « 31 » à la dernière demande, l'élève peut utiliser différentes procédures

- dénombrer par comptage-numérotage les éléments de la collection de 1 à 31
- sur-compter à partir d'un nombre d'éléments rapidement reconnu et dont l'équivalence dans la numération orale est disponible (par exemple « vingt » puis « 21, 22 31 »).
- utiliser une procédure arithmétique
 - additive avec repérants : l'élève fourni le résultat en disant « dix, vingt, trente, et un : trente-et-un).
 - multiplicative : l'élève fourni le résultat en disant « un, deux, trois (paquets de 10) trente, et un » = « trente-et-un ».

Pour réussir l'épreuve, l'élève doit donc réaliser successivement deux tâches : la première est une tâche de dénombrement (pour laquelle il existe comme nous venons de le voir différentes stratégies de résolution ne faisant pas appel aux mêmes connaissances), la deuxième est une tâche de transcription qui n'est pas prise en compte pour elle-même (on suppose que la traduction du nombre « 31 » obtenu dans la numération orale vers la numération écrite chiffrée ne pose pas de problème particulier).

c- la copie de l'élève :

Elle permet de constater que l'élève a su constituer des groupes de 10 en regroupant à chaque fois deux groupes de 5.

L'élève a correctement traduit le résultat du dénombrement en écriture chiffrée.

La validité de l'épreuve peut être discutée pour deux raisons :

Premièrement, elle ne permet pas de savoir si l'élève est capable de dénombrer une collection d'objets regroupés par dizaines car les collections ne sont pas organisées ainsi explicitement. Cette tâche mobiliserait des connaissances sur la décomposition additive (et/ou multiplicative) d'un nombre. $31 = 10+10+10+1$ ou $31 = 3 \times 10 + 1$. Elle ne permet pas non plus de savoir si l'élève est capable d'organiser lui-même la même collection d'objets présentée de façon non ordonnée, ce qui serait une tâche d'une autre nature dans laquelle les élèves utiliseraient probablement davantage une stratégie de comptage-numérotage. L'épreuve engage une tâche intermédiaire « regrouper les paquets de cinq pour former des dizaines » dont l'intérêt n'est pas remis en cause pour enseigner. Cette étape implicite conduit à remettre en cause l'usage de l'épreuve pour évaluer ce qui est annoncé à savoir la capacité des élèves à « dénombrer une collection organisée par dizaines »

Deuxièmement, l'épreuve va au-delà d'un seul item de dénombrement : il faut non seulement que l'élève dénombre mais aussi qu'il transcrive le résultat du dénombrement d'une numération (orale, intériorisée) dans une autre (écrite, positionnelle) ce qui implique la production d'une écriture chiffrée à deux chiffres implicitement supposée maîtrisée, ce qu'il n'est pas trivial à l'issue du premier trimestre de CP.

Épreuve n°12 : « Picbille »

a- l'item :

Au premier trimestre, l'enseignant C souhaite évaluer l'item « Dénombrer une collection inférieure à 20 ». Cet item ne figure pas dans les programmes et l'enseignant, au cours d'un entretien post-collecte, indique le référer à celui des programmes « connaître (savoir écrire et nommer) les nombres entiers naturels \leq à 100 ».

b- l'épreuve :

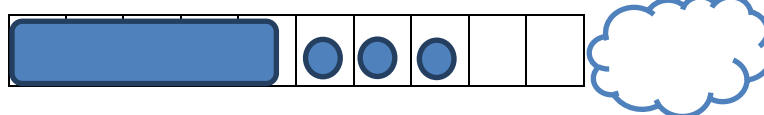
Pour ce faire, l'enseignant choisit d'utiliser l'épreuve reproduite ci-dessous :



Figure 15 : Épreuve MT1 (Annexe B, p. 159)

Cette épreuve comporte quatre demandes. Pour une meilleure compréhension nous explicitons ci-dessous la demande schématisée dans la partie supérieure gauche. Celle-ci, comme les trois autres, se compose de deux parties :

- 10 cases consécutives contenant ou non un « jeton » marquant une unité. Chaque case correspond en réalité à un logement d'une boîte en contenant 10 et séparable en deux parties égales, vendue en sus du manuel Picbille (Brissiaud⁵⁹). Lorsque 5 cases consécutives sont remplies, un rectangle les recouvre (il représente le couvercle de la boîte fermée). Un rectangle « vaut » donc 5. Dans l'exemple reproduit ci-dessous, il faut considérer que 5+3 soit 8 cases sont ainsi remplies.
- Un nuage dans lequel l'élève doit transcrire le nombre de cases remplies.



Il est attendu que l'élève dénombre les éléments de la collection proposée puis qu'il écrive le résultat du dénombrement dans la numération chiffrée. Dans cette tâche proche de tâches présentées dans le manuel, la figuration pseudo-stéréotypée engage l'élève à adopter une procédure de sur-comptage à partir de 5 (il ne peut pas compter les 5 premiers jetons) selon les principes de numération ordinale décrite par Mounier (2012), à considérer « 5 » comme un repérant (au moins partiel). Le comptage-numérotage, ni le dénombrement d'une collection en effectuant une correspondance terme à terme ne sont possibles. L'élève doit partir de 5 et poursuivre, soit en ayant recours à une stratégie de sur-comptage⁶⁰ « cinq, six, sept, huit », soit une stratégie arithmétique : « une barre c'est cinq plus un (« six »), plus un (« sept »), plus un (« huit ») : « huit » », ou « deux barres c'est dix, il reste deux places libres donc on peut soustraire deux de dix, et obtenir huit ».

La production du résultat du dénombrement engage une écriture chiffrée qui n'est pas explicite dans l'item du livret tel qu'il est rédigé. La tâche de dénombrement est évaluée au moyen d'une connaissance de l'écriture chiffrée des nombres.

La validité de cette tâche peut se discuter, malgré son intérêt mathématique. La tâche proposée dépasse la seule maîtrise du dénombrement par simple utilisation de la chaîne numérique orale. Le matériel contraint l'élève dans l'organisation et le traitement de la collection : il n'est de surcroît pas accessible et l'élève doit travailler à partir d'une représentation d'une collection (s'il utilise véritablement le matériel en classe rien ne l'empêche de débiter le comptage à 1, s'il ne l'utilise pas il faut qu'il associe le rectangle à 5).

⁵⁹ Brissiaud R. (2009). J'apprends les mathématiques avec Picbille - CP. Retz : Paris.

⁶⁰ Voire de sous-comptage : « dix, neuf, huit »

Épreuve n°13 : « Doigts »

a- l'item :

Au premier trimestre, l'enseignant G souhaite évaluer les items du livret « Je sais associer la désignation orale et chiffrée d'un nombre ». Cet item ne figure pas tel quel dans les programmes, et l'enseignant, au cours d'un entretien post-collecte, indique le référer à « connaître (savoir écrire et nommer) les nombres entiers naturels \leq à 100 ». Une dictée de nombre semblerait constituer une tâche de nature à correspondre à l'item.

b- l'épreuve :

L'enseignant choisit d'utiliser l'épreuve reproduite ci-dessous :

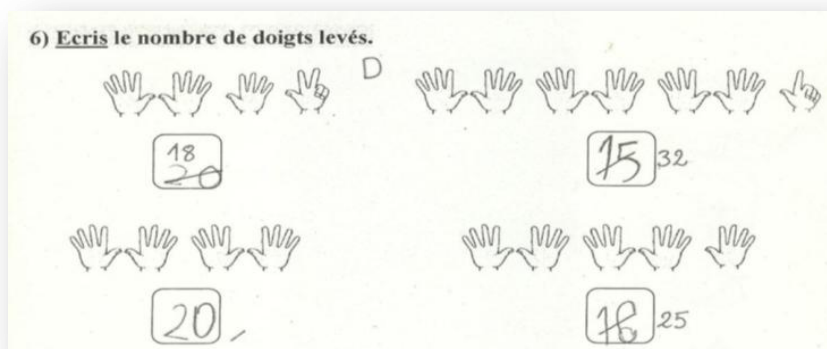


Figure 16 : Épreuve MT1 (Annexe B, p. 166) avec réponses d'élève

L'élève doit dénombrer les doigts levés des mains représentées, se donner oralement le résultat de ce comptage et écrire le résultat du dénombrement dans la numération chiffrée. La production du résultat engage une activité de dénombrement qui n'est pas explicitement formulé dans l'item.

c- la copie de l'élève :

Trois des quatre réponses que l'élève fournit sont des multiples de 5. Il fournit deux fois la même valeur (« 20 ») aux collections présentées à gauche, au moyen de 4 mains : on peut conjecturer qu'il a tenu compte des mains et non pas des doigts baissés. Les deux réponses, erronées, à droite (« 15 » et « 16 ») sont plus difficiles à analyser. Elles sont toutes deux numériquement inférieures à celles de gauche, quand bien même les collections de mains et de doigts levés sont supérieures, ce qui peut se concevoir sans avoir besoin de dénombrer. Nous n'avons pas pu élaborer d'hypothèse plausible.

A une tâche de transcription d'un système de numération à un autre s'est surajoutée une tâche de dénombrement non triviale parce qu'elle semble supposer acquise la capacité de l'élève à compter les groupes de 10, autrement dit à adopter une stratégie arithmétique (additive ou multiplicative), de nature à invalider du point de vue psycho-didactique.

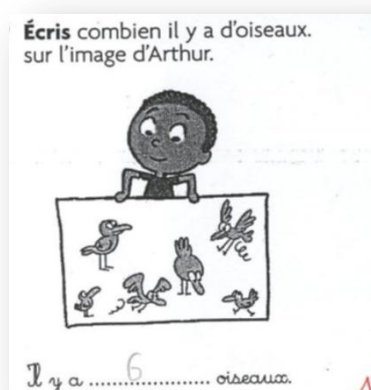
Épreuve n°14 : « Oiseaux »

a- l'item :

Au premier trimestre, l'enseignant E souhaite évaluer l'item du livret « connaître (savoir écrire et nommer) les nombres entiers naturels \leq à 100 ».

b- l'épreuve :

Pour ce faire, l'enseignant choisit d'utiliser l'épreuve reproduite ci-dessous :



Les spirales placées sous deux des oiseaux sont un ajout probablement estimé « ludique » par le concepteur de la tâche.

Figure 17 : Épreuve MT1 (Annexe B, p. 162) avec réponse d'élève

Il est attendu que l'élève dénombre les éléments de la collection d'oiseaux et traduise le résultat du dénombrement de la numération orale dans la numération chiffrée. La production du résultat du dénombrement engage une activité de dénombrement qui n'est pas explicite dans l'item tel qu'il est rédigé. La tâche d'association d'un nombre dans les numérations orale et chiffrée est évaluée au moyen d'une tâche de dénombrement.

Dans le cas de telles collections sans organisation spécifique, un cas moins fréquent que le cas précédent, on relève que la taille du nombre en jeu est souvent moindre que dans le cas des collections précédemment discutées : c'est probablement une interprétation ordinale du nombre qui est mobilisée pour accéder au cardinal, ce qui fait penser que dans les collections pseudo-organisées une interprétation arithmétique additive ou multiplicative du nombre est plutôt recherchée. Suivant le type d'analyse proposé par Mounier (2010, p. 5-27), pour dénombrer la collection présentée dans l'épreuve reproduite ci-dessous, l'élève mobilise un procédé modèle correspondant à une interprétation ordinale sans repérant puisque seule la série ordinale des comptants est employée (« un, deux, trois ... six»). Dans la deuxième partie de la tâche, il doit exprimer dans la numération chiffrée le nom du nombre. « Le travail mathématique dans la première partie de la tâche (exprimer le nom du nombre dans la langue parlée), n'est pas du même ordre que dans la deuxième (chercher à écrire un mot donné oralement) » (Mounier, 2010, p.20).

c- la copie de l'élève :

Elle n'apporte aucun élément nouveau à l'analyse.

La tâche proposée nous semble davantage être une tâche de dénombrement que d'écriture d'un nombre. En ce sens, elle ne paraît pas la plus indiquée pour évaluer l'item du livret choisi par l'enseignant.

Épreuve n°15 : « Cœurs »

a- l'item :

Au premier trimestre, l'enseignant C souhaite évaluer l'item du livret « dénombrer une collection <20 »

b- l'épreuve :

Pour ce faire, l'enseignant choisit d'utiliser l'épreuve reproduite ci-dessous :

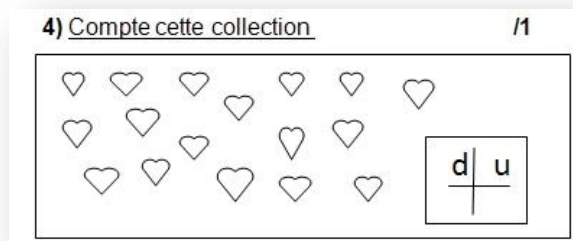


Figure 18 : Épreuve MT1 (Annexe B, p. 159)

Il est attendu de l'élève qu'il dénombre la collection de cœurs puis qu'il traduise le résultat du dénombrement de la numération orale vers la numération chiffrée. Ceci engage l'élève à adopter une procédure de dénombrement ordinaire avec repérant puisque deux séries sont utilisées (les comptants « un, deux... » et les repérants (« dix ... »)). La production du résultat du dénombrement engage une écriture chiffrée qui n'est pas explicite dans l'item tel qu'il est rédigé puisqu'il est simplement demandé à l'élève de « compter cette collection ». La tâche de dénombrement est évaluée au moyen d'une connaissance de l'écriture chiffrée des nombres en base 10 dont la maîtrise semble être un pré requis (d pour dizaine et u pour unité).

Du point de vue psycho-didactique, la validité de la tâche est remise en cause. A ce moment de l'apprentissage (fin de premier trimestre), il est tout à fait concevable qu'un élève puisse parvenir à dénombrer les 17 éléments de la collection de cœurs présentée mais ne puisse pas l'écrire dans la forme attendue.

Discussion

Dans les épreuves qui viennent d'être analysées, nous avons montré qu'il est toujours attendu que l'élève matérialise dans la numération écrite le résultat d'un dénombrement éventuellement entrepris au moyen de la numération orale (mais dont on peut faire l'hypothèse qu'elle est de toute façon utilisée mentalement). C'est une des raisons pour lesquelles la validité de toutes ces tâches ne semble pas assurée d'un point de vue psycho-didactique : dénombrement et écriture chiffrée sont toujours liés dans les tâches d'évaluation. On ne trouve pas d'épreuves demandant aux élèves de déterminer des collections comportant autant d'éléments qu'une collection témoin : elles permettraient d'éviter le recours à l'expression d'un nombre dans la numération chiffrée. Les tâches d'écriture de nombres sous la dictée sont plus directement liées à l'équivalence du mode de désignation d'un nombre dans les numérations parlées et chiffrées indépendamment de la question du dénombrement.

Ces tâches sont l'occasion d'une discussion au sujet du mode de présentation des collections. Elles engagent en effet une activité mathématique de l'élève différente selon les représentations des collections utilisées, et il semble bien que la pré-organisation des collections puisse influencer sur l'activité de l'élève au cours de la redéfinition de la tâche. « Rien n'assure qu'une bonne maîtrise du comptage un à un, dans le sens où la procédure conduit à l'expression correcte du nombre dans les situations les plus courantes, témoigne d'une compréhension suffisante du nombre » indique Mounier (2010, p. 15). La taille des collections est évidemment une autre variable des situations proposées : au premier trimestre quand certains enseignants demandent de dénombrer des collections de 6 objets (tâche 13 « oiseaux »), d'autres font travailler leurs élèves sur des collections nettement supérieures (tâche 14 « cœurs »).

À l'issue d'une recherche portant sur l'existence de connaissances nécessaires au comptage en CP, Briand, Lacave Luciani, Harvouët et coll. (2000, p.8) concluent que « lorsque le professeur commande une action de comptage, l'élève doit mettre en œuvre des connaissances (de nature spatiale) qui permettent d'explorer la collection à dénombrer afin de n'oublier aucun élément et de ne pas repasser deux fois sur le même. Ces connaissances ne font pas habituellement l'objet d'un enseignement. Leur dysfonctionnement entraîne pourtant des échecs dans le comptage ».

Dans le même article, ces auteurs ont analysé un ensemble d'épreuves de détermination du cardinal d'une collection dessinée et se sont posé la question des stratégies induites ou favorisées par le dessin. Ils en ont déduit (Briand et coll. 2000 p.9), que pour compter le nombre d'éléments d'une collection finie montrée, l'élève doit nécessairement⁶¹ :

« *Etre capable de distinguer deux éléments différents d'un ensemble donné.*

- 1- *Choisir un élément d'une collection.*
- 2- Enoncer un mot nombre (« un » ou le successeur du précédent dans une suite de mots-nombres).
- 3- *Conserver la mémoire de la collection des éléments déjà choisis.*
- 4- *Concevoir la collection des objets non encore choisis.*
- 5- *Recommencer (pour la collection des objets non encore choisis) 2-3-4-5 fois tant que la collection des objets à choisir n'est pas vide.*
- 6- *Savoir que l'on a choisi le dernier élément.*
- 7- Enoncer le dernier mot nombre.

Les étapes en italique (1, 2, 4, 5, 6, 7) constituent une tâche spécifique que nous appelons inventaire, au cours de laquelle il s'agit de passer en revue tous les éléments d'une collection finie une fois et une seule. Cette tâche caractérise une connaissance non enseignée que nous appelons énumération, faute d'un autre nom»

Selon le mode d'analyse de ces auteurs, on peut estimer que les présentations stéréotypées et pseudo-stéréotypées des collections utilisées dans les épreuves discutées suppriment, ou agissent fortement sur, les étapes 4 et 5 de l'énumération à savoir « la conservation de la

⁶¹ Cette suite d'étapes renvoie au schème du dénombrement dans la théorie des champs conceptuels de Vergnaud(1997, pp. 62-64) : ce schème comporte un but (associer à une collection discrète un nombre qui sera sa mesure), des règles (ne pas compter deux fois le même élément, les compter tous...), des invariants opératoires de type concepts-en-acte (bi-univocité, cardinal, successeur, itération...) et de type « théorèmes en acte » (le cardinal est indépendant de l'ordre dans lequel on compte les éléments).

mémoire de la collection des éléments déjà choisis » et « la conception des collections des éléments non encore choisis ». Dans le cas de l'usage de telles représentations pseudo-stéréotypées des collections, le dessin peut induire une stratégie de dénombrement : sur-comptage, décomposition multiplicative du nombre.

3.4. Synthèse du chapitre 3

En conclusion de ce chapitre, il apparaît que le dessin constitue un composant de la tâche susceptible d'intervenir dans la redéfinition de la tâche opérée par l'élève au cours de la réalisation de la tâche. Il peut intervenir en modifiant la nature du processus de réponse de l'élève tel qu'envisagé par l'enseignant au niveau de la détermination de l'item.

L'effet de ce composant de la tâche nous a conduit à remettre en cause, du point de vue de leur validité psycho-didactique, l'ensemble des épreuves analysées dans ce chapitre à des fins d'évaluation de divers items de Français ou de Mathématiques relevant de compétences complexes ou d'habiletés de base. Dans bien des cas, « l'habillage » de la tâche réalisée au moyen du dessin est susceptible de modifier la tâche elle-même.

Chapitre 4. Les questions écrites de compréhension en lecture

Nous avons montré que le dessin pouvait jouer un rôle important au cours de la redéfinition de la tâche prescrite à l'élève.

Nous allons maintenant nous attacher à analyser le composant le plus traditionnel et typique des épreuves d'évaluation de la compréhension en lecture que constitue le questionnaire (E. Fijalkow, 2003), et à en interroger l'usage.

Dans ce chapitre, nous proposons de considérer le questionnaire de lecture comme un texte en lui-même c'est-à-dire une « production signifiante »⁶² pouvant de surcroît éventuellement prendre le texte pour objet de discours. A ce titre, le questionnaire peut être assimilé à un méta-texte, « un texte sur un texte » (Daunay, 1993). Il participe à la représentation que l'élève élabore de la situation générée par la tâche. Nous nous proposons d'analyser particulièrement le fonctionnement des questions et des questionnaires (écrits à réponses écrites).

Dans le cas des mathématiques, la question en lien avec l'énoncé des données, problématise l'énoncé et le transforme en problème. Sans question il n'y a pas de problème. Dans le cas du français, les questions posées sur le texte lu se surajoutent au texte : le texte, support des questions, a une existence propre, totalement indépendante de questions qui orientent l'activité de lecture vers une « parcellisation du parcours de lecture » (cf. Cèbe et Goigoux, 2013, p.27). Pour paraphraser Daunay (opus cité), le questionnaire peut être considéré de ce fait comme une méta-tâche, une tâche (comprendre la question de compréhension et traiter le questionnaire) sur une tâche (comprendre un texte).

Nous formulons l'hypothèse que les épreuves contenant des questionnaires de lecture évaluent surtout la capacité de l'élève à répondre aux questions et à traiter le questionnaire que la compréhension du texte sur lequel les questions portent. Le cas échéant, cela constituerait un argument de nature psycho-didactique permettant d'invalidier l'usage de questionnaires écrits à réponse écrite pour évaluer la compréhension.

Cèbe et Goigoux soulignent bien le fait que les modalités de résolution (d'un questionnaire) sont rarement enseignées (2013, p.31). L'évaluation au moyen de questionnaire repose donc sur une mobilisation de connaissances stratégiques implicites.

Nous examinons dans un premier temps les traductions opérées par les enseignants de la prescription qui leur est adressée par les programmes en matière de compréhension. Comment les enseignants opérationnalisent-ils la prescription relative à l'enseignement de la lecture en matière de compréhension ? Dans un second temps nous analyserons quelques épreuves en caractérisant la relation entre des textes et les questions qui s'y rapportent et en examinant les processus de lecture impliqués.

⁶² Repéré à <http://www.cnrtl.fr/lexicographie/texte>

4.1. L'item relatif à la compréhension en lecture.

Analysons tout d'abord les traces de la redéfinition de la tâche prescrite aux enseignants au sujet de la compréhension en lecture : celle qui préside, en amont, à la détermination de l'item du livret.

a- l'item des programmes relatif à la compréhension.

Dans le texte des programmes de français du cycle 2, le mot « compréhension » n'est pas employé. Le mot, proche, « comprendre » l'est 3 fois : une fois dans le paragraphe consacré au Langage oral, une fois dans le paragraphe consacré au Vocabulaire et une fois dans le paragraphe consacré à la lecture et à l'écriture sous la forme suivante : « (les élèves) acquièrent le vocabulaire et les connaissances nécessaires pour comprendre les textes qu'ils sont amenés à lire. » (MEN 2008 p. 17). On peut être conduits à penser que concourent à la compréhension, le vocabulaire ainsi qu'un ensemble de « connaissances nécessaires » qui ne sont pas explicitement définies.

Le programme fait apparaître l'item de lecture « dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens ». C'est le seul des onze items du sous-domaine de la lecture qui a trait à la compréhension d'un texte. Ce qui est écrit sous la forme d'un item est en réalité une juxtaposition de trois composants amalgamés pour ne constituer qu'un seul item : « Dire de qui ou de quoi parle le texte lu (1^{er} composant) ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu (2^{ième} composant); reformuler son sens (3^e composant)».

Chacun de ces composants renvoie à des aspects bien distincts :

- Le premier composant renvoie précisément, selon Giasson (1990, p.76), à la détection du sujet du texte, lequel peut être résumé par un mot ou une expression. C'est une réponse à la question « de quoi parle le texte ? ».
- Le second composant renvoie à la production d'une réponse à une question concernant le texte. Il est précisé que la réponse se trouve dans le texte ou l'illustration⁶³ ce qui n'est pas du même niveau, ni n'engage la même activité. Dans le second cas (celui de l'illustration), on ne peut plus parler de lecture mais d'analyse d'image.
- Le troisième composant renvoie à la reformulation par l'élève du sens du texte. Toutefois il n'est pas dit ce que l'on souhaite voir reformuler par l'élève (le terme sens étant assez général). Il est ici visé, semble-t-il, la notion d' « idée principale » dont Giasson (1990, p. 75) indique que les caractéristiques varient selon le type de texte et qu'elle est à distinguer du sujet du texte. Ainsi, « dans un récit, l'idée principale concerne les événements et leur interprétation alors que dans les textes informatifs, ce qui est important peut être un concept, une généralisation, une règle etc. ». L'idée principale (celle que l'auteur a fournie pour expliciter le sujet) peut en outre être implicite et dans ce cas elle doit être inférée à partir des informations contenues dans le texte. Dans les récits, la grammaire de récit apporte les éléments nécessaires pour détecter l'idée principale et conduit certains auteurs à préconiser le rappel de récit comme stratégie d'enseignement et comme technique d'évaluation. Giasson (1990, p. 113-119) propose ainsi de mixer une approche quantitative (en référence aux travaux de Denhière) et qualitative (en référence aux travaux d'Irwin, 1991).

⁶³ Nous ne reviendrons pas sur les éventuels obstacles que le dessin représente en termes de représentation du texte par le lecteur. Cela a été discuté dans le chapitre précédent.

Cet item est donc une conjonction de plusieurs compétences telles que savoir détecter le sujet du texte lu, reformuler l'idée principale du texte lu (et qui engagent des macro processus selon Giasson (1990, p. 16) i.e. des processus orientés vers la compréhension globale du texte) et, produire une réponse à une question portant sur des éléments explicites. Cette dernière compétence étant plutôt une compétence de « répondeur » (nous nous en expliquerons dans le chapitre suivant). Nous laissons de côté dans ce chapitre l'indication « trouver dans l'illustration la réponse à des questions... », le chapitre précédent a permis de discuter de la place du dessin dans les épreuves d'évaluation. Toutefois, on peut simplement remarquer que la formulation de l'item invite les enseignants à utiliser des supports contenant des illustrations, et les incite à demander aux élèves de trouver la réponse à des questions portant sur le texte dans l'image.

Afin de guider les enseignants, le prescripteur a publié un document d'accompagnement des programmes⁶⁴ dans lequel on retrouve le terme « comprendre », dans la formule « apprendre à comprendre », associée à l'item discuté (MEN, 2012, p. 19). Des précisions sont données quant à la nature des textes donnés à lire devant « renvoyer à des univers connus et ne comportant pas de mots rares » (idem), et devant « susciter l'élaboration d'inférences » (idem) afin que les élèves prennent conscience de l'importance stratégique et linguistique d'en faire. Pour cela, les maîtres sont invités à ne pas faire de lecture préalable du texte : « c'est la complexité de l'acte de lire que l'on affronte » (idem). La formule « dire de qui ou de quoi parle le texte lu » de l'item des programmes est ainsi précisée : il s'agit d'un texte lu par l'élève et non pas par l'enseignant.

Les élèves sont « invités à mobiliser des connaissances en fonction du thème, du contexte ; ils font des hypothèses sur ce que le texte peut rapporter ou raconter » (idem). Une série de difficultés potentielles sont listées pour aider les enseignants à structurer leur enseignement : l'extraction du thème (que l'on peut assimiler au sujet) en est une, au contraire de la reformulation de l'idée principale ou de la difficulté à produire une réponse à une question posée qui ne sont pas mentionnées. De cette brève introduction, il nous apparaît que l'item du programme de CP constitue une tâche à réaliser très ouverte et susceptible de donner lieu à diverses interprétations, ce que nous avons cherché à observer en recueillant les livrets utilisés par les enseignants.

⁶⁴ Lire au CP.

b- l'item dans les livrets

Nous reproduisons dans le tableau ci-dessous les seize traductions de l'item collectées dans les livrets des enseignants (pour les 7 premières lignes, les items du livret du premier et du troisième trimestre sont identiques).

Formulations de l'item dans les livrets :

« Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens ». (Rédaction des programmes, un seul item)	Enseignants L
« Dire de quoi et de qui parle le texte lu » (livret SCEREN)	Enseignants A ; D ; F
« Répondre à des questions sur un texte lu » et « Comprendre le sens global d'un texte. » (Deux items différents)	Enseignant B
« Donner, après lecture, des renseignements sur un texte (compréhension) » et « Comprendre un texte et montrer qu'on l'a compris » et « Comprendre une consigne. » (Trois items différents)	Enseignant C
« Comprendre un texte lu. »	Enseignant E ; N ; O
« Lire silencieusement un mot, une phrase, un texte et en manifester sa compréhension. » et « Lire une consigne, un énoncé et comprendre ce qui est attendu. » (Deux items différents)	Enseignant P
« Comprendre une phrase puis un texte lu silencieusement »	Enseignant G
Au premier trimestre : « - Je sais lire et comprendre des phrases simples. » et « Je sais associer un mot et une illustration. » et « Je sais repérer un intrus dans une liste. » et « Je sais remettre une phrase en ordre. » (Quatre items différents) Au troisième trimestre : « - Je sais dire de quoi ou de qui parle un texte lu. » et « -Je trouve, dans un texte, les réponses à des questions simples. » (Deux items différents)	Enseignants H et I
Au premier trimestre : « - Être capable de dire de qui ou de quoi parle le texte lu à haute voix. » Au troisième trimestre : « - Comprendre les informations d'une phrase. » et « Comprendre les informations explicites d'un texte en répondant à des questions. » (Deux items différents)	Enseignants J et K
Au premier trimestre « Dire de qui ou de quoi parle le texte lu. » et « Compléter une phrase ». (Deux items différents) Au troisième trimestre : « - Comprendre un petit texte et répondre à des questions » et « Compléter une phrase pour qu'elle ait du sens. » et « Savoir reconstituer une phrase avec des mots / groupes de mots. » (Trois items différents)	Enseignant M

Tableau 16 : Formulations de l'item « dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens » dans les livrets utilisés par les enseignants.

Analyse des formulations :

Nous pouvons établir le premier constat selon lequel tous les enseignants font figurer un item relatif à la compréhension en lecture dans leurs livrets, quel que soit le trimestre considéré. C'est le seul item de Français qui rencontre une telle unanimité.

Suite à cela, nous pouvons formuler un deuxième constat : on trouve 10 versions différentes (autant de lignes dans le tableau) du même item des programmes dans les livrets utilisés par

les enseignants. Presque tous les enseignants se détachent de la rédaction officielle (sauf l'enseignant L). Parmi eux, notons que, pour 3 d'entre eux (cf. deuxième ligne du tableau), la redéfinition de la prescription officielle est d'une certaine manière exogène : elle est le fait du SCEREN dont ils utilisent la publication. Un choix a été fait par cet éditeur : seul le premier composant de l'item a été conservé (que nous pouvons considérer comme le but et non les moyens), comprendre un texte semble se résumer à être capable d'en dégager le thème, ce qui est très réducteur. Pour les autres, on souligne que les traductions opérées aboutissent à un, deux, trois voire quatre items différents matérialisés par autant de lignes différentes dans leurs livrets pourtant déjà fournis (enseignants H et I au premier trimestre). Ces choix sont parfois des choix d'équipes, parfois des choix personnels (lorsque par exemple il n'y a qu'une seule classe de CP dans l'école).

Un troisième constat peut être ensuite dressé : les 12 enseignants se démarquant des programmes et d'une traduction institutionnelle (le livret du SCEREN) utilisent pour leur part les termes de « compréhension » ou « comprendre », de manière tantôt particularisée (en renvoyant à un texte lu), tantôt générique (comprendre le sens global d'un texte). Cette subtilité renvoie à la représentation que les enseignants ont de la compréhension : dans le sens où la compréhension réfère à un texte lu, elle n'est pas universelle, indépendante du texte sur laquelle elle est évaluée.

Quatrième constat : les composants de l'item du programme sont diversement représentés dans les traductions opérées par les enseignants.

- « Dire de qui ou de quoi parle le texte lu », « Reformuler son sens » sont deux composants de l'item des programmes qui fonctionnent comme une invitation à envisager la compréhension comme une restitution orale (rappel de récit) puisqu'il s'agit de « dire » (quelque chose au sujet du texte) et de « reformuler » (le sens du texte). 9 enseignants⁶⁵ sur 16 conservent cette idée dans la rédaction même de leur livret.
- « Trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu » est un composant de l'item qui semble inviter à poser aux élèves des questions sur le texte (ou l'illustration qui l'accompagne, ce qui n'est pas à proprement parler une tâche de lecture comme nous l'avons souligné). 8 enseignants⁶⁶ conservent cette idée dans la rédaction de leurs livrets sans que la nature des questions ne soit toujours indiquée. Ainsi, aucun enseignant ne précise si les questions sont posées par écrit ou si elles sont données oralement.
- 6 enseignants⁶⁷ ne reprennent aucun des termes des trois composants de l'item des programmes. Ils utilisent une formulation assez générale mobilisant fréquemment le terme « comprendre (le sens global d'un texte ; un texte) » parfois sans plus de précisions. Parmi eux, deux mentionnent que l'élève doit « montrer qu' (il) a compris » ou « manifester sa compréhension » sans apporter davantage de précisions sur les moyens de le faire.

Il faut donc considérer que les rédactions de cet item n'étant pas identiques, les attentes risquent de ne pas être les mêmes. Plusieurs enseignants attendent que leurs élèves « comprennent » une phrase et/ou un texte sans que l'on sache comment se manifeste cette

⁶⁵ Les enseignants L ; A ; D ; F ; H ; I ; J ; K et M utilisent le terme « dire » ou celui de « reformuler » au moins une fois.

⁶⁶ Les enseignants L ; B ; H ; I ; J ; K et M utilisent le terme « question » au moins une fois.

⁶⁷ Les enseignants C ; E ; N ; O ; P et G.

compréhension. La modalité de restitution orale sous-entendue par les termes « dire » et « reformuler » n'est plus aussi explicite dans leurs livrets qu'elle pouvait l'être dans la rédaction de l'item des programmes.

Un cinquième constat porte sur la marge d'interprétation laissée aux termes de « texte lu » : s'agit-il d'un texte lu par un tiers ou d'un texte lu silencieusement ? Nous venons d'en voir dans le chapitre précédent une manifestation (cf. tâche 6 « le chien Hector »). Dans le sous-domaine « Langage oral » figure un item intitulé « manifester sa compréhension d'un récit ou d'un texte documentaire lu par un tiers en répondant à des questions le concernant : reformuler le contenu d'un paragraphe ou d'un texte, identifier les personnages principaux d'un récit ». Ceci invite à considérer, comme cela est précisé dans le document d'accompagnement des programmes, que le texte lu est en réalité un texte lu par l'élève. Cette divergence d'interprétation quant à celui qui lit le texte est tangible dans la rédaction choisie par les enseignants B ; J et K : si le texte est lu à haute voix, on ne sait pas par qui. La fusion opérée entre l'item précité relatif au langage oral et celui dont nous discutons ici en lecture par l'enseignant B, indique cette confusion autour du sens attribué à la formulation « texte lu ». Les enseignants G ; O et P évacuent cet obstacle précisant que la lecture est silencieuse.

Ceci permet de considérer qu'une même tâche à réaliser (évaluer un item des programmes) peut être l'objet de diverses interprétations, lesquelles risquent bien d'avoir pour effet de donner lieu à une palette assez large de tâches prescrites. Nous avons indiqué que le SCEREN avait réduit l'item des programmes au premier de ses composants (le but). Nous pouvons ajouter que le SCEREN a choisi de ne pas mentionner dans le livret qu'il édite un item des programmes très proche attaché au sous-domaine Langage oral : « manifester sa compréhension d'un récit ou d'un texte documentaire lu par un tiers en répondant à des questions le concernant : reformuler le contenu d'un paragraphe ou d'un texte, identifier les personnages principaux d'un récit ». Cette compétence fait pourtant partie des compétences de lecteur : il y a une différence entre comprendre un énoncé oral et comprendre un énoncé oral comportant de l'écrit en filigrane. Cet oral, « scriptural », fonctionne davantage selon les règles de l'écrit (par exemple absence de redondances). Les enseignants qui se réfèrent à la publication du SCEREN peuvent être tentés d'intégrer cet item à l'item de lecture dont nous discutons ici.

Ce préambule nous invite à considérer qu'une grande variété de redéfinition de la tâche à réaliser par l'enseignant prévaut autour du seul item des programmes ayant trait à la compréhension en lecture, sur lequel tous les enseignants s'accordent néanmoins quant à la nécessité de le faire figurer. Les diverses publications institutionnelles (livret SCEREN ou document d'accompagnement des programmes) ne permettent pas de dégager une base d'interprétation partagée. J. Fijalkow (2000, p. 23) avait déjà formulé une remarque assez proche « la multiplication de ce type de textes (qui n'ont de « non officiels » que l'absence de la mention « officielle ») au cours de la dernière décennie est sans doute l'indication de la difficulté croissante que rencontre l'institution à assumer un discours sur la lecture ».

De ce fait, il semble prévisible que les enseignants n'évaluent pas la même compétence de compréhension chez leurs élèves, ni de la même façon. Pour la moitié d'entre eux, l'évaluation de la compréhension de ce que les élèves ont lu passe par la réponse à des questions. Pour l'autre moitié, rien ne permet de savoir, à la lecture des livrets, la forme que l'évaluation prendra.

4.2. Les questionnaires pour évaluer la compréhension en lecture :

À l'issue d'une enquête conduite sur les pratiques (déclarées) d'enseignement/apprentissage de la lecture auprès de 1 253 enseignants de CP, E. Fijalkow indique que, demander aux élèves de répondre à des questions après lecture d'un texte constitue une pratique d'évaluation qui fait consensus (2003, p.85). La « lecture silencieuse » (un texte suivi de questions) instaurée par Thorndike en 1917, et qui, d'après Lafontaine (2003, p.1) est une « pratique dominante de référence » jusque dans les années 1970, semble l'être encore en 2003 (cf. Giasson, 1990, p.26 ; Viola, 1998, p.40). Giasson (1990, p.237) prévient que, « même si les questions en lecture ont été l'objet de plusieurs critiques, aucun pédagogue ne prône leur disparition de la classe ». Surtout en situation d'évaluation, pour des raisons évidentes de commodité de passation, distribuer un questionnaire (écrit) semble relever de la tradition. Nous venons de voir que les mêmes programmes en requièrent l'usage, laissant aux enseignants le soin de déterminer si les questions sont orales ou écrites à réponse orale ou écrite.

En tout début d'apprentissage de la lecture, tout au moins, on pourrait imaginer que les enseignants n'aient pas un recours massif à ce type de tâche, à considérer les conclusions de la conférence de consensus organisée par le PIREF en 2003 : « les questionnaires (et la lecture à voix haute) sont aujourd'hui des moyens imparfaits d'évaluation, qui n'ont pas d'efficacité avérée dans la maîtrise de la lecture » (2003, p.4). Lafontaine (2003, p.9) indique dans sa contribution à cette conférence de consensus que, « souvent, parce que la lecture relève de l'écrit, on cantonne son évaluation au canal écrit, par une sorte d'homologie qui n'a pas de fondement. En procédant de la sorte, on se prive d'informations extrêmement précieuses ». Notamment parce qu'il serait « intéressant d'observer un élève en train de répondre à un questionnaire, et de recueillir un complément d'information en discutant ses réponses avec l'élève, particulièrement pour les élèves en difficulté » (idem). L'usage abusif de questionnaires (écrits) de lecture pour l'enseignement de la compréhension pourrait avoir pour effet de privilégier l'évaluation du résultat des lectures plutôt que des activités portant sur le processus de compréhension (Giasson, 1990, p. 224 ; Vanhulle & Dufays, 2010, p.239), ce qui conférerait au lecteur un rôle de « répondeur » résumé par Viola (1998) en ces termes : « le questionnement traditionnel fait de la tâche de lecture une tâche externe plutôt qu'interne ». Mais l'ensemble des experts sollicités par les groupes chargés de l'élaboration des projets de programmes 2016⁶⁸ s'accorde pour reconnaître qu'aucune alternative n'a été sérieusement proposée.

De ce fait, lorsque l'on parle d'évaluation de la (compréhension en) lecture, il convient d'analyser le média par lequel elle se réalise.

4.2.1. Le questionnaire « Caroline »

a- l'item :

Au premier trimestre, l'enseignant F souhaite évaluer l'item « dire de quoi et de qui parle le texte lu ». Il utilise le livret du SCEREN. Comme nous venons de l'évoquer précédemment, une sélection est effectuée dans ce qu'il y a à évaluer ce qui constitue un premier décalage : on n'attend pas explicitement que l'élève soit capable de reformuler le sens du texte qu'il vient de lire ni même qu'il trouve dans le texte réponse à une question concernant le texte lu.

⁶⁸ Cf. <http://www.education.gouv.fr/cid82307/le-conseil-superieur-des-programmes-contributions-des-experts-sollicités-par-les-groupes-chargés-de-l-elaboration-des-projets-de-programmes.html> consulté en sept. 2015.

b- l'épreuve :

L'enseignant choisit d'utiliser l'épreuve ci-dessous :

A Réf.16. Dire de quoi et de qui parle le texte lu.

Consigne : Lis plusieurs fois le petit texte, puis réponds aux questions que pose la maîtresse.

Caroline est malade

Caroline ne va pas à l'école.
Elle est dans son lit, car elle est malade.
Elle tousse et elle a mal à la tête.
Ses joues sont toutes rouges.
Maman lui donne du sirop.
Demain elle mangera une mandarine, puis elle ira dans le jardin où elle jouera avec Nicolas.

1. Pourquoi Caroline ne va-t-elle pas à l'école?
Caroline ne va pas à l'école car... *elle est malade.* ✓

2. Que donne maman à Caroline pour soigner sa toux?
Elle... *lui donne du sirop.* ✓

3. Avec qui Caroline veut-elle jouer dans le jardin ?
Elle jouera avec Nicolas. ✓

4. Coche les bonnes réponses

<input checked="" type="checkbox"/> Caroline a mal à la gorge.	<input type="checkbox"/> Caroline a les joues fraîches.
<input type="checkbox"/> Caroline va très bien.	<input checked="" type="checkbox"/> Caroline a les joues toutes rouges.
<input checked="" type="checkbox"/> Caroline a mal à la tête. ✓	<input checked="" type="checkbox"/> Caroline est dans son lit.

Consigne : **Colorie** la bonne réponse.

1 Caroline va à l'école mais elle tousse toute la journée et elle a mal à la tête.

2 ~~Caroline est malade, elle ne va pas à l'école, elle est dans son lit.~~ ✓

3 Nicolas a mal à la tête, Caroline joue dans le jardin.

Figure 19 : Épreuve FT3 (Annexe B, p. 139) avec réponses d'élève

Il est attendu que l'élève lise plusieurs fois le petit texte puis qu'il réponde aux questions que pose (« par écrit »⁶⁹) la maîtresse. La consigne rapproche texte et questions. Analysons ce lien.

a-Déchiffrage et recodage phonologique du texte et du questionnaire

Observons d'abord le texte. Il ne présente pas de difficultés majeures de recodage phonologique. Globalement, dans une première approche quantitative, on peut remarquer que ce texte est court, composé de 54 mots dont 42 différents. Les phrases sont plutôt courtes (54 mots pour 7 phrases), composées elles-mêmes de mots courts (3,8 lettres par mot en moyenne, avec un écart-type de 1,96 lettres). Il contient 27 mots monosyllabiques

⁶⁹ Précision obtenue lors de l'entretien post-collecte.

(nous ne considérons que les syllabes écrites), 18 mots de deux syllabes, 6 mots de trois syllabes et 3 mots de quatre syllabes, « mandarine, Caroline »⁷⁰.

Nous avons par ailleurs cherché à estimer le niveau de difficulté que pose l'identification des mots du texte, en tenant compte de deux facteurs⁷¹ :

Le premier facteur, basé sur l'exposition à l'écrit via le recours à la base Manulex, permet d'estimer si les élèves ont une chance plus ou moins élevée de reconnaître « directement » un mot, donc de l'identifier rapidement. En fonction des fréquences (SFI, Standard Frequency Index⁷²), nous avons attribué 0 ; 0,25 ; 0,50 ; 0,75 et 1 point, respectivement, aux mots hyper fréquents (dans), fréquents (maman), rares (mal), très rares (mandarine) et extrêmement rares ou absents de la base Manulex au niveau du CP (jouera).

Le deuxième facteur fait appel à plusieurs caractéristiques linguistiques des mots. D'abord, il est admis que les mots longs sont plus difficiles à décoder : 0,25 point a été attribué aux mots de plus de 7 lettres. Ensuite, sur le plan du recodage phonologique, les élèves doivent recourir à des procédures grapho-syllabiques et/ou graphophonologiques (Ecalte et Magnan, 2002). Concernant les premières, un ensemble de travaux a montré que le « traitement syllabique semble fonction du degré de complexité des syllabes » (Ecalte et Magnan, 2002, p. 133). Nous avons attribué 0,5 point (voire 1 point en cas de difficultés multiples) aux mots qui comportent des structures syllabiques complexes, c'est-à-dire autres que les syllabes simples CV⁷³ : par exemple CCV (aucune ici), CVC (a/vec), CVnasale (jar/din). Il faut ici signaler que la majorité des mots comportent des syllabes simples du type CV et que les syllabes les plus complexes contiennent simplement pour la majorité des voyelles nasales nécessitant deux lettres. Pour ce qui relève des procédures graphophonologiques, nous avons attribué 0,25 point de difficulté pour la présence de graphèmes complexes autres que les archigraphèmes (Catach, 1980), et ce uniquement aux mots n'ayant auparavant reçu aucun point au titre du deuxième facteur. Le texte ne s'y prêtant pas, nous n'avons pas ajouté un facteur de facilité lié à la possibilité de recourir aux contextes (« texte, phrase, monde de référence », Goigoux et Cèbe, 2006) lors de l'identification.

En rassemblant ces éléments, nous avons constitué 4 catégories de mots, de difficulté croissante en termes d'identification pour les 42 mots différents qui constituent le texte :

- La première catégorie (score = 0 ou 0,25 point) regroupe 12 mots différents très courts et faciles, tels « le, la, une, école ... » ;
- la deuxième catégorie (score = 0,5 ou 0,75 point) regroupe 18 mots différents, tels « est, lui, maman, dans, toutes, ses, ... », qui ne devraient pas poser de problème à la plupart des élèves ;

L'ensemble des mots constituant ces deux catégories représente 76% des mots du texte.

⁷⁰Les éléments retenus sont proches de ceux qui interviennent dans le calcul d'indices de lisibilité, tel celui de Fleche-De Landsheere (nombre de mots par phrase, nombre de syllabes pour cent mots). Nous ne recourons pas à cet indice car celui-ci, outre les critiques dont il a fait l'objet, ne serait pas valide pour les textes courts de l'école primaire (Cf. Henry, G. 1980. Lisibilité et compréhension. *Communication et langage*, 7-16).

⁷¹ Nous reprenons une proposition que nous adoptons pour analyser ce nouveau texte (cf. C. Blanc & Vantourout &, soumis).

⁷² Sans présenter l'intégralité de cette analyse, on peut distinguer deux pôles, avec, d'une part 10 mots (lemmes et/ou formes orthographiques), que nous qualifions d'hyper fréquents, qui seraient rencontrés au moins tous les 10 mots et au plus tous les 100 mots (le, la, une, est, et, à, son, l', dans, a), et, d'autre part, 10 mots, que nous qualifions de rares et de très rares qui seraient au mieux rencontrés tous les 10 000 mots (joues, sirop, rouges, ira, toussé, mandarine, Nicolas) voire tous les 100 000 (mangera, Caroline, jouera).

⁷³ C pour consonne graphique, V pour voyelle graphique

- la troisième catégorie (score = 1 ou 1,25 points) regroupe 9 mots qui pourraient occasionner des difficultés d'identification « demain, rouges, Nicolas ... » ;
- enfin, la quatrième catégorie (1,25 < score ≤ 1,75 points) regroupe 3 mots, « mandarine, jouera, mangera », susceptibles de poser des problèmes d'identification à une partie des élèves.

Le score total (35,5 points pour les 54 mots soit, ramené à un texte de 100 un indice de 65,74) pourrait entrer dans le calcul d'un indice de difficulté et permettre une comparaison entre textes⁷⁴.

Si l'on élargit à présent cette caractérisation au questionnaire, il apparaît qu'outre ces mots, de nouveaux mots apparaissent dans les catégories les plus difficiles « pourquoi, soigner (1,5) » voire forment une nouvelle catégorie « fraîches (2,25) ».

Le texte constitué par le questionnaire comporte des mots susceptibles de poser davantage de difficulté de déchiffrement que le texte qu'il questionne.

b- Compréhension du texte et du questionnaire :

En référence aux modèles de la compréhension en lecture, le texte ne présente pas de difficultés majeures de compréhension :

- Il contient assez peu d'inférences : une inférence causale implicite (« Caroline est malade (*donc*) elle ne va pas à l'école), une inférence causale explicite (« elle est dans son lit *car* elle est malade »), une inférence de lieu explicite (« elle ira dans le jardin *où* elle jouera ») auxquelles il faut ajouter une inférence liée aux connaissances sur le monde (« maman donne du sirop (pour la toux) »). Il faut encore ajouter plusieurs emplois de l'anaphore « elle » pour désigner le même personnage « Caroline ».
- Il mobilise des connaissances partagées sur le monde (situation de maladie). On peut d'ailleurs voir dans les réponses formulées par l'élève dont nous avons reproduit la copie fonctionner l'effet de ces connaissances : il coche « Caroline à mal à la gorge » (ce n'est, étonnamment, pas la bonne réponse attendue mais il est possible que, pour cet élève, le fait que Caroline soit malade, qu'elle tousse et que sa maman lui donne du sirop renvoie à un mal de gorge).

Le texte constitué par le questionnaire comporte des difficultés propres parmi lesquelles :

- des inversions sujet-verbe absentes du texte support (des structures spécifiques de l'écrit dont l'étude n'est pas au programme de CP : si l'on se reporte aux progressions annexées aux programmes (MEN, 2008 p. 28) on constate que « savoir transposer une phrase affirmative en phrase négative ou interrogative » n'est au programme que du CE1 et de façon orale),
- des phrases interro-négatives non employées dans le texte support (« pourquoi *ne* va-t-elle *pas* à l'école ? »),
- de nombreux recours à des substituts anaphoriques à l'intérieur des questions « pourquoi caroline ne va-t-elle pas à l'école ? », « avec qui caroline veut-elle jouer dans le jardin ? »,
- des substituts lexicaux ou synonymes (la *toux* (questionnaire)/ elle *tousse* (texte)) absents du texte support.

⁷⁴ Cf. Vantourout & C. Blanc, soumis. Dans cet article, le texte support « Léa » obtient une valeur tout à fait comparable (72,2).

A ceci s'ajoute la prise en compte des amorces de réponse à produire aux premières questions : pour réussir l'élève doit, après l'avoir conçue, aménager sa réponse en tenant compte des amorces « Caroline ne va pas à l'école car... » ou « Elle... »

Du point de vue du récit, Barthes (1977, p. 39) rappelle qu' « il ne peut y avoir de récit sans narrateur et sans auditeur (le lecteur) ». Le récit donné à lire au lecteur met en scène une petite fille (Caroline) et sa maman (Maman). Ces personnages sont évoqués par un narrateur qui s'adresse à un destinataire par le biais d'un auteur. Mais, celui qui parle dans le récit n'est pas celui qui écrit dans la vie indique Barthes (1977, p. 40) : auteur et narrateur ne sont pas confondus dans la réalité. Le texte « Caroline » fonctionne comme une somme de faits présentés de façon totalement objective, sans affect, proche d'une description, sans narrateur. L'auteur (l'enseignant qui semble avoir construit le texte à des fins d'évaluation) réduit le narrateur au rang de descripteur d'une scène qu'il donne à voir au destinataire sans donner son point de vue, faisant de lui une sorte de narrateur empêché. Le texte fonctionne comme un prétexte au questionnaire.

L'auteur introduit une dimension affective dans le questionnaire lorsqu'il pose une question sur la volonté d'un personnage : « avec qui Caroline veut-elle jouer dans le jardin ? » On peut considérer cela comme la trace d'une dérive méta textuelle telle que celle évoquée par Daunay (1993) et que Genette (1972, p. 244, cité par Daunay, 1993), appelle une « métalepse » (dans le domaine des œuvres narratives fictionnelles, c'est le passage d'un niveau narratif à un autre). La compréhension évaluée (par cette question) ne concerne pas le texte mais le monde de référence considéré comme commun à l'auteur et au lecteur : le texte, le narrateur, ne dit pas que Caroline veut jouer avec Nicolas, il l'affirme. En revanche, l'auteur impose au lecteur sa projection sur le personnage : il est implicitement admis que lorsqu'on est malade on souhaite jouer dans le jardin au plus vite.

Le questionnaire peut induire une telle dérive méta-textuelle.

c-Structure du questionnaire

Le questionnaire comprend 5 questions :

-Les trois premières questions suivent l'ordre du texte si bien qu'une mémoire (linéaire et littérale) du texte est davantage sollicitée que le modèle de situation construit par l'élève, ou la mémoire de la compréhension du texte.

-Les deux dernières questions portent sur l'ensemble du texte. Cependant, les réponses apportées aux trois premières questions permettent de répondre et de révéler un phénomène particulier agissant dans la relation entre questions. Des éléments fournis par le concepteur du questionnaire dans l'énoncé des trois premières questions on déduit que « Caroline ne va pas à l'école », que « sa maman lui donne quelque chose pour soigner sa toux » et qu' « elle veut jouer dans le jardin ». Ainsi la seule réponse convenable à la dernière question (« Caroline est malade, elle ne va pas à l'école, elle est dans son lit ») peut être trouvée (en excluant les modalités de réponse proposés ou en choisissant directement la bonne réponse) en n'ayant seulement lu que les questions.

La lecture du titre suffit à comprendre le sens global du texte et la lecture du questionnaire permet de répondre à une grande partie des questions en procédant à des inférences. Aucune question ne porte sur la temporalité des éléments du texte, ni sur l'état émotionnel du personnage principal. La nature de l'information collectée par les questions amène à privilégier les informations suivantes sur lesquelles l'évaluateur estimera satisfaisante la compréhension du texte « Caroline est malade. Sa maman lui donne du sirop. Elle veut jouer avec Nicolas ». Le modèle de situation que s'est construit l'élève n'est pas interrogé : qu'est-

ce que le sirop en question ? Pourquoi pourra-t-elle sortir jouer demain ? Quel jour sera-t-on : un jour non travaillé ou un jour au cours duquel Caroline se reposera ? etc.

Le questionnement proposé est parcellaire et engage la mémoire d'éléments factuels. La troisième question demande à l'élève de nommer avec qui Caroline jouera (bonne réponse : Nicolas). Cette précision est de moindre importance que le lien de cause à effet entre le sirop et la toux. On pourrait ne pas avoir retenu le prénom Nicolas (et de ce fait produire une réponse erronée) et avoir très bien compris le sens du texte.

c- la copie de l'élève :

- L'élève répond « elle jouera avec Nicolas » à la troisième question « Avec qui Caroline veut-elle jouer dans le jardin ? ». Cette réponse modifie, à propos, la temporalité imposée de façon erronée par le concepteur du questionnaire avec l'emploi du présent : le texte indique que Caroline jouera demain (i.e. quand elle sera soignée).

- La réponse considérée comme erronée à la quatrième question met en lumière un effet intéressant qui ne relève ni de l'« intrusion textuelle » ni de l'« intrusion d'un schéma », correspondant, respectivement, à un recours inadéquat au texte ou à l'activation d'un schéma inapproprié. En effet, « Caroline a mal à la gorge » renvoie à un schéma du lecteur qui s'appuie sur une connaissance du monde telle que dans une situation où l'on est malade et que l'on prend un sirop pour se soigner alors on a mal à la gorge.

Cette réponse est considérée comme non-valide alors qu'elle signerait plutôt une très bonne compréhension de la situation.

En résumé :

Le traitement du questionnaire engage une tâche de lecture supérieurement difficile à celle de la lecture du texte dont on évalue la compréhension au moyen du questionnaire. Cette difficulté est de plusieurs ordres : déchiffrement, compréhension, prise en compte d'un point de vue méta- sur le texte, mémoire de la tâche.

L'évaluation repose sur les traces écrites de l'activité laissée par l'élève. Rien ne permet d'affirmer que l'absence de phrases réponses signe automatiquement un défaut de compréhension du texte. L'épreuve induit éventuellement une dérive méta-textuelle dont le traitement est laissé à la charge du lecteur.

Si l'on considère stricto-sensu l'intitulé de l'item « dire de quoi et de qui parle le texte lu » la lecture du titre (trois mots) suffit. De plus, aucune question ne permet de « dire de qui parle le texte lu », toutes portent sur « Caroline » de façon explicite. Dire « de quoi parle le texte lu » oblige à formuler une phrase. Pour réduire la difficulté de rédaction de la réponse, le concepteur du questionnaire a rédigé le début de la réponse « Caroline ne va pas à l'école car... ». A vrai dire, cette partie de la réponse suffit à répondre à la question de savoir de quoi parle le texte. Peuvent se placer à la suite du mot « car » un ensemble assez important de bonnes réponses : « car elle est dans son lit », « car elle est malade », « car elle tousse », « car elle a mal à la tête ».

L'analyse du questionnaire, considéré comme un tout et non plus comme une suite de relations entre une question et le texte, permet de considérer des effets internes : on peut élaborer une représentation du texte assez fine, simplement à partir des informations sur le texte fournies par les questions, questions qui comportent en outre des éléments de réponse à d'autres questions.

La tâche proposée est donc discutable du point de vue psycho-didactique, en regard de ce qu'elle est censée permettre d'évaluer, à savoir : « dire de qui et de quoi parle le texte lu ».

Elle permet principalement d'évaluer la compétence des élèves à répondre à des questions portant sur un texte ce qui est fondé sur l'implicite discutable selon lequel bien répondre à des questions est consécutif à avoir bien compris le texte. Cette compétence s'exerce sur un couple texte-questionnaire et évalue de ce fait la compréhension du questionnaire et du lien texte-questionnaire.

Nous nous intéressons à présent aux processus mobilisés dans le traitement des questionnaires.

4.2.2. Processus engagés par les questionnaires

1. Processus attentionnels :

Camus (1996) rapporte que les processus attentionnels sont liés à la mémoire de travail par la nécessité de maintenir continuellement disponibles des représentations différentes dont le nombre risque d'excéder les capacités de la mémoire à court terme ou qui doivent être maintenues un laps de temps excédant la capacité temporelle de cette mémoire. Traiter un questionnaire lié à un texte mobilise des représentations différentes : celles rattachées au texte, celle rattachées aux questions portant sur le texte. Estimons en cela que comprendre une question engage l'élaboration d'une représentation qui doit cohabiter avec la représentation que l'élève s'est construite du texte. Deux sortes de représentations doivent donc cohabiter en mémoire de travail : celle émanant de la lecture du texte, celle émanant de la lecture des questions.

Barrouillet, Camos, Morlaix et coll. (2008, p.31) font état des liens étroits entre mémoire de travail et lecture « des premières acquisitions comme le nom des lettres jusqu'aux plus complexes comme l'intégration du sens des phrases et la production d'inférences en vue de la compréhension de texte ». Ils précisent que « les enfants présentant des déficits en mémoire de travail sont handicapés dans toutes ces activités » (idem). Ils concluent un chapitre présentant « un panorama des connaissances actuelles sur le développement de la mémoire de travail, les déterminants de ce développement, sa nature, son impact sur les apprentissages et principalement les apprentissages scolaires, et plus généralement ses liens avec le développement cognitif au sens large » (idem, p. 3) en indiquant qu' « étant le lieu du maintien de l'information nécessaire aux traitements en cours, les limites de la mémoire de travail contraignent une grande partie de nos activités de pensée et l'évolution de ces limites avec l'âge joue un rôle essentiel dans le développement intellectuel. » (Idem)

L'ensemble « texte-questionnaire » représente dans le cas de l'épreuve analysée ci-dessus un volume de mots environ trois fois supérieur au texte seul. Le texte comporte 54 mots quand le questionnaire destiné à évaluer la compréhension du texte en comporte 112. On peut concevoir que le traitement du questionnaire mobilise fortement la mémoire de travail de l'élève. On pourra objecter que, fréquemment, l'enseignant lit les questions (ce qui peut décharger l'élève d'une part de l'activité de lecture du questionnaire). Soit il les lit avant que les élèves s'engagent dans la lecture (cf. travaux de Devidal, Fayol et Barrouillet sur le placement de la question, 1997), soit il les lit après la lecture du texte. Deux modalités qui ne sont pas sans influence sur l'activité de l'élève. Dans le premier cas, l'activité de lecture de l'élève est orientée : certaines informations sont à privilégier dans le but de répondre plus que dans le but de comprendre le texte (l'accent est mis sur la réponse au questionnaire, l'élève est susceptible de répondre en sélectionnant dans les questions mêmes des éléments de réponse). Dans le deuxième cas, l'accent est mis sur la compréhension du

texte. Mais durant la lecture des questions, l'élève doit mettre le texte en mémoire, puis mobiliser des processus attentionnels sur le questionnaire (auquel il peut de ce fait répondre en sélectionnant dans les questions des éléments de réponse).

L'emploi des questionnaires serait de nature à désavantager les élèves présentant des déficits en mémoire de travail. L'emploi de tâches de rappel de texte semblerait de nature à limiter ce fait, l'élève n'ayant pas à traiter le questionnaire en plus du texte.

2. Processus de contrôle de l'activité et stratégie de lecture

Dans un document intitulé « Enseigner la compréhension : principes didactiques et exemples de tâches », Cèbe, Goigoux et Thomazet (2004) soulignent le problème du contrôle de l'activité par l'élève au cours de la lecture ainsi que le fait que certains élèves se méprennent sur les procédures à utiliser. Ils rapportent un résultat décrit par Goigoux (2000b) : « les faibles lecteurs procèdent essentiellement à un contrôle de la compréhension au niveau propositionnel mais peu au niveau local (inter-phrastique) et global (textuel) ».

Nous en reproduisons un large extrait. Les auteurs indiquent que ces « faibles lecteurs » :

« utilisent massivement des stratégies de lecture mot à mot et traitent chacune des phrases comme autant de phrases isolées ce qui les prive de pouvoir contrôler la cohérence de l'information tout au long du texte. Ils ne comprennent pas l'utilité des processus d'intégration sémantique en cours de lecture ni la nécessité de procéder à des inférences pour mettre en relation les diverses données du texte. Ils ne remettent pas en cause les représentations ou les interprétations qu'ils se font dès le début du texte. Les élèves les plus faibles ont aussi du mal à repérer l'information importante d'un passage, à saisir en quoi une information nouvelle est liée à ce qu'ils savent déjà ou à ce qu'ils ont déjà lu plus haut, à détecter les incohérences d'un passage, à faire attention aux contraintes syntaxiques et sémantiques, à évaluer s'ils ont bien compris le contenu. Il leur est donc difficile de comprendre « l'utilité des processus d'intégration sémantique en cours de lecture et la nécessité de procéder à des inférences pour mettre en relation les diverses données du texte ». Quand on leur demande : « comment je peux savoir si tu as compris ce texte », ils répondent « il faut me poser des questions » et si on leur demande « et si je veux faire autrement », ils ne trouvent pas de réponse (résumer, raconter avec ses mots, paraphraser...). Enfin on observe que les lecteurs peu efficaces, parce qu'ils s'attendent à ne pas comprendre, étudient moins longtemps le texte que les autres. Lorsqu'on les interroge sur leurs piètres performances, ils incriminent le texte (pas intéressant, trop difficile), leurs faibles capacités (je suis nul en lecture) mais ne citent jamais l'effort qu'ils consacrent (ou plutôt ne consacrent pas) à l'étude : pour eux, tout effort est vain. Là encore, ils s'en remettent au contrôle de l'adulte pour qu'il les aide à se centrer sur l'activité et pourquoi pas qu'il leur donne les bonnes réponses aux questions qu'ils ne se posent pas. »

Les auteurs poursuivent en indiquant que les faibles lecteurs « essaient souvent de se rappeler la forme littérale des énoncés : ils cherchent à apprendre par cœur les phrases et procèdent là à l'inverse des lecteurs plus habiles qui s'attachent à centrer leur attention sur les informations importantes du texte (la forme littérale faisant l'objet d'un oubli rapide)».

L'emploi des questionnaires est de nature à favoriser des stratégies peu efficaces, ou insuffisantes pour comprendre un texte. Par exemple en s'appuyant sur un seul contrôle de la compréhension au niveau propositionnel, les élèves ne sont pas incités à utiliser de

procédures de rétrocontrôle en cours d'activité. Les élèves n'ont pas à faire état de leur traitement de l'ensemble du texte. Celui-ci est pris en charge par le questionnaire. Cela accentue la conception erronée selon laquelle le texte n'est qu'une suite de phrases, ce qui est typique, selon les auteurs précités, des élèves les plus faibles lecteurs.

Ces travaux portent sur des élèves plus âgés mais on peut voir dès le début de l'apprentissage de la lecture de telles stratégies de réponse se mettre en place : elles sont susceptibles d'être survalorisées par l'usage de questionnaires.

3. Processus de lecture

Diverses classifications de questions existent dont Giasson a fait état (1990, p. 226 et suivantes). Aux classifications de questions basées sur le lien entre la question et la réponse attendue par l'enseignant (taxonomie de Smith et Barrett, 1974) ont succédé des classifications axées sur la relation entre les questions et les réponses données par l'élève (taxonomie RQR de Pearson et Johnson 1978).

Celles-ci nous intéressent particulièrement dans le cadre de notre approche psychodidactique en ce qu'elles prennent en compte les réponses des élèves : « les lecteurs jouent un rôle beaucoup plus actif-constructif dans leur propre compréhension que nos conceptions antérieures passives-réceptives affirmaient⁷⁵ » (Pearson, 1984, p.8)

La taxonomie RQR de Pearson et Johnson (Pearson, 1984, p. 22) comprend trois niveaux :

1. La relation entre la question et la réponse est explicite et textuelle si la question et la réponse découlent toutes deux du texte ET que la relation entre la question et la réponse est clairement indiquée par des indices dans le texte même.
2. La relation est implicite et textuelle si la question et la réponse découlent encore toutes deux du texte MAIS qu'il n'y a pas dans le texte d'indice grammatical qui relie la question à la réponse. Cette catégorie exige l'utilisation d'au moins une inférence par le lecteur.
3. La relation est implicite et fondée sur les schémas du lecteur si la question seule découle du texte ET que le lecteur utilise ses propres connaissances pour répondre à la question.

A ces niveaux, il faut ajouter les intrusions textuelles constituées de réponses empruntées au texte mais sans rapport avec la question posée, et les intrusions liées aux schémas du lecteur : des réponses issues des connaissances du lecteur mais qui sont sans lien avec la question (cf. Joole, 2008, p. 39).

Dans le même temps, Durkin (1978, p.11) a proposé de distinguer les questions portant sur le produit des questions portant sur le processus.

Nous adoptons la proposition d'Irwin (1991) qui tente de prendre en compte la nature des questions portant sur les textes et la nature des processus engagés par le lecteur. Elle fait la synthèse d'une approche centrée sur le produit et d'une approche centrée sur le processus de réponse : « les questions sur le produit serviront à évaluer l'acquisition de connaissances et les questions sur le processus serviront à évaluer la maîtrise de cette stratégie » indique Giasson (1990, p.235). Cette proposition comprend les trois niveaux de la classification R-Q-R de Pearson et Johnson (relation explicite et textuelle, relation implicite et textuelle, relation implicite fondée sur les schémas du lecteur) qu'elle croise avec les processus engagés pour y répondre.

⁷⁵ "Readers play a much more active-constructive role in their own comprehension than our earlier passive-receptive views dictated."

Le tableau ci-dessous synthétise ces éléments :

Relation question-réponse	Type d'information			
	Microprocessus	Processus d'intégration	Macro processus	Processus d'élaboration
Rappel des connaissances antérieures avant la lecture				Questions sur les connaissances préalables
Textuelle-explicite	Questions sur les détails d'un texte	Questions sur les connecteurs explicites	Question sur les idées principales explicites et les résumés explicites	
Textuelle-implicite	Questions sur le langage figuré	Questions sur les connecteurs implicites, les anaphores et les inférences implicites	Questions sur les idées principales implicites et les résumés implicites	
Implicite et basée sur les schémas du lecteur				Questions sur les prédictions, sur les images mentales, sur les réponses affectives, sur le raisonnement.

Tableau 17 : Taxonomie de la relation question-réponse (RQR), d'après Irwin (1991, p.182)

Giasson (1990) précise qu'il faut entendre par microprocessus « les processus qui servent à comprendre l'information contenue dans une phrase », par processus d'intégration, « les processus qui ont pour fonction d'effectuer des liens entre les propositions ou les phrases », par macro processus « les processus orientés vers la compréhension globale du texte, vers les liens qui permettent de faire du texte un tout cohérent », et par processus d'élaboration, « les processus qui permettent au lecteur de dépasser le texte, d'effectuer des inférences non prévues par l'auteur » (p. 16)

Nous avons classé les questions des questionnaires recueillis suivant cette taxonomie.

Résultats :

- Au premier trimestre, dans notre corpus, les questionnaires portant sur un texte lu (seul ou par l'enseignant) sont rares. Considérons les quatre textes suivis de questions de notre corpus de données (figurant en pages 66 ; 77 ; 83 et 105 du volume d'annexe B - ce sont les seuls pour lesquels nous disposons du texte source et du questionnaire). Sur un total de 19 questions :

- 15 questions sont de type textuelle-explicite engageant *a priori* des microprocessus de traitement : la réponse et la question portent sur un détail du texte immédiatement accessible.
- 3 questions seulement sont de type textuelle-implicite. Parmi celles-ci, des microprocessus sont *a priori* engagés dans le traitement de la question [comprendre le sens figuré d'une expression (« La fillette est haute comme trois pommes ». Q5 : « Oui/non » - texte p. 66)] ou des processus d'intégration [comprendre une anaphore

(« Téo est un pirate. Il vit sur un navire ». Q4 : « Où vit Théo ? » texte p. 77), établir une inférence causale (« Soline est affamée, son amie Zélie lui donne une tomate ». Q3 « Pourquoi Zélie donne une tomate à Soline ? » - texte p. 87)]

- Le cas de la question 4 du texte p. 83 est assez intéressant en ce qu'il introduit une discussion sur ce qu'est un texte (nous y reviendrons plus loin). Pour répondre à la question « Où se passe cette histoire ? », les élèves ont trois choix matérialisés par des dessins. Mais aucun n'est pleinement satisfaisant du point de vue de l'histoire. Sans doute est-ce lié à la structure du texte qui n'en est pas vraiment un (les phrases du texte sont plutôt juxtaposées).

Le type de relation question-réponse *a priori* le plus rencontré est de type textuel-explicite. *A priori* les questions n'engagent pas de macro processus de traitement qui sont toutefois susceptibles d'être mobilisée dans l'activité de lecture du texte.

- Au troisième trimestre, 15 questionnaires sont utilisés par 11 des 16 enseignants qui constituent notre population pour évaluer la compréhension de texte.

Cela occasionne 84 questions que nous avons répertoriées *a priori* suivant la taxonomie d'Irwin :

Relation question-réponse	Type d'information			
	Microprocessus	Processus d'intégration	Macro processus	Processus d'élaboration
Connaissances antérieures				1 (1%)
Textuelle-explicite	50 (60%) Questions sur les détails d'un texte	2 (2%) Questions sur les connecteurs explicites	13 (15%) Question sur les idées principales explicites et les résumés explicites	
Textuelle-implicite	1 (1%) Questions sur le langage figuré	12 (14%) Questions sur les connecteurs implicites, les anaphores et les inférences implicites	3 (4%) Questions sur les idées principales implicites et les résumés implicites	
Schémas				2 (2%)

Tableau 18 : Répartition des questions selon taxonomie de la RQR présentée dans le tableau 17

Analyse des résultats :

Les questions posées sont, *a priori*, majoritairement de type textuel-explicite et, parmi celles-ci, la grande majorité engage des microprocessus de traitement. Si l'on se reporte à la définition qu'en donne Giasson (1990, p. 16, 37), ces processus permettent de comprendre une phrase. On trouve peu de questions mobilisant (*a priori*) des processus de lecture plus large (d'intégration ou macro) ou, pour reprendre les termes de Cèbe, Goigoux et Thomazet (2004), peu de questions obligeant un contrôle de compréhension au niveau local (interphrastique) et global (textuel). Les questions de type textuel-implicite sont plus rares et, si elles engagent des processus d'intégration, cela est massivement dû à une formulation de la question faisant appel à une anaphore (le personnage du texte est désigné par un pronom

comme par exemple la question 3 du questionnaire accompagnant le texte « Le lapin sauvage » (p. 129 du volume d'annexe) : « où le lapin se cache-t-il ? »)

Le reproche que l'on peut toutefois adresser à ces typologies est de ne concevoir que la seule relation que la question entretient avec le texte et les processus de lecture mais pas la relation que chaque question entretient avec les autres questions, ce qui fait précisément du questionnaire un méta-texte. Elles nous apparaissent donc incomplètes pour décrire les questionnaires⁷⁶. Cette discussion permet de considérer que les modèles psychologiques fondés sur la mémoire et les processus cognitifs engagés dans la lecture ne peuvent suffire et qu'il est nécessaire de croiser leurs apports avec des approches didactiques qui s'intéressent aux propriétés des textes et des questionnaires (comme celle de Daunay (1993) par exemple).

⁷⁶ Cf.C. Blanc et Goasdoué (2016).

4.3. Synthèse du chapitre 4

L'ensemble des analyses conduites dans ce chapitre permet de considérer que l'emploi de questionnaires, utilisés dès le début de l'apprentissage de la lecture, renvoie à une série de problèmes relatifs au fait que les questionnaires constituent de véritables textes sur les textes dont ils sont censés évaluer la compréhension. Ils prennent le texte pour objet, ce qui peut avoir pour effet une dérive méta-textuelle telle que dénoncée par Daunay (1993).

Outre qu'ils signent une évaluation de la mise en mémoire du texte support, davantage qu'un véritable travail de compréhension selon Joole (2008, p.11), ils occasionnent une charge attentionnelle accrue, engagent une difficulté de lecture supplémentaire (difficulté de déchiffrement, difficultés de compréhension, changement de type de texte, difficulté liées à la syntaxe, à la polysémie des termes utilisés), une mobilisation de stratégies variées (de contrôle de l'activité par exemple) non triviales.

Ils engagent en outre une activité de lecture dont on peut penser *a priori* qu'elle est peu fondée sur l'utilisation de processus d'intégration ou de macro processus. Les questionnaires permettent de considérer la référence privilégiée à deux modèles (« essentialiste⁷⁷ » et « technique⁷⁸ ») qui selon Jorro (1999, p.19) « arriment les lecteurs au texte, les polarisent sur sa logique interne, sur son explication, de sorte que « l'interaction lecteur - texte » se limite alors à l'extériorisation des connaissances formelles de l'écrit ». Il semble (Jorro, 1999, p.27), que l'enseignant, « expert textuel, impose de manière implicite un modèle du discours » et que lui seul puisse « faire le lien entre les théories du texte et les prescriptions adressées aux lecteurs » (Jorro, 1999, p.19). Les tâches de lecture engagent des actions parcellisées que l'on peut comprendre rapportées à la nécessité d'élaborer un score ou d'attribuer des points. Selon Jorro (1999, p.27) : « c'est le maître qui parcellise le savoir en questions successives mais la globalité de cette programmation n'est connue que de lui seul ». Les tâches de lecture consistent pour la plus grande partie en des tâches techniques ou d'analyses mais rarement complexes telles que « schématiser un texte, résumer un texte, interpréter un texte » (199, pp.23-24). Au final, selon Jorro (1999, p.27) : « les questions, au lieu de constituer une aide, augmentent la difficulté car elles placent les lecteurs dans l'analyse logique du texte ».

L'ordre des questions suit généralement le texte de sorte que l'élève peut lire une question, y répondre en cherchant la réponse dans le début du texte, lire une deuxième etc. sans avoir au final jamais lu le texte en entier. Pearson et Gallagher (1983, p.8) relatent un résultat semblable décrit par Beck et coll. (1979) à l'issue d'une analyse des questions de compréhension dans des manuels : « les questions sur les histoires représentent souvent un quiz accumulé au hasard de détails indépendants plutôt qu'une séquence soigneusement planifiée de questions visant à élucider les liens de causalité entre les éléments et les événements majeurs de l'histoire »⁷⁹.

⁷⁷ Dans le modèle « essentialiste » de la lecture, on désigne selon Jorro, le fait que « l'information textuelle est étudiée sans qu'intervienne l'interprétation du lecteur. A partir du sujet du texte, l'activité du lecteur est centrée sur le discernement des idées principales et des idées secondaires. Alors, l'activité de compréhension porte exclusivement sur le corps du texte, sur son message intrinsèque » (1999, p. 16)

⁷⁸ Dans le modèle, « technique » de la lecture, on désigne, selon Jorro, un mode d'interrogation programmé : « Qui ? Quand ? Quoi ? Où ? Pourquoi ? Pour quoi ? » (1999, p. 19) parfois appelé « Why questions » (pour Who ? When ? What ? Where ? Why ? How ? When ? »).

⁷⁹ Questions for stories often represented a randomly accumulated quiz of unrelated detail rather than a carefully planned sequence of questions designed to elucidate the causal connections between major story elements and events."

Déchiffrer, comprendre, répondre à l'ensemble texte-questionnaire est donc en ce sens une tâche d'une difficulté supérieure à celle qui consiste à déchiffrer et comprendre le texte objet principal du questionnement (puisque'il faut comprendre le texte ET comprendre le questionnaire sur le texte), qui n'est que peu prise en compte en tant que telle. Au final, on peut considérer que le questionnaire constitue en lui-même une épreuve surajoutée amenant les élèves à différencier plusieurs sortes d'informations : réussir cette épreuve revient à avoir bien compris et traité le questionnaire sur le texte. Ceci est trivial mais permet de considérer l'inverse : ne pas avoir bien traité le questionnaire engage nécessairement ne pas avoir réussi l'épreuve, indépendamment de la compréhension du texte ou du problème. Il faudrait pouvoir observer l'effet du questionnaire au niveau de la représentation que se construit l'élève : le niveau de représentation que l'évaluateur infère a-t-il été activé lors de la compréhension ou résulte-t-il du questionnement et se situe-t-il après traitement ? En effet, si l'on considère le questionnaire comme un second texte dans la tâche, comprendre le questionnaire revient à élaborer une représentation de chacune des questions. Cette représentation (elle aussi influencée par des « prior knowledge » (connaissances antérieures sur les questionnaires)) est susceptible d'entrer en conflit avec la représentation construite du texte.

Les épreuves d'évaluation au moyen de questionnaires permettent aux enseignants de ne répondre que partiellement à la demande institutionnelle telle qu'elle est formulée par l'item des progressions que nous avons examiné en début de chapitre : « trouver dans le texte la réponse à des questions concernant le texte lu ». L'usage de questionnaires est implicitement préconisé mais en revanche, il n'est pas explicitement dit que les réponses et les questions soient écrites. Il y a sans doute, par tradition scolaire mais aussi par commodité de mise en œuvre, une extrapolation qui est réalisée : le terme « questions » semble être traduit par « questions écrites nécessitant des réponses écrites ». De telles épreuves d'évaluation de la compréhension en lecture ne permettent d'assurer la validité curriculaire de l'évaluation que de façon partielle.

Nous avons également détaillé un ensemble de raisons qui conduisent à remettre en cause la validité psycho-didactique de telles épreuves : elles ne permettent qu'imparfaitement d'évaluer la compréhension en lecture des élèves. Ce résultat nous conduira à observer si d'autres modalités sont choisies par les enseignants pour évaluer la compréhension en lecture.

Chapitre 5. Le système de réponse

Au cours de la redéfinition de la tâche, l'élève s'empare des éléments de la tâche prescrite pour agir et la résoudre. Parmi ces éléments, on considère maintenant ce qui lui est demandé de faire pour répondre, et pour apporter la preuve de son activité. Le cadre de réponse suggéré, voire imposé par la tâche structure une part de la tâche effective de l'élève : celle qui consiste à produire une réponse à la question posée. Au niveau de la tâche, l'activité de l'élève se matérialise par la réponse qu'il fournit, et l'enseignant infère la réussite de l'élève de cet observable. En effet, dans le cas de passations d'épreuves simultanées et collectives, l'enseignant se fonde principalement sur les observables que constituent les réponses écrites des élèves pour évaluer. Se pose la question du crédit que l'enseignant peut accorder à ces observables.

Deux types de réponses peuvent être distingués : la réponse correcte et la réponse erronée. Bien souvent, analyse Astolfi, la réponse erronée possède un statut négatif tandis que l'on pourrait lui attribuer le statut « d'indicateur et d'analyseur des processus intellectuels en jeu. Au lieu d'une fixation sur l'écart à la norme il s'agit plutôt de décortiquer la logique de l'erreur, à en tirer parti pour améliorer les apprentissages » (1997, pp.9-28).

La typologie des erreurs dressée par cet auteur (opus cité, pp.96-97) est féconde pour l'analyse. Toutefois, elle ne fait pas apparaître de type d'erreur particulier lié à l'expression de la réponse.

Certaines réponses sont parfois très décalées en regard de la question posée. Elles ont amené Binet et Simon à évoquer en leur temps le terme de « nimportequisme » pour qualifier cette réaction à la question. Piaget (1926/1947, p.12) commenta ce terme en indiquant que « lorsque la question posée ennuie l'enfant, ou, d'une manière générale ne provoque aucun travail d'adaptation, l'enfant répond n'importe quoi et n'importe comment, sans même chercher à s'amuser ou à construire un mythe ». En accord avec Bastien et Bastien-Tonniazzo (2004, p. 21), nous pensons que les élèves ne font pas, dans la plupart des cas, n'importe quoi et écartons l'idée de « nimportequisme » dans les réponses que fournit l'élève. D'un point de vue général, nous concevons avec Astolfi (1997) l'erreur de l'élève comme un indicateur et non comme un défaut, et d'un point de vue particulier, nous estimons, à l'image de Joshua et Dupin (1993, p.123) que les erreurs peuvent être « liées aux rapports profonds entretenus avec un savoir dans des conditions données ».

Certaines réponses sont erronées sans qu'elles traduisent pour autant fidèlement une connaissance erronée : cette dernière peut être juste mais l'expression de la réponse peut conduire à une erreur sous l'effet de conditions inhérentes à la tâche.

De façon corollaire, on peut aussi questionner la réponse juste : résulte-t-elle de connaissances ou de procédures « justes » ? Astolfi (1997, p. 66-67) mais aussi Bastien et Bastien-Tonniazzo (ibidem) ont parfaitement illustré le fait qu'il convient de ne pas se laisser abuser par « des réponses justes (qui peuvent) résulter de connaissances fausses, et de réponses fausses (qui peuvent) être produites à partir de connaissances justes ».

Nous laisserons de côté ce qui est imputable à l'élève dans la production d'erreur pour observer ce qui dans la tâche elle-même pourrait être susceptible de le conduire à en commettre. Le cadre de réponse imposé ou suggéré par la tâche nous apparaît en effet être

susceptible d'avoir une influence au cours du processus de redéfinition de la tâche qui conduit l'élève à produire une réponse.

La demande, adressée à l'élève, de rédiger une phrase complète de réponse constitue une condition de la tâche prescrite que nous observons. L'élève de CP est, rappelons-le, un débutant en matière de maîtrise de la langue écrite : répondre au moyen d'une phrase complète rédigée l'oblige (indépendamment des contenus mis en jeu dans la tâche) à mobiliser une compétence rédactionnelle que l'on ne peut considérer comme allant de soi. Mais au-delà, Cèbe, Goigoux et Thomazet (2004, pp. 9-10) indiquent qu'il faut enseigner aux élèves à savoir comment traiter les questionnaires de lecture en tant que tâches spécifiques de l'univers scolaire, sans quoi ils seront évalués sur des compétences qui n'auront pas été enseignées explicitement. Ces auteurs insistent par exemple sur le fait qu'il est important de « distinguer les compétences qui permettent de comprendre les textes de celles qui sont sollicitées par le traitement des questionnaires » (idem).

Dans le cas où il n'est pas demandé à l'élève de rédiger une réponse, diverses modalités de réponse peuvent être choisies (QCM, dessin, réponse orale...). Elles posent la question de l'influence du cadre de réponse sur le processus de réponse⁸⁰. Nous interrogerons particulièrement les QCM de ce point de vue. Pour cela, nous nous référerons amplement aux travaux de Leclercq (2006) qui, s'intéressant aux QCM, en a proposé une classification de nature à outiller les concepteurs.

5.1. Rédiger une (phrase) réponse complète

Tauveron (2002, p. 93), se préoccupant de lecture littéraire, critique l'aspect chronophage des questionnaires employés en lecture : la rédaction des réponses « occupe l'essentiel du temps inscrit comme temps de « lecture » dans la grille d'activités de la semaine ». Au niveau du CP, rédiger une réponse risque bien d'être plus que chronophage : les élèves sont des lecteurs-scripteurs débutant, et à ce titre, n'ont pas nécessairement suffisamment d'expertise pour utiliser l'écrit pour répondre à une question posée. En effet, répondre par une phrase complète à une question posée nécessite pour l'élève de respecter des contraintes inhérentes à tout écrit (construction de phrase, choix des marques d'énonciation, adéquation du lexique, lien sémantique entre les phrases (entre la question et la réponse), morphologie verbale, respect des règles du code graphophonologique, orthographe, ponctuation etc. cf. EVA-INRP, 1991, p.57). On ne peut considérer que les compétences mobilisées aillent de soi.

On retrouve pourtant trace, à de très nombreuses reprises, de demandes de réponses écrites dans les épreuves de lecture en français.

Prenons par exemple l'épreuve suivante « Après l'école » pour n'en observer qu'un seul aspect au niveau du système de réponse.

Épreuve n°17 : "Après l'école "

a- l'item :

L'enseignant O souhaite évaluer l'item du livret « Lire seul et comprendre un texte » au troisième trimestre.

⁸⁰ Dans la typologie de réactions observables à une question qu'il a élaborée, Piaget (1926/1947, p. 13) souligne que l'une d'elle (la croyance déclenchée) est « influencée par l'interrogatoire puisque la manière même dont la question est posée à l'enfant le force à raisonner dans une certaine direction et à systématiser son savoir d'une certaine façon ».

b- l'épreuve :

Il utilise pour ce faire la seule épreuve reproduite ci-dessous :

2. Lis le texte suivant puis réponds aux questions par une phrase :

Après l'école

Il est cinq heures du soir, les enfants sortent de l'école. Pierre, sa sœur et son petit frère rentrent à la maison. Ils suivent le trottoir. De temps en temps, ils s'arrêtent pour parler ou pour regarder quelque chose. En arrivant devant la maison de monsieur Léon, ils se mettent à courir car ils ont peur du chien qui aboie derrière la grille.

a. Quelle heure est-il ?
Il est cinq heures. ✓
ait.

b. Où étaient les enfants avant cinq heures ?
Ils étaient à l'école. ✓
étaient.

c. Où vont les enfants ?
Ils vont à la maison. Ils rentrent à la maison.

d. Combien y a-t-il de garçons dans cette histoire ?
Il y a 3 garçons.

e. Combien y a-t-il de filles dans cette histoire ?
Il y a une fille.

f. Pourquoi les enfants se dépêchent-ils en passant devant l'une des maisons ?
Parce qu'il y a un chien devant la maison. parce qu'il y a un chien devant la maison. ✓

Lire un texte seul et le comprendre.

Figure 20 : Épreuve FT3 (Annexe B, p. 154) avec réponses d'élève

La tâche prescrite nécessite que l'élève réponde à des questions sur un texte après lecture du texte et du questionnaire, par écrit, en concevant des phrases entières. Ceci ne correspond pas strictement à l'énoncé de l'item « lire seul et comprendre un texte » : la compétence de compréhension d'un texte lu de façon autonome formulée par l'item est évaluée au moyen d'une compétence rédactionnelle « écran ».

c- la copie de l'élève :

Les réponses produites par l'élève manifestent bien la difficulté que peut représenter une telle tâche :

-difficulté de concevoir une phrase qui correspond à la fois à ce que l'élève veut dire et à la question. Toutes les questions sont formulées avec une inversion sujet-verbe qu'il faut rétablir. Certaines mobilisent un pronom « y » dont la référence « dans cette histoire » est post-positionnée. La question « Combien y a-t-il de filles dans cette histoire ? » sous-entend implicitement une réponse devant débiter par « dans cette histoire, il y a ». La réponse de l'élève « Je vois une fille » est corrigée par l'enseignante en « Il y a (une fille) ». Elle est donc pour partie perçue comme erronée puisqu'une correction lui est apportée. Pourtant, l'élève ne dit pas autre chose que le fait qu'il a vu qu'il y avait une fille dans l'histoire, donc dans le texte puisqu'il l'a lu seul. Le marqueur d'énonciation choisi (« je ») renvoie davantage au discours qu'au récit, ce que le marqueur corrigé « il » semble manifester.

-difficulté de produire une phrase orthographiquement correcte. Ceci est bien visible dans les réponses aux questions a et b : « il et té » pour « il était » ou dans la réponse à la question f : « pase que » mis pour « parce que » et « clein » mis pour « chien ».

L'enseignant utilise une bande à 4 cases pour matérialiser la réussite des élèves en bas de page. La case de droite est cochée, elle semble indiquer une réussite maximale. L'ensemble de réponses fournies par l'élève dont nous avons reproduit la copie a été évalué positivement. Pour cet élève performant, on peut se demander quel est l'intérêt de porter autant de corrections sur la copie. Aucune ne porte sur la compétence évaluée (« lire un texte et le comprendre ») : nous avançons l'hypothèse d'une attente implicite en matière de correction de l'expression de la réponse susceptible d'entrer en ligne de compte dans l'évaluation de la compétence de lecture évaluée.

Nous pouvons considérer que réussir cette épreuve renvoie à une certaine expertise de la part de l'élève. On pourrait tout aussi bien concevoir le cas contraire d'élèves qui seraient à même de « lire le texte et de le comprendre » (ce qui est la compétence en jeu) sans être capables de répondre par écrit aux questions. Cela nous conduit à remettre en cause la validité psycho-didactique de l'épreuve.

La situation dans laquelle l'élève doit rédiger une réponse complète se retrouve à plusieurs reprises y compris en mathématiques comme dans l'exemple suivant (épreuve « Les biscuits »).

Épreuve n°18 : « Les biscuits »

a- l'item :

L'enseignant B souhaite évaluer l'item du livret « Résoudre un problème simple » au troisième trimestre. Il utilise pour ce faire trois problèmes dont celui reproduit ci-dessous.

b- l'épreuve :

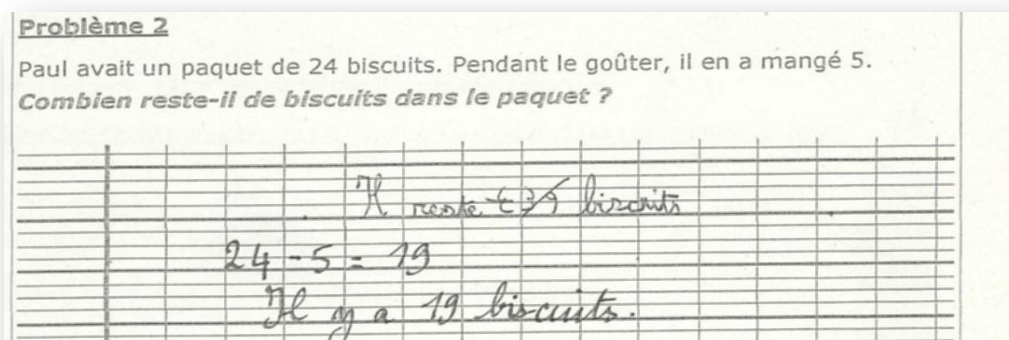


Figure 21 : Épreuve MT3 (Annexe B, p. 307) avec réponses d'élève

Pour réussir ce problème additif avec transformation négative de l'état initial, l'élève doit mathématiser la situation qui correspond à soustraire 5 de 24 pour produire la réponse numérique « 19 ». L'opérateur sémantique⁸¹ « MANGER » permet de confirmer ce que l'usage des temps verbaux permettait de construire : il y avait une situation initiale qui a été modifiée (notons que l'étude de la valeur des temps verbaux n'est pas un objectif de fin de CP, une connaissance implicite est donc mobilisée). Nous discuterons plus loin de la référence de ce problème à une typologie de problèmes additifs. Nous ne revenons pas non plus ici sur la difficulté de déchiffrement de l'énoncé et de la question, ceci a déjà été évoqué précédemment.

Arrêtons-nous seulement sur le cadre de réponse imposé à l'élève. La présentation de l'épreuve est intéressante sur ce point : le lignage de type Seyès reproduit sous le problème ne laisse aucune place à l'élève pour qu'il puisse produire des traces intermédiaires. Tout semble se passer comme si un problème simple impliquait une réponse simple et immédiatement accessible à l'élève, et formulable sous la forme d'une soustraction en ligne qu'il n'y aurait plus qu'à transcrire comme elle vient, puis d'une phrase réponse synthétique. Les lignes ne favorisent pas l'utilisation d'une stratégie moins experte.

c-la copie de l'élève:

L'élève dont nous avons recueilli la copie propose une réponse erronée. Il a répondu « il reste 39 biscuits » en produisant une graphie du « 3 » mal assurée.

Mais il a tenté de produire une phrase réponse condensée dans laquelle figure le résultat du calcul entrepris, dans l'unité adéquate. L'enseignant a repris le problème avec l'élève et a proposé une bonne réponse modèle⁸² : il a écrit une soustraction en ligne et il a produit une phrase réponse. La réponse de l'élève est pourtant plus proche de la question posée que ne l'est celle de l'enseignant : la question était de déterminer un reste de biscuits, ce à quoi

⁸¹ Nous détaillons cette notion dans le chapitre 2.3 de la Partie IV.

⁸² On pourrait probablement estimer que cela constitue une forme d'apprentissage des modalités de réponse à une tâche d'évaluation.

l'élève répond bien. Par l'emploi du terme précis « il reste », il signe le fait que la valeur produite est le résultat de la transformation négative d'une valeur initiale alors que la bonne réponse de l'enseignant ne mentionne pas la référence à la situation initiale et ne porte que l'état final (« il y a 19 biscuits »).

Les valeurs données ne permettent pas d'être certain de l'usage d'une procédure additive ou d'une procédure soustractive : $24-5=19$ et $24+5=29$ sont deux réponses ayant 9 pour chiffre des unités, de sorte que l'enseignant ne peut recueillir d'indices sur le type de procédure utilisée. On pourrait concevoir que l'addition des données numériques étant plus courante (et aussi plus facile ici) que, malgré la présence du terme « reste », l'élève aurait additionné les données et effectué « $24+5= (39)$ »

L'intervention de l'enseignant sur la copie de l'élève révèle des attentes multiples :

- L'écriture du calcul en ligne correspondant à la solution du problème
- La détermination du résultat de ce calcul (une soustraction obligeant le passage d'une dizaine difficile d'accès)
- L'expression rédigée du résultat associant la valeur numérique résultant du calcul et l'unité dans laquelle il faut l'exprimer.

Le cadre de réponse que cette épreuve contient apparaît constituer un important ensemble de contraintes implicites. Le lignage est orienté vers l'écriture d'une phrase réponse. On peut postuler que l'enseignant pense qu'il facilite l'écriture. La suppression de celui-ci semblerait pourtant de nature à ouvrir davantage le type de stratégie que les élèves pourraient mobiliser dans la résolution du problème. En l'état, le cadre de réponse est de nature à invalider la tâche alors que l'usage de ce problème en lui-même est parfaitement fondé du point de vue conceptuel.

5.2. Trouver la bonne réponse parmi plusieurs propositions de réponse

Près de 40% des épreuves d'évaluation de la lecture recueillies au 3^e trimestre nécessitent de répondre à des questions à choix multiples ou QCM⁸³.

Ce type de question est suffisamment employé par les enseignants pour évaluer leurs élèves pour que nous l'étudions dans ce chapitre. Considérant l'obstacle que représente l'écriture de phrases réponses pour les élèves de CP, les QCM présentent l'avantage de ne pas engager l'élève dans la rédaction de la réponse mais dans le choix d'une proposition de réponse.

5.2a Travaux de références en matière de QCM

Leclercq (1986, p.25) souligne que les QCM contribuent « en principe à améliorer la validité, la fidélité et la sensibilité d'un examen parce qu'elles constituent un moyen d'obtenir et de corriger un grand nombre de réponses en peu de temps ». Avant de poursuivre la présentation des travaux de cet auteur, il est nécessaire de clarifier le point de vue adopté par cet auteur au sujet de la validité. Pour Leclercq, différentes critères concourent à définir la qualité d'un système d'évaluation. Plusieurs formes de validités sont considérées, résumées par l'auteur (2006) sous le sigle « ETIC PRAD » :

⁸³ Selon Leclercq (1986, p.15) une QCM est « une question à laquelle l'étudiant répond en opérant une sélection (au moins) parmi plusieurs sélections proposées, chacune étant jugée (par le constructeur de l'épreuve et par un consensus entre spécialistes) correcte ou incorrecte indépendamment de l'étudiant qui doit y répondre ».

- E désigne la validité Ecologique (selon l'auteur, cette validité est d'autant plus grande que la situation d'évaluation réfère à une situation de vie réelle).
- T désigne la validité Théorique, laquelle se décompose en validité de contenu (assimilé par l'auteur à la validité de couverture : tout ce qu'il faut tester l'est-il ?) et de construct (le système d'évaluation est-il fondé sur un modèle crédible ?).
- I désigne la validité Informatrice (si possible diagnostique).
- C désigne la validité Conséquentielle.
- P désigne la validité Prédictive de l'évaluation (capacité à prédire d'autres mesures ultérieures).
- R désigne la validité de Réplicabilité de l'évaluation (autrement dit la fidélité).
- A désigne l'Acceptabilité ou praticabilité de l'évaluation. Cela renvoie pour l'évaluateur, à l'adhésion aux principes et à l'applicabilité des méthodes utilisées et, pour l'élève, à l'adhésion et/ou la familiarité du dispositif.
- D désigne enfin la validité Déontologique (ce qui renvoie à une nécessité de recalculabilité des scores selon l'auteur).

Le questionnement autour de la qualité d'un dispositif d'évaluation enjoint l'auteur à se porter davantage sur la manière d'interroger les connaissances (et sur les conséquences de l'évaluation) que sur les connaissances elles-mêmes ce qui constitue un point de démarcation avec notre approche.

Toutefois la prise en compte de la manière dont on interroge et recueille les réponses des élèves constitue un point de convergence entre ces deux approches de la validité.

Selon Leclercq (1986), les QCM permettent de poser un grand nombre de questions et offrent la possibilité d'une bonne représentativité des aspects questionnés. A ces avantages quantitatifs, s'ajoutent des avantages qualitatifs liés à la concision du codage des réponses, à la précision des exigences (l'auteur cite Noizet et Caverni (1978) : « les QCM ne permettent pas l'esquive ») et à l'objectivité de la correction. Toutefois, s'appuyant sur les travaux de Noizet et Caverni abondamment cités, Leclercq indique que ces avantages sont mis en balance par la difficulté de concevoir une QCM (cela nécessite de s'appuyer sur les erreurs fréquemment commises par les élèves, ce qui n'est pas toujours possible. Par ex. en lecture, l'évaluation dépend fortement du texte lui-même), l'impossibilité d'une présentation orale, la tentation de poser des questions de détail, la facilité de copier les réponses d'un voisin ou encore la possibilité de fournir des réponses correctes en répondant au hasard. Il souligne particulièrement quatre désavantages fondamentaux : l'impuissance à mesurer certaines capacités (les QCM sont plus adaptés pour évaluer des performances isolables), la présentation de solutions erronées (particulièrement en orthographe, un risque existe de mémoriser des formes incorrectes), l'incapacité à exercer l'habileté à formuler des réponses et la contraction du champ cognitif qui pourrait tendre vers une forme d'uniformisation des individus.

Suite à la présentation de cet ensemble d'avantages et d'inconvénients, Leclercq propose une taxonomie de consignes et un ensemble de 20 règles à considérer lors de la construction des QCM (chapitre 3 de son ouvrage). Ces règles se répartissent en règles d'adéquation aux objectifs, règles sur la valeur diagnostique, règles générales de rédaction, et, règles de rédaction des solutions proposées.

Au sujet des corrections, Leclercq montre (2006) que la modalité de correction classique⁸⁴ est injuste dans le sens où « elle pénalise aveuglement les personnes à qui on a interdit d'exprimer leur degré de doute » (2006, p. 143). Dans le texte de 2006, Leclercq reprend une partie des critiques théoriques sur les processus mentaux mesurés et non mesurés des QCM. Ceux-ci portent premièrement sur le fait que les QCM ne mesurent pas l'évocation de connaissances provenant de la mémoire. Les QCM ne peuvent prétendre à mesurer la capacité d'évoquer des connaissances, mais bien celle de les reconnaître. Deuxièmement, les QCM invitent à un raisonnement à rebours : avec un QCMC⁸⁵ l'élève a tendance à d'abord considérer et éliminer les solutions proposées au lieu d'en évoquer une personnelle. Troisièmement, les QCMC renforcent le curriculum caché de l'école (ce que personne n'enseigne mais que tout le monde apprend à l'école) qui pourraient selon nous renvoyer à un aspect du contrat pédagogique (toute question a une réponse, on ne remet pas en question la question posée etc.). D'après ce chercheur, une attention portée au degré de certitude des réponses permettrait de palier certains manques de validité. Ce type d'attention fait l'objet de travaux en particulier pour l'évaluation de la compréhension en lecture (cf. De Croix et Ledur, 2014).

Nous ne trouvons pas dans notre corpus d'épreuves dont les modalités de réponse sont assorties de la prise en compte de l'expression d'un degré de certitude.

5.2b Analyses de QCM en lecture

Fondamentalement, une QCM engage l'élève à choisir une bonne réponse parmi un ensemble de modalités de réponse. De ce fait, comme le rappellent justement Chartier et Loarer (2008, p.200), « la possibilité d'obtenir une bonne réponse au hasard n'est pas négligeable », sans présumer des relations inter questions ou de la possibilité de répondre au questionnaire en inférant les réponses à partir des autres questions. Il est donc nécessaire, selon ces auteurs, « de proposer un nombre significatif de modalités de réponse et de s'assurer de leur « égale attractivité » » (idem)

Nous questionnons cette recommandation formulée dans une optique différente de la nôtre (test SPM de Raven) à travers l'analyse d'épreuves collectées : sur quoi porterait l'«égale attractivité » des modalités de réponses ?

5.2b-1 Proximité graphique ou graphophonologique des solutions

Épreuve n°19 : «Pirate»

a- l'item :

L'enseignant H souhaite évaluer l'item « Je sais lire et comprendre des phrases simples ». C'est une adaptation personnelle de l'item des programmes « Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens. ». Elle trouve sa justification du côté de la quantité d'écrit à lire : l'enseignant estime que lire un texte en fin de 1er trimestre n'est pas accessible pour une part des élèves, il propose donc une épreuve fondée sur la lecture de phrases.

⁸⁴ 1 point par réponse correcte, 0 par omission. Pour une réponse erronée, il faut retirer $1/(k-1)$ point (pour k désignant le nombre de solutions proposées). Par exemple il faut retirer -1/3 pt si la réponse choisie par l'élève est l'une des deux modalités de réponse ne correspondant pas à la bonne réponse (qui serait alors la troisième modalité proposée). (Mc Call, 1920)

⁸⁵ Questions à Choix Multiple fonctionnant avec la consigne Classique (QCMC), i.e. : « Une seule des solutions proposées est correcte et vous avez droit à une seule réponse »

b- l'épreuve :

Pour cela, il propose la tâche suivante :

1. Lis le texte.
Téo est un pirate. Il vit sur un navire. Son amie Lila est une tortue.
Lila mange de la salade et des radis. Téo joue aux dés avec Lila.

2. Colorie la bonne réponse.


2a Téo est : un papa un pirate une pie

2b Lila est : un tapis une tarte une tortue

2c Lila mange : de la soupe et des radis de la salade et des radis

2d Téo et Lila jouent : aux dominos aux dés aux cartes

3. Colorie les personnages de l'histoire.

Q3 

Q4 **4. Où vit Téo ? Colorie.**


E 

Figure 22 : Épreuve FT1 (Annexe B, p. 77) avec réponses d'élève

Comme l'indique la consigne « lis le texte », il s'agit de lire non pas « des phrases simples » (i.e. isolées) ce qui est formulé dans l'item, mais un ensemble de phrases assimilables à un texte (des reprises anaphoriques lient les phrases), plutôt construit pour répondre à un besoin d'évaluation des compétences d'identification de mots et de maîtrise des règles de correspondances phonographiques. Ceci constitue un premier écart entre la tâche à réaliser et la tâche prescrite : l'élève doit pouvoir mobiliser des processus d'intégration (repérage et compréhension de la chaîne anaphorique, cf. Giasson, 1990) au cours de sa lecture qui ne sont typiquement pas nécessairement mobilisés dans la lecture de chacune des phrases prises isolément les unes des autres (dans ce cas Lila peut être un personnage humain féminin plutôt qu'une tortue dans la phrase n°4 « Lila mange de la salade et des radis »). Ces processus de lecture ne sont pas de même nature que les microprocessus mis en œuvre pour décoder un (ou une suite de) mot(s).

Répondre à la question « Lila est » (Q2b) oblige l'élève débutant en lecture (nous sommes en fin de premier trimestre, c'est-à-dire environ fin novembre) à un ensemble d'opérations dont la complexité est fonction de ses ressources et de leur état de disponibilité.

Après la lecture du texte et des questions, l'élève doit tout d'abord reconstruire successivement les phrases qui lui sont incomplètement présentées (« Lila est » + « un tapis », « Lila est » + « une tarte », « Lila est » + « une tortue »).

Il doit intégrer la solution candidate au début de phrase et évaluer le sens de la phrase ainsi construite : les deux premières propositions de solution ne sont pas vraisemblables (on peut estimer très faible le risque que les élèves mobilisent des connaissances liées au registre de langue les conduisant à estimer que « tarte » pourrait être synonyme d'« idiot »).

Il doit ensuite comparer le sens de la réponse potentielle avec la mémoire du texte lu ou la représentation du modèle de situation construit pendant la lecture ou encore la reconnaissance d'un mot dans le texte contenu dans la réponse candidate (« tapis » et « tarte » ne sont pas inclus dans le texte contrairement à « tortue » ce qui est susceptible d'orienter le choix de cette réponse).

Il doit décider de choisir cette réponse ou recommencer cet enchaînement avec la proposition de solution suivante. Il évite un certain nombre d'enchaînements inutiles s'il connaît certaines des règles implicites qui régissent ce type d'épreuve : chaque question a obligatoirement une et une seule solution, ceci participe des connaissances antérieures convoquées au cours de la lecture (et invite à considérer que plus l'élève est familier de ce format d'épreuve, plus il est à même de mieux y répondre).

Répondre à la question « où vit Téo ? »(Q4) oblige l'élève à associer le mot du texte « navire » (probablement choisi en raison de sa plus grande facilité de déchiffrement comparé au mot « bateau ») au dessin d'un bateau, ce qui nécessite une connaissance lexicale précise. Sans quoi l'élève peut associer « pirate » à « bateau » ce qui est plus simple et dépend de ses connaissances antérieures. Les questions n'ont pas le même niveau de généralité que le texte (plus précis).

Les questions de l'épreuve ne sont pas rédigées de la même manière que le texte. Dans le texte les lettres muettes (souvent des morphèmes grammaticaux) sont écrites dans une autre couleur (le « e » de « amie » est écrit en rouge). Ce n'est pas le cas dans les questions : de sorte que ce qui était considéré comme source de difficulté dans la lecture du texte ne l'est plus dans la lecture du questionnaire.

c- la copie de l'élève :

L'élève se méprend dans la sélection de la réponse à la question 2b (« Lila est : »). Il sélectionne la réponse « une tarte » plutôt que « une tortue ».

L'élève colorie la réponse attendue à la question suivante « Lila mange » : « de la salade et des radis »

Parallèlement, il sélectionne le dessin représentant la tortue dans la question 3 portant sur les personnages de l'histoire.

Les réponses produites ne semblent pas cohérentes entre elles : l'histoire ne parle pas de « tarte » et l'élève n'a pas eu l'occasion d'entourer un dessin de tarte (nous ne savons pas s'il l'aurait fait).

On peut donc penser que l'élève a été attiré par la modalité de réponse « tarte », non pas pour sa valeur sémantique mais pour sa proximité graphique avec « tortue », alors qu'il a peut-être reconnu que l'histoire présentait une tortue (ce que la sélection du dessin de tortue dans Q3 « colorie des personnages de l'histoire » invite à penser).

La proximité graphique des modalités de réponse avec la bonne réponse ne semble pas être le meilleur choix s'il s'agit d'évaluer la compréhension plus que le décodage.

On voit bien que l'objectif de l'enseignant qui était d'évaluer « la compréhension de phrases simples » se réalise au moyen d'une tâche qui constitue pour l'élève à écarter des modalités de réponse présentant de grandes similitudes graphiques avec la bonne réponse. De sorte

que l'élève peut très bien avoir compris ce qu'il a lu (ce qui correspondrait à ce que souhaite évaluer l'enseignant) mais que le cadre de réponse de l'épreuve peut en empêcher la manifestation.

La validité psycho-didactique de cette tâche est discutable. La tâche proposée ramène la compréhension au plan de la comparaison de mots qui n'ont *a priori* aucune raison de se trouver confrontés (pie, papa et pirate ou tapis, tarte et tortue) et pourrait être analysée comme une tâche de vérification de mots présents dans le texte. Dans des tâches destinées à évaluer la compréhension, il n'est pas certain que la proximité graphique ou graphophonologique des solutions proposées soit un critère à privilégier. Les élèves peuvent se méprendre sur le choix de la bonne réponse en sélectionnant une modalité de réponse graphiquement proche sans que cela signifie qu'il n'ait pas compris ce qu'il a à lire.

Une autre possibilité consiste à proposer des solutions proches sur le plan thématique. Cette option est souvent utilisée par les enseignants dans des tâches d'évaluation de la compréhension d'un texte lu par un tiers, surtout en début d'année.

5.2b-2 Proximité thématique des solutions

Épreuve n°20 : « Compréhension orale »

a- l'item :

L'enseignant B souhaite évaluer l'item « Comprendre le sens global d'un texte lu par l'enseignante ». C'est une adaptation personnelle de l'item des programmes « Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens. ».

b- l'épreuve :

Pour cela, il utilise l'épreuve ci-dessous :

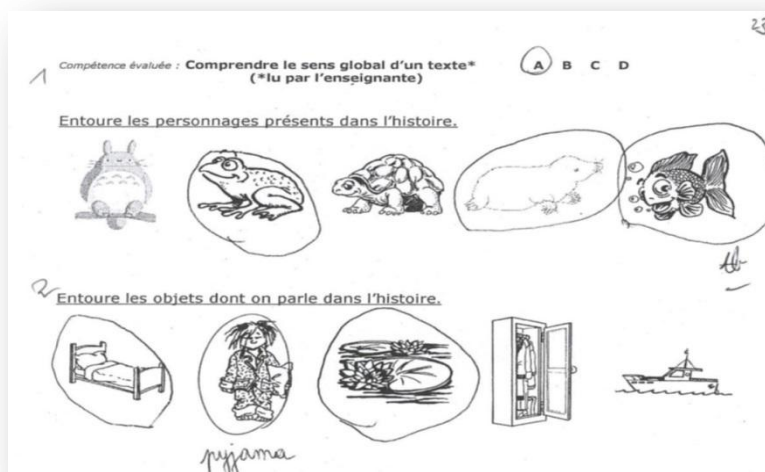


Figure 23 : Épreuve FT1 (Annexe B, p. 67) avec réponses d'élève

Cette tâche fait suite à la lecture par l'enseignant de l'album « Scritch, Scratch, Dip, Clapote » (K. Crowther, éd. Ecole des loisirs).

Nous reproduisons le texte de présentation qu'en fait l'éditeur sur son site pour mieux situer la tâche : « *Comme chaque soir, la nuit se couche sur l'étang. Et, comme chaque soir, Jérôme a peur. Papa lui lit une belle histoire puis l'embrasse. Maman vient lui faire un dernier câlin. Jérôme se retrouve tout seul dans sa chambre. Tout seul dans son lit, tout seul dans son cœur. Je crois que j'ai entendu un bruit... Qui fait "Scritch scratch dip clapote" sous mon lit ?* »

Dans le même but, nous reproduisons également une page de l'album :

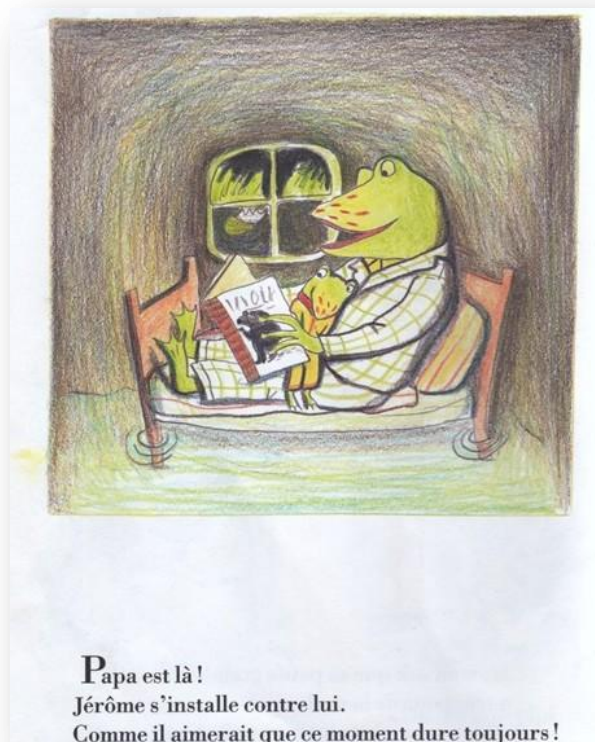


Figure 24 : Extrait de l'album "Scritch, scratch, dip, clapote" (Crowther, Ecole des loisirs)

La tâche n'est pas une tâche de QCMC à proprement parler : il y a plusieurs bonnes réponses par question, elle s'apparente plutôt à un vrai/faux. Produire la bonne réponse engage de choisir non pas une mais plusieurs bonnes réponses parmi des solutions thématiquement proches ce qui contraste avec le contrat implicite du questionnaire précédent.

La question 1 ne permet pas de distinguer les personnages principaux (la petite grenouille et son père), ni ceux-ci des personnages secondaires. Pour bien répondre⁸⁶, l'élève doit avoir mémorisé tous les animaux intervenant dans l'histoire plutôt que leur rôle dans l'histoire (par exemple le bruit qu'ils émettent et qui conduisent l'éditeur à choisir le titre). Un oubli peut donc conduire l'élève à sélectionner un autre animal en raison de sa proximité avec un autre. Tous les animaux partagent en effet un caractère commun : ils peuvent vivre au bord d'un étang. Une telle redéfinition de la tâche prescrite ne conduirait pas l'élève à répondre à la demande de l'enseignant.

⁸⁶ Il faudrait sélectionner : la grenouille, la taupe qui émet le son « Scritch, Scratch », l'oiseau de nuit qui émet le son « Dip » (si tant est que le dessin de gauche de la première ligne en soit un) et le poisson qui fait « Clapote ».

Pour répondre à la deuxième question, il faut entourer les objets dont parle l'histoire. Les modalités de réponses renvoient à différentes catégories : celle des objets, explicitement nommée (lit, armoire, bateau) et celle, par opposition, des êtres vivants (enfant, nénuphar). L'illustration du texte dans l'album ne montre pas d'armoire. Pourtant, l'élève pourrait entourer cet élément mobilisant pour cela une connaissance sur le monde telle que « dans toutes les chambres on trouve une armoire (ou une étagère) pour ranger ses vêtements ». Pour une raison similaire (une connaissance sur le monde inverse relative à son expérience personnelle), il pourrait tout aussi bien ne pas entourer l'armoire si dans sa propre chambre il n'y en a pas.

Le nénuphar est important dans l'histoire, il permet de localiser l'histoire (près d'un étang). Pourtant ce n'est pas un objet. La réponse de l'élève devrait donc exclure cette réponse alors qu'il est sans doute intéressant de vérifier que l'élève a compris le cadre dans lequel se déroule l'histoire.

c- la copie de l'élève :

L'élève a réussi l'épreuve (il obtient A) pourtant il a omis d'entourer « l'oiseau de nuit » dans la première question : cette erreur n'est pas relevée et l'élève obtient le crédit maximal.

Il a entouré le dessin d'une grenouille, d'une taupe et d'un poisson pour répondre à la demande d'entourer les personnages présents dans l'histoire.

Il a entouré le dessin d'un lit, d'un (enfant en) pyjama et d'un nénuphar pour répondre à la demande d'entourer les objets dont on parle dans l'histoire.

La validité psycho-didactique de l'épreuve est remise en cause, elle ne permet pas d'évaluer « la compréhension globale du sens d'un texte » annoncée. Rien ne permet de savoir à l'issue de la passation de l'épreuve si l'élève a compris que la petite grenouille a fait un cauchemar et que son père, en la rassurant, la conduit à découvrir l'origine des bruits qui l'effrayaient.

La tâche de l'élève pourrait se résumer à une tâche de catégorisation (reconnaître les animaux vivant au bord de l'étang, reconnaître les objets que l'on trouve dans une chambre) le conduisant à entourer toutes les modalités de réponses (le bateau peut être considéré comme un jouet) sans que cela ne constitue pour autant une mauvaise réponse. Cela nous semble imputable à la trop grande proximité thématique des modalités de réponse proposées.

5.2b-3. Proximité sémantique des solutions

Épreuve n°21 : « Petit Gaston »

L'épreuve « Petit Gaston » est une épreuve qui a été utilisée à plusieurs reprises par la DEPP dans le cadre des évaluations diagnostiques à l'entrée en CE2.

Nous l'avons reconnue, aménagée, à deux reprises parmi les épreuves utilisées par les enseignants pour évaluer la compréhension en lecture en CP.

Elle a été utilisée telle qu'elle par un enseignant auprès de qui nous avons collecté des épreuves lors d'un premier travail exploratoire (cf. première version reproduite ci-dessous). Elle provient d'un site de circonscription (circonscription de Cluses, Haute-Savoie).

Nous l'avons également collecté auprès de l'enseignant C (cf. deuxième version de l'épreuve reproduite ci-dessous). Cela souligne la nécessité de s'interroger sur le choix des épreuves, ce dont nous discuterons plus loin dans le chapitre 6.2.

2 Comprendre un texte et montrer que l'on a compris : donner après lecture des informations ponctuelles contenues explicitement dans le texte.

A Prélever une information ponctuelle (items 4, 5 et 6)

A Comprendre le sens général d'un texte en choisissant parmi plusieurs phrases celle qui résume le mieux le sens général (item 7)

RG

Petit Gaston

Cette nuit, cinq hérissons sont arrivés dans le jardin. Le père, la mère et leurs trois fils. Ils se dirigent droit vers les salades, qu'ils adorent. Surtout Petit Gaston, le plus gourmand.

« Petit Gaston est si vorace, dit la maman, qu'un jour il mangera les étoiles ! »

Aussitôt Petit Gaston lève le nez. Les étoiles sont sûrement meilleures que les salades.

Il n'y a qu'un problème : les attraper ! Comment faire ? Petit Gaston trouve une échelle appuyée au cerisier. Il ne lui reste qu'à monter.

1 Cherche les bonnes réponses dans le texte et colorie les cases correspondantes.

L'échelle est appuyée :

it 4

à un mur

à un cerisier

à un buis

à une étoile

Petit Gaston est :

it 5

coléreux

timide

gourmand

méchant

La famille hérisson adore :

it 6

les tomates

les salades

les poireaux

les cerises

2 Choisis la phrase qui raconte le mieux l'histoire et colorie la case correspondante.

it 7

C'est l'histoire d'un petit hérisson qui veut monter au sommet d'une montagne.

C'est l'histoire d'une famille hérisson qui va manger des salades.

C'est l'histoire d'un petit hérisson qui croit qu'on peut manger les étoiles.

C'est l'histoire d'un petit hérisson qui veut attraper des cerises.

Petit Gaston

La nuit, cinq hérissons sont arrivés dans le jardin. Le père, la mère et les trois enfants. Ils marchent vers les salades, qu'ils adorent. C'est surtout Petit Gaston qui aime les salades, il est très gourmand.

Maman dit : « Il est si gourmand, Petit Gaston, qu'un jour il ira même manger les étoiles ! »

Aussitôt Petit Gaston réfléchit. Les étoiles c'est sûrement mieux que les salades. Il s'approche d'une échelle posée contre un cerisier, il ne lui reste plus qu'à monter.

Cédric Selmir

1 **Note ici :**

Le nom de l'auteur :

Le titre du texte :

2 Cherche les bonnes réponses dans le texte et colorie les cases correspondantes.

L'échelle est posée contre

<input type="checkbox"/>	un mur
<input type="checkbox"/>	un cerisier
<input type="checkbox"/>	un but
<input type="checkbox"/>	une étoile

Petit Gaston est

<input type="checkbox"/>	colérique
<input type="checkbox"/>	timide
<input type="checkbox"/>	gourmand
<input type="checkbox"/>	méchant

3 La famille hérisson adore

<input type="checkbox"/>	les tomates
<input type="checkbox"/>	les salades
<input type="checkbox"/>	les poireaux
<input type="checkbox"/>	les cerises

Dans la famille hérisson il y a :

<input type="checkbox"/>	les parents et deux enfants
<input type="checkbox"/>	la mère et trois enfants
<input type="checkbox"/>	les parents et trois enfants

Choisis la phrase qui correspond le mieux à l'histoire.

- | | |
|--------------------------|--|
| <input type="checkbox"/> | C'est l'histoire d'un petit hérisson qui veut monter au sommet d'une montagne. |
| <input type="checkbox"/> | C'est l'histoire d'une famille hérisson qui va manger des choux. |
| <input type="checkbox"/> | C'est l'histoire d'un petit hérisson qui croit qu'on peut manger les étoiles. |
| <input type="checkbox"/> | C'est l'histoire d'un petit hérisson qui veut attraper des cerises. |

Figure 25 : Épreuves « Petit Gaston » (site de circonscription de Cluses pour la 1^{ère} version avec réponses d'élève, et Annexe B, p. 134 pour la 2^e version)

Les aménagements de l'épreuve « Petit Gaston » présentés contiennent deux questions communes.

-La première question (sur le premier exemple) porte sur l'item « prélever une information ponctuelle ». Elle n'est pas à proprement parler une question. L'élève n'a donc pas à trouver stricto-sensu une réponse. Cela sous-entend que la question ne se trouve pas intégralement dans la consigne mais que l'élève doit la reconstruire. Ainsi la consigne « cherche les bonnes réponses dans le texte et colorie les cases correspondantes » pourrait-elle être reformulée par l'élève comme une demande visant à déterminer le groupe de mots qui permettrait de compléter la phrase selon le sens du texte.

Il est aisé de constater qu'il ne s'agit pas pour l'élève de prélever une information ponctuelle dans le texte mais bien de sélectionner parmi plusieurs possibilités de réponses, la bonne réponse à la question posée portant sur un aspect du texte. Les propositions de réponse sont construites de manière à minimiser les redondances entre elles : ainsi on propose des éléments de phrase (« Petit Gaston est ... ») et non des phrases entières (« Petit Gaston est coléreux »). C'est à l'élève de reconstruire les phrases entières, en associant le début à une fin proposée, avant d'en juger le caractère plausible puis, le cas échéant, conforme au texte dont elles se rapprochent. Ceci constitue une difficulté de lecture non négligeable.

Plusieurs possibilités de réponses sont proposées aux élèves à la question « L'échelle est appuyée à ... ». Les modalités de réponse « à un mur », « à un but », « à une étoile » n'ont pas de proximité graphophonologique ni de proximité thématique avec la bonne réponse « à un cerisier ». Le choix d'une modalité de réponse autre que la « bonne réponse » pourrait plutôt relever d'une erreur de compréhension. Analysons les différentes possibilités de répondre :

- Dans l'épreuve de gauche, la bonne réponse est « l'échelle est appuyée à un cerisier », et dans l'épreuve de droite « l'échelle est posée contre un cerisier ». Le texte diffère légèrement de façon à contenir littéralement la bonne réponse. L'élève peut donc choisir la bonne réponse sur la base de la reconnaissance de segments de texte identiques dans la question et le texte. Il peut aussi construire sa réponse sur la base de connaissances sur le monde dont il disposerait : un cerisier est un arbre, on cueille les cerises dans un cerisier, une échelle peut être nécessaire car elle permet d'accéder aux parties les plus inaccessibles de l'arbre.

Cela permet de mettre en évidence le fait que les catégorisations de question envisagées par les concepteurs pensées *a priori*⁸⁷ ne traduisent pas nécessairement les procédures utilisées par les élèves pour répondre. La bonne réponse peut être produite par simple reconnaissance d'une partie de texte contenue dans la question et dans le texte, elle peut aussi faire l'objet d'une véritable construction.

- La modalité de réponse « à un mur » est sémantiquement équivalente à la bonne réponse « à un cerisier » du point de vue des connaissances sur le monde relative à l'usage d'une échelle. L'élève peut sélectionner la modalité de réponse « à un mur » sur la base de son expérience du monde et de ce type de connaissances. La phrase

⁸⁷ Pour les évaluations CEDRE (situées en fin de CM2) réalisées par la DEPP celles-ci sont : Prélèvement d'une information explicite ; Inférence de niveau 1 (mise en relation d'information du texte pour répondre à une question) ; Inférence de niveau 2 (mise en relation d'information du texte avec des connaissances extérieures pour répondre à une question) ; Synthèse de niveau 1 (trouver le thème, choisir un titre ou un résumé) ; Synthèse de niveau 2 (analyser un document : tirer les éléments essentiels qui aident à la compréhension d'une situation...) Cf. DEPP, NI n°20, juillet 2016.

ainsi reconstituée est sémantiquement acceptable : « l'échelle est appuyée à un mur ». En revanche, cette réponse n'est pas articulée au texte. L'élève répond bien à une partie implicite de la question portant sur l'emboîtement d'une modalité de réponse avec l'entame de réponse.

- La modalité de réponse « à un but » se comprend difficilement. S'agit-il du concept de but ? S'agit-il de l'usage sportif du terme « but » (zone de jeu gardée par un gardien de but) ? La phrase réponse reconstituée « l'échelle est appuyée à un but » n'a pas de sens dans le premier cas. Elle ne correspond pas à une erreur typique d'élève (telle que par exemple le choix du dernier mot du texte), et peut de ce fait être rapidement écartée par l'élève indépendamment du texte lu. Il est donc fort plausible que les élèves ne sélectionnent pas cette proposition de réponse. Il serait intéressant de pouvoir questionner les élèves ayant choisi cette réponse sur les raisons de leur choix.
- La modalité de réponse « à une étoile » mériterait un examen attentif des motifs qui conduiraient les élèves à la sélectionner. La phrase réponse reconstituée, « l'échelle est appuyée à une étoile », renvoie éventuellement à un récit de fiction (mais un animal qui parle est également caractéristique d'un récit de fiction). Il existe de nombreux récits faisant état de tels rêves (par exemple « La tour jusqu'à la lune⁸⁸ »). L'élève pourrait avoir mobilisé des connaissances intertextuelles sur les récits de fiction pour répondre (ce qui renvoie à une réponse fondée sur une intrusion, cf. classifications RQR de Pearson et Johnson présentée dans le chapitre 4.2.2). Mis à part cette référence intertextuelle, il est peu probable que les élèves choisissent cette réponse si ce n'est parce que le mot « étoile » est contenu dans le texte.

Les modalités de réponses n'ont pas ici la même vraisemblance (Leclercq, 1986, p. 105). Deux d'entre elles (dont la bonne réponse) entretiennent une grande proximité sémantique entre elles deux, et entre chacune d'elle et le texte. Les deux autres n'ont pas de véritable vraisemblance sémantique en elle-même, pas plus qu'elles n'ont de proximité sémantique avec les deux autres. Le choix de l'une ou l'autre modalité (« à un but », « à une étoile ») risque bien de ne pas renseigner utilement l'enseignant sur la compréhension que l'élève a du texte (Leclercq utilise le terme de défaut de « validité informative » (2006, p.143)). Tout au plus pourra-t-il estimer voir fonctionner un effet de contrat (il faut cocher une proposition).

En 1995 et en 2000, 85 % d'élèves de CE2 sélectionnent la bonne réponse (MEN DEPP 1996, p.25 et MEN DEPP 2001, p. 34). Ce qui signifie que 15% des élèves de CE2 sélectionnent une autre réponse⁸⁹ sans que les tentatives d'explications de ce fait en soient très convaincantes. Colmant et Jeantheau, rapporteurs de l'étude de 1995 indiquent que la moindre réussite sur le premier item⁹⁰ est difficile à expliquer « peut-être est-elle due au fait que la phrase permettant de répondre est loin dans le texte et nécessite donc une relecture rapide » (MEN DEPP 1996, p. 27). Robin, rapporteur de l'étude de 2001, explique que les mauvaises réponses sont imputables à de « multiples raisons : situation habituelle d'une

⁸⁸ Quarante petits Contes. Bayard. 1984 ou bien encore <http://blog.bayam.fr/le-petit-theatre-des-contes-la-tour-jusqua-la-lune/>

⁸⁹ Malheureusement il ne nous a pas été possible d'obtenir le détail des réponses des élèves en fonction des choix proposés. Les corrections étaient à cette époque directement prises en charge par les enseignants et seules les bonnes réponses, non réponses et réponses erronées étaient codées.

⁹⁰ employé ici au sens classique du terme dans un test.

échelle contre un mur ; méconnaissance pour certains du « cerisier » ; détail secondaire par rapport à l'ensemble du texte, etc.). » (MEN DEPP, 2001, p. 35)

On peut considérer en utilisant la typologie de questions de Leclercq (1986, chapitre III) que le problème vient du style de rédaction de la question « L'échelle est appuyée... ». Si l'on considère cette question relevant du style direct (« à quoi l'échelle est appuyée ? ») on peut y répondre sans avoir à lire les réponses proposées (« l'échelle est appuyée à un mur » est la plus plausible). Si l'on considère que cette question est rédigée dans un style indirect, i.e. partant des solutions, il faut alors que l'élève « soit obligé d'examiner les solutions proposées » (Leclercq, 1986, p. 83). Pour ce faire il faudrait qu'il soit impossible de répondre à la question sans avoir lu les réponses, ce qui du fait des schémas des lecteurs (considérés comme des connaissances provenant de l'expérience) n'est pas certain. L'effet des connaissances antérieures risque bien d'être important.

-La deuxième question (sur le premier exemple) porte sur l'item « comprendre le sens général d'un texte en choisissant parmi plusieurs phrases celle qui résume le mieux le sens général ».

A considérer les propositions de réponses, deux réponses sont possibles : la deuxième et la troisième. L'élève a d'ailleurs coché les deux :

- La deuxième réponse cochée est centrée sur le but poursuivi par la famille hérisson. Elle est de ce point de vue parfaitement correcte, ce qu'environ un tiers des élèves de CE2 considère également⁹¹. Elle ne tient pas compte de la deuxième partie du texte concernant l'un des membres de la famille (Gaston).
- La réponse attendue est la troisième proposition, ciblée sur la compréhension du but poursuivi par Petit Gaston, ce qu'environ 40%⁹² des élèves de CE2 considère. Cette réponse ne tient pas compte de la première partie du texte. Ne pas avoir coché cette bonne réponse n'indique pas que l'on n'a pas, pour autant, « compris » le texte « d'une façon générale » (cf. rédaction de l'item).

22% des élèves n'ont pas produit de réponse ce qui n'est pas analysé par les rapporteurs de l'enquête. Est-ce imputable à la trop grande proximité des deux réponses plausibles et à l'impossibilité de déterminer LA bonne parmi les deux ?

Les autres choix n'ont été sélectionnés que par moins de 5% des élèves de CE2. Cela n'est pas commenté dans les documents de synthèse de la DEPP :

- La première proposition de réponse n'a pas de lien avec l'histoire. En ce sens elle n'est pas équivalente aux autres. Son éventuelle égale attractivité se situe au niveau de sa construction (correcte sémantiquement et syntaxiquement) et de sa longueur. Le mot « montagne » n'apparaît pas, et on trouve simplement le terme « monter » en fin de texte. La proposition est sémantiquement correcte si bien qu'un élève de CP, ayant quelques difficultés à parvenir jusqu'au bout de cette longue phrase comportant de réelles difficultés de décodage (« montagne »), pourrait la sélectionner. Ceci pourrait s'expliquer par un effort de déchiffrement tel qu'il aurait oublié le texte auquel la question se rapporte et/ou qu'il pourrait décourager l'élève de lire les autres propositions et le conduire à sélectionner la première qui aurait un sens acceptable.

⁹¹ 32% des élèves de CE2 ont sélectionné cette réponse en 1995 et 30% en 2000 (MEN DEPP 1996 p. 26 et MEN DEPP 2001 p. 35).

⁹² 42% des élèves de CE2 ont sélectionné cette réponse en 1995 et 44% en 2000 (MEN DEPP, 1995, p. 26 et MEN DEPP, 2001 p. 35).

Surtout, quel renseignement tire l'enseignant du fait que l'élève a coché cette première proposition de réponse ?

- La quatrième proposition de réponse n'a pas de lien avec l'histoire. Son éventuelle égale attractivité se situe au niveau de sa construction (correcte sémantiquement et syntaxiquement) et de sa longueur. Le mot « cerise » n'apparaît pas, et on trouve simplement le terme « cerisier » en fin de texte, ce qui pourrait conduire l'élève à sélectionner cette proposition, non pas sur la base du sens mais sur la base d'une mémorisation du mot. Cette proposition de réponse est assez proche en termes d'attractivité de la première.

Cette analyse permet de considérer que les distracteurs n'ont pas d'égale attractivité avec la bonne réponse. Deux des trois distracteurs n'ont pas de lien avec l'histoire du texte, le troisième est si proche qu'il ne permet pas d'invalidier la compréhension du texte par l'élève : celle-ci serait en effet correcte.

La validité de cette épreuve est d'autant plus remise en cause que si plus d'un élève de CE2 sur deux ne répond pas (ibidem) à cette question, on peut conjecturer qu'en CP la proportion risque d'être notablement plus élevée, sans pouvoir en déduire quoique ce soit de fondé. Connaissances sur le monde, connaissances lexicales, connaissances hypertextuelles, compétence à reconstruire des phrases complètes à partir d'éléments fractionnés, capacité à lier la réponse au texte (car pendant que l'élève essaie les diverses combinaisons possibles, qu'il sature sa mémoire de travail de ces éléments qu'il peut avoir du mal à déchiffrer, il peut oublier le texte source) ... sont autant d'éléments susceptibles d'interférer dans le processus de sélection de la bonne réponse.

La question supplémentaire que pose ce type de QCM est celle de la (non-)prise en compte d'éléments plausibles, ne provenant pas du texte mais des connaissances du lecteur. L'épreuve permettrait autant d'évaluer les connaissances antérieures et le raisonnement verbal des élèves que la compréhension du texte lu : « reading comprehension test scores may be assessing prior knowledge or some verbal reasoning ability as much as the ability to understand the passage » (Keenan et Betjemann, 2006, p.364).

Discussion :

Nous venons d'analyser trois QCM dans lesquels les choix de distracteurs ne renvoient pas nécessairement à une compréhension erronée et, pour partie, sont imputables à un non-respect de règles de construction des questions et des réponses proposées. Dans les QCM analysés, il apparaît de surcroît que :

- situer l'égale attractivité sur le plan graphophonologique pour des QCM de compréhension ne permet pas véritablement à l'enseignant de recueillir une information quant à ce que l'élève a compris du texte donné à lire.
- situer l'égale attractivité sur le plan thématique engage une tâche de catégorisation qui n'est pas évaluée pour elle-même et qui est susceptible d'interférer avec la réponse de l'élève.
- situer l'égale attractivité sur le plan sémantique risque d'entraîner les élèves vers une tâche de détection de la plausibilité de la réponse en tant que proposition isolée, indépendante du texte. Ceci renvoie à la mobilisation des connaissances antérieures (« prior knowledge »).

Dans les épreuves collectées, nous ne trouvons pas de traces de l'usage de Questions à Choix Multiples en mathématiques. Il se pourrait pourtant qu'elles puissent être plus adaptées dans ce domaine en raison de la connaissance (issue de l'expérience, de travaux antérieurs) de procédures classiquement utilisées à tort par les élèves dans une situation précise. En lecture, de telles connaissances ne semblent pas être disponibles ou aisément mobilisables dans des QCM. En effet, le risque de voir des élèves ne se soucier que de la cohérence de la phrase réponse pour la valider indépendamment du texte questionné semble important.

Les rapporteurs de l'étude menée par la DEPP en 1996 indiquent, dans la synthèse de leur analyse des résultats de l'épreuve Petit Gaston, qu'« il semble possible de dire qu'une véritable pédagogie de la compréhension est maintenant à l'œuvre sur l'ensemble du territoire national » (MEN DEPP 1996, p. 27). Il est sans doute prudent de ne pas s'avancer aussi loin sur la seule base de l'exploitation des données chiffrées des résultats et des typologies de questions sur lesquelles elles reposent.

5.3. Trouver la bonne réponse et rédiger la réponse :

La tâche « Niche » que nous analysons ci-après combine deux difficultés précédemment évoquées de nature à invalider l'épreuve : il s'agit dans cette épreuve de répondre à une variété de QCM et de produire une réponse écrite. S'ajoute ici le fait de produire une réponse écrite qui s'insère dans un début de réponse imposé.

Épreuve n°22 : « Niche »

a- l'item :

Au premier trimestre, l'enseignant B souhaite évaluer les items du livret « Lire aisément les mots étudiés » et « Lire aisément les mots les plus fréquemment rencontrés (dits mots-outils) ».

Ces deux items figurent dans les programmes mais nécessitent que l'on précise ce que sont ces différents types de mots.

- Le mot-outil s'oppose au mot plein (ou mot lexical) : le sens de ce mot est aussi important que son rôle syntaxique. Les mots-outils regroupent des catégories grammaticales variables (déterminant, pronom) et invariables (mots de liaison coordonnants et subordonnants).

-Le mot fréquent est un mot que les élèves rencontrent fréquemment dans les textes qu'ils lisent. Une liste rassemblant près de 1500 mots, les plus fréquents de la langue française, a été constituée par le lexicologue Étienne Brunet. Elle rend compte « de la langue que lisent les élèves francophones. Ces mots, extraits de textes littéraires ou non, ont été ramenés à leur base lexicale. Cette liste de fréquence lexicographique permet aux enseignants, à partir du corpus de vocabulaire de base du français, de bâtir des progressions qui répondent aux objectifs d'apprentissage fixés pour leurs élèves ». (Source : Eduscol. Listes de fréquences lexicales).

Nous reproduisons ci-dessous les 45 mots les plus fréquents⁹³ (le mot est donné suivi de sa catégorie grammaticale et de sa fréquence d'occurrence dans les textes) : ils intègrent les mots-outils précédemment définis :

⁹³ Table hiérarchique (Eduscol)

Le	dét.1050561		Son	dét. 181161		De	dét.119106		Plus	adv. 75499		Avec	prép. 55081
De	prép.862100		Que	conj.176161		Ce	pron.107074		Dire	verbe 72134		Tout	adj.47221
Un	dét.419564		Se	pron.168684		Le	pron. 105873		Me	pron.71086		Y	pron. 46031
Être	verbe 351960		Qui	pron. 148392		Pour	prép.104779		On	pron. 70246		Aller	verbe 41702
Et	conj. 362093		Ce	dét. 141389		Pas	adv.103083		Mon	dét.70121		Voir	verbe 39659
À	prép. 293083		Dans	prép.139185		Que	pron. 99412		Lui	pron. 65988			
Il	pron. 270395		En	prép. 143565		Vous	pron. 89623		Nous	pron.62554			
Avoir	verbe 248488		Du	dét.127384		Par	prép. 82277		Comme	conj.59902			
Ne	adv. 186755		Elle	pron. 126397		Sur	prép. 80180		Mais	conj. 57690			
Je	pron. 184186		Au	dét. 123502		Faire	verbe 77608		Pouvoir	verbe 55394			

Tableau 19 : Liste des 45 mots les plus fréquents de la langue française écrite (avec mention de leur catégorie grammaticale et de leur fréquence d'occurrence)

Ce type de ressource attire quelques commentaires en ce qui concerne l'identification des mots en lecture :

- si l'on met de côté les verbes en raison des variations liées aux désinences, on constate que tous ces mots sont des mots monosyllabiques. Les reconnaître par la voie directe n'est donc pas véritablement dissociable de les reconnaître par voie indirecte (décodage).

-aucun substantif ne figure dans cette partie de la liste. Ces mots (verbes présentés à l'infinitif mis à part) ne prennent de sens que dans l'interaction avec d'autres mots ce qui est fortement réaffirmé par Lehman (2011) dans le document d'accompagnement des programmes portant sur le vocabulaire et son enseignement : « les mots entretiennent entre eux des relations de sens et des relations de formes et sont liés entre eux par des relations syntaxiques. On ne saurait donc étudier un mot de façon isolée ».

-les travaux de Brunet sont des travaux statistiques qui ont permis à l'auteur d'indiquer qu'environ 90% du corpus (90 millions d'occurrences décrites dans le Trésor de la langue française⁹⁴ et de la banque Frantext, représentant 70 000 vocables) est couvert par 907 mots dits hyper fréquents (i.e. d'une fréquence supérieure à 7000 cf. Picoche et Rolland, 2002, p. 8). Ces mots constituent selon Picoche et Rolland (idem) « le noyau dur du lexique français ». Ces travaux établissent aussi que 8% du corpus est couvert par environ 5 800 mots (de fréquence inférieure à 7000 et supérieure à 500) et que 64 000 mots permettent de couvrir les 2% restants.

Picoche et Rolland (2002) ont poursuivi ce travail pour établir un dictionnaire du français usuel : des recoupements ont été effectués (rapprochement d'antonymes, de paronymes, de mots ayant entre eux une relation de réciprocité etc.) leur permettant d'établir une liste de mots, dits eux aussi, hyper fréquents desquels ont été isolés des tables de mots pour chaque niveau d'enseignement primaire sélectionné en fonction de leur « productivité en termes linguistique » sur le site Vocanet (une quarantaine de mots pour le CP).

-la base Manulex (Lété, Sprenger-Charolles, Colé ; 2004) est une base de données lexicales qui fournit les fréquences d'occurrences de 23.900 mots (lemmes) et 48.900 formes orthographiques extraits d'un corpus de 54 manuels scolaires de lecture de laquelle, indiquent les auteurs, pour l'enseignement du vocabulaire, les enseignants du primaire

⁹⁴ Ce dictionnaire existe en version informatisé (TLFi) à l'adresse suivante : <http://www.atilf.fr/spip.php?rubrique77>

pourront extraire des listes de mots adaptées aux enfants de cycle 2 et 3. La fréquence des mots est alors établie en fonction d'un corpus de mots que les élèves sont les plus à mêmes de rencontrer lorsqu'ils sont confrontés à un texte à lire.

Pour résumer, les mots de la langue les plus fréquents (i.e. les plus fréquemment lus par les élèves) sont des mots outils dont nous avons souligné qu'ils étaient monosyllabiques.

L'item « lire aisément les mots les plus fréquemment rencontrés (mots-outils) » porte donc sur la capacité de l'élève de CP à identifier rapidement ces mots par la voie orthographique.

L'autre item évoque des mots étudiés. Cela renvoie à des mots qui ne figureraient pas dans la liste précédente mais que les élèves rencontrent souvent au cours de l'apprentissage. Par exemple, le mot « maman » que les élèves rencontrent très tôt. Ayant été très souvent rencontré, l'élève est susceptible de l'identifier directement sans passer par le décodage.

Évaluer ces deux items en même temps semble un choix judicieux : on souhaite évaluer le lexique orthographique de l'élève.

b- l'épreuve :

Pour évaluer ces items, l'enseignant utilise l'épreuve suivante :

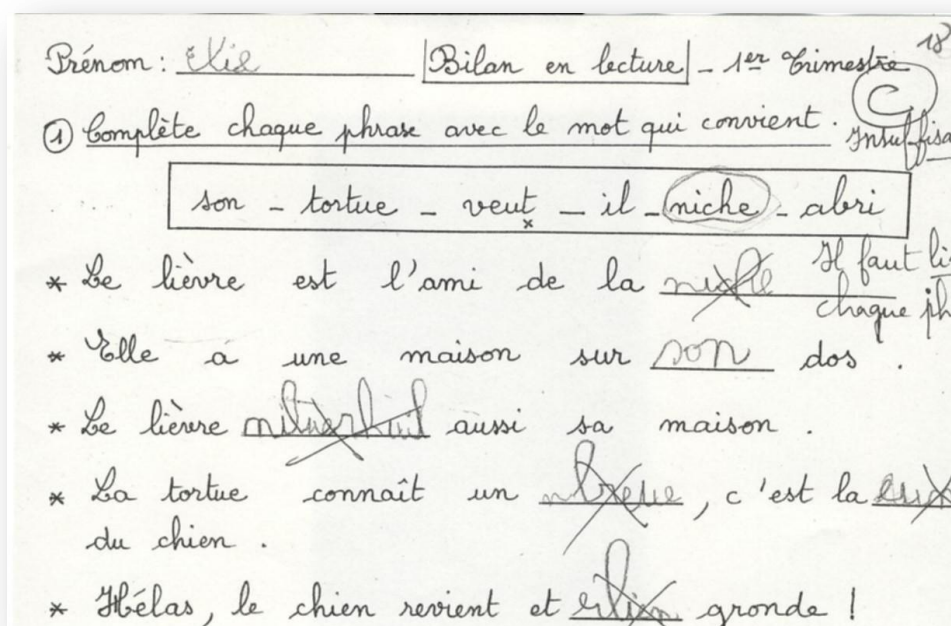


Figure 26 : Épreuve FT1 (Annexe B, p. 40) avec réponses d'élève

Le premier constat, évident, est l'écart entre les mots figurant dans la tâche, et les mots fréquents précédemment listés. L'épreuve relève d'un très haut niveau d'exigence (la variété des degrés d'exigence des enseignants dans les évaluations internes était déjà pointée par Grisay, 1989, 1992).

Pour réussir cette épreuve, l'élève doit d'abord lire une liste de mots présentés dans une police cursive manuscrite. Cette police est celle que l'on utilise pour écrire mais pas pour lire : l'enseignant nous a indiqué avoir donné ces mots sous cette forme pour faciliter le fait

de les recopier ensuite : une tâche de copie est donc implicitement contenu dans la tâche, elle constitue une partie implicite du référent de l'enseignant pour évaluer.

L'élève doit conserver en mémoire le sens des mots candidats (« son », « tortue » etc.) qu'il a reconnu directement ou déchiffré- ce qui n'a pas le même coût cognitif (« abri » est ainsi un mot difficile à déchiffrer pour les élèves en raison d'une structure syllabique complexe). Il doit ensuite lire chaque phrase. Les mêmes remarques peuvent être faites. Parvenu à l'espace laissé vacant, il doit alors réactiver la mémorisation effectuée (ou non) des mots candidats, les relire si besoin jusqu'à trouver le mot « qui convient » i.e. « dont l'adjonction à la phrase lacunaire permet de bâtir la phrase cible ». Enfin, il doit recopier le mot choisi. Une tâche multiple, complexe à ce titre, d'un coût cognitif fort élevé.

c- la copie de l'élève :

L'élève dont nous avons recueilli la copie a manifestement largement échoué sans que cela ne puisse renseigner l'enseignant sur la nature des erreurs. Il peut tout juste constater que c'est « insuffisant » et engager l'élève à « lire chaque phrase ».

La validité psycho-didactique de l'épreuve est remise en cause. L'épreuve nous semble fort éloignée de l'item que l'enseignant souhaite évaluer. Elle n'évalue pas seulement le lexique orthographique de l'élève, elle évalue également des compétences de lecture (lecture de phrases, réalisation d'inférences, choix de mots) et d'écriture (copie) qui ne sont pas des compétences encore acquises par les élèves.

5.4. Dessiner en guise de réponse

Les paragraphes précédents ont permis de considérer que le système de réponse inclus dans la tâche posait un certain nombre de problèmes surtout lorsqu'il imposait la production écrite d'une réponse ou la détermination de la bonne réponse parmi un ensemble de modalités de réponse. Une des solutions imaginées par les enseignants pour éviter le recours à l'écrit dans les réponses consiste à demander aux élèves de produire leurs réponses en recourant au dessin.

5.4-1 Le dessin en lecture

Épreuve n°23 : « Tortue »

a- l'item :

Au premier trimestre, l'enseignant A souhaite évaluer l'item de son livret « Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens ».

b- l'épreuve :

Pour évaluer cet item, il prescrit la tâche suivante :

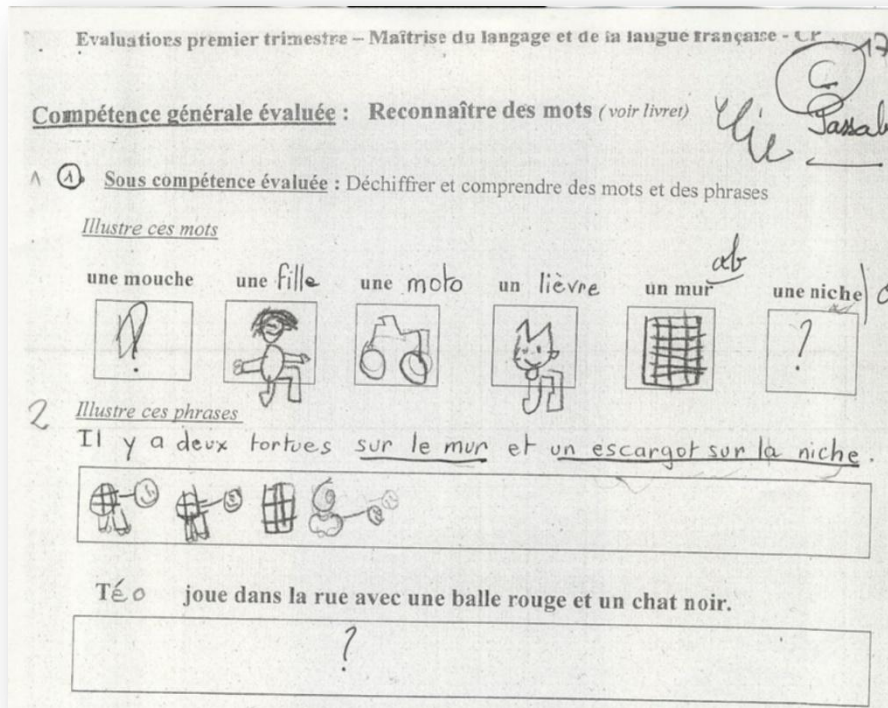


Figure 27 : Épreuve FT1 (Annexe B, p. 65) avec réponses d'élève

Un écart entre l'item du livret évalué, la « compétence générale évaluée - reconnaître des mots » et la « sous compétence évaluée - déchiffrer comprendre des mots et des phrases » est observable : ce que l'enseignant désigne par « compétence générale » est en réalité au service de ce qu'il désigne par « sous-compétence ».

Pour réussir cette épreuve, l'élève doit illustrer. Prenons la tâche dans laquelle il doit illustrer les phrases : il est attendu que l'élève dessine les éléments « une tortue » ET « une tortue » ET « un mur » ainsi que « un escargot » ET « la niche ». Ces dessins doivent en outre matérialiser la position relative des tortues par rapport au mur ainsi que la position relative de l'escargot par rapport à la niche. Ces positions relatives sont indiquées dans la phrase par le mot « sur ». On voit donc que l'élève ne peut se contenter de dessiner chaque élément successivement, il doit aussi les intégrer les uns aux autres.

c- la copie de l'élève:

L'élève dont nous avons recueilli la copie a manifestement échoué selon le jugement évaluatif de l'enseignant visible en haut à droite de la copie.

En ce qui concerne la première tâche, il n'a pas dessiné « une mouche » ni « une niche », ce que l'enseignant considère trivial si l'on se réfère au commentaire porté dans la marge : « oh ! ». L'attribution d'une valeur correcte aux dessins de « lièvre » et de « mur » renvoie à la part d'interprétation qui reste à la charge de l'évaluateur dans ce type d'épreuve.

En ce qui concerne la deuxième tâche, il est intéressant de noter que l'élève dessine bien « une tortue » ET « une tortue » ET « un mur » ET « un escargot ». En revanche, il ne traduit pas la position des tortues sur le mur (ni celle de l'escargot sur la niche mais pour une autre raison : il n'a pas dessiné la niche). Ceci laisse à penser que l'élève procède à une lecture linéaire, discrète, fondée sur l'utilisation de microprocessus d'identification de mots. Il ne

mobilise pas la lecture par groupes de mots. Il ne semble pas rechercher à établir une cohérence locale et reste à un niveau de surface du traitement des mots, lesquels sont correctement lus. En ce sens, cette tâche est tout à fait intéressante pour évaluer la compréhension en début d'apprentissage : elle permet d'apprécier que l'élève prend appui (ou pas encore) sur des groupes de mots pour construire du sens. Pour dessiner « sur un mur » après avoir dessiné les deux tortues, l'élève a-t-il, se donne-t-il, par exemple le droit de sortir du cadre pour produire sa réponse ? Il aurait fallu que l'élève dessine d'abord le mur, ce qui souligne une compétence de lecteur assez experte : la phrase est en effet longue.

Cette tâche illustre bien que « lire n'est pas que décoder » (J. Fijalkow, 2000, p. 133) même dans des situations *a priori* conçue pour cela.

La tâche prescrite (illustrer) oriente l'activité de l'élève vers l'utilisation d'une procédure linéaire de traitement des éléments de la phrase. Rien ne permet d'affirmer qu'il a compris ce qu'il a lu, ni le contraire. Le fait que l'élève n'a pas pu illustrer la deuxième phrase ne permet pas de conjecturer la nature de l'empêchement (obstacle lié à l'usage du dessin ou obstacle de déchiffrage ou obstacle de compréhension voire manque de temps ou de capacité à « dessiner » un mot). Le point d'interrogation porté par l'enseignant sur la copie de l'élève témoigne de cela.

De ce fait, la validité psycho-didactique de l'épreuve est discutable : elle mobilise des compétences supérieurement expertes à la seule reconnaissance ou à la capacité à déchiffrer des mots.

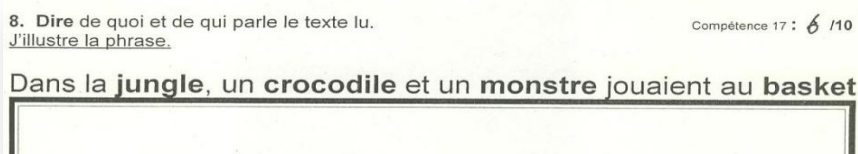
Épreuve n°24 : « Jungle »

a- l'item :

Au premier trimestre, l'enseignant D souhaite évaluer l'item de son livret « Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens ».

b- l'épreuve :

Pour évaluer cet item, il prescrit la tâche suivante :



8. Dire de quoi et de qui parle le texte lu. J'illustre la phrase. Compétence 17 : 6 /10
Dans la **jungle**, un **crocodile** et un **monstre** jouaient au **basket**.

Figure 28 : Épreuve FT1 (Annexe B, p. 69)

Cette tâche pose très clairement le problème de la réponse attendue. Nous n'avons malheureusement pas de trace d'activité d'élève, mais ce qui est demandé contient de nombreux obstacles que l'on peut essayer d'anticiper :

- Le premier obstacle est lié au temps verbal utilisé : comment représenter une scène qui s'est déroulée dans le passé ? Rien ne dit que si les animaux jouaient, ils jouent encore. On retrouve de tels obstacles dans d'autres épreuves. Par exemple dans l'épreuve p. 90 du volume d'annexe B, l'enseignant P demande aux élèves de

dessiner « Milo a avalé une pomme », « Le chat a léché le sirop », « Léa a avalé sa soupe ».

- Le deuxième obstacle est lié aux représentations couramment admises des différents éléments (jungle, crocodile, monstre). Le manuel de lecture utilisé mobilise une certaine représentation (de la jungle) dont il est implicitement suggéré que l'élève fasse référence dans son dessin (en dessinant des palmiers par exemple). Or la jungle est un terme qui n'a pas de signification biogéographique précise, qui désigne bien souvent la forêt tropicale humide et dont la densité exclut plutôt les palmiers.
- Le troisième obstacle est lié à la syntaxe de la phrase. Le déplacement du complément de lieu en fin de phrase rendrait la tâche plus simple : l'élève pourrait dessiner tous les éléments dans l'ordre. Placé en en tête, comment ferait-il pour dessiner, par-dessus le dessin de la jungle (on peut l'imaginer relativement touffue), « le crocodile et le monstre qui jouaient au basket » ? Cette remarque est à rapprocher de la remarque faite au sujet de l'illustration des phrases de l'épreuve précédente.

A moins de questionner ensuite l'élève sur le dessin produit (auquel cas lui demander de dire ce qu'il a compris de la phrase serait plus simple), la tâche prescrite ne peut être considérée *a priori* comme valide du fait des nombreux obstacles précités.

5.4-2. Le dessin en mathématiques

Il est très rarement prescrit aux élèves de dessiner en mathématiques.

Deux cas peuvent toutefois être isolés :

1/ matérialiser une collection au cours de tâches de dénombrement

Dans la mesure où ces prescriptions sont des épiphénomènes nous ne nous y attarderons pas. Au premier trimestre une seule épreuve⁹⁵ (utilisée par deux enseignants différents) demande aux élèves de compléter, par le dessin de ronds, une collection de 10 ronds dont 4 sont déjà matérialisés. Au troisième trimestre, une seule épreuve⁹⁶ (utilisée par deux enseignants différents) demande aux élèves de dessiner « 47 billes en faisant des groupements par 10 ». Cette demande, assortie d'une demande de groupement par 10, oriente la tâche vers une organisation de la collection qui serait justement très intéressante à observer.

2/ laisser des traces de l'activité de résolution de problèmes.

Il est également rarement proposé, ou suggéré, aux élèves de recourir au dessin au cours de la résolution de problèmes. Nous trouvons trace de quelques épreuves que nous reproduisons. Dans les deux premières (tâches « Bonbons » et « Fruits ») le dessin est prescrit : il doit être un élément de la réponse. Dans la troisième épreuve en revanche (tâche « Zoé »), il est laissé au libre choix de l'élève de dessiner ou pas.

⁹⁵ Epreuves MT1 p.162 et p. 169 du volume d'annexe B.

⁹⁶ Epreuves MT3 p.228 et p. 236 du volume d'annexe B.

Épreuve n°25 : « Bonbons »

a- l'item :

Au premier trimestre, l'enseignant A souhaite évaluer l'item de son livret « Résoudre un problème simple à une opération ». Pour ce faire, il utilise l'épreuve que nous reproduisons ci-dessous.

b- l'épreuve :

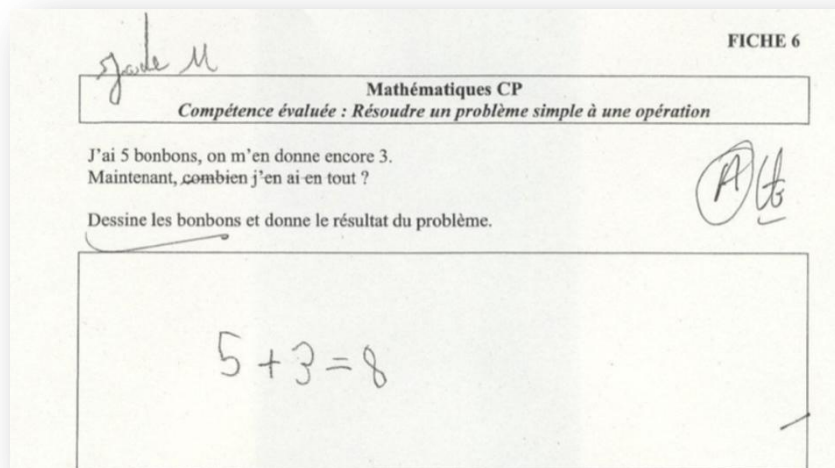


Figure 29 : Épreuve MT1 (Annexe B, p. 220) avec réponse d'élève

Le problème présenté est un problème additif avec accroissement de la valeur initiale. Plusieurs stratégies sont utilisables par les élèves qui vont du calcul au dénombrement en passant par le sur-comptage. La consigne de la tâche nous semble de nature à favoriser la deuxième voire la troisième.

Il est prescrit à l'élève de « dessiner les bonbons » puis de « donner le résultat du problème ». Si l'on essaie d'appliquer stricto-sensu cette consigne, on dessine 5 bonbons puis encore 3. En fin de premier trimestre, on peut concevoir comme probable qu'un élève dénombre les éléments de la collection représentée pour donner le résultat. Il produirait alors la suite orale « un ... huit » et traduirait le dernier nombre prononcé « huit » dans son écriture chiffrée : « 8 ». On peut aussi concevoir que l'élève utilise une stratégie de sur-comptage à mesure qu'il dessine les 3 éléments supplémentaires à la première quantité donnée 5 : « six (dessin du 6^{ième} bonbon), sept (dessin du 7^{ième} bonbon), huit (dessin du 8^{ième} bonbon) ».

c- la copie de l'élève :

La réponse de l'élève dont nous avons recueilli la copie satisfait les exigences de l'enseignant indiquée par une appréciation doublement positive « A » et « tb » (très bien). Pour autant, l'élève n'a pas dessiné, ce que l'enseignant a souligné dans le texte du problème, et, en guise de réponse l'élève a fourni une somme en ligne qu'il a calculée. Il n'a pas non plus interprété le résultat qu'il a produit : il n'a pas explicitement indiqué dans quelle « unité » s'exprimait la mesure déterminée par le calcul de $5+3$. De sorte qu'il est tout à fait concevable que l'élève ait additionné les deux valeurs numériques de l'énoncé sans plus se soucier du contexte dans lequel elles étaient fournies. Il fournit donc une réponse numérique parfaitement juste mais que l'on ne peut assimiler à la réponse du problème posé

laquelle implique d'inférer de la demande « combien j'en ai en tout ? » qu'il s'agit de bonbons. La réussite de l'élève est attestée sur la base d'inférences qui sont à la charge de l'évaluateur.

En conclusion, la validité de l'épreuve est discutable : le dessin est requis comme devant être un élément de la réponse. Nous avons indiqué qu'il nous semblait que cette demande était susceptible d'engager l'élève à utiliser une stratégie de dénombrement plutôt qu'une autre, plus experte comme le calcul par exemple. L'élève dont nous avons reproduit la copie n'a pas produit de dessin. Il a cependant utilisé une stratégie efficace de calcul. Le respect du contrat (« dessiner ») ne semble pas compatible avec la stratégie qu'il a utilisée et qui est une stratégie experte tout à fait adaptée à cette classe de problèmes, ce que l'enseignant reconnaît implicitement en validant la réponse donnée par l'élève (avec la réserve que nous avons formulée liée à la part d'interprétation du résultat qui reste à la charge de l'évaluateur).

Épreuve n°26 : « Fruits »

a- l'item :

L'enseignant F utilise l'épreuve ci-dessous au troisième trimestre pour évaluer l'item de son livret « Résoudre un problème simple à une opération ».

b- l'épreuve :

8. Résous le problème suivant :

Au marché, maman a acheté 5 roses, 10 pommes, 4 bananes et 3 poires.
Combien maman a-t-elle acheté de fruits ?
Dessine les fruits puis écris une opération.

Maman a acheté 17 fruits.

13
+ 4
—
17

10 pommes, 4 bananes et 3 poires dessinées à la main.

Figure 30 : Épreuve MT3 (Annexe B, p. 314) avec réponse d'élève

Le problème présente une donnée « surabondante » (5 roses) ; l'élève doit ainsi effectuer, lors de l'extraction des données pertinentes, une tâche de catégorisation : il faut qu'il reconnaisse les différents éléments de l'énoncé comme appartenant, ou n'appartenant pas, à la catégorie des fruits. Il est cependant peu probable que les élèves, en fin de CP, échouent à cette tâche de catégorisation, tous les éléments pouvant être considérés comme familiers. En revanche la donnée surabondante peut intervenir lors de l'élaboration du modèle de situation et alourdir la charge de la mémoire de travail ; Certains élèves peuvent être tentés, par un effet de contrat, d'utiliser tous les nombres figurant dans l'énoncé. Cela constitue des sources de difficultés potentielles qui se posent avec tous les problèmes à donnée(s) surabondante(s), problèmes que l'on hésite d'ailleurs généralement à considérer comme « simples ».

Cela étant dit, si l'on fait abstraction de la donnée surabondante ainsi que de la prescription de dessiner les fruits, le problème se présente comme un problème classique de composition de mesures (Vergnaud, 1981)⁹⁷. Cela s'inscrit en cohérence avec l'item auquel l'épreuve est référée et la tâche attendue par l'enseignant serait alors une modélisation mathématique de la situation par une opération additive. Tout élève qui traiterai directement le problème comme tel, serait ainsi amené à écrire l'opération $10+4+3=17$ (en ligne, voire même en colonne en fin de CP), en calculant $10+4$: « 14 », $14+3$: « 17 », et à conclure « Maman a acheté 17 fruits ».

Quel effet pourrait avoir la nécessité de produire en premier lieu un dessin comme le prescrit la consigne ? L'élève qui procéderait ainsi « transformerait » le texte, dans ce premier temps, sous une forme (le dessin produit) qui lui permettrait de se focaliser sur l'information critique et « d'oublier » l'information surabondante, ce qui pourrait avoir un effet bénéfique sur la résolution. Mais en revanche, le risque serait grand que, ce faisant, pour répondre à la question posée « combien maman a-t-elle acheté de fruits ? » l'élève réalise un simple dénombrement de la collection représentée dans son dessin, manifestant ainsi une redéfinition de la tâche en une tâche de dénombrement.

c- la copie de l'élève :

L'élève dont nous avons recueilli la copie a dessiné tous les fruits conformément à la prescription de l'énoncé. Mais il est intéressant de remarquer qu'il n'y a pas congruence entre l'ordre des données de l'énoncé (les pommes, les bananes et enfin les poires) et celui selon lequel elles apparaissent dans son dessin : une lecture de celui-ci de gauche à droite fait en effet apparaître, dans l'ordre, les pommes puis les poires, enfin les bananes⁹⁸. L'apparition du nombre « 13 » dans l'opération posée par l'élève (« $13+4=17$ ») indiquerait alors que, selon toute vraisemblance et comme nous le mentionnions plus haut, l'élève pour répondre s'est appuyé sur son dessin (en déterminant en premier le nombre d'éléments « pommes, poires ») et non directement sur l'énoncé. Ce faisant, il a pu obtenir « 13 » par le calcul (« $10+3=13$ ») ou, le dessin l'y engageant, par un comptage ou un sur-comptage en énumérant « 10, 11, 12, 13 ». Il a pu réitérer la procédure avec les 4 bananes à partir de 13 : « 14, 15, 16, 17 » et enfin reporter le dernier mot-nombre prononcé dans la phrase réponse. L'opération pourrait alors avoir été posée par cet élève a posteriori du processus de résolution du problème afin de respecter toutes les prescriptions de l'énoncé, parmi lesquelles figure « écris une opération ». Il est ainsi possible que l'élève ait utilisé une procédure de dénombrement pour résoudre le problème et reconstruit a posteriori une réponse qui convienne à la demande (ce qui serait très expert).

En définitive, si la tâche attendue par l'enseignant était la résolution d'un problème à données surabondantes (comme peut le faire supposer l'énoncé du problème-hors consigne-), l'exigence d'un dessin serait susceptible de transformer la tâche attendue en une tâche de résolution d'un problème « simple » de composition de mesures. Mais la présence de ce même dessin serait susceptible de conduire à une redéfinition de la tâche par l'élève en une tâche de dénombrement. Ainsi, quelle que soit l'attente de l'enseignant, il

⁹⁷ Ce type de problème additif est appelé « problème de combinaison » dans la classification de Riley et coll. (1983), voisine de celle de Vergnaud. Il s'agit ici de trouver le cardinal d'un ensemble connaissant les cardinaux des parties.

⁹⁸ Il est vraisemblable que l'élève, après avoir dessiné les pommes puis les bananes, ait dessiné les poires dans le seul espace disponible sur la feuille, c'est à dire à gauche des bananes.

y a tout lieu de penser que pour certains élèves, la tâche redéfinie sera une tâche de dénombrement.

Épreuve n°27 : « Zoé »

a- l'item :

Au troisième trimestre, l'enseignant F souhaite évaluer l'item de son livret « Résoudre un problème simple à une opération ». Pour ce faire il utilise trois épreuves. Nous en reproduisons une ci-dessous.

b- l'épreuve :

Zoé a reçu 25€ pour son anniversaire. Elle s'achète un livre qui coûte 12€.
Combien lui reste-t-il d'euros ?

Il lui reste ...13... euros.

Figure 31 : Épreuve MT3 (Annexe B, p. 310) avec réponses d'élève

Il s'agit d'un problème additif avec transformation négative de l'état initial. L'énoncé est lu par l'enseignant. Pour répondre, l'élève doit inscrire dans une phrase lacunaire une valeur chiffrée (« 13 ») : la phrase réponse prend en charge l'expression de l'unité (« euros »). Contrairement aux deux épreuves précédemment analysées (« Fruits » et « Bonbons »), il ne lui est pas demandé de laisser trace des procédures employées. De cette façon, l'élève n'est pas orienté vers une stratégie plutôt qu'une autre : il peut résoudre le problème en utilisant une stratégie de dénombrement ou de calcul.

c- la copie de l'élève :

L'élève dont nous avons recueilli la copie a produit la réponse attendue. L'épreuve n'exige pas de dessin mais l'élève y a eu recours : la zone de réponse est suffisamment grande pour le permettre. Il a matérialisé la mesure initiale sous forme de carrés (un par unité) et de barres (une par dizaine). Il s'est ensuite probablement servi du dessin pour matérialiser la mesure à soustraire (deux carrés et une barre) : la différence s'obtenant ensuite par le calcul du reste (trois carrés et une barre : 13). Il semble relever du contrat que l'élève produise une opération en ligne avant de reporter le résultat dans une phrase réponse dont il ne manque que la donnée numérique, de sorte que c'est le cadre de réponse qui prend en charge le sens de l'opération demandée (en reliant la valeur numérique et l'unité dans laquelle elle s'exprime).

Si l'élève a parfaitement bien modélisé la situation du point de vue mathématique il n'est pas certain qu'il sache ensuite interpréter le résultat. Rien ne permet d'être sûr qu'il sache que « 13 » est une mesure qui s'exprime en « euros ».

La réponse produite par le dessin ou le calcul n'est qu'un élément de la réponse qui doit ensuite être rédigée en associant la valeur numérique obtenue et l'unité. Dans cette épreuve cette liaison est prise en charge par le cadre de réponse. Le cadre de réponse prend en charge le sens du résultat ce qui est de nature à invalider l'épreuve du point de vue psychodidactique.

Discussion :

Dans les tâches de résolution de problèmes « simples à une opération », la prescription du dessin est plutôt de nature à invalider les épreuves en ce sens qu'elle risque de les transformer en des tâches de dénombrement, qu'elles peuvent conduire les élèves à devoir produire un dessin alors qu'ils ont utilisé une procédure plus experte (le calcul par exemple). L'épreuve « Zoé » laisse à l'élève la possibilité de dessiner. Mais dans ce dernier cas, la présence d'une phrase à compléter renvoie à l'interprétation du résultat numérique : celle-ci est alors prise en charge par le cadre de réponse qui lie la donnée numérique trouvée et l'unité dans laquelle elle s'exprime.

En français comme en mathématiques, la prescription du dessin pour manifester de la compréhension ou la solution d'un problème est susceptible d'orienter l'élève vers l'utilisation de procédures linéaires dans le traitement des données (selon l'ordre dans lequel elles sont présentées en lecture, ou de procédures de dénombrement d'éléments figurés (plutôt que de calcul par exemple)).

La prescription de dessiner pour répondre soulage l'élève de la rédaction de la réponse. Toutefois un certain nombre de nouvelles critiques sont susceptibles d'être formulées à ce type de cadre de réponse. Parmi elles, citons : la part d'interprétation laissée à l'enseignant qui doit inférer du dessin la réussite de l'élève, le fait que le dessin est souvent cumulatif et peu intégrateur (cf. épreuves de compréhension en lecture), ou bien encore que le dessin risque bien de privilégier l'utilisation de certaines procédures dont par ailleurs il n'est pas certain qu'elles soient les plus expertes pour traiter les situations données (cf. dénombrer vs calculer).

Lors de la résolution de problèmes en mathématiques, certains cadres de réponse exigent la production d'un dessin pour justifier l'obtention d'une donnée numérique. Lorsque ces mêmes cadres de réponse demandent ensuite de compléter une phrase réponse dont il ne manque que la donnée numérique, le reproche leur est adressé de prendre en charge le sens du résultat.

5.5. Lire à haute voix

Dans de très rares cas, les enseignants ont recours à des passations individuelles, orales. Cette modalité très intéressante représente une alternative aux modalités précédemment étudiées. Elle évite tout recours à l'écrit pour l'expression de la solution (rédaction ou dessin) et évite également le fait de sélectionner la bonne réponse parmi un ensemble de propositions.

Elle n'est toutefois jamais employée pour évaluer la compréhension en lecture ni aucun autre item en mathématiques, mais uniquement pour évaluer les items du programme suivants correspondants à la maîtrise du code graphophonologique :

- « Lire aisément les mots étudiés »,
- « Déchiffrer des mots réguliers inconnus »,

- « Lire aisément les mots les plus fréquemment rencontrés (dits mots-outils⁹⁹) » et
- « Lire à haute voix un texte court dont les mots ont été étudiés ».

Pour cela, les enseignants utilisent des tâches de lecture à haute voix « traditionnelles » (cf. E. Fijalkow, 2003) qui renvoient au « savoir-déchiffrer ». Cela correspond assez bien au large consensus établi autour du rôle central que joue la reconnaissance visuelle rapide et automatique de mots dans la capacité de lecture (Lecocq, Casalis, Leuwers et coll., 1996). La compréhension des textes écrits dépend pour partie du degré d'automatisation des processus d'identification des mots écrits (Goigoux, 2000a, p. 235 ; Martinet et Rieben, 2010, p. 222).

La question que nous nous posons dans ce paragraphe est celle des supports sur lesquels il est demandé à l'élève d'exercer ses compétences de décodeur (ou plutôt de recodage phonologique puisqu'il est attendu qu'il oralise le décodage). Un élève « bon décodeur » est généralement, selon Martinet et Rieben (2010, p. 211), « un enfant qui a automatisé les différentes procédures d'identification des mots¹⁰⁰ et qui se révèle capable de décoder rapidement et sans erreurs le texte qu'il a en face de lui. » Les auteurs de cette définition ont choisi d'évoquer le terme de texte pour qualifier ce sur quoi s'exerce la compétence de l'élève. Aussi prennent-ils la précaution d'indiquer qu'« un bon décodage n'assure cependant pas que le rythme de lecture soit équilibré et que la prosodie soit adaptée » (idem). Selon ces auteurs, on devrait ainsi différencier un « lecteur fluide » (qui correspond à une lecture fluide du texte) d'un « bon décodeur » (qui serait plus associé à la capacité à lire correctement des mots isolés) (idem).

La fluence en lecture est définie comme la capacité à lire avec aisance, rapidement, sans erreurs et avec une intonation adaptée. L'équipe de Zorman a élaboré un protocole d'évaluation de la lecture en fluence avec le Laboratoire des Sciences de l'éducation Université Pierre Mendès France Grenoble (juin 2008). En deçà d'un score de 40 mots lus correctement par minute (MCLM¹⁰¹) au mois de juin en CP, les élèves « doivent participer aux ateliers de fluence ». Cette valeur constitue de fait une sorte de norme. Le test E.L.FE¹⁰² des mêmes auteurs établit le MLCM moyen à 68 en CE1. La prise en compte du lien entre vitesse et compréhension reste significative dans deux cas minoritaires et opposés dont celui du CP affirme Mesnager¹⁰³ : « pour nombre d'élèves de fin de CP, ou encore faibles en CE2, la lenteur est corrélée à une mauvaise compréhension parce que, ne possédant que peu de mots identifiables « directement », ils déchiffrent la plupart de ceux qu'ils rencontrent et ne peuvent, trop enlisés sur chaque mot, construire de façon efficace des ensembles de sens. ». D'après Giasson (2012, p. 331), 28, 53 et 82 MLCM seraient respectivement les valeurs du premier quartile, de la médiane et du troisième quartile au printemps de la première année d'apprentissage de la lecture.

⁹⁹ Nous avons discuté dans le paragraphe 5.3 de la distinction mot-outil / mot-fréquent. Les 50 mots les plus fréquemment rencontrés par un élève lorsqu'il lit sont principalement des mots-outils.

¹⁰⁰ Nous renvoyons à Chauveau (2007, p. 82-92) qui décrit, du point de vue du fonctionnement cognitif du lecteur, sept procédures pouvant être mobilisées par des élèves de CP pour identifier des mots écrits.

¹⁰¹ Le premier est un score, le MCLM (Mots Correctement Lus par Minute, cf. Zorman, 2008). Il correspond à une moyenne relative à la durée de lecture : (nombre de mots correctement lus/durée de la lecture en secondes) x 60 et peut être considéré comme un indicateur de la « vitesse de lecture ». Le nombre de mots correctement lu est le nombre de mots contenus dans le texte diminué du nombre total d'erreurs (y compris les omissions, ajouts et substitutions). Si l'élève hésite ou se reprend pour finalement lire correctement le mot, l'erreur n'est pas comptée ; il est parvenu à déchiffrer et sera pénalisé uniquement par le chronomètre.

¹⁰² Test Evaluation de la Lecture en Fluence (laboratoire Cognisciences- Grenoble)

¹⁰³ cf. Mesnager (non daté) article Vitesse de lecture, TFL

Bien évidemment, ces valeurs ne sont que des repères car le MCLM dépend du texte lu. Aussi, certains auteurs à l'instar de Giasson (2012, p.166) évoquent-ils des barèmes pour déterminer le niveau de lecture d'un texte se basant sur le pourcentage de mots reconnus par l'élève. Elle en distingue trois : le niveau d'autonomie (96% au moins de mots correctement reconnus), le niveau d'apprentissage (de 90% à 95%), et le niveau de frustration (moins de 90%), invitant à considérer que si le texte place l'élève dans un niveau de frustration, « il y aurait avantage à choisir un autre texte pour l'élève » (idem).

Selon Giasson, évaluer la lecture orale des lecteurs débutants vise « à vérifier si le niveau de difficulté du texte est approprié à l'élève et à procéder à l'analyse des méprises¹⁰⁴ afin d'observer les stratégies du lecteur débutant » (2012, p. 169).

Tous les enseignants qui constituent la population étudiée dans cette thèse mentionnent un item de lecture à voix haute dans les livrets qu'ils utilisent. Nous présentons l'analyse de 5 épreuves destinées à évaluer cette compétence.

Épreuve n°28 : « Fluence »

a- l'item :

Nous avons choisi de présenter parallèlement deux épreuves. Elles sont respectivement employées par l'enseignant D pour évaluer l'item du livret et du programme « Lire à haute voix un texte court dont les mots ont été étudiés » et par l'enseignant M pour évaluer l'item du livret « Lire un texte composé de mots connus ».

b- l'épreuve :

Dans les deux cas, les élèves doivent lire le texte support (une phrase pour l'une des épreuves, un texte pour l'autre).

¹⁰⁴ Giasson (2012, pp. 169-170) propose de considérer les substitutions de mots, les omissions, les ajouts, les mots fournis par l'enseignant et les autocorrections comme des méprises.

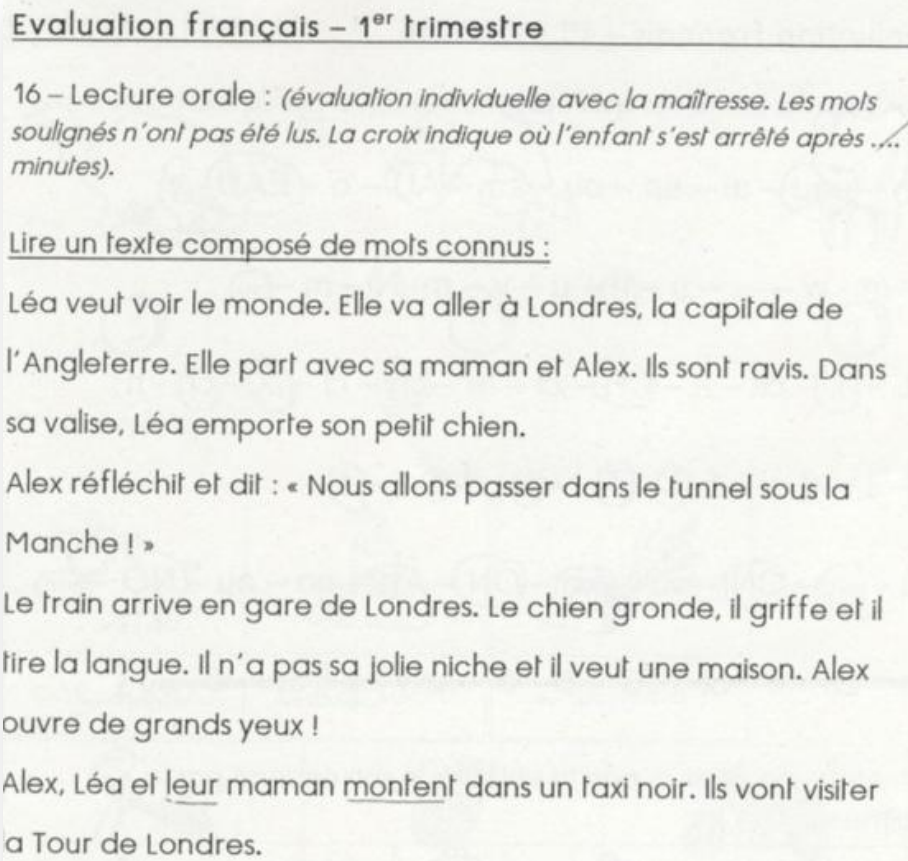


Figure 32 : Épreuves FT1 (Annexe B, p. 58 (à gauche) et p. 61 (à droite))

Le premier constat effectué est celui de la diversité des exigences des évaluateurs : l'élève placé dans la première situation est soumis à une épreuve de complexité assez peu comparable à celle de la deuxième épreuve de droite, ne serait-ce que par le volume de texte à lire (ceci renvoie à la diversité des exigences des enseignants décrite par Grisay (1992) et discutée en II-3.1.3)).

Dans les deux épreuves, les enseignants indiquent que les mots sont « connus » ce qui sous-entendrait qu'ils ont été déjà rencontrés (dans le manuel de lecture Grindelire CP (éd. Bordas) pour l'épreuve de droite) donc déjà déchiffrés, et qu'ils ont fait l'objet d'un travail lexical i.e. leur reconnaissance du point de vue graphophonologique et du point de vue sémantique serait assurée. La tâche comporte une part de re-connaissance, de re-lecture. Dans la première épreuve, les mots « au », « bord » et « du » peuvent être connus séparément, ils ne garantissent pas la lecture fluente de l'expression « au bord du (fleuve) » qui fait appel à la compréhension du lien existant entre ces trois mots.

Dans la deuxième épreuve, le verbe « monte » (avant dernière phrase) peut être connu mais la lecture du verbe conjugué au pluriel « montent » pose de nouveaux problèmes qui requièrent des connaissances grammaticales hors de portée pour un lecteur débutant (i.e. 1er trimestre de CP). Le texte comporte en outre la quasi-totalité des archi-graphèmes de la langue (A, E, I, O, U, EU, OU, AN, IN, ON, UN, OI soit tous les archi-graphèmes voyelles et semi-voyelles sauf IL(L) ; et P, B, T, D, C, G, F, V, S, Z (intervocalique), X, CH, J, L, R, M, N, soit tous les archi-graphèmes consonnes sauf GN). Autrement dit, être capable de lire à haute voix ce texte requiert une compétence experte de décodage qui semble hors de portée des élèves dans le cas du second texte.

c- la copie de l'élève :

L'élève n'a pas su lire « leur » et « montent » dans la deuxième épreuve : ce sont deux erreurs prévisibles en fin de premier trimestre. Il est tout à fait surprenant d'ailleurs de constater l'absence d'autres erreurs. Ceci ne peut que renforcer l'hypothèse d'un texte déjà lu et connu (mémorisé).

La mise en parallèle de ces deux épreuves conduit à l'appréciation évidente que les évaluateurs n'évaluent pas véritablement la même chose bien que cela soit formulé de manière analogue.

Dans la première épreuve, le texte donné à lire est nouveau (i.e. jamais rencontré). Il est composé de mots « connus », mais dont l'assemblage est susceptible de ne pas l'être.

Dans le cas du texte de la deuxième épreuve, plus que les mots, c'est le texte même qui est déjà connu, auquel cas la mémoire de la compréhension du texte étudié entre en ligne de compte. Ce qui est évalué n'est donc pas seulement la lecture d'un texte (fut-il composé de mots connus) mais également la mémoire d'une compréhension antérieure.

De ce fait, la validité psycho-didactique de ces deux épreuves n'est pas assurée. D'un point de vue didactique, l'enjeu n'est pas véritablement de savoir si l'élève sait lire un texte composé de mots déjà connus mais davantage de savoir si l'élève sait lire des mots quels qu'ils soient (soit qu'il les identifie par la voie d'assemblage soit qu'il les reconnaît orthographiquement par la voie d'adressage soit encore qu'il mixe diverses stratégies (les deux précitées mais aussi des connaissances morphologiques pour lire le mot « montent » par exemple)). Sans doute l'intitulé de l'item choisi génère-t-il, pour ces enseignants, une confusion sur l'enjeu de l'évaluation.

Une épreuve retient particulièrement notre attention dans le corpus des épreuves destinées à évaluer l'item en question en ce sens qu'elle associe lecture à voix haute et compréhension. Nous la reproduisons ci-dessous :

Épreuve n°29 : « une tâche pour deux items »

a- l'item :

L'enseignant P utilise la même épreuve pour évaluer deux items différents de son livret au premier trimestre. Il s'agit des items « lire à haute voix un texte court, un dialogue » et « lire silencieusement un mot, une phrase, un texte et en manifester sa compréhension ».

b- l'épreuve :

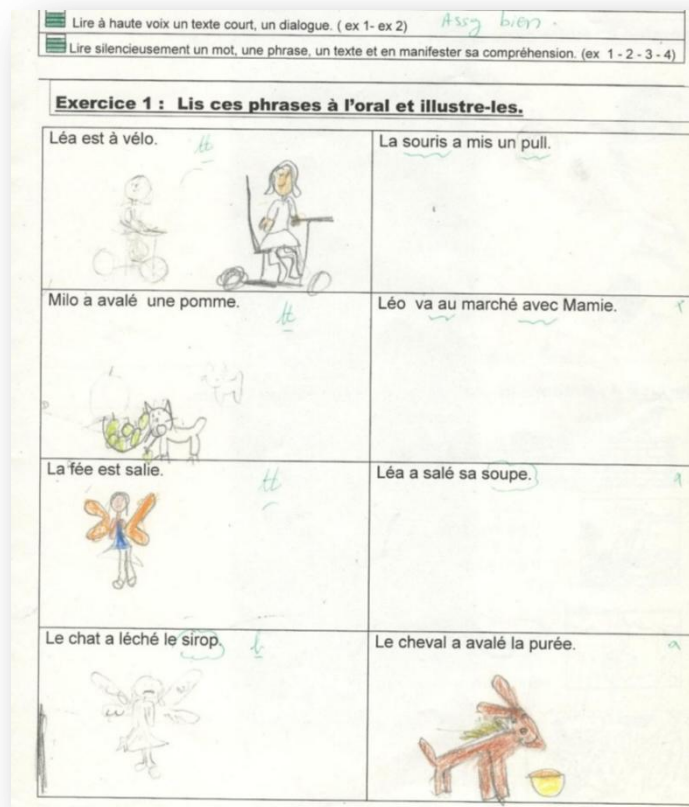


Figure 33 : Épreuve FT1 (Annexe B, p. 63) avec réponses d'élève

Dans cette épreuve l'élève doit tout d'abord « lire à l'oral » (passation individuelle, lecture à l'enseignant qui évalue ainsi le premier item) puis illustrer. Nous ne savons pas comment ces deux temps de l'épreuve s'articulent, l'organisation choisie laisse penser qu'un élève pourrait illustrer une phrase qu'il n'aurait pas su lire oralement.

c- la copie de l'élève :

L'élève a produit une illustration dans 5 cas sur 8, dont l'une est erronée.

Discussion :

Comment un élève pourrait illustrer quelque chose qu'il n'aurait pas su déchiffrer ? Ou, autrement dit, quel crédit accorder à un dessin correct de la situation représentée par la phrase si l'élève n'a pas su lire les mots qui la composent ?

Comment préjuger de la compréhension de l'élève de certaines phrases au moyen du seul dessin dont l'exécution n'est pas simple : comment dessiner « la fée est salie » ou « le chat a léché le sirop » ? (cf. analyse de la tâche n°24 « Jungle », III-5.4-1)

Cette tâche serait valide si elle était utilisée pour évaluer la lecture de phrases à voix haute (le premier item ciblé). Elle pourrait être valide pour évaluer la compréhension de phrases lues seul (si les inférences à la charge de l'évaluateur lors de l'analyse des réponses étaient explicitées par un entretien avec l'élève au sujet des dessins par exemple).

En revanche, elle n'apporte pas de preuves de validité pour évaluer les deux items en même temps : la réussite à un item conditionne la réussite au second et inversement puisque Lecocq et coll. (1996, p.93) indiquent que « les performances recueillies en compréhension

dans les deux modes de lecture (silencieuse et oralisée) sont équivalentes » dans les premières années de l'apprentissage de la lecture. Ce que l'on semble constater dans les réponses données par l'élève confirme ce résultat.

L'épreuve suivante fait mention d'un chronométrage. Nous en avons trouvé trace à différents endroits sur le web. Toutefois, à l'issue de nos requêtes nous n'en avons pas trouvé une utilisation telle que décrite ci-après. Le texte (parfois modifié comme dans les deux épreuves reproduites dans les figures 33 et 34) est utilisé comme support pour évaluer la compréhension d'un texte lu seul¹⁰⁵ et non la fluence (ou le décodage textuel).

Épreuve n°30 : « Le petit oiseau »

a- l'item :

Les enseignants A et N utilisent ces épreuves pour évaluer respectivement l'item du livret « Lire à haute voix un texte court dont les mots ont été étudiés, en articulant correctement et en respectant la ponctuation » et l'item du livret « Je sais lire de manière fluide et expressive »

b- l'épreuve :

Pour évaluer ces items, les enseignants utilisent les épreuves reproduites ci-dessous :

tu peux lire plus vite maintenant.

(C+) ab

Le petit oiseau.

Un petit oiseau jaune et vert est entré dans la maison. Il se pose, sur un fauteuil. Il se cogne contre les vitres de la fenêtre.

Papa ouvre la porte. Le petit oiseau s'envole. Il passe au-dessus d'un grand marronnier et disparaît.

gache M

Lecture individuelle à haute voix: Note: ___ / 18

Aide	Temps	Note
Lecture à haute voix :		
Jusqu'à 30 s :	18	de 100 à 110 s : 6
de 30 à 40 s :	16	de 110 à 2 mn : 5
de 40 à 50 s :	14	de 2 mn à 2 mn 30 s : 4
de 50 à 60 s :	12	de 2 mn 30 à 3 mn : 3
de 60 à 70 s :	10	de 3 mn à 3 mn 30 s : 2
de 70 à 80 s :	9	de 3 mn 30 à 4 mn : 1
de 80 à 90 s :	8	plus de 4 minutes : 0
de 90 à 100 s :	7	

Figure 34 : Épreuve FT3 (Annexe B, p. 119)

¹⁰⁵ Sites consultés en avril 2015 :

http://peronne.ia80.ac-amiens.fr/circo/images/stories/Evaluations/eva_CP_2_1_.pdf

<http://www.etab.ac-caen.fr/ecauge/MaitriseLangueC2/LivretCP.pdf>

http://pedagogie.ac-toulouse.fr/circ-montauban-1/EDV/Eval_diagnostique_04_COMPRENDRE.doc

<http://bonpointetbonnetdane.eklablog.fr/evaluation-de-francais-2eme-trimestre-a78985091>

8) Lecture chronométrée 20 / 20

Le petit oiseau

Un petit oiseau jaune et vert est entré dans la maison. Il se pose sur la table, sur un fauteuil. Il se cogne contre les vitres de la fenêtre. Papa ouvre la porte. Le petit oiseau s'envole. Il passe au-dessus d'un grand marronnier et disparaît.

Temps mis : 27 s = 17 points

- jusqu'à 40 s 17 points	- 91 à 100 s 11 points	- 151 à 160 s 5 points
- 40 à 50 s 16 points	- 101 à 110 s 10 points	- 161 à 170 s 4 points
- 51 à 60 s 15 points	- 111 à 120 s 9 points	- 171 à 180 s 3 points
- 61 à 70 s 14 points	- 121 à 130 s 8 points	- 181 à 190 s 2 points
- 71 à 80 s 13 points	- 131 à 140 s 7 points	- 191 à 200 s 1 point
- 81 à 90 s 12 points	- 141 à 150 s 6 points	- 201 et plus 0 point

0 erreur : + 3 points de 1 à 5 erreurs : + 2 points au-delà de 5 erreurs : + 1 point

Figure 35 : Épreuve FT3 (Annexe B, p. 124)

L'élève doit lire ce texte très descriptif à voix haute. Il est composé de 44 mots pour la première épreuve et de 47 mots pour la deuxième. Le texte a été amputé de « sur la table » dans la première épreuve ce qui supprime une difficulté de compréhension (l'élève doit comprendre que l'oiseau se pose successivement sur la table et sur le fauteuil) susceptible d'avoir une influence sur la vitesse de lecture. La deuxième épreuve est de ce fait légèrement plus difficile que la première.

10 mots peuvent être suspectés de présenter certaines difficultés de lecture au moment du recodage phonologique. A titre de comparaison, le texte « Caroline » (cf. III-4-2a) ne contenait que 3 mots obtenant un score de difficulté supérieur ou égal à 1,25. Dans le texte du petit oiseau, 10 mots entrent dans cette catégorie : « grand », « cogne », « maison » – « oiseau », « fenêtre », « contre », « ouvre » – « entré », « fauteuil », « marronnier »- « disparaît ». Le texte donné à lire est donc d'une difficulté non négligeable du point de vue du recodage phonologique.

Lorsque les élèves lisent, ils sont chronométrés.

c- la copie de l'élève :

L'enseignant mentionne le temps de lecture de l'élève.

L'élève de la classe de l'enseignant A obtient un score de 12 points : il a mis entre 50 et 60s pour lire le texte. Cela permet d'obtenir un score de Mots Lus Correctement par Minute (MLCM) situé entre 44 et 52, ce qui correspond d'après Giasson ou Zorman (cf. début de ce paragraphe) à une valeur acceptable (Giasson obtient une médiane à 53 et Zorman place à 40 un seuil d'alerte). Ces auteurs s'intéressent toutefois à la lecture de listes de mots ce qui est un cas différent, mais au vu de ces données, l'élève en question semble manifester une vitesse de décodage textuelle tout à fait satisfaisante. Elle ne semble pas justifier l'appréciation portée (lettre C (+ ?) sur une échelle de 4 lettres A étant synonyme de réussite, D d'échec).

Dans le second cas, l'élève de la classe N obtient un score de 17 points : il a mis 27 secondes pour lire le texte de 47 mots sans erreur. Son score MLCM est ainsi de 104.

Discussion :

Cette épreuve dispose d'une certaine validité psycho-didactique dans la mesure où l'on recherche à évaluer explicitement la fluence de la lecture i.e. la rapidité du décodage, mais aussi un certain nombre de paramètres comme par exemple la prosodie.

Toutefois, le système de cotation de la réponse de l'élève doit être considéré avec prudence. La grille de cotation proposée laisse supposer que les temps de lecture sont étalonnés pour le texte, mais la référence n'est pas accessible aux enseignants qui l'ont utilisé (lors de l'entretien post-collecte, ils n'ont pas su en faire part), et nous n'en avons pas trouvé trace dans nos recherches. Nous constatons que cet étalonnage n'est pas fiable car il n'est pas homogène. Dans le premier cas l'élève met entre 50s et 60s pour lire le texte et obtient 12/18 pts. Il obtiendrait un score voisin de 15/17 pts avec la deuxième échelle pour les mêmes performances de lecture¹⁰⁶.

Se pose aussi la question de l'exploitation de ces scores : sont-ils traduits en note comme suggéré sur le premier document ? Manifestement l'enseignant A ne sait pas interpréter correctement la valeur du MCLM obtenu par l'élève dont la copie est reproduite (d'après les résultats de recherche consultés et précédemment cités elle semblerait plutôt correcte). Il est fort probable que cet outil de mesure lui soit d'ailleurs inconnu.

Enfin, une lecture fluide et expressive d'un texte mobilise une compétence de compréhension : elle serait implicitement évaluée dans ce type d'épreuve et pourrait conduire à estimer qu'un lecteur non fluent ne serait pas capable de comprendre ce qu'il a lu avec peine ce qui peut être remis en cause (cf. C. Blanc et Vantourout, soumis). Il est sans doute plus difficile de lire pour un élève de CP « il se pose sur la table, sur un fauteuil » que « il se pose sur un fauteuil » suivi de « il se pose sur la table » du fait de la présentation successive des lieux sur lesquels se pose l'oiseau.

Discussion

L'analyse des épreuves conduites dans ce paragraphe permet de souligner que la formulation de la compétence de lecture à voix haute dans les programmes « lire à haute voix un texte court dont les mots ont été étudiés » invite les enseignants à utiliser des épreuves de natures différentes selon que le texte (et pas seulement les mots qu'il contient) est déjà connu ou non (évalue-t-on la mémoire de la compréhension du texte ou la mémoire du texte ?), suivant la taille du texte, mais aussi selon que l'on souhaite évaluer la fluence textuelle ou les compétences de décodage/recodage ce qui renvoie à deux compétences que l'on peut distinguer (la première compétence mobilise davantage la compréhension du texte lu).

Nous avons pu souligner que la mesure de la fluence pose un certain nombre de problèmes au rang desquels le choix du support sur lequel elle est effectuée (une liste de mots (outils ou et/ou fréquents) ou un texte n'engagent pas la même activité), l'étalonnage des temps de lectures et l'exploitation de la mesure obtenue (comment la transformer en note ?)

¹⁰⁶ Pour le propos général, nous ne tenons pas compte du fait que le second texte contient trois mots de plus que le premier (« sur la table »). Cela n'a pas d'incidence sur l'estimation de score rapportée (qui pourrait aussi être de 14pts sur 17, ce qui produit également un écart important comparé au score obtenu dans la première épreuve).

5.6. Synthèse du chapitre 5

Les éléments soulignés dans ce chapitre permettent de dire que les choix de cadres de réponse opérés sont susceptibles d'avoir une influence sur l'activité de réponse de l'élève :

- le cadre de réponse imposé peut avoir une incidence sur les procédures (cas du dessin en résolution de problème qui favorise plutôt le dénombrement que le calcul),
- le cadre de réponse imposé peut favoriser la mobilisation de compétences qui ne sont pas l'objet principal de l'évaluation (par exemple la tâche mobilise des compétences rédactionnelles pour évaluer la lecture),
- le cadre de réponse imposé est susceptible d'influencer la représentation que l'élève se fait de la tâche qui le sollicite (par exemple lire c'est répondre à des questions ou encore lire c'est trouver la bonne réponse parmi un ensemble de propositions fausses).
- le cadre de réponse peut laisser à l'enseignant une large part dans l'interprétation des réponses (dans le cas de réponse au moyen de dessin).
- le cadre de réponse oral est rarement privilégié, il pourrait pourtant limiter certains défauts précités.

On peut ajouter que lorsqu'il faut répondre dans un tableau par exemple, il est implicitement supposé que l'élève sait traiter le mode de présentation des données (lire un tableau à double entrée). L'activité de réponse mobilise alors des compétences précises des programmes (par exemple l'item « lire ou compléter un tableau dans des situations concrètes simples » du sous-domaine Organisation et gestion des données en Mathématiques) dont on ne peut pas attendre des élèves qu'elles soient maîtrisées au point de servir de support à l'expression de réponse sans évaluation préalable.

Le cadre de réponse imposé ou suggéré par la tâche est une composante majeure de la tâche que l'élève doit prendre en compte. Un crédit trop important est susceptible d'être attribué par l'enseignant à la réponse écrite de l'élève, indépendamment du processus de réponse lequel s'accommode plus ou moins bien du cadre de réponse imposé ou suggéré par la tâche. Dans une large majorité des cas, dès le début du CP et pour des raisons liées à la mise en œuvre de l'évaluation, les élèves sont soumis à des épreuves nécessitant des réponses écrites, analysées par l'enseignant a posteriori de la réalisation de la tâche. Cette caractéristique de l'évaluation interne est source de difficultés pour l'élève lorsqu'il traite la tâche prescrite mais aussi pour l'enseignant lorsqu'il analyse la tâche effectivement réalisée par l'élève a posteriori de la réalisation de celle-ci, de nombreuses inférences (sur le processus de réponse des élèves) restent à sa charge.

Chapitre 6. Le choix des tâches d'évaluation

Au cours de la redéfinition de la (double) tâche qui leur est prescrite - évaluer et informer - les enseignants déterminent les items du livret puis les épreuves pour évaluer.

Les progressions guidant l'évaluation (cf. Partie II), les livrets des enseignants contiennent un grand nombre d'items. Par conséquent, les tests contiennent un grand nombre d'épreuves.

Pour constituer les tests, les enseignants élaborent eux-mêmes (ou à plusieurs) les épreuves ou les « empruntent », partiellement ou totalement, à d'autres concepteurs (manuels, sites web, ...).

Dans ce chapitre, nous nous intéressons à l'analyse de tâches « empruntées ». Nous nous questionnons sur l'adéquation de leur usage aux fins d'évaluation poursuivies par les enseignants. En particulier, nous questionnons l'apparente validité que leur conférerait le fait d'être publiée.

6.1. Tâches empruntées au web

Au cours d'entretiens post-collecte, nous avons interrogé les enseignants sur la provenance des épreuves qu'ils utilisent.

L'enseignant H confie « *bricoler mais (tout) fabriquer* » tandis que l'enseignant P reconnaît « *(prendre) des choses sur le web. Tu tapes « éval CP t3 » (évaluation troisième trimestre). En fait il y a quand même beaucoup de (choses) qui viennent d'internet.* »

La requête « éval CP t3 » trouve de nombreuses réponses. Les sites ou blogs *ekldata*, *ipotame*, *alencreviolette*¹⁰⁷, entre autres, sont des sources desquelles sont extraites certaines des épreuves collectées dans notre corpus. La mention « *Professeur des écoles dans le sud de la France ... 10 ans de CP* » sur le troisième de ces sites ou blog semble valider le contenu hébergé. Les commentaires laissés par les visiteurs renforcent les propos tenus par les enseignants que nous avons sollicités. Ainsi peut-on lire le commentaire suivant laissé par un internaute sur le site *alencreviolette* : « *Je lis tout, je pioche et m'inspire, je suis loin de France et cela m'aide beaucoup. Surtout pour les évaluations. Ça m'aide de ne pas les faire complètement moi-même, je me sens moins juge et partie.* »¹⁰⁸ ou bien encore, « *Merci pour ce superbe travail. J'ai récupéré des CP en cours d'année en plus de mes CE1. J'ai dû m'absenter pour des raisons de santé et j'étais dans la m... pour les évaluations. Merci encore. Il ne me restera qu'à faire une petite évaluation pour compléter les mathématiques.* »¹⁰⁹

L'utilisation de l'épreuve « Petit Gaston » pour évaluer la compréhension en lecture au troisième trimestre (cf. épreuve 21, III.5.2b-3) nous apparaît emblématique d'une utilisation insuffisamment distanciée de tâches que l'enseignant n'a pas conçues lui-même mais a

¹⁰⁷

<http://ekldata.com/wy8opj-Um53hGeu7iWrk3288E88/evals-maths-3eme-trimestre-perso.pdf> ; <http://ipotame.blogspot.fr/2014/05/evaluations-CP-maths-periode-5.html> ; <http://www.alencreviolette.fr/evaluations-CP-a24726688> consultés en novembre 2014.

¹⁰⁸ Commentaire posté le samedi 12 Avril à 18:51

¹⁰⁹ Commentaire posté le mercredi 5 Juin 2013 à 18:03

empruntées. Au troisième trimestre, 7 enseignants, sur les 16 qui forment la population étudiée, évaluent la compréhension en lecture au moyen d'une seule épreuve (un enseignant ne propose même aucune épreuve). On mesure ainsi l'importance du choix de l'épreuve : si l'on admet une moyenne de 25 élèves par classe, 50 élèves n'ont été évalués en compréhension en lecture en fin de CP qu'au seul moyen de l'épreuve « Petit Gaston ».

Nous avons saisi (en février 2014) la requête « *évaluation lecture petit Gaston* » dans un moteur de recherche. Les résultats sont édifiants : le texte est utilisé en CP, en CE1 ou en CE2 comme le prouve la capture d'écran reproduite ci-dessous. Les hébergeurs de cette ressource sont variés : sites personnels, sites de groupes académiques, site ministériel laissant à l'enseignant la responsabilité de choisir.

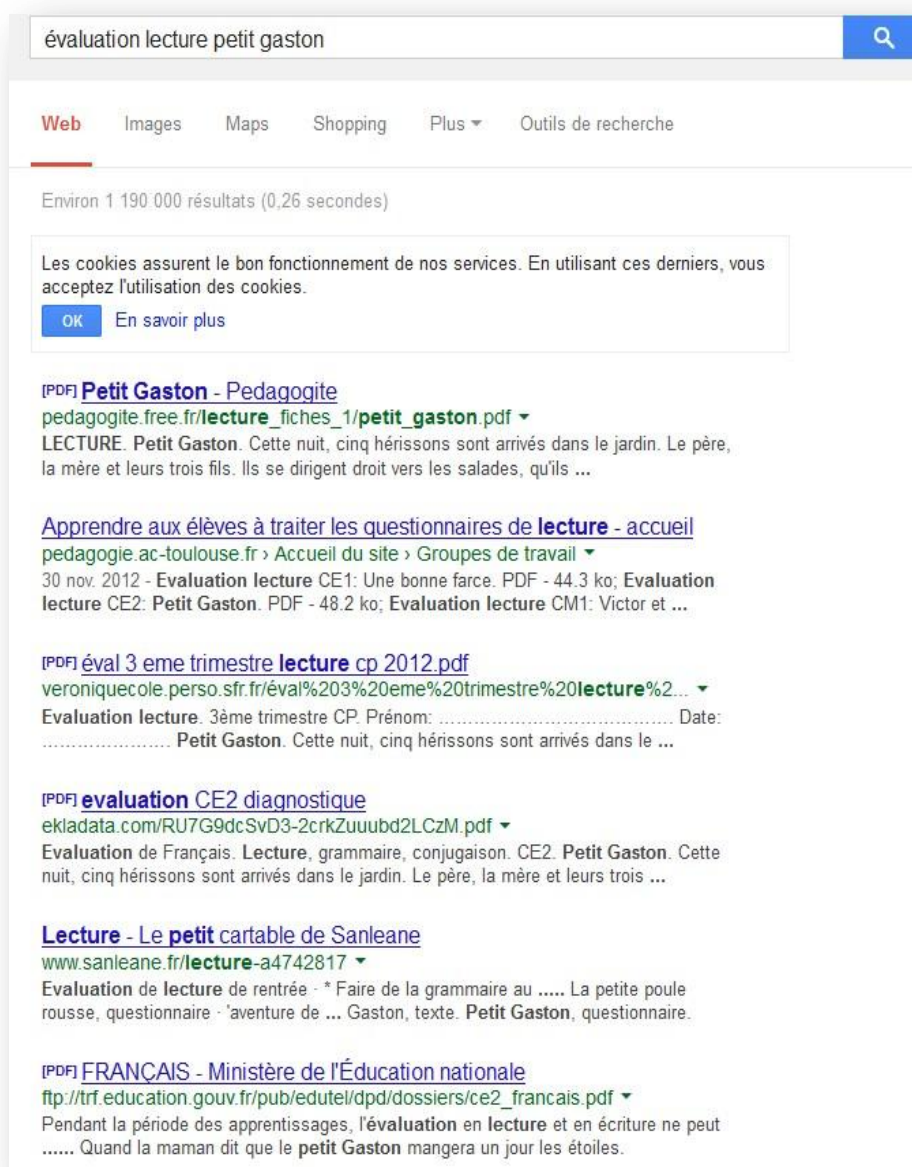


Figure 36 : Capture d'écran de la requête « évaluation lecture petit Gaston »

L'épreuve « Petit Gaston » a été utilisée par plusieurs enseignants pour évaluer la compréhension en fin de CP. Nous avons reproduit les deux adaptations de l'épreuve que nous avons collecté (cf. épreuve 21 Partie III paragraphe 5.2b-3). La première provient d'un

site de circonscription¹¹⁰. Elle figure dans un dispositif d'évaluation de fin de CP du département de Haute-Savoie, pour les années 2005-2006 et 2007-2008. Elle a été utilisée telle qu'elle par un enseignant auprès de qui nous avons collecté des épreuves lors d'un premier travail exploratoire. L'enseignant C l'utilise avec quelques aménagements également.

Nous les reproduisons de nouveau pour rendre la référence plus fluide :

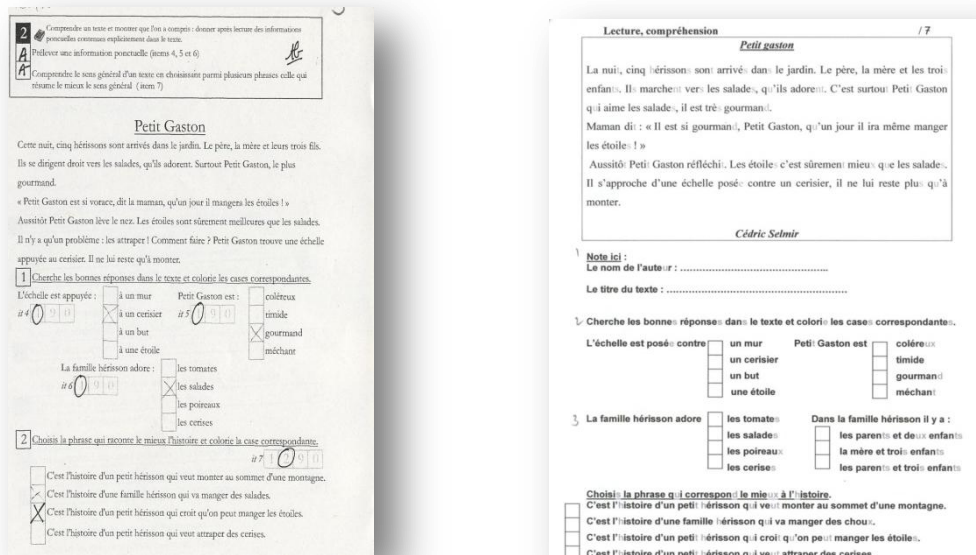


Figure 37 : Épreuves « Petit Gaston » (site Cluses (gauche) et Annexe B, p. 134 (droite)) avec réponses d'élève

Dans un rapport de 2013, les inspecteurs généraux¹¹¹ déplorent le manque d'analyse de ce type de ressources par les enseignants :

« La dépendance par rapport aux manuels, fichiers ou, aujourd'hui de manière croissante sites internet, est d'autant plus forte que les connaissances des maîtres ne constituent pas d'appuis solides pour construire un enseignement. Les limites de ces outils ou les erreurs qu'ils renferment ne sont pas perçues. Les inspecteurs regrettent le manque d'appropriation véritable et d'esprit critique face aux supports empruntés dont certains n'ont reçu aucune validation (sites de pairs par exemple). Ils déplorent également le caractère décousu des pratiques ainsi construites, en puisant à des sources multiples. On ne saurait leur donner tort. »

Cette conclusion renvoie les enseignants à leur professionnalité (ils devraient disposer de connaissances suffisamment solides pour être capables de percevoir les limites ou les erreurs des outils qu'ils utilisent).

¹¹⁰ Site : http://www.ac-grenoble.fr/ien.cluses/IMG/pdf/livret_maitre-fin-CP_2007.pdf

¹¹¹ Bilan de la mise en œuvre des programmes issus de la réforme de l'école primaire de 2008 - Rapport - n° 2013-066, juin 2013.

Les données collectées nous conduisent à nuancer ce propos : tout d'abord, nous avons relevé que certaines épreuves ne sont pas valides d'un point de vue psycho-didactique, pour des questions qui ne relèvent pas du bien-fondé des connaissances mises en jeu mais d'autres paramètres, sous-estimés parce que peu étudiés et mis en lumière, puis diffusés, dans les (très rares) formations d'enseignants à l'évaluation. Plus encore, le reproche adressé produit un effet de boomerang dans le cas de l'épreuve étudiée. En effet, l'un des enseignants de notre population s'est référé à un site de circonscription qui propose lui-même cette ressource pour l'évaluation de la compréhension en lecture en CP. Cette ressource est validée par l'institution. Nous avons pour autant conclu à son manque de validité psycho-didactique pour évaluer la compréhension en lecture.

Cette ressource n'a même pas été épurée des formats de cotation utilisée par l'opérateur externe qui l'a produite : 1 (bonne réponse) ; 9 (autre réponse) ; 0 (absence de réponse). Ce type de cotation n'est jamais utilisé par les enseignants. On retrouve pourtant pareil type d'épreuves pour évaluer la compréhension en lecture à plusieurs reprises dans les données collectées (cf. épreuves reproduites pp. 138 ; 145 ; 146 du volume d'annexe B concernant 3 enseignants) avec parfois une adaptation du codage à un système de cotation propre à l'enseignant.

La présentation de l'épreuve augmente la densité informationnelle (Rouet, Germain, Mazel, 2007) et le risque d'erreur, puisque le barème de codage (1, 9, 0) figure au milieu de l'épreuve, sous forme de cases, induisant une difficulté supplémentaire : il faut colorier la case face à une proposition de réponse et non pas une case du codage de réponse.

Rappelons que l'épreuve « Petit Gaston », est une épreuve qui a été utilisée dans le cadre des évaluations à l'entrée en CE2 en 1995, 2000 et 2004¹¹². Ces évaluations nationales n'ont pas d'autre visée qu'une visée diagnostique : il est ainsi précisé¹¹³ que « l'information apportée par l'évaluation nationale de CE2 et de 6^{ième}, en français et en mathématiques, ne constitue pas un bilan de cycle 2 ou de cycle 3 mais un diagnostic, au début d'un nouveau cycle scolaire, des réussites, des erreurs et des difficultés éventuelles de chaque élève par rapport à son avancée dans les apprentissages ». Hormis la question que pose l'intérêt d'utiliser en fin de CP une épreuve conçue pour les élèves entrant en CE2, se surajoute une réaffectation de la finalité de l'évaluation : de diagnostique elle devient sommative.

Dans l'épreuve proposée par la DEPP, 7 items servent à évaluer la compréhension. Nous reproduisons ci-dessous l'analyse détaillée effectuée par la DEPP après la passation 2000.

¹¹²Voir Évaluation à l'entrée au CE2, Cahier de l'élève, DEPP, 2004, exercice 3, pp.6-7 et voir également sur le site educ.gouv.fr le texte d'analyse des évaluations CE2. Consultable à l'adresse <http://media.education.gouv.fr/file/66/3/4663.pdf> (BREZILLON, 2000)

¹¹³ Source : Site MEN <http://test.evaCE26.education.gouv.fr/sitearchive/archive2004/> consulté en septembre 2014.

Exercice 4(Exercice 2 de 1995)⁶

2000

Support d'activité
vuConsignes
vues**Compétence 4**

Comprendre un texte et montrer qu'on l'a compris :

- donner après lecture des informations ponctuelles contenues explicitement ou implicitement dans le texte ;
- dans un récit, identifier les personnages et les retrouver, quels que soient les procédés utilisés (nom, pronom, périphrase, etc.) ;
- construire la signification générale d'un texte simple en prenant en compte ses énoncés successifs.

Résultats (en %)

Sept prises d'information pour cet exercice :

Q1	<i>L'échelle est appuyée... (item 14) :</i>	Réponse exacte : à un cerisier 85,2
		Autres réponses 13,4
		Absence de réponse 1,4
Q2	<i>Petit Gaston est... (item 15) :</i>	Réponse exacte : gourmand 92,8
		Autres réponses 4,9
		Absence de réponse 2,3
Q3	<i>La famille hérisson adore... (item 16) :</i>	Réponse exacte : les salades 91,8
		Autres réponses 5,4
		Absence de réponse 2,8
Q4	<i>Combien y a-t-il d'enfants dans cette famille de hérissons ? (item 17) :</i>	Réponse exacte : trois ou 3 (avec ou sans un substantif qui convient) 60,8
		Autres réponses 29,5
		Absence de réponse 9,7
Q5	<i>Petit Gaston a réussi / n'a pas réussi... (item 18) :</i>	Réponse exacte : a réussi à arriver en haut de l'échelle ET n'a pas réussi à attraper les étoiles 63,9
		Autres réponses 25,8
		Absence de réponse 10,3
Q6	<i>Quand la maman dit que le petit Gaston mangera un jour les étoiles... (item 19) :</i>	Réponse exacte : elle veut dire qu'il est très, très gourmand 66,2
		Réponse : elle veut dire qu'il n'est pas très malin, car c'est impossible 10,9
		Autres réponses 8,8
		Absence de réponse 14,1

⁶ Pour les résultats détaillés de l'exercice en septembre 1995, voir : Exercice 2 de 1995 dans Les dossiers d'Éducation et Formations, numéro 65, « Évaluation CE2-6ème Résultats nationaux - septembre 1995 », M.E.N.-D.E.P., mars 1996, pages 24-27.

Q7	<i>Choisis la phrase qui raconte le mieux l'histoire... (item 20) :</i>	Réponse exacte : C'est l'histoire d'un petit hérisson qui croit qu'on peut manger des étoiles 44,0
		Réponse : C'est l'histoire d'une famille hérisson qui va manger des salades 30,4
		Autres réponses 8,4
		Absence de réponse 19,2

Figure 38 : Résultats détaillés à l'épreuve « Petit Gaston ». Évaluation CE2, DEPP 2000.

Céard, Remond et Varier (2003, p.17), indiquent que les résultats de 2000 confirment ce qui avait été constaté en 1995 : « les enfants ne sont pas déroutés par ce type de questions (le taux d'absence de réponse est seulement de 1,4 %) et les résultats sont très bons. »

Nous reproduisons leurs analyses :

« Les *items 15 et 16 qui sont les mieux réussis* supposent pourtant des transformations linguistiques différentes (« Surtout Petit Gaston, le plus gourmand. » → « Petit Gaston est gourmand. » ; « Ils se dirigent droit vers les salades, qu'ils adorent. » → « La famille hérisson adore les salades. »).

L'*item 14, bien que demandant lui aussi une prise d'information directe* est un peu moins bien réussi (85 % en 1995 et en 2000) pour de multiples raisons : situation habituelle d'une échelle contre un mur ; méconnaissance pour certains du « cerisier » ; détail secondaire par rapport à l'ensemble du texte, etc.).

Les items suivants sont plus difficiles (les taux de réussite obtenus à ces items en 1995 et en 2000 l'attestent. Ils supposent tous des traitements de l'information plus importants. Cet item fait apparaître un écart particulièrement important en défaveur des élèves de ZEP).

Pour l'*item 17 (« Combien y a-t-il d'enfants dans cette famille de hérissons ? »)*, il s'agit d'une question pour laquelle il faut écrire la réponse, ce qui est toujours plus difficile à réussir qu'un QCM. De plus, la bonne réponse impose que l'équivalence lexicale « fils / enfant » soit parfaitement maîtrisée, dans ce contexte. En outre, la mention en tout début de texte de « cinq hérissons » peut fonctionner comme un piège. Cependant 61 % des élèves donnent la réponse exacte.

L'*item 18 (« Petit Gaston a réussi / n'a pas réussi »)* tend à vérifier la compréhension globale du texte. Les deux propositions que l'élève doit choisir en résumé en effet le ressort narratif. Le choix doit être fait, chaque fois, entre la forme affirmative ou la forme négative d'un même énoncé. C'est là, pour beaucoup d'enfants, une tâche difficile qui relève d'un calcul logique délicat. Échouer à cet item ne signifie donc pas systématiquement que l'élève n'a pas compris le sens global du texte, mais en tous cas qu'il n'est pas parvenu à l'exprimer sous la forme de la conjonction d'une proposition affirmative et d'une proposition négative : « Petit Gaston a réussi à arriver en haut de l'échelle. » et « Petit Gaston n'a pas réussi à attraper les étoiles. ». C'est le cas d'un élève sur quatre en 1995 comme en 2000.

L'*item 19 (« Quand la maman dit que le petit Gaston mangera un jour les étoiles... »)* porte sur le sens d'une expression métaphorique, qui ne met pas en cause la compréhension globale du texte. En 2000, deux enfants sur trois montrent qu'ils sont capables d'en percevoir le sens. Ce bon score est encourageant. On peut certainement encore améliorer ce résultat en faisant expliciter systématiquement dès le cycle 2 ce type d'énoncé.

Le dernier item (*item 20 : « Choisis la phrase qui raconte le mieux l'histoire... »*) demande de choisir une phrase parmi quatre, qui résume l'ensemble du texte et non pas seulement son début. En 2000 comme en 1995 près d'un tiers des enfants choisissent une réponse non contradictoire avec le texte mais partielle (réponse : « C'est l'histoire d'une famille hérisson qui va manger des salades »). Davantage (44 %) sont capables de choisir celle qui raconte le mieux (« C'est l'histoire d'un petit hérisson qui croit qu'on peut manger des étoiles. »).

Les résultats obtenus cette année montrent que la progression qui se dessinait dans les évaluations antérieures se maintient. »

On voit bien que si deux élèves sur trois comprennent le sens de la métaphore « manger les étoiles », précisant bien que ce type de compétence faisait partie de « compétences de haut niveau pour des élèves (de CE2 ndlr) qui viennent d'apprendre à lire » (2003, p.17), il est aisé de concevoir que cette proportion sera plus faible en fin de CP. Lecocq et coll. après avoir indiqué que « l'accès aux métaphores par l'enfant est relativement tardif » (1996, p. 281), ont montré que « les enfants dès l'âge de 8 ans ont une compétence telle que certaines métaphores conceptuelles complexes deviennent accessibles à la faveur d'une consigne et d'un matériel qui, manifestement, permettent une exploitation des connaissances et une gestion de celles-ci » (1996, p ; 229). Notons que ce que l'on peut attendre, après enseignement, d'un élève de 8 ans ne peut être attendu d'un élève de 6 ans, sans que son attention ait été a minima attirée. Il n'est pas raisonnable de penser que tous les élèves puissent en fin de CP obtenir des scores supérieurs à des élèves entrant en CE2

Dans les versions utilisées avec des élèves de CP, l'épreuve initiale a été épurée (pour la rendre plus facile ?). Dans les deux cas, le nombre de « prises d'informations » - pour reprendre les termes de l'étude DEPP- sur le texte est moindre : 4 sur 7 pour le premier (site IEN Cluses) ; 5 sur 7 pour le second. Aucune des deux épreuves ne questionne les élèves sur la métaphore « Quand la maman dit que le Petit Gaston mangera un jour les étoiles ... (it. 19) ». La question n'est pas posée mais ce passage important pour la compréhension reste tout aussi difficile à comprendre. Et si 44% des élèves entrant en CE2 trouve le bon résumé, ce qui revient à dire que 56% n'y arrive pas, il est prévisible que le taux d'échec sera plus important (ceci étant potentiellement renforcé par le fait que l'attention de l'élève n'est pas attirée sur cette difficulté particulière au moyen d'une question). L'enseignant s'efforce de faciliter la tâche des élèves dans l'épreuve qu'il donne à ses élèves (cf. page précédente, épreuve de droite). Il écrit en rouge les lettres muettes des mots : « hérissons » est ainsi décomposé en « érisson » (en noir) encadré par les lettres « h » et « s » en rouge. Cette aide présumée à la lecture des mots est en décalage avec la difficulté du texte choisi et l'« aide » n'est pas reprise dans le questionnaire cf. III.5-2b-3.

La responsabilité du mésusage de l'épreuve ne peut être totalement attribuée aux enseignants : ils l'ont trouvé sur un site institutionnel de ressources pour l'évaluation en fin de CP ce qui lui conférait une forme de validité.

6.2. Tâches empruntées à des manuels

L'exemple discuté précédemment nous a permis de mettre au jour la problématique suivante : sur quelles ressources publiées, « valides », un enseignant peut-il s'appuyer ? Les enseignants empruntent massivement des tâches pour évaluer aux manuels qu'ils utilisent, ou qui se trouvent dans leur classe.

De façon pragmatique, les enseignants utilisent souvent un matériel « déjà là ». Or, les manuels ne sont pas nécessairement renouvelés chaque année. Les éditeurs sont confrontés à la nécessité de proposer des ouvrages conformes aux programmes en vigueur. Ils ont souvent peu de temps pour adapter une collection aux exigences d'un nouveau programme. Souvenons-nous qu'aux programmes de 2002 ont succédé le socle commun en 2005, une adaptation des programmes en 2007 et de nouveaux programmes de 2008 : tout ceci représente un défi de taille pour un éditeur.

Nous analysons la tâche « jetons » au filtre de cette remarque :

Épreuve n°31 : « Les jetons »

a- l'item :

La même tâche sert d'épreuve d'évaluation de deux items différents pour deux enseignants différents. Pour l'enseignant L, elle sert à évaluer l'item du livret « calculer mentalement des sommes et des différences » quand il sert à l'enseignant D à évaluer l'item du livret « résoudre des problèmes simples à une opération ».

b- l'épreuve :

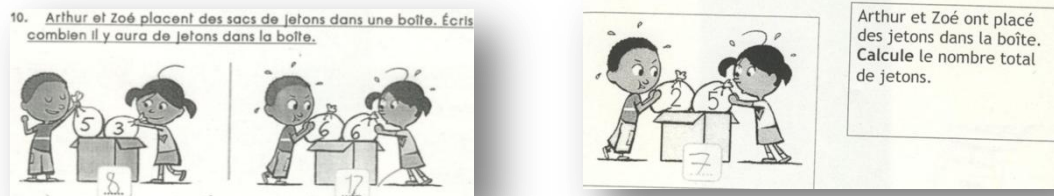


Figure 39 : Épreuves MT1 utilisées par les enseignants L (à gauche) et D (à droite) (Annexe B, p. 208 et 221) avec réponses d'élève

Cette tâche permet sans doute d'évaluer les deux items ce qui en soi ne pose pas de problème particulier.

c- les copies d'élèves :

Elles n'apportent aucun élément nouveau.

Autour de la question de savoir si l'épreuve est valide ou non, nous discutons de l'utilisation qu'il en est fait à des fins qui n'étaient pas celles initialement prévues par l'éditeur. Un décalage est susceptible de se produire dont nous analysons les effets.

Le guide du maître indique (2005, p. 39) qu'« il est probable que beaucoup de procédures de résolution ne donneront pas lieu à une trace écrite : usage des doigts, de la file numérique pour trouver l'écriture du nombre-réponse ou même pour trouver la réponse, en avançant ou en reculant sur la file (pour certains élèves) ». Cette tâche intitulée « combien

de jetons dans la boîte » est conçue pour être entreprise au moyen d'une file numérique (ladite frise n'est fournie dans aucune des situations d'évaluation présentées ici). L'observation des programmes de 2002¹¹⁴ fournit une piste de réponse à la question posée de savoir pourquoi ces deux items sont mis en concurrence pour la même épreuve. Le troisième sous-domaine des mathématiques intitulé « Calcul » se décline dans ces programmes en trois groupes de compétences dont l'un est intitulé « calcul réfléchi ». Ce dernier groupe de compétences comprend la compétence suivante : « résoudre mentalement des problèmes à données numériques simples. ». Le manuel, s'il a été apparemment réactualisé, date en réalité (la consultation du guide du maître l'indique) de ces précédents programmes lesquels envisagent autrement le calcul mental. Nous avons consulté les dates de parution de ces outils : le manuel affiche « nouveaux programmes ». Il a été bien été mis en conformité avec les nouveaux programmes (en 2009) mais ... pas le guide du maître, qui date de 2005. Dans les programmes de 2002, résoudre un problème est décliné selon deux type de procédures : en utilisant une procédure experte, en utilisant une procédure personnelle. La tâche proposée visait l'objectif d' « anticiper le résultat d'un ajout (ou d'un retrait) » et celui de « préparer l'addition (et la soustraction) » : des objectifs au programme en 2002 mais qui ne le sont plus tels quels depuis 2008.

L'utilisation de ressources telles que web, manuels ... représentent autant d'emprunts à risques. Faire coïncider l'item du livret avec la compétence mise en avant dans l'exercice emprunté n'est pas sans poser quelques obstacles parmi lesquels celui de faire entrer dans une case quelque chose qui n'a pas été pensé pour cela.

L'emploi de la tâche suivante (« Léon ») nous permet de poser un nouveau problème : celui du risque d'évaluer la mémoire d'une situation étudiée dans le manuel au cours de l'étude d'un texte plutôt que d'évaluer la compréhension du texte en elle-même.

¹¹⁴<http://www.education.gouv.fr/bo/2002/hs1/cycle2.htm>

Épreuve n°32 : « Léon »

a- l'item :

Pour évaluer l'item du livret « Comprendre un texte lu » en fin de premier trimestre, l'enseignant E conçoit et utilise deux épreuves. Nous reproduisons ci-dessous la page 30 du manuel utilisée en classe :

2

Aide-mémoire
À / il y a / un / qui / dans / aussi / garçon

Je lis ce texte.
D'après *Léon et son croco* (1)

- À Grand Poco, il y a un crocodile qui vit dans l'eau.
À Grand Poco, il y a aussi un petit garçon qui s'appelle Léon. Il a un bâton.
- À Grand Poco, il y a Léon avec son bâton sur le ponton et le crocodile dans l'eau.
À Grand Poco, il y a Léon avec son bâton qui tombe à l'eau.



Recherche
« Il y a Léon avec son bâton qui tombe à l'eau. »
Que va-t-il se passer ?

30

Figure 40 : Extrait de manuel (A l'école des albums, CP Retz)

Le texte évoque un garçon nommé Léon, un bâton, un ponton (la double page 30 et 31 du manuel porte sur le « son on »), et un crocodile. L'illustration du texte mentionne ces quatre éléments.

b- l'épreuve :

Analysons à présent les épreuves conçues par l'enseignant.

La première, reproduite ci-dessous, demande à l'élève de dessiner « Léon est sur le ponton. Il a un bâton. Il y a un crocodile dans l'eau ».

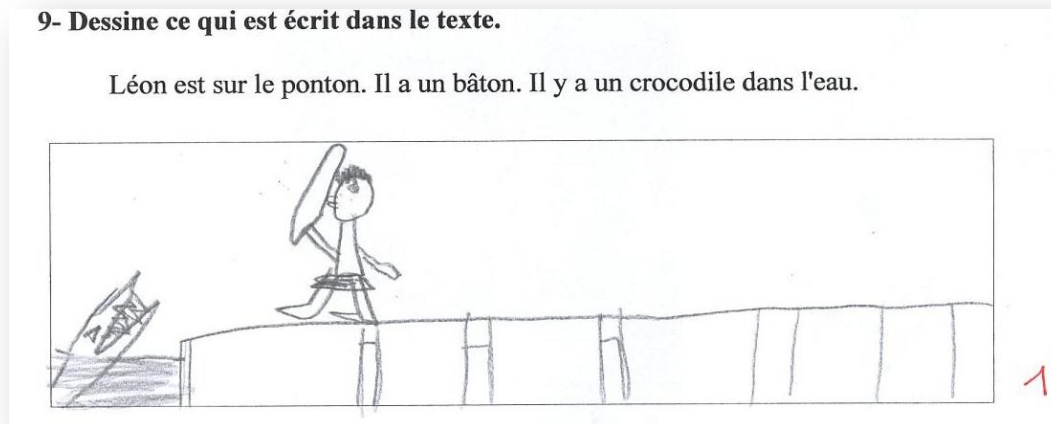


Figure 41 : Épreuve FT1 utilisée par l'enseignant E (Annexe p. 70) avec réponse d'élève

La deuxième épreuve, reproduite ci-dessous, demande à l'élève de dessiner un roi, Zoé sur un ponton et un crocodile dans l'eau.

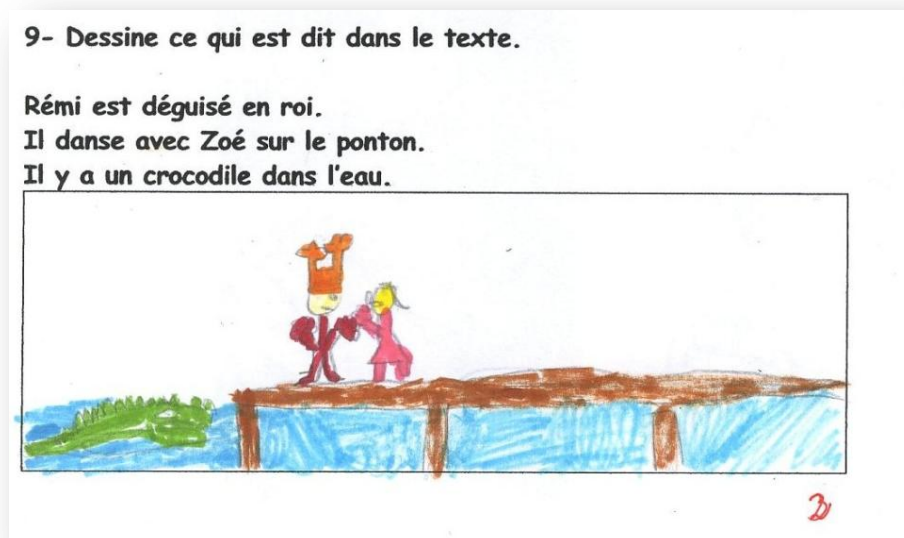


Figure 42 : Épreuve FT1 (Annexe B, p. 72) avec réponses d'élève

Pour réussir l'un ou l'autre de ces épreuves, l'élève doit dessiner un personnage sur un ponton et un crocodile dans l'eau. Il doit ajouter un bâton dans le premier cas et une couronne (il est archétypique qu'un roi soit couronné) et une fille.

c- la copie de l'élève :

Les deux copies reproduites indiquent simplement que les dessins sont conformes aux attentes.

Comparons le manuel avec les textes au sujet desquels l'élève doit manifester de la compréhension par le dessin : la proximité graphique des éléments figurés est importante. Elle peut conduire à penser que les réponses produites dépendent pour une large part de la mémoire de la situation présentée par le manuel et non de la compréhension en elle-même de ce qu'il y a à lire. Ceci est de nature à invalider la tâche prescrite : elle ne permet pas d'évaluer la compréhension du texte mais plutôt la mémoire de la situation générale de l'histoire du texte déjà étudié (même si les mots du texte ont changé).

Le fait que les tâches utilisées soient publiées ne garantit en rien qu'elles soient valides, indépendamment de la notoriété ou de la qualité de l'éditeur ou du concepteur. Ce résultat paraît trivial mais doit être affirmé à considérer les sources analysées dans ce chapitre : des sites de circonscription, des manuels conformes aux programmes d'éditeurs reconnus.

Le fait que les tâches soient publiées n'impliquent pas non plus qu'elles soient correctement employées lorsqu'elles sont utilisées pour évaluer : la responsabilité de l'emprunteur est engagée s'il modifie le contexte de la tâche tel qu'il a été conçu (évaluation diagnostique, fin d'une unité d'apprentissage, situation de découverte, tâche outillée, référence à des programmes antérieurs ...).

Synthèse de la partie III

Dans cette partie, nous avons présenté et analysé un certain nombre d'épreuves dont la validité, au sens psycho-didactique, peut être contestée. Nous avons mis en évidence l'influence de certains composants de la tâche prescrite à l'élève susceptibles d'engendrer un écart entre les attentes de l'enseignant et l'activité de l'élève au cours du traitement de la tâche et de la production de la réponse.

Parmi ces composants, nous avons mis en exergue :

- la présentation des données : nous avons montré le rôle ambigu que le dessin pouvait tenir à ce niveau, particulièrement dans les tâches de dénombrement, mais aussi dans les tâches de lecture ou de résolution de problèmes. Il est susceptible d'intervenir au niveau de la représentation du texte lorsqu'il l'illustre ; il est susceptible d'engager l'élève dans une procédure de dénombrement alors qu'une procédure de calcul est attendue.
- la rédaction des demandes : l'élément le plus prescripteur dans une tâche est sans nul doute la consigne. Au-delà de la (ou des) consigne(s), le format du questionnement peut être mis en cause. L'usage très répandu des questionnaires par les enseignants pour évaluer la compréhension en lecture pose le problème de la « transparence » qui semble leur être accordée au niveau du traitement de la tâche par l'élève. Le questionnaire, considéré comme un texte sur le texte, surajoute de la complexité (en termes de lecture, en termes de compréhension, en termes de stratégie de réponse). Cela ne semble jamais être pris en compte : aucun enseignant n'a indiqué avoir évalué la compréhension d'un texte lu par l'élève au moyen de questions-réponses orales.
- le système de réponse sous-tendu par la tâche : il constitue un élément au moins aussi prescriptif que la consigne. Il est susceptible de modifier le sens de la tâche. Manifester sa réponse est en réalité une compétence complexe qui apparaît peu prise en compte. Cela prend d'autant plus d'importance en CP car les élèves sont tous, *a priori*, des lecteurs-rédacteurs débutants. Ils sont confrontés à une utilisation experte de la langue écrite tant en réception (consignes et questions) qu'en production (rédaction des réponses). Lorsque, à raison sans doute, des aménagements quant à ce système sont pensées (QCM et autres), ils ne sont pas sans poser de nouveaux problèmes.

Nous avons enfin soulevé le problème du choix des épreuves. En particulier, nous avons souligné le défaut d'adéquation aux objectifs de l'enseignant de certaines tâches, conçues et publiées par d'autres pour un certain usage. La validité, supposée ou réelle, de ces tâches n'est, dans les exemples analysés, pas remise en cause par l'enseignant lorsqu'il les utilise à des fins d'évaluation d'items du livret. Cela permet de réaffirmer que la validité d'une tâche est dépendante des objectifs d'évaluation qui lui sont assignés : ce n'est pas une qualité endogène.

Pour deux items fondamentaux des sous-domaines du Français et des Mathématiques auxquels renvoient particulièrement les épreuves étudiées dans cette partie (comprendre en lecture, connaître et désigner les nombres entiers naturels), l'analyse que nous avons menée a permis de montrer que les enseignants sont susceptibles d'émettre un jugement évaluatif imparfait quant aux compétences de l'élève dû à un manque de validité psycho-didactique des épreuves utilisées.

Nous avons pu établir à de nombreuses reprises qu'il était possible d'obtenir une bonne réponse sur la base d'un raisonnement qui pourrait être erroné, ou que ce que la tâche évaluait pouvait ne pas correspondre à ce que l'on attendait que l'élève mobilise. Ce résultat n'est pas nouveau, ce qui semble prouver que ce phénomène est mal connu et/ou mal pris en compte par les enseignants.

Ces résultats plaident en faveur d'une analyse psycho-didactique des tâches utilisées à des fins d'évaluation qui prenne conjointement en compte : ce que l'enseignant souhaite évaluer au moyen de l'épreuve (son référentiel), des résultats de recherches disponibles, des composants des tâches non directement liés *a priori* aux concepts évalués et des réponses d'élèves.

La méthodologie d'analyse des épreuves que nous avons utilisée dans cette partie correspond à un mode d'analyse local des épreuves d'évaluation.

Dans la dernière partie de la thèse, nous croisons cette modalité d'analyse des épreuves avec une analyse des tests. Nous nous posons la question de savoir si un test contient des épreuves valides (du point de vue psycho-didactique) permettant d'évaluer l'ensemble des principaux composants d'un domaine. Le cas échéant on pourrait alors conclure à la validité de couverture (didactique) du test.

Partie IV – Étude de la validité globale et locale des tests et des épreuves

L'analyse de la validité de l'évaluation peut être globalement conduite au moyen d'une approche curriculaire qui permet d'apprécier la couverture des programmes réalisée par l'évaluation. Une telle analyse est fondée sur les items des programmes que les enseignants déclarent évaluer au moyen des épreuves (cf. partie II).

Une autre approche globale de la validité de l'évaluation peut être conduite au moyen de référentiels d'ordre didactique. Elle est fondée sur l'analyse du contenu des tests relativement aux composants réellement mis en jeu, elle permet d'apprécier la couverture didactique des tests, i.e. des principaux composants des domaines étudiés.

Nous avons détaillé dans la partie précédente un mode d'analyse local des tests consistant à étudier les épreuves dans le but d'apprécier si la tâche prescrite mobilise une activité de réponse de l'élève correspondante à ce pourquoi elle est donnée.

Dans la partie qui s'ouvre, nous relient ces deux entrées et analysons d'un double point de vue les données collectées. Du point de vue global, les tests couvrent-ils les principaux composants d'ordre didactique des champs concernés ? Du point de vue local, les épreuves qui forment les tests évaluent-elles bien la maîtrise des composants qu'elles prétendent évaluer ?

La conduite de cette analyse étant étroitement liée à la détermination des principaux composants des différents champs étudiés qui constituent le référentiel du chercheur, nous y consacrons, pour chaque domaine, un chapitre préalable. Cela renvoie au processus de référentialisation (du chercheur) évoqué en Partie I, chapitre.2.3

Chapitre 1. Analyse des tests et des épreuves d'évaluation des compétences relevant des nombres et du calcul

Dans ce chapitre nous nous intéressons à la validité globale et locale des tests et épreuves du premier et du troisième trimestre relatifs au sous-domaine des Nombres et du calcul

1.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences du domaine des nombres et du calcul en CP.

Préalablement à l'analyse des données collectées, nous cherchons à définir ce que constitueraient, en CP, les principaux composants du domaine des nombres et du calcul.

Nous fondant sur les propositions épistémologiques et didactiques de Margolinas et Wozniak (2012, p. 10), on peut estimer qu'au cours des années de primo-scolarisation en école maternelle le nombre entier naturel a été abordé comme :

- un représentant de la quantité abordée comme une grandeur,
- un représentant d'une grandeur au travers de sa mesure (le cas générique est celui des quantités discrètes),
- un représentant de la position d'un point sur une droite.

Des connaissances ont en outre été enseignées aux élèves qui concourent à la construction du dénombrement par comptage (par exemple les connaissances en énumération, cf. Briand, 2000) ou de la désignation du nombre (numération parlée, numération chiffrée, numération écrite chiffrée de position). Ces fondements didactiques pourraient constituer un référentiel de nature à permettre l'analyse des épreuves portant sur le nombre en fin de grande section mais, ils doivent être actualisés à la lumière des enseignements/apprentissages réalisés en CP. En effet, d'après Mounier (2010, p.7), en arrivant en CP, « les élèves ont une approche des nombres essentiellement en lien avec leur désignation orale, associée le plus souvent à une activité de dénombrement un par un ».

Mercier, Bodin et Quilio (2015, p. 45) affirment qu'il n'y a eu dans le monde francophone « que deux programmes de recherches cherch(a)nt l'administration d'une preuve d'efficacité mesurable par les meilleurs apprentissages des élèves ». Ces deux programmes, ACE et Arithmecole¹¹⁵, sont coordonnés par Sensevy pour l'Ifé. Ils impliquent une centaine de classes de CP.

Cinq « points d'appuis » sont utilisés par les chercheurs qui participent à ce collectif :

- Le premier point d'appui est constitué par les situations visant la maîtrise de la numération décimale de position dans ses formes de notation chiffrée et de système d'adjectifs numériques.
- Le deuxième point d'appui est l'entraînement explicite et systématique du calcul mental appuyé sur la construction de répertoire de résultats qui peu à peu deviennent disponibles dans le travail écrit comme dans l'évaluation des grandeurs discrètes.
- Les troisième et quatrième points d'appui portent sur la résolution de problèmes lesquels sont subdivisés en problèmes concrets et en problèmes additifs abstraits.
- Le cinquième point correspond à la numération orale et écrite et à la traduction de l'un dans l'autre système de désignation.

¹¹⁵ Cf. <http://python.espe-bretagne.fr/ace/>

Nous référant aux travaux de ces chercheurs, trois composants du domaine du nombre et du calcul peuvent être isolés : la connaissance des différentes numérations (parlée et écrite, chiffrée et/ou littérale) et des traductions opérant des unes vers les autres, le calcul mental, la résolution de problème. En effet, le premier et le cinquième point d'appui de l'approche précédemment décrite ont trait aux différents modes de représentation des nombres et aux traductions opérant entre eux et peuvent à ce titre être fusionnés. De plus, nous référant à la typologie de Vergnaud concernant le champ conceptuel des structures additives, les troisième et quatrième points d'appui nous semblent pouvoir fusionner : en situation d'évaluation en CP, il est difficile de distinguer un problème concret d'un problème additif abstrait, laquelle est par ailleurs une distinction non formulée de la sorte dans les programmes.

Le cadre théorique utilisé par Ricco et coll. (2008) dans leur étude de l'hétérogénéité des rapports des élèves au domaine numérique au début et à la fin du CP permet de préciser et de rendre explicite le dénombrement. Pour leur recherche, ces auteurs utilisent cinq « champs » ou types de tâches formant système :

- « énoncer une suite numérique » une épreuve qui prend appui sur les trois premières propriétés du modèle de Fuson (1998),
- « dénombrer une collection discrète » qui prend appui sur les travaux portant sur l'énumération de Briand (2000),
- « comparer des collections discrètes » qui prend appui sur les travaux de Fuson (1998) particulièrement la troisième tâche (« passer d'un cardinal d'objets donné à la construction d'un ensemble »),
- « résoudre un problème de type additif » qui prend appui sur les travaux de Vergnaud (1991)
- « déterminer les ostensifs permanents dans le contexte » qui prend appui sur les travaux de Numa-Bocage (1997) lesquels visent à évaluer l'appropriation de la numération en tant que système aussi bien écrits qu'oraux.

L'un des résultats de cette recherche est d'indiquer que les rapports des élèves à la numération sont hétérogènes ce qui s'exprime particulièrement dans la tâche de résolution de problème. Les auteurs soulignent également que les difficultés des élèves « relèvent de la sécabilité de la chaîne, de la généralisation des techniques de dénombrement à n'importe quel type de matériel et enfin de la maîtrise de la situation fondamentale de dénombrement » (2008).

Nous complétons nos sources concourant à l'édification d'une typologie des composants du domaine par l'étude de Grapin (2015, p.98) qui définit dans sa thèse quatre composants du domaine « numération décimale et arithmétique des entiers » : les types de représentation du nombre, les opérations sur les nombres entiers, les relations arithmétiques entre les nombres et la résolution de problèmes additifs et/ou multiplicatifs. Cette typologie contient un composant « relations entre les nombres » absent des précédentes.

Le croisement de ces différentes sources nous amène ainsi à considérer quatre principaux composants concourant à déterminer d'un point de vue didactique le domaine des nombres et du calcul en CP. Il constitue une proposition d'opérationnalisation de différentes approches du nombre (cf. Roditi, 2005) :

- Les modes de représentation des nombres et les traductions d'un mode à l'autre (numération orale, parlée, écrite).

- Les relations entre les nombres, les situations de comparaison et les relations d'ordre qui les sous-tendent.
- Le calcul, en général, qui peut être découpé en sous-composants (dénombrement, calcul réfléchi (mental, en ligne), calcul posé). Selon Vergnaud (1991, p.272) « sans addition il n'y a pas de nombre, puisque c'est l'addition qui donne au nombre des propriétés distinctes ».
- La résolution de problèmes de type additif.

Les paragraphes suivants nous permettent de présenter certains résultats de recherche et de détailler les trois premiers composants retenus pour analyser les données recueillies. Le quatrième composant (problèmes additifs) fera l'objet d'un chapitre particulier (cf. chapitre suivant)

1.1.1. Les modes de représentation des nombres (ou de désignations) et les traductions

Deux grandes questions se posent de manière centrale. La première est liée à la connaissance de chacun des modes de représentation des nombres et à la possibilité de traduire une représentation d'un nombre d'un mode à un autre et au sujet de laquelle Mercier et coll. (2015, p.37) soulignent que « le nom des nombres en français standard est d'une inutile complexité qui fait obstacle à la compréhension de la numération décimale de position ». La deuxième est liée à la connaissance de la structure décimale du nombre.

Dans le modèle de Fuson (1998), des liens sont établis entre les quantités, les mots-nombres et les chiffres écrits. Collet et Grégoire (2008, pp. 85-86) en proposent une interprétation et isolent trois grands types de connaissances respectivement nommées « traductions¹¹⁶ », « liens de structure » et « procédures de dénombrement en base 10 ».

-Les traductions sont, selon ces auteurs, les connaissances impliquées dans les procédures de passage d'un mode d'expression du nombre dans un autre de ses modes d'expression parmi les six possibles. Ainsi la traduction entre le mot-nombre vingt-trois et son expression écrite est établie si l'enfant qui entend le mot-nombre vingt-trois peut écrire 23.

-Les liens de structures concernent les connaissances relatives à la structure décimale (partie décimale et unitaire) de chaque mode d'expression du nombre et impliquées dans la mise en relation des structures décimales de différents modes d'expression du nombre. Le lien de structure entre la quantité composée de vingt-trois cubes et le mot-nombre vingt-trois est établi si l'enfant peut séparer la quantité en vingt cubes et trois cubes et qu'il sait qu'il s'agit des quantités correspondantes respectivement aux mots-nombres vingt et trois.

-Les procédures de dénombrement, au nombre de trois :

- Une procédure unitaire dans laquelle l'enfant dénombre une quantité constituée en comptant les objets un à un (un trente-deux)
- Une procédure de comptage par dix dans laquelle l'enfant compte des objets par dix puis un par un (dix, vingt, trente, trente et un, trente-deux)
- Une procédure de comptage des dizaines par un (une, deux, trois dizaines, trente-deux)

¹¹⁶ On pourra se reporter à Janvier, 1993 pour une discussion du rôle des traductions dans les processus de modélisation et d'interprétation

Nous délaissions pour l'instant la question des procédures de dénombrement (nous y reviendrons dans le paragraphe consacré au calcul) et approfondissons les deux premiers types de connaissances : « traductions » et « structures ».

Les travaux de Mounier (2010) offrent cette opportunité. Dans sa thèse, il préfère le terme générique de « désignation » à celui de « mode de représentation », et suggère d'adopter le terme numération parlée plutôt que celui de numération orale (2012) « de façon à bien marquer le fait que cette dernière n'est pas l'homologue de la numération écrite chiffrée (avec la graphie des chiffres arabes) qui s'est universellement imposée ». Le terme « chiffré » indique que la numération utilisée est positionnelle. Cette précision permet de considérer comment peuvent opérer les traductions évoquées plus haut par Collet et Grégoire (2008). Mounier (2010, p.70) distingue alors deux suites intervenant dans la structuration de la numération :

- la suite des comptants (« un », « deux », « trois », « quatre », « cinq », « six », « sept », « huit », « neuf », « dix », « onze », « douze », « treize », « quatorze », « quinze », « seize », « dix-sept », « dix-huit », « dix-neuf ») et,
- la suite des repérants (« vingt », « trente », « quarante », « cinquante », « soixante », « quatre-vingts », « cent »). L'obtention de « trente-deux » compris comme « deux après trente » est la traduction d'une interprétation ordinale fondée sur l'usage de repérants. Le premier mot énoncé est toujours un repérant, il est concaténé à un comptant. C'est le principe de postériorité qui est toujours utilisé.

Au terme d'une analyse syntaxique de la numération parlée fondée sur les travaux de Cauty (1980), il dégage (2010, p. 94) des principes mathématiques généraux qui définissent trois types de numération :

1. une numération arithmétique multiplicative fondée sur des décompositions mixtes utilisant des appuis multiplicatifs et additifs,
2. une numération arithmétique additive fondée sur des décompositions additives utilisant des appuis additifs et,
3. une numération ordinale fondée sur l'obtention de cardinaux par des principes ordinaux et utilisant des repérants.

Par exemple, « quarante » a été analysé comme « quatre fois dix » (appuyant multiplicatif/appui multiplicatif), comme « trente plus dix » (obtention de l'appui additif par ajout de dix) et comme « quatrième repère » (succession des repérants). Chacune des interprétations est mieux adaptée à une certaine catégorie de nombres, ce qui met en relief des aspects différents des nombres entiers selon le champ numérique concerné indique Mounier (2010, p.112) : pour les nombres inférieurs à seize, une interprétation ordinale sans repérant ; pour les nombres inférieurs à cent, un choix entre arithmétique additif ou ordinal avec repérant ; pour les nombres inférieurs à mille un choix entre arithmétique additif ou multiplicatif ; au-delà, une interprétation arithmétique multiplicative. Dans la numération parlée, tout se passe comme si deux principes coexistaient : l'ordinal pour les nombres inférieurs à cent (i.e. avant de passer à un deuxième type de groupement/appui/repérant), le cardinal pour les nombres supérieurs à seize. Une zone mixte, de seize à cent, permet à une interprétation arithmétique additive d'être plus pertinente qu'une interprétation arithmétique multiplicative (l'ordinal avec repérant le restant aussi), de fait puisqu'il n'y a qu'un seul type de groupement en jeu. Cet auteur estime que la structure pour les nombres inférieurs à mille (et plus encore cent) a permis d'une certaine manière d'initier des groupements réguliers, qui, entre autres pour des raisons linguistiques que nous avons indiquées, a permis de structurer la numération en base mille.

On notera que d'autres types de représentation (sémiotique) des nombres sont utilisées : au moyen d'un matériel de numération ou de files numérique. Elles sont susceptibles de se retrouver dans les tâches d'évaluation.

Afin d'apprécier la validité de couverture didactique d'un test d'évaluation concernant le domaine des Nombres et du calcul, il est donc nécessaire de rechercher la présence d'épreuves engageant différents modes de représentations des nombres et les traductions permettant de passer d'un mode à l'autre :

- passage (en début de CP) d'une expression d'un nombre dans la numération écrite chiffrée à une expression dans la numération parlée (et réciproque),
- passage (lorsque les élèves sont suffisamment avancés en lecture-écriture, donc plutôt en fin d'année de CP) de l'expression d'un nombre dans la numération écrite littérale à la numération écrite chiffrée (et réciproque).

1.1.2. Les relations entre les nombres

Les relations entre nombres s'appuient sur les relations entre les objets, et « l'activité de comparaison entre objets est évidemment à l'origine du développement des notions d'équivalence et d'ordre qui sont nécessaires au développement de la notion de nombre » souligne Vergnaud (1981, p.83), que celle-ci porte sur des ensembles continus ou discrets (comme par exemple l'ordre d'arrivée dans une course).

Nous rappelons brièvement que les relations d'ordre permettent d'ordonner les éléments de telle sorte qu'il n'y ait pas deux éléments à la même place (dans ce cas on parle de relation d'ordre strict) : « est contenu dans », « est plus petit que » sont de telles relations. Ces relations satisfont trois propriétés mathématiques que nous rappelons.

- Elles sont antisymétriques. Cela signifie que pour tout élément x en relation avec y alors y n'est pas en (même) relation avec x (ou est en non-relation avec x). Si 15 est "plus petit que" 30, 30 n'est pas "plus petit que" 15.
- Elles sont transitives. Cela signifie que pour tout élément x en relation avec y et y en relation avec z alors x est en relation avec z . 15 est "plus petit que" 30 et 30 est "plus petit que" 45 alors 15 est plus petit que 45.
- Elles sont antiréflexives. Cela signifie que tout élément x n'est pas en relation avec lui-même. 15 n'est pas plus petit que 15.

Les relations d'équivalence partagent la propriété de transitivité satisfaite par les relations d'ordre et se distinguent de ces dernières par le respect de propriétés de réflexivité (un élément x est en relation avec lui-même) et de symétrie (tout élément x en relation avec y entraîne que y est en relation avec x). La relation d'égalité est une relation d'équivalence.

Vergnaud affirme (1981, p. 85) que la notion de nombre « s'appuie nécessairement sur les propriétés antérieurement acquises par l'enfant des relations d'équivalence et d'ordre » (sur des ensembles discrets pour commencer comme les cardinaux d'ensembles).

La connaissance de la suite des mots-nombres (comptine ou chaîne numérique) est au fondement des relations arithmétiques entre les nombres. D'après l'étude de Ricco et al. (2008), en début d'année de CP, « la majorité des élèves ont acquis le chaîne numérique au moins jusqu'à 12 » ce qui témoigne selon ces auteurs « de la maîtrise des propriétés mathématiques d'ordre, d'exhaustivité et de successeur » (idem). En revanche, ces auteurs précisent que « les aspects conceptuels liés à la structuration de la chaîne numérique posent

problème et plus particulièrement la sécabilité » (idem). La connaissance de la sécabilité constitue un troisième niveau de connaissance de la chaîne numérique dans le modèle de Fuson (1991)¹¹⁷. A ce niveau, le comptage peut commencer à partir de n'importe quel mot-nombre de la séquence : Ricco et al. (2008) observent plus de 40% d'échec en début d'année à la question « peux-tu compter à partir de 4 ? ». Ces auteurs indiquent qu'en fin d'année, « compter à partir d'une borne pose encore problème à un tiers des élèves » (idem).

La maîtrise des relations entre les nombres, ce qui passe premièrement par la connaissance de la chaîne numérique verbale en début d'année et de sa sécabilité, constitue ainsi un point essentiel de l'évaluation en CP. Les tests d'évaluation devraient contenir au moins une épreuve portant sur ces relations pour couvrir le domaine d'un point de vue didactique.

1.1.3. Le calcul

Plusieurs points sont à distinguer pour ce qui a trait au calcul. Le premier concerne des aspects relatifs aux premiers apprentissages réalisés en maternelle (autour du nombre cardinal et du dénombrement d'éléments d'une collection), le deuxième est lié au calcul, hors de toute représentation de collections.

1.1.3a. Le nombre cardinal, mesure de quantités discrètes

Le cardinal d'une collection est le représentant de toutes les collections qui lui sont équipotentes. Il permet de pouvoir associer un nombre à une propriété d'une collection : la quantité, dont Margolinas et Wozniak (2012, p. 13) précisent que cette propriété n'est pas caractéristique d'un objet isolé mais d'une collection d'objets. Ceci permet de considérer que l'accession à cette propriété sous-entend de pouvoir regarder la pluralité comme une nouvelle entité. Deux ensembles ont le même cardinal lorsqu'ils sont équipotents, c'est-à-dire en bijection. Ceci définit une relation d'équivalence sur les ensembles. C'est une conception « globale » qui est ici convoquée indique Mounier (2010, p.27) puisque « tous les nombres (assimilés à des cardinaux), ainsi que les opérations, sont définis d'emblée sans passer par une construction de proche en proche, sans notion de successeur *a priori*. » Cette conception ensembliste diffère de la définition du nombre établie par Peano : le second axiome indiquant que tout entier naturel n à un successeur $S(n)$.

Margolinas et Wozniak identifient un certain nombre de variables didactiques constituant des paramètres de la situation sur lesquels on peut agir et qui affectent la hiérarchie des stratégies utilisées par les élèves du point de vue de leur coût, de leur validité ou de leur complexité. Parmi elles les trois suivantes sont mises en exergue :

- Le caractère déplaçable ou non des objets constitue une première variable didactique. La correspondance terme à terme des éléments de deux collections peut passer par une action physique que l'on peut empêcher : un lien symbolique devra alors être établi.
- Le caractère plus ou moins identifiable des collections constitue une deuxième variable relevée par ces auteurs. Il est plus aisé d'associer deux collections distinctes

¹¹⁷ Ce modèle en contient quatre : 1. *le chapelet* (les noms nombres ne sont pas différenciés au sein de la suite) et pour produire un mot-nombre particulier, les enfants doivent réciter l'ensemble de la séquence. 2. *La chaîne non sécable* : à ce niveau les mots-nombres sont différenciés mais la chaîne ne peut être produite qu'à partir du début de la séquence. 3. *La chaîne sécable* que nous avons décrite dans le texte. 4. *La chaîne dénombrable* : à ce niveau, la signification de la suite du comptage et de la cardinalité fusionnent.

homogènes (des carottes // des lapins) plutôt que des collections constituées d'objets hétéroclites.

- Une troisième variable concerne l'association des deux collections. Les auteurs citent l'exemple du milieu matériel « œufs-coquetiers » qui oblige et valide l'appariement des objets des collections. « Le milieu d'une situation n'est pas seulement constitué des objets qui permettent le jeu (ou la réalisation de la tâche) » indiquent Margolinas et Wozniak (2012, p. 16), « il est également constitué des mots et des expressions qui permettent de parler du jeu, ainsi que des actions qui sont permises dans le jeu ». Margolinas (1989) a montré l'existence d'interaction entre le milieu et les connaissances antérieures des élèves.

Une situation fondamentale est alors proposée (en maternelle) dans le milieu des œufs et des coquetiers (Margolinas et Wozniak, 2012, p. 18) : « Voici des coquetiers. Vous ne devez pas toucher aux coquetiers ni les utiliser avant d'avoir rempli votre panier. Pour gagner, il faut mettre dans le panier juste assez d'œufs pour pouvoir mettre un œuf dans chaque coquetier. Il ne doit pas rester d'œufs dans le panier. Quand vous aurez rempli votre panier, nous vérifierons si vous avez gagné ou perdu ».

Ricco et al. (2008) soulignent le fait que « la réussite des élèves au dénombrement d'une collection discrète ne dépend pas uniquement de leurs capacités à cardinaliser mais également de la robustesse de leur technique de comptage », ce qui correspond à l'une des tâches du modèle de Fuson (1998). Pour cet auteur, la maîtrise du principe de cardinalité est attestée par trois tâches différentes et hiérarchiques, faisant appel au concept de cardinalité :

1. Savoir passer d'un comptage au cardinal associé
2. Savoir comparer un cardinal donné avec le cardinal issu du comptage
3. Passer d'un cardinal donné à la construction d'un ensemble.

La situation fondamentale des coquetiers oblige l'élève à garder en mémoire le résultat du comptage (ou du dénombrement) des coquetiers (étape 1 du modèle de Fuson). Cela peut être fait au moyen d'une collection équipotente (doigts de la main) ou d'une représentation parlée et/ou chiffrée du cardinal. Un comptage ou dénombrement a ensuite lieu avec les œufs dont l'élève doit comparer le produit au cardinal de la collection des coquetiers (étape 2) jusqu'à obtenir un ensemble correspondant au cardinal qu'il s'est donné à l'issue de l'étape 1 (étape 3).

1.1.3b. Le nombre comme résultat d'un dénombrement et/ou d'un comptage

Hormis le cas de très petites collections d'objets à dénombrer (et qui peuvent favoriser le subitizing¹¹⁸), l'évaluation d'une quantité d'objets requière l'emploi d'une opération de dénombrement et/ou de comptage. Compter est une stratégie de dénombrement qui consiste à utiliser un compteur. La comptine numérique des désignations parlées en est un qui permet de compter un à un. Elle s'adapte à une perception ordinale du nombre. On associe à l'énumération des objets de la collection à dénombrer, les items du compteur (la suite des mots-nombres). Le dernier item utilisé est une désignation du cardinal de la collection.

Le comptage conduit, selon Lemoine et Sartiaux (2005, p.154), à un « codage absolu de l'information numérique » par la récitation de la suite des mots-nombres. Toutefois, Brissiaud

¹¹⁸ Selon Gelman, le subitizing se définit comme « la récitation d'un nom de nombre appris à associer à des modèles visuels donnés » : dans un tel cas, il n'y pas d'énumération des objets de la collection.

(1989, p. 105) souligne le caractère incomplet de cette définition. Selon lui, cette définition « n'évoque que les objets d'une collection comme entités qu'on puisse compter. Or, on peut également compter des groupements d'objets ». Il y a donc lieu selon cet auteur de distinguer comptage-numérotage et dénombrement.

Gelman et Gallistel (1978) et Gelman, (1991, pp. 217-218) ont défini plusieurs principes de dénombrement par comptage largement repris (Fayol, 1990, p.66 ou encore Lemoine 2005, p.154) qui sont à respecter :

1. « Le principe de bijection » : une correspondance terme à terme est établie entre objets décomptés et nom des nombres (appelés *numérons*)
2. « Le principe d'ordre stable » selon lequel la suite des nombres correspond à une séquence fixée. Quelques soient les numérons ceux-ci doivent être répétés dans le même ordre.
3. « Le principe de cardinalité » : le dernier numéron utilisé a le statut spécial de représentant de la valeur cardinale de la collection.
4. « Le principe d'abstraction » : l'hétérogénéité des collections n'a pas d'incidence sur le dénombrement.
5. « Le principe de non pertinence de l'ordre » : le résultat du dénombrement ne se trouve pas modifié selon l'endroit par où on le débute.

Ces travaux ont permis d'indiquer que les enfants respectent très tôt ces principes isolément « mais ont des difficultés à les coordonner lorsque la tâche l'exige » (Fayol, 1990, p. 67) ou « ne comprennent pas que le comptage est lié par certains principes à la valeur cardinale des collections » (Gelman, 1991, p.219). Une célèbre expérience piagétienne (Piaget, 1941) portant sur l'établissement d'une bijection entre deux ensembles d'objets spatialement écartés indique que la mise en correspondance terme à terme de deux ensembles soulève des difficultés jusqu'assez tard dans le développement de l'enfant, « ce qui interdit de considérer que la grandeur d'un ensemble est pour l'enfant indépendante de la configuration spatiale prise par cet ensemble » (Fayol, 1990, p. 83).

Briand (1993) s'est particulièrement intéressé à la question du traitement d'une collection au cours du dénombrement et/ou du comptage et souligne que des compétences non numériques (qu'il nomme énumération) entrent en jeu. Pour cet auteur, le contrôle d'une situation de comptage renvoie pour l'élève à « faire fonctionner une connaissance (l'énumération) qui se réfère à l'exploration de la collection et qui conditionne complètement le bon déroulement de l'activité. [...] une absence de connaissance (l'énumération) se manifeste par une absence de synchronisation effective entre une connaissance numérique et une organisation conjointe de la collection et empêche l'inventaire de la collection » (1993, pp.52-53). Pour qu'un élève compte le nombre d'élément d'une collection finie montrée, Briand (idem) distingue huit phases que l'élève suit au cours de l'activité : « 1. Etre capable de distinguer deux éléments différents d'un ensemble donné ; 2. Choisir un élément de la collection ; 3. Enoncer un mot nombre ; 4. Conserver la mémoire de la collection des éléments déjà choisis ; 5. Concevoir la collection des objets non encore choisis ; 6. Recommencer pour la collection des objets non encore choisis tant que la collection des objets à choisir n'est pas vide ; 7. Savoir que l'on a choisi le dernier élément ; 8 énoncer le dernier mot nombre. ». Si, selon cet auteur, les étapes 3 et 8 concernent la comptine numérique et le comptage proprement dit, les autres correspondent à des connaissances autres, dites d'énumération.

Cela permet de remettre en cause et de dépasser des résultats anciens produits par l'équipe ERMEL (1991) (lesquels sont bien connus des enseignants : ils accordent une trop grande importance au seul produit du numérotage. Ces auteurs indiquaient (1991, p. 23) que « pour 3 élèves sur 4 (« ayant eu une activité numérique limitée en GS »), la connaissance des mots-nombres permet de dénombrer correctement des quantités d'objets (à condition que la quantité soit adaptée à leur propre champ numérique). Pour eux, la capacité à dénombrer est donc, reliée à celle de dire les nombres ».

Différents auteurs se sont attachés à souligner le rôle incomplet des stratégies de comptage numérotage et à insister sur l'importance des décompositions-recompositions de nombres.

Parmi ceux-ci on peut citer

- Margolinas et Wozniak (2012, p.95) qui indiquent que la construction explicite des différents espaces (intervenants dans l'énumération) « a une fonction épistémologique importante : appréhender la pluralité des objets comme une succession d'unités qu'on ajoute. [...] Sans cette prise de conscience l'activité de dénombrement pourrait être perçue comme un simple numérotage des objets pointés. »

-Brissiaud (1989/2003, p.13 ; 1993, p.106) pour qui « l'enseignement du comptage-numérotage est surtout dangereux du fait qu'il permet de réussir (au sens de l'obtention du « bon résultat ») pratiquement tous les problèmes utilisés à l'école, du moins ceux qui mettent en jeu les différentes fonctions du nombre, il est dangereux parce qu'il permet aux élèves d'accéder à la solution en traitant seulement les quantités, pas les nombres ».

Les travaux de Van Nieuwenhoven (1996) permettent de signaler l'aspect lent et progressif de l'acquisition des principes relatifs au comptage et à la cardinalité. Le passage « depuis la signification du comptage quand le mot est dit en comptant vers la signification cardinale de ce mot quand il est donné comme la réponse à la question « combien en tout ? » » nécessite selon Van Nieuwenhoven (1996) un regroupement conceptuel appelé « intégration cardinale » par Steffe et Cobb (1988).

1.1.3c. Calcul réfléchi et calcul posé

En classe de CP, les élèves sont confrontés à un type de calcul reposant sur un ensemble de connaissances exigeant de dépasser les stratégies de dénombrement et/ou de comptage¹¹⁹. Le calcul constitue dans les programmes un domaine d'apprentissage particulier qui s'appuie sur des connaissances relevant du fonctionnement de la numération et sur des connaissances relevant du fonctionnement d'algorithmes de traitement des nombres et des chiffres composant les nombres.

Dans une présentation des enjeux du calcul à l'école primaire effectuée lors de la conférence de consensus sur la numération (CNETSCO, Paris, 2015) Roditi (2015)¹²⁰ indique que les

¹¹⁹ Les élèves peuvent obtenir des résultats numériques en utilisant diverses procédures de comptage fondées sur l'utilisation de la comptine numérique. Ayant deux quantités « à ajouter » les élèves peuvent déterminer l'union des deux collections en recomptant tous les éléments de la nouvelle collection (« recomptage »), ils peuvent aussi « sur-compter » à partir du cardinal de la première collection, « décompter » ou « sous-compter » (dans le cas de soustraction) ou avoir recours au « double comptage » (i.e. utiliser une deuxième suite de nombre permettant de contrôler « de combien il faut avancer » par exemple »)

¹²⁰ Repéré à <http://www.cnetesco.fr/numeration/paroles-dexperts/calcul-et-operations/>

difficultés majeures de calcul rencontrées par les élèves se situent autour de la mémorisation de faits numériques¹²¹, de la mémorisation des algorithmes et de leur compréhension.

L'utilisation des techniques opératoires nécessite des adaptations en fonction des chiffres présents dans l'écriture des nombres et cet auteur insiste sur le fait que ces adaptations sont contrôlées par des connaissances, et non des automatismes. Butlen (2007) a particulièrement montré que les élèves disposant de peu de connaissances numériques (notamment sur les décompositions additives et multiplicatives des nombres) sont ceux qui mobilisent spontanément les procédures de calcul les plus coûteuses (« poser l'opération dans sa tête ») autrement dit qui adaptent le moins leurs procédures de calcul en fonction de la particularité des nombres mobilisés dans le calcul.

Lors de cette même conférence de consensus, Verschaffel (2016) souligne ainsi la nécessité d'enseigner les atouts et l'efficacité de toutes les méthodes de calcul afin que l'élève puisse adapter aux diverses situations rencontrées. Cela fait référence à la distinction effectuée dans l'enseignement entre calcul « posé » (ou « effectué ») et calcul « réfléchi ». Le calcul réfléchi, indique Ivora¹²², désigne « tout calcul pour lequel l'élève n'a pas recours à une procédure automatisée. Autrement dit, un calcul réfléchi est un calcul pour lequel l'élève met en œuvre une procédure, une stratégie personnelle adaptée (au calcul) pour aboutir au résultat. Le calcul réfléchi est basé sur l'utilisation, en général implicite, des propriétés des opérations et de connaissances relatives à la numération décimale ». Le calcul réfléchi est en ce sens plus large que le terme « calcul mental » qui est le plus souvent synonyme d'un calcul qui ne ferait appel qu'à un résultat mémorisé. De ce fait, on pourrait situer le calcul réfléchi dans une zone d'exclusion du calcul posé (en colonne) et de la seule restitution de faits numériques mémorisés (type tables de Pythagore).

Le jury du CNESEO¹²³ (2016, p.10) reprend dans ses conclusions la différence de nature des activités cognitives impliquées dans le calcul mental et le calcul effectué (posé) exposée par les différents chercheurs invités à s'exprimer. Il souligne que « le calcul posé repose sur des procédures standards, il fonctionne sur les chiffres, et il exige de travailler «de droite à gauche», alors que le calcul mental a un fondement heuristique, il opère sur les nombres au lieu des chiffres, et va «de gauche à droite » et recommande de privilégier l'enseignement du calcul mental et du calcul en ligne (2016, p.17)

Il apparaît donc que pour couvrir le domaine du calcul les tests d'évaluation doivent, en classe de CP, comporter des épreuves engageant différentes méthodes de calcul (réfléchi et posé), sans négliger, tout au moins en début d'année, le dénombrement.

¹²¹ Ceci rejoint un constat effectué par Grapin (2015, p.83) « la connaissance des répertoires apparaît comme un élément préalable à la mise en œuvre des différentes techniques (calcul posé ou réfléchi) »

¹²² Article sur la notion de calcul réfléchi exact sur le site TFM. Repéré à <http://www.uvp5.univ-paris5.fr/TFM/AC/AffFicheT.asp?CleFiche=3040&Org=QUTH>

¹²³ <http://www.cnesco.fr/wp-content/uploads/2015/11/Recommandations-du-jury.pdf>

1.2. Analyse des tests et des épreuves

1.2.1. Les tests d'évaluation du 1er trimestre

Nous avons recensé 170 épreuves d'évaluation relatives au nombre et au calcul au premier trimestre¹²⁴. Elles sont comptabilisées dans le tableau ci-dessous en fonction du principal composant du domaine qu'elles évaluent prioritairement. Le composant « calcul » est subdivisé en trois sous-composants (dénombrement d'ensembles figurés entièrement accessibles, calcul réfléchi, calcul posé) comme nous l'avons présenté dans le chapitre précédent.

Ens.	Désignations Représentations	Relations	Calcul		
			Dénombrement	Calcul réfléchi	Calcul posé
A		4		9	
B	2	3		3	
C	1	3	5	2	
D	2	3	1	2	
E	1	4	1	3	
F	2	5		6	
G	3	1	3	10	
H	2	6	2	1	
I	3	5	3	4	
J	4	6	3	3	
K	2	6	2	2	
L	1	3	3	1	
M	1	3	1	1	
N	1	4	1	2	
O	1	3	2	2	
P	2	3	2		
Total	28	62	29	51	0

Tableau 20 : Nombre d'épreuves d'évaluation par aspects du nombre évalué et par enseignant au 1^{er} trimestre.

Excepté l'enseignant A, les enseignants évaluent explicitement les connaissances des élèves liées aux représentations des nombres et aux traductions entre les modes de représentation des nombres au premier trimestre. Dans la majorité des cas l'élève doit traduire un nombre n ($n < 40$) de la numération parlée vers la numération chiffrée.

Les désignations chiffrées des nombres sont ensuite mobilisées dans toutes les autres épreuves sans que les connaissances qui en relèvent ne soient alors évaluées en elles-mêmes. On peut donc concevoir qu'un élève échouant dans une épreuve de désignation échoue dans les autres épreuves. Cela tend à renforcer l'importance d'une telle évaluation des différents modes de représentations des nombres en début d'année.

Le seul composant du domaine étudié à être unanimement évalué est l'aspect ordinal au travers d'épreuves de comparaison de nombres (détermination du prédécesseur, du

¹²⁴ Cf. Annexe C pp. 134-180.

successeur, rangement dans un ordre donné) ; de placement de nombres sur une file orientée débutant à 0, à 1 ou à une autre origine ou dans un tableau à double entrée (ce qui pose par ailleurs le problème de savoir si l'élève sait suffisamment lire un tel tableau en début de CP).

Le composant « calcul » est diversement évalué.

Nous relevons que le calcul posé est une dimension absente de l'évaluation dans les tests du 1^{er} trimestre : il n'est pas enseigné en début de CP.

Nous observons que la plupart des enseignants évaluent la capacité des élèves à dénombrer une collection bien que cet aspect soit absent du programme. Ce résultat important peut être analysé de plusieurs manières : on pourrait y voir l'existence d'une connaissance de métier qui prendrait en compte des connaissances plus spécifiquement travaillées en maternelle, on pourrait aussi y voir l'influence des manuels qui, à l'image d'Euro Maths par exemple, proposent des situations de dénombrement en CP.

Nous constatons enfin que les compétences des élèves en calcul sont évaluées au moyen d'épreuves de calcul réfléchi par presque tous les enseignants. Pour certains tests (tests A et G par exemple), cela constitue même un volume d'épreuve important.

Dans une première approche, on peut estimer que la couverture didactique du domaine numérique i.e. des principaux composants du nombre est correctement réalisée par 12 tests sur 16 soit 75% tests du premier trimestre. On peut surtout exclure de l'étude de validité globale le test A¹²⁵ qui n'évalue pas de façon explicite les connaissances des désignations des nombres suivant les systèmes de numération utilisés ; le test P qui n'évalue pas les connaissances en calcul des élèves et les tests B et F qui n'évaluent pas les compétences de dénombrement des élèves.

¹²⁵ Il correspond au test proposé par l'enseignant A

1.2.2. Les tests d'évaluation du troisième trimestre

217 tâches d'évaluation portent sur ce domaine au troisième trimestre¹²⁶. Cela nous permet d'obtenir, par enseignant et par type de tâche prioritairement visée une carte de la couverture du domaine des nombres et du calcul par les épreuves d'évaluation du troisième trimestre.

Ens.	Désignations Représentations	Relations	Calcul		
			Dénombrément	Calcul réfléchi	Calcul posé
A	4	5	2	6	1
B		4		4	1
C	4	4		5	1
D	2	4		4	1
E	2	7	2	14	1
F	5	5	2	10	2
G		2			2
H	1	2		2	1
I	1	5			2
J				5	1
K	4	10		6	3
L	4	5	3	7	2
M	2	4	1	4	1
N	2	5	2	3	2
O	3	2	1	4	1
P	3	3	2	3	1
Total	37	67	15	77	23

Tableau 21 : Nombre d'épreuves d'évaluation par aspects du nombre évalué et par enseignant au 3^e trimestre.

En fin d'année de CP, les enseignants évaluent dans la plupart des cas les différentes représentations et traductions des modes de représentations du nombre (passer de la numération parlée à la numération chiffrée voire écrite (écrire les nombres en lettres)). Ils évaluent également dans la plupart des cas les relations d'ordre et la comparaison de nombres. Globalement, la part des tâches de dénombrement a diminué. Elle passe de 29 à 15, soit, si on estime en pourcentage de l'ensemble des épreuves de 16,5 % à 7,8 % au profit d'un nouveau type de calcul : le calcul posé (plutôt des additions, mais pas seulement, avec retenue). Les élèves sont massivement évalués en fin de CP en calcul au moyen d'opérations à poser « en colonne » pour lesquelles la technique employée constitue souvent un obstacle.

Le tableau ci-dessous permet de faire état de cette variation : les tâches de calcul sont prédominantes dans l'évaluation en début d'année et le restent en fin d'année. Leur nature

¹²⁶ Cf. Annexe C pp. 190-257

se modifie toutefois suite à l'introduction du calcul posé sans pour autant diminuer la part du calcul réfléchi.

	Désignations Représentations	Relations	Calcul
T1	16,5 %	36,5 %	47 %
T3	17 %	30,9 %	52,1 %

Tableau 22 : Pourcentages d'épreuves de chaque type de tâche d'évaluation du nombre par trimestre

De façon synthétique, et dans une première approche la couverture globale du domaine du nombre et du calcul est susceptible d'être assurée au troisième trimestre par les tests A ; C ; D ; E ; F ; H ; K ; L ; M ; N ; O et P, soit dans 75% des cas.

1.2.3. Validité de couverture didactique des tests au premier et au troisième trimestre

Les tests C, D, E, H, K, L, M, N et O (environ 56% des 16 tests collectés) contiennent au moins une épreuve d'évaluation portant sur chacun des composants du domaine étudié au premier et au troisième trimestre. Dit autrement, ce résultat indique que près d'un test sur deux ne couvre pas, à l'échelle du 1^{er} et du 3^e trimestre, l'ensemble des principaux aspects didactique du domaine des nombres et du calcul. La recherche de validité de ces tests est de ce fait sans objet.

Nous nous intéressons à présent à l'analyse des épreuves constituant ces 9 tests précédemment considérés comme valides au premier et au troisième trimestre au sens où les épreuves qu'ils contiennent couvrent les principaux composants des domaines considérés. Nous recherchons à affiner le résultat précédent concernant l'appréciation de la validité des tests en nous intéressant à leur validité locale i.e. à la validité psycho-didactique des épreuves qu'ils contiennent.

1.2.4. Validité psycho-didactique des épreuves composants les tests couvrant les principaux composants du domaine des nombres et du calcul¹²⁷ au 1er trimestre

1.2.4a Les épreuves portant sur le composant Désignations-Représentations

Nous considérons les épreuves destinées à évaluer les connaissances ayant trait aux différents modes de désignations des nombres et à leur équivalence, présentes dans les tests précédemment considérés comme susceptibles d'être valides. Elles représentent un total de 12 épreuves au premier trimestre et de 24 épreuves au troisième trimestre ce qui montre que ces connaissances sont évaluées au moyen d'un plus grand nombre d'épreuves en fin d'année.

11 épreuves sur 12 portent sur un seul type de traduction : il s'agit de traduire l'expression d'un nombre de la numération parlée dans la numération écrite chiffrée. Pour cela les enseignants procèdent à une « dictée de nombres ». Les nombres en jeu diffèrent suivant

¹²⁷ Tests C, D, E, H, K, L, M, N et O

les enseignants ce qui témoigne de la diversité des exigences : par exemple, l'élève dont la copie est reproduite en Annexe C p. 140 (épreuve 64) doit écrire en chiffre une série de 5 nombres dictés tous inférieurs à 10 (« 8 ; 6 ; 5 ; 2 ; 4 ; 7 ») tandis que l'élève dont la copie est reproduite en Annexe C p. 139 (épreuve 56) doit écrire une série de 9 nombres dictés tous supérieurs à 10 (« 12 ; 25 ; 15 ; 20 ; 16 ; 30 ; 28 ; 38 ; 19 »).

Seul l'enseignant D propose aux élèves une épreuve mobilisant des traductions d'une numération dans une autre et leurs réciproques. Dans l'épreuve reproduite ci-dessous, l'élève doit traduire, comme précédemment évoqué, l'expression d'un nombre de la numération parlée dans la numération écrite chiffrée mais aussi, réciproquement, traduire l'expression chiffrée d'un nombre dans la numération parlée.

A-Nommer les nombres

Code a su / a hésité X n'a pas su ou s'est trompé

10/10

7	25	11	32	17	18	26	13	4	35	31	14
16	8	24	12	33	9	39	28	15	20	6	19

b - Ecrire les nombres dictés

10/10

1	2	3	4	5	6	7	8	9	10
12	5	10	16	20	32	25	4	36	18

Figure 43 : Épreuve MT1 (Annexe C, p.135) avec réponse d'élève

Le test proposé par l'enseignant D couvre ainsi le mieux le composant examiné.

Du point de vue de la tâche de l'élève, on peut estimer que la passation individuelle orale de la première partie de l'épreuve permet de compenser l'éventuelle difficulté de repérage occasionnée par la présentation des nombres à lire dans un tableau (éventuels sauts de case...). Elle permet (probablement) en outre à l'enseignant de s'assurer que l'élève sait établir la correspondance entre « 7 » et « sept » ce qui sera utile pour écrire dans la deuxième partie de l'épreuve le nombre dicté dans la case n°7. La validité psycho-didactique de cette épreuve est attestée : elle engage bien une tâche de désignation-représentation du nombre dans les numérations écrite chiffrée et parlée ainsi que de traduction d'une numération dans l'autre sans que d'autres paramètres n'interfèrent avec la tâche demandée.

Cela constitue un ensemble de preuves de validité, globale et locale, de l'évaluation du composant « Désignations-Représentations » du test D : le test couvre le composant examiné au moyen d'épreuves valides. Ce cheminement permet de souligner que la validité du test ne peut être établie que lorsque celle des épreuves est prouvée.

Nous poursuivons nos analyses au sujet des autres composants du domaine des nombres et du calcul selon le même fil conducteur i.e. vérifier dans un premier temps que les composants étudiés sont bien évalués par les différentes épreuves de chaque test puis

analyser chaque épreuve des tests de façon à nous assurer qu'elles engagent une activité de l'élève correspondante à celle que l'on souhaite qu'il ait au cours de la réalisation de la tâche prescrite. Cela permet alors de revenir au niveau global du test (si l'on retire les épreuves jugées non valides le composant est-il toujours couvert par le test ?) et de conclure quant à la validité globale et locale du test.

1.2.4b Les épreuves portant sur le composant Relations entre les nombres

Les tests étudiés dans ce chapitre contiennent principalement des épreuves qui requièrent l'application d'une relation d'ordre (stricte) « est plus grand que » (ou « est plus petit que ») à deux ou plusieurs valeurs. Dans ces tests¹²⁸, aucune épreuve ne mobilise le symbolisme usuellement attaché aux comparaisons entre nombres (symboles « < », « > » ou « = »).

Deux types d'épreuves peuvent être distingués selon le nombre de valeurs à comparer. 4 enseignants (E ; H ; K et O) utilisent des épreuves dans lesquelles les élèves ont à déterminer le plus petit (ou plus grand) nombre parmi deux. Elles constituent des épreuves que l'on pourrait qualifier de simple dans la mesure où, pour répondre correctement à la tâche de comparaison de deux nombres proposée, l'élève s'appuiera probablement sur sa connaissance plus ou moins automatisée de la place des nombres dans la file numérique. Dans l'exemple ci-dessous extrait du test O : l'élève peut avoir répondu que 6 est plus grand que 5 parce que 5 se situe avant 6 sur la file numérique ou est prononcé avant lorsque l'on récite la comptine numérique.

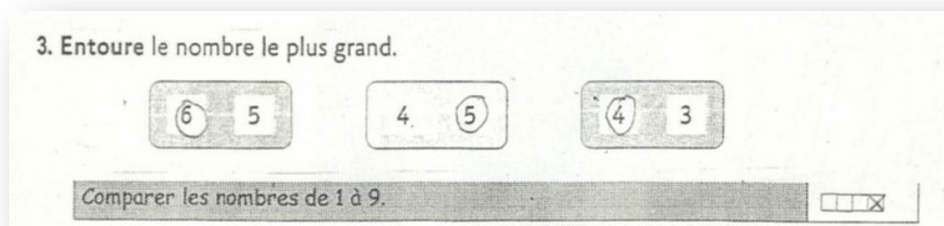


Figure 44 : Épreuve MT1 (Annexe C, p. 155) avec réponse d'élève

Mais dans trois autres cas (tests des enseignants D ; L et M) les épreuves portent sur la comparaison de plusieurs nombres dont la taille est supérieure à ceux contenus dans les épreuves précédentes. Dans l'épreuve ci-dessous (cf. test M¹²⁹), les nombres à comparer sont plus grands et éventuellement supérieurs à 10.

¹²⁸ Le test F ne fait pas partie des tests examinés dans ce chapitre mais on relève que l'épreuve 24-5 qu'il contient (cf. Annexe C, p. 145) mobilise l'utilisation de ces signes de comparaison ce qui serait de nature à invalider l'épreuve.

¹²⁹ L'enseignant L propose une épreuve similaire (cf. Annexe C, p. 152).

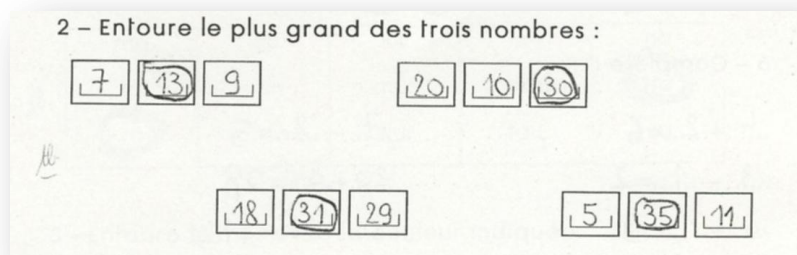


Figure 45 : Épreuve MT1 (Annexe C, p. 153) avec réponse d'élève

Ceci constitue une variable importante de la situation et modifie la tâche de l'élève. Pour réaliser correctement la tâche de comparaison prescrite, l'élève peut énumérer la suite des mots-nombres à partir de 1 (ou du nombre estimé comme étant le plus petit) comme dans le cas précédent, puis éliminer deux des trois nombres pour déterminer le plus grand. Mais cette procédure est relativement coûteuse. L'élève peut plutôt s'appuyer sur des connaissances relatives au fonctionnement positionnel du système de numération pour répondre : « un nombre à deux chiffres est nécessairement supérieur à un nombre à un chiffre » (tâche « comparer 7 ; 13 et 9 ») ou encore « lorsque des nombres ont le même nombre de chiffres, on s'intéresse au chiffre du groupe de base le plus grand » (tâche « comparer 18 ; 31 ; 29 »). Dans le cas de l'épreuve ci-dessous, il semble manifeste que ces connaissances sont exigées des élèves. Ce type d'épreuve est assurément d'une difficulté conceptuelle supérieure à celle précédemment présentée.

Les réponses d'élèves que nous avons collectées n'apportent pas d'information autre que le fait qu'ils aient parfaitement réussi. Mais on peut toutefois s'intéresser au cas théorique où l'élève répondrait correctement à une partie de l'épreuve (isoler que 13 est supérieur à 7 et 9 mais pas que 31 est supérieur à 18 et 29 par exemple). On pourrait alors interpréter la non-réussite de l'élève comme se situant du côté d'une maîtrise insuffisante des règles de fonctionnement du système de numération et non pas comme se situant du côté d'une compréhension insuffisante de la relation d'ordre existant entre deux nombres. C'est la raison pour laquelle la validité psycho-didactique de ces épreuves du 1^{er} trimestre nous semble pouvoir être remise en cause.

A l'exception du test O, les tests étudiés contiennent un autre type d'épreuve mobilisant de façon implicite une relation d'ordre stricte. Ces épreuves requièrent le placement d'un nombre sur une file numérique ou encore la localisation du n-ième objet d'une collection orientée, et mobilisent explicitement la connaissance de la comptine numérique (la suite des mots-nombres). Dans ces épreuves, les élèves ont pour tâche de compléter une file (ou un tableau) dans laquelle certaines valeurs sont déjà positionnées. Cela renvoie à la maîtrise de la sécabilité de la chaîne numérique au sujet de laquelle Ricco et al. (2008) ont souligné qu'en début de CP, elle était imparfaitement réalisée¹³⁰. C'est donc une tâche qui comporte une certaine difficulté pour les élèves au premier trimestre. Le cadre de réponse imposé par l'épreuve, particulièrement de l'usage de tableaux à plusieurs lignes dans des épreuves du type de celle reproduite ci-dessous utilisée par l'enseignant H, a pour effet d'augmenter singulièrement la difficulté de la tâche prescrite en obligeant l'élève à gérer le mode de présentation des données.

¹³⁰ Ces auteurs rapportent que 40% des élèves de CP échouent à la tâche « compter à partir de 4 »

2) Complète le tableau des nombres.

0	1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18	19
20	21	22	23	24	25	26	27	28	29
30	31	32	33	34	35	36	37	38	39

Figure 46 : Épreuve MT1 (Annexe C, p. 148) avec réponse d'élève

L'élève doit compléter les cases du tableau laissées libres avec le successeur du nombre situé dans la case immédiatement placée à sa gauche, ou dans la dernière case de la ligne précédente. Il peut aussi compléter la case libre avec le 10^{ième} successeur du nombre situé dans la case positionnée immédiatement au-dessus (à partir de la deuxième ligne), voire le 10^{ième} prédécesseur. Le zéro présent dans la première case ne modélise pas le point d'origine d'une graduation mais renvoie plutôt à l'absence d'une unité dans un groupement de la base. Cela a pour effet de créer un décalage entre le comptage des cases et celui des nombres : la n-ième case comporte le nombre n-1.

Il est hasardeux d'inférer la stratégie utilisée par l'élève dont la copie est reproduite ci-dessus. On constate simplement qu'il produit toutes les bonnes réponses jusqu'à la 21^{ième} case : il place correctement les 8 valeurs exigées. Des méprises (ou absence de réponse) sont ensuite constatées pour cinq des sept demandes suivantes bien que la cinquième soit correcte (« 31 »). Les erreurs commises (« 23 » ; « 34 » ; « 45 » ; « 76 » ; absence de réponse pour la case devant comporter la valeur 37 ; « 90 ») pourraient avoir pour origine l'environnement graphique très dense de la case réponse lequel serait susceptible de provoquer une surcharge cognitive. Sans entretien avec l'élève, il ne nous est pas possible de nous prononcer davantage, mais il est surprenant de constater que les 8 premières réponses sont correctes quand 1 seule des 7 suivantes l'est alors que l'écriture des nombres ou leur prononciation ne pose pas particulièrement plus de problèmes qu'avant. On peut donc penser que la deuxième partie de la tâche (mathématiquement équivalente) est échouée pour d'autres raisons qu'un défaut de connaissances au rang desquelles figurerait la maîtrise de la lecture d'un tableau à double entrée.

Si l'on place de côté les épreuves dont la validité psycho-didactique est discutable, seuls les tests C, E, H, K et O parmi les tests étudiés contiennent encore des épreuves occasionnant une tâche de comparaison de nombres et des épreuves engageant une tâche de placement d'un nombre sur une file ou dans un tableau. Ils couvrent donc le mieux le composant évalué. Nous n'avons pas détecté d'obstacles particuliers liés à la présentation des données, au système de réponse imposé par l'épreuve ou de compréhension de consigne toutes données oralement dans des termes accessibles dans les autres épreuves que proposent ces tests. La validité globale et locale des tests C, E, H, K et O nous semble assurée pour le composant « Relations »

1.2.4c Les épreuves portant sur le composant Calcul

-Dénombrement

La traduction dans la numération écrite chiffrée du cardinal de la collection obtenu est requise dans tous les cas. A moins d'épreuves portant spécifiquement sur les désignations/représentations, ces épreuves évalueraient donc plusieurs composants en même temps sans les différencier (désignations/représentations des nombres et dénombrement).

Quelle que soit l'épreuve destinée à évaluer la capacité de l'élève à dénombrer une collection d'objets, les collections d'objets à dénombrer possèdent deux caractéristiques communes : elles sont représentées (au moyen de dessin ¹³¹ ou de configurations stéréotypées, aucune situation ne porte sur le dénombrement d'une collection d'objets réels) et elles sont homogènes (il n'y a aucune collection d'objets hétéroclites) ce que l'on peut assimiler à un cas simple. Ricco et al. (2008) soulignent toutefois que « la réussite (au dénombrement d'une collection discrète) ne dépend pas uniquement des capacités à cardinaliser de l'enfant mais également de la robustesse de sa technique de comptage » laquelle est selon ces auteurs, influencée par le contexte de présentation de la collection à dénombrer. Les contextes variés proposés par les épreuves influencent la tâche de dénombrement des élèves.

Certaines épreuves valorisent l'utilisation d'une procédure de sur-comptage en présentant des collections tronquées : l'épreuve reproduite ci-dessous provenant du test C en est un exemple déjà discuté dans la partie III (cf. III 3.3.2, tâche n°12). Elle présente une collection d'éléments dont une partie est masquée. Les élèves savent (c'est un effet de contrat) que la bande cache 5 « billes ». Ils sont donc orientés vers une procédure de sur-comptage (partir de 5 pour arriver à 8). Cette épreuve mobilise une connaissance de sécabilité de la chaîne numérique précédemment évoquée mais oriente fortement l'activité de l'élève



Figure 47 : Extrait d'Épreuve MT1 (Annexe C, p. 157)

Certaines épreuves n'engagent pas cette connaissance et le dénombrement de la collection peut alors être entrepris comme un comptage : à chaque élément de la collection est associé un mot nombre, et le dernier mot nombre prononcé correspond au cardinal de la collection. C'est par exemple le cas de l'épreuve ci-dessous dans laquelle les élèves doivent dénombrer les bougies d'un gâteau. L'alignement des objets à dénombrer favorise d'ailleurs cette procédure linéaire et ordinale.

¹³¹ Cf. Partie III chapitre 3, paragraphe 3.1 pour une analyse de l'impact du dessin dans l'activité de l'élève au cours de l'accomplissement d'une tâche de dénombrement.

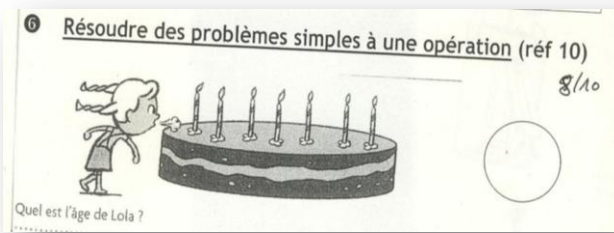


Figure 48 : Épreuve MT1 (Annexe C, p.1 58)

Les épreuves qui présentent des collections éparses d'objets (comme par exemple l'épreuve ci-dessous extraite du test C présentée dans la partie III (cf. III.3.3.2 tâche 15) suscitent outre l'usage d'une stratégie de comptage-numérotage, la mobilisation d'une stratégie que Briand et coll. (2000, p.9) ont qualifiée d' « inventaire » (cf. III-3.3.3) : les élèves s'appuient sur des connaissances spatiales d'énumération (passer en revue tous les éléments d'une collection finie une fois et une seule). Dans l'épreuve ci-dessous, l'enseignant C demande aux élèves de « compter la collection » ce qui sous-entend l'utilisation d'un compteur donc plutôt l'emploi d'une stratégie de comptage-numérotage. Le cadre de réponse présente toutefois un tableau de numération et invite à penser que la tâche de dénombrement de l'élève est évaluée au moyen de la réalisation d'une autre tâche, de numération, consistant à écrire un nombre dans un tableau « de numération ». Pour répondre, les élèves peuvent procéder par comptage-numérotage des éléments de la collection « un, deux... dix-sept » et écrire « 17 » dans le tableau (ou à côté). Ils peuvent aussi (il semble que cela soit la procédure attendue par l'enseignant) constituer un groupe de 10 éléments et compter les éléments restants. L'élève doit alors reporter dans la colonne « d » le nombre de groupes de 10 (« 1 ») puis dans la colonne « u » le nombre d'éléments isolés (« 7 »). Le cardinal de la collection n'apparaîtrait qu'après cette opération.

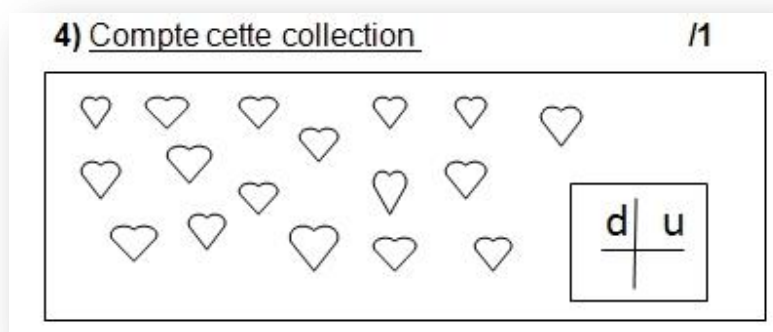


Figure 49 : Épreuve MT1 (Annexe C, p. 157)

L'exigence d'une réponse formulée dans le tableau nous conduit à remettre en cause la validité de l'épreuve : l'élève peut parfaitement être en mesure de compter les éléments de la collection et ne pas être capable d'écrire le résultat dans un tel tableau.

Les épreuves présentant des collections stéréotypées ou pseudo stéréotypées (5 est représenté comme sur une face de dé, 10 comme une combinaison de 2 représentations stéréotypées de 5) semblent offrir davantage de choix de procédures aux élèves pour parvenir au résultat. Dans l'épreuve ci-dessous utilisée par l'enseignant H, dénombrer une collection de 9 doigts peut se faire en comptants les doigts un à un, en sur-comptant à partir de 5, en décomptant depuis 10, voire en calculant et en mobilisant un résultat mémorisé ($5 + 4 = 9$).

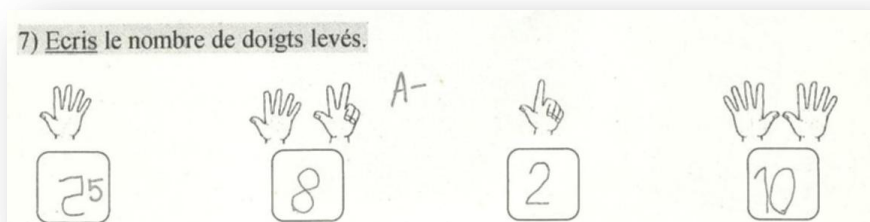


Figure 50 : Épreuve MT1 (Annexe C, p. 160) avec réponse d'élève

Dénombrer les points des cartes (cf. épreuve ci-dessous) peut se faire de la même manière ou encore en groupant les paquets de 5 pour en faire des paquets de 10.

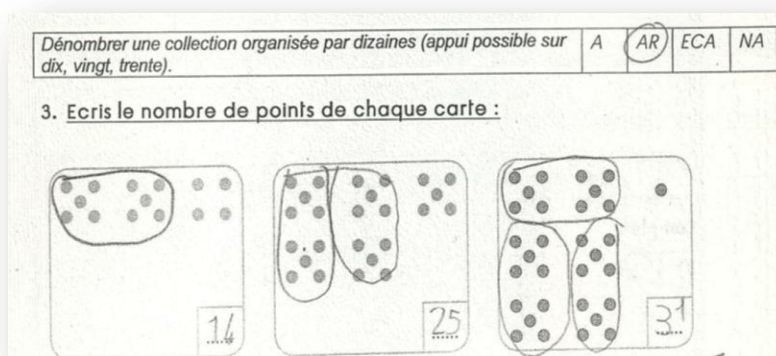


Figure 51 : Épreuve MT1 (Annexe C, p. 164)

Toutefois, comme nous l'avons indiqué dans la partie III (cf. III.3.3.2 tâche n°11) la constitution des groupes de 10 est pour partie prise en charge par le mode de présentation pseudo-stéréotypée des éléments de la collection, si bien que l'on ne sait pas si l'élève saurait dénombrer la même collection d'objets non organisée. D'autre part, l'épreuve va au-delà d'un seul item de dénombrement : il faut non seulement que l'élève dénombre mais aussi qu'il transcrive le résultat du dénombrement d'une numération (parlée, intériorisée) dans une autre (écrite, positionnelle) ce qui implique la production d'une écriture chiffrée à deux chiffres implicitement supposée maîtrisée, ce qu'il n'est pas raisonnable de considérer comme trivial à l'issue du premier trimestre de CP.

Nous pouvons dès lors revenir sur la validité globale des tests estimée grossièrement dans le point 1.2.3.

L'enseignant C utilise des épreuves (cf. Annexe C pp. 157-158) dans lesquelles les élèves doivent déterminer le cardinal de collections au moyen de stratégies de surcomptage (épreuve 72-2), de comptage-numérotage (épreuves 73) ou s'appuyant sur la constitution de groupements par 10 (épreuve 72-4). Pour des raisons développées ci-dessus et dans la partie III (cf. III.3.3.2) la validité psycho-didactique de la plupart des épreuves est remise en cause : le test ne peut donc être considéré comme valide du point de vue global et local.

Le test D n'évalue la capacité à dénombrer des élèves qu'au moyen d'une seule épreuve de comptage-numérotage très sommaire (compter les 6 bougies alignées surmontant le gâteau) qui ne permet pas d'apprécier les compétences d'énumération de l'élève au cours du dénombrement.

Le test E ne contient qu'une seule épreuve reproduite ci-dessous et discutée dans la partie III (cf. III 3.3.2 tâche 14) :

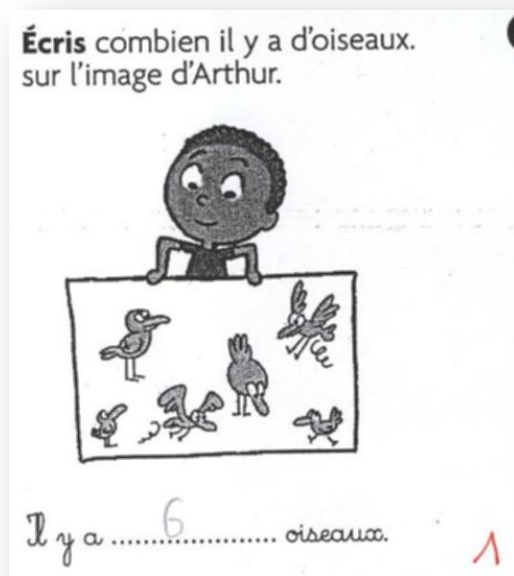


Figure 52 : Épreuve MT1 (Annexe C, p. 159) avec réponse d'élève

Pour produire la bonne réponse, l'élève doit dénombrer les oiseaux représentés sur le dessin puis produire une traduction du nombre obtenu dans la numération parlée (intériorisée) dans la numération écrite chiffrée. La collection présente des « objets » hétérogènes : les oiseaux ne sont pas identiques. Le dessin peut être source de confusion : la ressemblance avec un oiseau n'est pas évidente (cf. dessin de l'oiseau en bas à gauche) et les spirales dessinées pour indiquer un mouvement pourraient être comptées à tort : elles densifient le dessin. Nous avons indiqué en partie III que cette épreuve n'avait pas été utilisée comme épreuve de dénombrement mais pour évaluer l'item du livret « connaître (savoir écrire et nommer) les nombres entiers naturels \leq à 100 ». A condition de réaffecter l'épreuve à un item de dénombrement l'épreuve pourrait malgré les défauts constatés être considérée comme valide elle engage bien une tâche de dénombrement ouverte (sans procédure imposée).

Considérer que le test est valide est sans doute grossier car le test ne contient que cette épreuve portant de surcroît sur des collections de petite taille (inférieures à 10 objets) ne permettant que la mobilisation d'une stratégie de comptage-numérotage. Sans être véritablement mise en défaut la validité globale et locale du test concernant l'évaluation des compétences liées au dénombrement ne nous semble pas être suffisamment assurée.

Nous avons donné plusieurs arguments permettant d'invalider l'une des deux épreuves utilisées par l'enseignant H (cf. épreuve 39 discutée dans la partie III, III.3.3.2 tâche 13). A une tâche de transcription d'un système de numération à un autre s'est surajoutée une tâche de dénombrement non triviale parce qu'elle semble supposer acquise la capacité de l'élève à compter les groupes de 10, autrement dit à adopter une stratégie arithmétique (additive ou multiplicative).

En revanche, la deuxième épreuve du test engage l'élève dans des tâches de dénombrement qui offrent plusieurs stratégies possibles : appui sur une représentation stéréotypée du nombre 5, surcomptage (à partir de 5 pour aller jusqu'à 8) ou comptage-numérotage, subitizing (pour 2), stratégie arithmétique (deux mains de 5 doigts = 10 doigts).

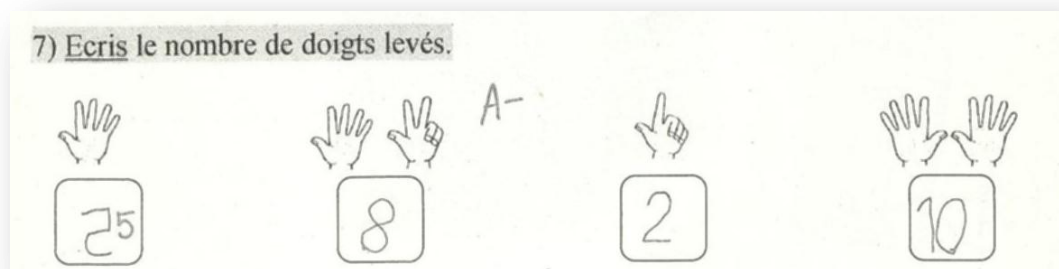


Figure 53 : Epreuve 37-3 (Annexe C, p. 160) avec réponse d'élèves

Le test H peut donc être considéré comme valide sur le plan global et local au niveau de l'évaluation des compétences de dénombrement.

Les tests K, L et O contiennent des épreuves¹³² qui exigent le dénombrement de collections de tailles variées, représentées au moyen de stratégies variées comme nous venons de les décrire pour le test H : subitizing, comptage-numérotage, surcomptage (voire « sous comptage ») groupements. Toutefois, le test L contient des épreuves dont la difficulté a été discutée en partie III (cf. III 3.3.2 tâche 13) parce qu'elles impliquent la production d'une écriture à deux chiffres implicitement supposée maîtrisée, ce qu'il n'est pas raisonnable de considérer comme trivial à l'issue du premier trimestre de CP. La validité du test L est ainsi remise en cause.

Les tests M et N utilisent une seule épreuve mais proposant plusieurs tâches : l'élève doit dessiner des collections dont les cardinaux sont donnés (ou compléter la collection présentée). Pour représenter une collection de 8 objets, l'élève peut s'appuyer sur des connaissances spatiales (configurations pseudo-stéréotypées des nombres), sur des connaissances arithmétiques (8 c'est 4 et encore 4) ou sur un comptage-numérotage (« un

¹³² Cf. épreuves situées en pages 163, 164 et 166 de l'Annexe C

.... huit »). Les tests couvrent le sous-composant « Dénombrement » étudié et les épreuves engagent une activité de dénombrement.

-Calcul réfléchi

Au premier trimestre, que l'élève doive déterminer une somme ou une différence, en ligne ou mentalement, toutes les épreuves proposées portent sur la connaissance de faits numériques mémorisés (tables) généralement désignées par les enseignants comme relevant du calcul « mental » et la capacité à calculer des sommes ou des différences mentalement (qui peut aussi s'appuyer sur des décompositions de nombres) désignées par les enseignants comme étant du calcul « réfléchi ». Dans la mesure où notre mode de collecte ne permet pas de savoir si, pour déterminer la somme de 4 et 5 par exemple, l'élève mobilise un résultat mémorisé ou s'il utilise une autre procédure (surcomptage, décomposition en 4+4+1..) nous avons regroupé toutes ces épreuves sous le terme générique de calcul réfléchi.

On peut souligner deux caractéristiques communes : ces épreuves ne portent que sur l'addition (ou soustraction) de deux termes, et elles ne portent que sur des nombres (les élèves ne peuvent pas adopter de procédures de dénombrement).

L'épreuve utilisée par l'enseignant H reproduite ci-dessous permet de constater que suivant l'opération en jeu et le format de la demande la difficulté de la tâche varie :

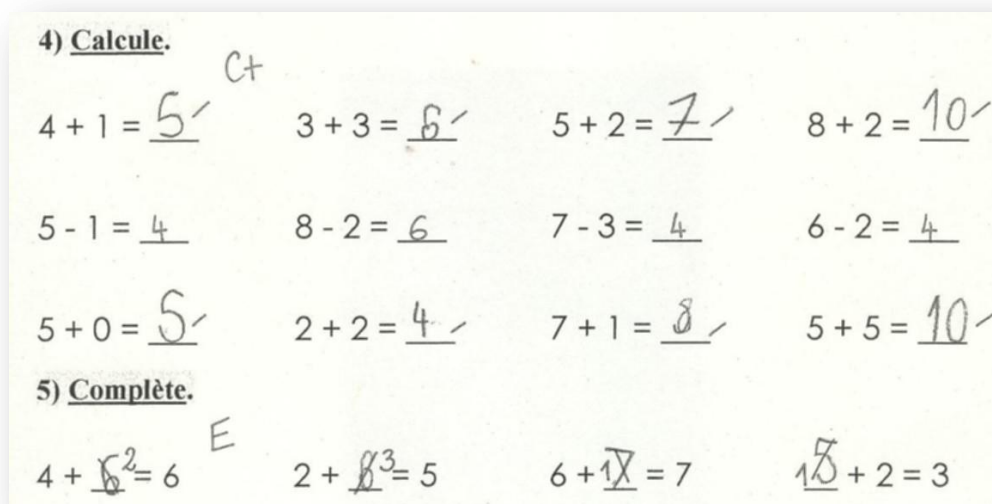


Figure 54 : Épreuve MT1 (Annexe C, p. 176) avec réponse d'élève

Tous les résultats des calculs additifs de la première et de la troisième ligne sont corrects. L'élève a su additionner les deux termes de chaque égalité et en trouver la somme. Il n'a en revanche pas su produire de résultats aux différences proposées (l'enseignant a complété les zones non remplies). L'élève réussit 100% des 8 additions demandées et 0% des 4 soustractions demandées.

Il est possible que le formalisme du calcul constitue un obstacle à l'accomplissement de la tâche : l'élève pourrait ne pas savoir que faire du signe « - » qu'il a, en début de CP, moins

fréquenté que le signe « + », mais être capable de trouver mentalement (sans support de l'écrit) le résultat du retrait de 1 à 5.

La tâche de « complètement » des égalités du type $a+\dots=b$ ou $\dots + a = b$ pose également problème à cet élève : il ne parvient pas à donner une bonne réponse alors que manifestement il sait additionner deux termes.

La validité de cette épreuve est remise en cause : la première épreuve (trouver la somme ou la différence de deux termes) ne prend pas en compte le fait que la tâche de calcul n'est pas homogène suivant l'opération en jeu. Pour un élève de début de CP il est plus aisé de déterminer la somme de deux nombres (cela peut prendre appui sur une mémoire des tables) que de déterminer une différence de deux nombres (qui ne peut pas prendre appui dans les mêmes proportions sur la connaissance de résultats mémorisés). L'appréciation portée par l'enseignant confirme le fait que cette difficulté n'est pas prise en compte : « C+ » est une appréciation globale qui constitue une sorte de moyenne entre une tâche de calcul d'additions en ligne parfaitement accomplie (qui équivaldrait à A dans ce système de cotation), et une tâche de calcul de différences en ligne non exécutée (qui équivaldrait à D dans son système de cotation)

Cette épreuve constituant à elle-seule le test H, la validité du test est également remise en cause ce qui permet de discuter d'un fait évaluatif important : il est fréquent de constater que quel que soit le domaine et le composant évalué (cela est encore plus vrai pour des tâches dont la résolution demande du temps comme répondre à un questionnaire de lecture ou résoudre un problème de mathématique), un test ne comporte qu'une seule épreuve. De ce fait, l'invalidation de l'épreuve conduit à l'invalidation du test.

La présence de calculs additifs (susceptibles de permettre à l'élève de mobiliser des faits numériques mémorisés) et de calculs soustractifs dans la même épreuve se retrouve dans de nombreux cas au niveau de la seule épreuve qui compose le test ou de toutes les épreuves du test relatives à l'évaluation du sous-composant « calcul » étudié : épreuves du test C, H comme nous venons de le voir, K, L, M et N.

Il est intéressant de relever que les mêmes constats que ceux précédemment établis peut être faits au sujet de l'épreuve du test K reproduite page 178 de l'Annexe C : les additions sont correctement réussies, les soustractions ne sont pas traitées (absence de réponse).

La validité de ces tests est ainsi remise en cause.

La taille des nombres en jeu a évidemment un impact en termes de stratégies de calcul mobilisée par l'élève pour répondre. Le test D comporte deux épreuves (cf. Annexe C, p. 171 et 172). La première demande de donner par écrit la réponse à un calcul donné oralement dont nous ne connaissons pas la nature (addition ou soustraction ? deux ou plusieurs termes ?) mais dont nous constatons que le résultat n'excède pas 14. Nous conjecturerons que cette épreuve porte ainsi plutôt sur la connaissance de résultats mémorisés. La seconde épreuve comporte des calculs additifs en ligne pouvant mobiliser ce type de stratégie, mais pas exclusivement, et dont certains calculs nécessitent l'usage d'une autre procédure (calcul de $12+3$). La validité du test au point de vue global et local nous semble assurée : différentes modalités de calcul sont possibles mais la procédure la plus sollicitée repose sur la mémoire de faits numériques. Des constats similaires peuvent être dressés par l'analyse des tests E et O.

1.2.5. Validité psycho-didactique des épreuves composant les tests couvrant les principaux composants du domaine des nombres et du calcul ¹³³ au troisième trimestre

1.2.5a Les épreuves portant sur le composant Désignations-Représentations

Au troisième trimestre, les tests proposés par les enseignants C ; E ; K ; L ; M ; N ; O proposent des épreuves portant sur l'ensemble des différents modes d'expression d'un nombre n ($n < 100$), i.e. dans la numération parlée, dans la numération écrite chiffrée, dans la numération écrite littérale, et nécessitant d'effectuer des traductions d'un mode à l'autre.

On ne trouve pas d'épreuves mobilisant l'écriture de nombres dans la numération écrite littérale dans les tests D et H.

Le test C contient 4 épreuves (Annexe C, p. 191). Les trois premières épreuves engagent successivement l'élève dans des traductions de la numération parlée vers la numération écrite chiffrée (épreuve 31-1), de la numération écrite chiffrée vers la numération parlée (épreuve 31-2) et enfin de la numération écrite chiffrée vers la numération écrite littérale (épreuve 31-3). Une quatrième épreuve (31-4) engage l'élève dans une tâche de traduction multiple. L'élève doit proposer une décomposition additive d'un nombre (figurée ou chiffrée) et/ou une écriture du nombre dans la numération écrite chiffrée et littérale. La validité de ce test nous semble totalement avérée. De telles qualités se retrouvent dans les autres tests étudiés.

4. Observe et complète le tableau suivant : /5


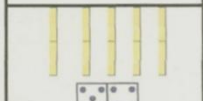
	37	$10 + 10 + 10 + 7$ trente-sept
	42	
		$10 + 10 + 10 + 10 + 10 + 10 + 5$
		
		vingt-trois

Figure 55 : Épreuve MT3 (Annexe C, p. 191)

Cette dernière épreuve comporte toutefois un obstacle orthographique lié à l'écriture des nombres dans la numération écrite littérale qui renvoie non pas à la validité de l'épreuve mais à la validité de la correction.

¹³³ Tests C, D, E, H, K, L, M, N et O

L'une des épreuves proposées par l'enseignant K et reproduite ci-dessous permet d'éviter le recours aux compétences orthographiques des élèves, toujours mobilisée dans les épreuves portant sur l'expression d'un nombre dans la numération écrite littérale :

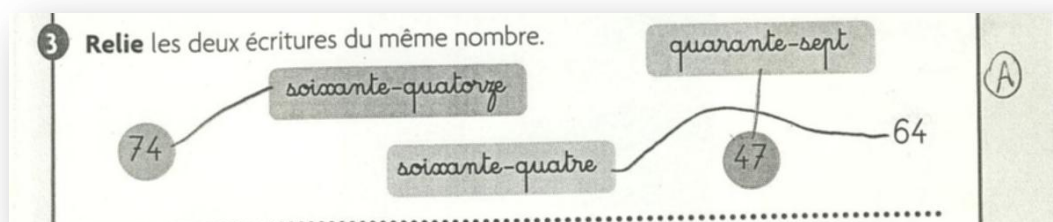


Figure 56 : Épreuve MT3 (Annexe C, p. 195) avec réponses d'élève

Dans cette épreuve, les élèves n'ont pas à écrire les nombres dans la numération écrite littérale. Au lieu de cela, ils ont à effectuer un lien entre deux modalités d'expressions différentes du même nombre dans la numération écrite. La tâche d'écriture est modifiée en tâche de lecture ce qui n'est pas une tâche de la même nature. Considérant la haute difficulté d'écriture des adjectifs numéraux, on peut estimer que cela facilite la tâche de l'élève par rapport au fait de devoir écrire lui-même les nombres et évite ainsi le risque d'une évaluation de la maîtrise de compétences orthographiques que l'on peut estimer expertes¹³⁴.

Pour comparaison, nous reproduisons ci-dessous une épreuve utilisée par l'enseignant F destinée à évaluer l'item d'Écriture du programme : « choisir et écrire de manière autonome des mots simples »

¹³⁴ Hormis cette épreuve, les données collectées ne nous permettent pas d'affirmer ou d'infirmer que les tâches de traduction de l'expression d'un nombre dans les numérations orales et/ou écrite chiffrée dans la numération écrite littérale comportent ou non une part d'évaluation des compétences orthographiques des élèves. Cela renvoie à une question de correction non explorée dans notre travail.



Figure 57 : Épreuve FT3 (Annexe A, p. 242) avec réponses d'élève

Cette comparaison d'épreuves permet de souligner qu'au même moment de l'année (fin de CP) l'écriture des adjectifs numéraux est souvent requise au sein d'une épreuve visant à évaluer la connaissance et la maîtrise des désignations des nombres et des traductions entre les différents modes de représentation, quand sont requises des compétences orthographiques nettement moins expertes (assurer simplement la correspondance graphophonologique) dans une épreuve spécifique d'écriture. Cela nous conduit à estimer que la validité psycho-didactique des épreuves du troisième trimestre requérant l'écriture des adjectifs numéraux dépend éventuellement de la correction : une non-réussite de l'élève pourrait être due à des difficultés orthographiques indépendantes des connaissances proprement mathématiques en jeu portant sur l'équivalence de la désignation d'un nombre dans les différentes numérations considérées.

Un deuxième point relatif aux nombres nécessite d'être observé précisément dans les épreuves. En fin de CP, la lecture/écriture des nombres au-delà de 69 constitue une difficulté bien connue des enseignants. Nous avons constaté que la connaissance de tels nombres était évaluée dans tous les tests.

En conclusion, nous estimons que la validité des tests C, E, K, L, M, N et O est assurée à un niveau global et à un niveau local pour le composant étudié ici.

1.2.5b Les épreuves portant sur le composant Relations entre les nombres

Dans les tests considérés, la majorité des 67 épreuves porte sur la mise en ordre d'une série de valeurs situées dans l'empan numérique] 0, 100], sur la recherche du successeur et/ou du prédécesseur d'un nombre donné ou sur la comparaison de deux nombres éventuellement exprimés sous forme d'une décomposition additive.

Une part des épreuves¹³⁵ nécessite l'usage des signes de comparaison : la réussite de la tâche est alors évaluée au moyen de la maîtrise de l'emploi des signes. Une procédure ordinale peut être entreprise pour comparer deux nombres à un chiffre : 9 se situe plus loin sur la file numérique que 5 ou encore 9 se situe plus loin dans la suite des mots nombres. Une procédure cardinale peut être utilisée qui consiste à estimer qu'une collection de 9 objets (billes par exemple) est supérieure en taille à une collection de 5 mêmes objets. Lorsque les nombres sont plus grands, ce qui est le cas dans ces 8 épreuves, la comparaison de deux nombres peut prendre appui sur les règles de numération. L'aspect cardinal du nombre est mobilisé par référence aux groupements sous-jacents à l'écriture chiffrée (62 est plus grand que 59 parce que le chiffre des dizaines est supérieur dans 62) ou à la désignation verbale. Malgré l'intérêt mathématique de ces épreuves (autour de la multiplicité des procédures de réponse qu'elles occasionnent), leur validité psychodidactique est remise en cause : la maîtrise de l'écriture symbolique avec l'utilisation des signes « < » et « > » n'est pas un objectif du CP : il n'en est d'ailleurs pas fait mention dans les programmes ni dans les progressions qui leur sont annexées.

Dans tous les tests examinés, les enseignants ont recours à d'autres épreuves de comparaison nécessitant l'écriture ordonnée d'une série de 5 à 8 nombres (ordre croissant et/ou décroissant) sans exiger l'usage des signes conventionnels de comparaison. Pour cela les expressions littérales « plus grand » ou « plus petit » sont préférées aux signes qui les représentent. Pour répondre, l'élève doit prendre connaissance de l'ensemble des nombres à ordonner (par exemple dans l'ordre croissant), détecter le nombre le plus petit (dans cet exemple), chercher son successeur dans la liste, le comparer avec le premier pour s'assurer qu'il est bien en relation, écarter de la liste ce nombre (en le barrant par exemple) puis renouveler ce processus avec les nombres restants à comparer et à ordonner. Au fur et à mesure de l'avancée dans la résolution de la tâche la liste des nombres à comparer diminue. La recherche du successeur peut être effectuée en disant la suite des mots-nombres (procédure ordinale) ou en observant la valeur positionnelle des chiffres.

④ Ecrire une suite de nombres dans l'ordre croissant ou décroissant. (réf 4)
Range ces nombres du plus petit au plus grand. 10/10

76	35	53	12	59	97	21
----	----	----	----	----	----	----

→

12	21	35	53	59	76	97
----	----	----	----	----	----	----

Figure 58 : Épreuve MT3 (Annexe C, p. 203) avec réponses d'élève

¹³⁵ 8 épreuves situées en Annexe C : épreuve 32-1 page 202 (test C), épreuve 38-4 page 203 (test D), épreuve 44-2 page 204 (test E), épreuve 78-1 page 210, 82-2 et 83-1 p.211 (test K), épreuve 87-1 page 212 (test L) et épreuve 95-1 page 213 (test M).

Dans cette épreuve, après avoir trouvé que 12 est la plus petite valeur à partir de laquelle il faut ordonner les nombres dans l'ordre croissant, l'élève peut réciter la suite des mots-nombres à partir de « douze » jusqu'au prochain nombre figurant dans la liste : « vingt et un ». Il peut aussi s'appuyer sur les chiffres des dizaines des différents nombres pour répondre (« 21 » est le deuxième plus petit nombre de la liste car « 2 » est le plus petit chiffre des dizaines des nombres restants). Il peut également panacher l'usage de ces stratégies pour répondre (particulièrement si sa maîtrise de la chaîne numérique n'est pas assurée au-delà de 69).

L'utilisation d'une procédure ordinale pour comparer des nombres repose sur connaissance de la sécabilité de la chaîne numérique dont Ricco et al. (2008) ont montré qu'en fin de CP elle « n'est pas complète sur l'empan] 0,16] pour un tiers de la population étudiée ». Hormis l'enseignant K, tous les enseignants proposent des épreuves mobilisant de façon privilégiée cette connaissance. Dans celles-ci, les élèves doivent reporter (dans un tableau à double entrée ou sur une file numérique) un certain nombre de valeurs (prédécesseurs et/ou successeurs d'un nombre donné, nombres situés dans un empan donné...).

Seuls les enseignants E et N proposent des épreuves qui nécessitent de compter de deux en deux voire de 10 en 10 : des tâches qui occasionnent un taux d'échec voisin de 33% dans l'expérience précitée de Ricco et al. (2008). Les réponses reproduites ci-dessous permettent de constater que l'écriture des prédécesseurs ou l'écriture du 5^{ième} successeur d'un nombre donné représente une difficulté non négligeable :

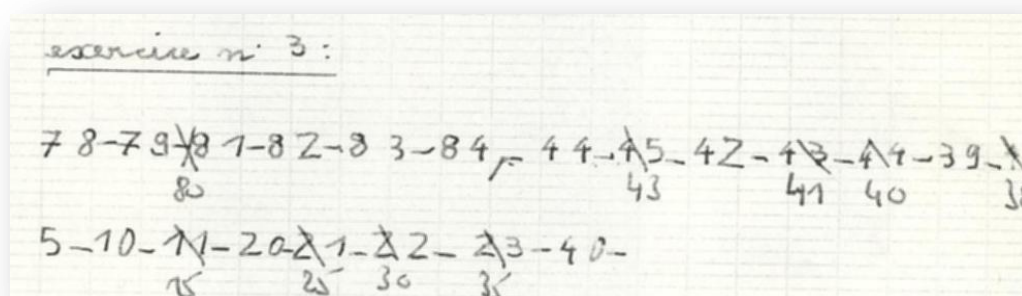
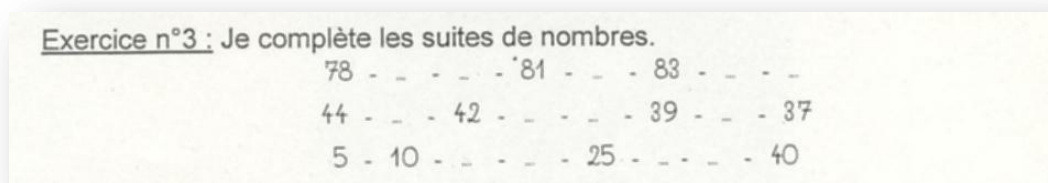


Figure 59 : Extrait de l'épreuve 95 (Annexe C, pp. 213-214) avec réponse d'élève

Ce même élève réussit les épreuves qui exigent la production d'une suite ordonnée de nombres à deux chiffres (cf. épreuves 95-5 et 95-6 Annexe C, p. 213). Il réussit également l'épreuve qui nécessite l'emploi des signes de comparaison (cf. épreuve 95-1, Annexe C, p. 212). L'échec à l'épreuve reproduite ci-dessus n'est donc pas imputable à un défaut de connaissance ou de maîtrise concernant la comparaison de nombres ou la production d'une

suite ordonnée. Il est plutôt lié à la difficulté conceptuelle de la tâche : les exigences de l'enseignant M qui utilise ces épreuves sont plus élevées que celles de l'enseignant O qui n'utilise que des épreuves dans laquelle l'élève doit ordonner une série de nombres dans l'ordre croissant ou l'ordre décroissant (cf. épreuves 108 p. 216, Annexe C).

On peut enfin souligner le fait que la réussite à toutes ces tâches est exclusivement constatée par l'écriture des nombres dans la numération écrite chiffrée de sorte qu'un élève n'ayant pas réussi une tâche d'écriture d'un nombre dans la numération écrite chiffrée (cf. 1.2.4a) ne pourrait pas manifester ses connaissances de la sécabilité de la chaîne numérique. Cela pourrait¹³⁶ constituer un point de discussion de la validité des tests dans le cas où ces connaissances relatives aux modes de désignations et aux traductions entre ces modes ne seraient pas évaluées pour elles-mêmes dans des épreuves spécifiques.

Tous les tests examinés contiennent au moins une épreuve valide portant sur la comparaison de nombres (et/ou la mise en ordre d'une série de nombres). Pour réaliser cette tâche l'élève est susceptible de mobiliser des connaissances relatives à la sécabilité de la chaîne numérique et/ou des connaissances liées à la valeur positionnelle des chiffres dans l'écriture d'un nombre. Les données collectées ne nous permettent pas de savoir le type de procédure utilisée par les élèves. On peut simplement ajouter que les enseignants C, D, E, H, L, M et N proposent d'autres épreuves (compléter une suite en ayant préalablement trouvé la raison qui permet de passer d'un terme n au terme $n+1$; ou positionner des nombres sur une file numérique) portant plus spécifiquement sur la connaissance de la sécabilité de la chaîne : ces tests semblent donc mieux couvrir le composant examiné ce qui ne remet toutefois pas en cause la validité des tests K et O.

1.2.5c Les épreuves portant sur le composant Calcul

-Dénombrement

Parmi les tests étudiés, on ne trouve plus au troisième trimestre, que 11 épreuves réparties dans 5 tests (tests E¹³⁷ ; L¹³⁸ ; M¹³⁹ ; N¹⁴⁰ et O¹⁴¹) dans lesquelles l'élève doit dénombrer ou dessiner des collections d'objets de n éléments (avec $39 < n < 83$).

Selon les données fournies par la DGESCO et la DEPP¹⁴², la tâche de dénombrement des objets d'une collection représentée (pour $n=43$), sans stratégie imposée, occasionne en début de CE2 un taux de réussite situé entre 50% et 80%. Cette tâche serait donc une tâche d'une difficulté non négligeable pour une partie des élèves en fin de CP.

Dans les épreuves analysées, les élèves sont invités de façon plus ou moins explicite à grouper les éléments puis (après avoir dénombré le nombre de groupes de 10 et les unités restantes) à fournir le résultat du dénombrement. L'épreuve comporte ainsi une tâche de dénombrement et/ou de comptage orientée (dans le sens où la consigne demande de réaliser des groupements de 10) et une tâche de numération mobilisant plus ou moins explicitement des connaissances relatives au fonctionnement du système de numération et à la décomposition additive des nombres.

¹³⁶ Les données collectées ne nous permettent pas de savoir si l'enseignant interroge les élèves oralement en cas d'hésitation.

¹³⁷ Annexe C p. 219

¹³⁸ Annexe C p. 221

¹³⁹ Annexe C p. 222

¹⁴⁰ Annexe C p. 223

¹⁴¹ Annexe C p. 224

¹⁴² Cf. Evaluation diagnostique. Livret 3 page 23. <http://eduscol.education.fr/evaluations-CE2>

L'élève dont la copie est reproduite ci-dessous produit une écriture du cardinal de la collection (par ailleurs juste¹⁴³) à l'issue d'un dénombrement entrepris au moyen de groupements par groupes de 10 objets comme demandé. Il effectue l'inverse dans l'épreuve du dessous en produisant une collection organisée pseudo-stéréotypée (dans le sens où elle mobilise plusieurs représentations stéréotypées de 5). Les appréciations (« ab » et « b ») indiquées par l'enseignant laissent entendre que la consigne n'a pas été correctement exécutée. Nous ne disposons pas d'éléments permettant de l'expliquer mais il est évident que les « manques » (sans quoi l'élève obtiendrait « tb » pour « très bien ») ne sont pas de l'ordre du défaut de connaissance relative au dénombrement.

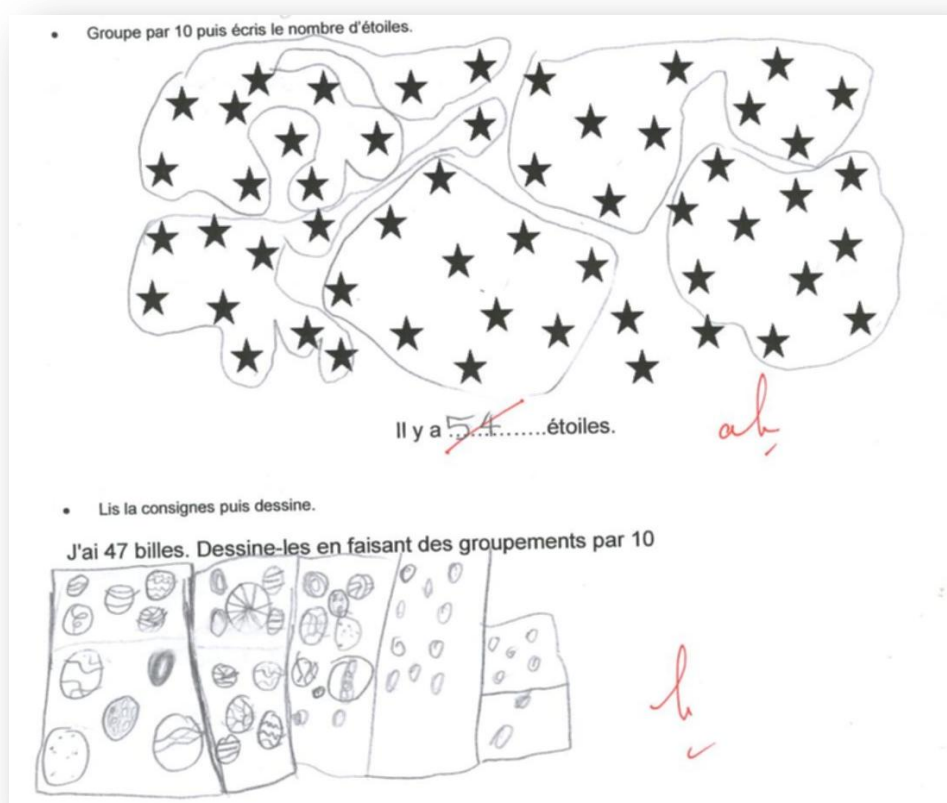


Figure 60 : Épreuve MT3 (Annexe C, p. 219) avec réponses d'élève

Dans certaines de ces épreuves, le système de réponse suggère ou impose l'utilisation d'un « tableau de numération » (une colonne « d » pour dizaine et « u » pour unité). Cela n'est pas neutre dans la réalisation de la tâche (cf. Partie III chapitre 5) et peut pour certains élèves constituer une gêne importante. Une tâche de numération formelle (dans le sens où l'on peut très bien se passer du fait de remplir le tableau pour produire la bonne réponse comme dans le cas des épreuves utilisées par l'enseignant N (cf. Annexe C, p. 223)) se surajoute à la tâche principale de dénombrement et d'écriture du résultat dans la numération écrite chiffrée.

¹⁴³ L'élève groupe bien les étoiles par 10 mais le groupement situé en haut à gauche isole une étoile. Toutefois elle est prise en compte dans le résultat final : à cette étoile l'élève ajoute 5 groupes de 10 plus 3 étoiles restants en bas à droite pour répondre 54 (étoiles).

Les deux exemples suivants permettent de le constater :

1/ L'élève dont la copie est reproduite ci-dessous transcrit « 6-10 » pour exprimer le résultat du dénombrement « 70 ». Ce résultat placé dans le tableau imposé se lit « 6 dizaines et 10 unités » (ce qui est un résultat correct¹⁴⁴). Le système de réponse imposé par l'épreuve modifie la tâche commandée à l'élève qui est de dénombrer les éléments de la collection présentée et d'écrire le résultat de ce dénombrement dans la numération écrite chiffrée. Il surajoute une tâche d'expression d'un nombre dans un tableau de numération dont on constate qu'elle contient des difficultés spécifiques (équivalence 10 unités = 1 dizaine). Sur la base des données recueillies nous ne pouvons pas savoir si l'élève aurait écrit « 610 ». Plus que la tâche de dénombrement et d'écriture du résultat du dénombrement, la tâche d'écriture du nombre dans le tableau implicitement exigée par le système de réponse pose problème : la validité de cette épreuve utilisée par l'enseignant O est remise en cause.

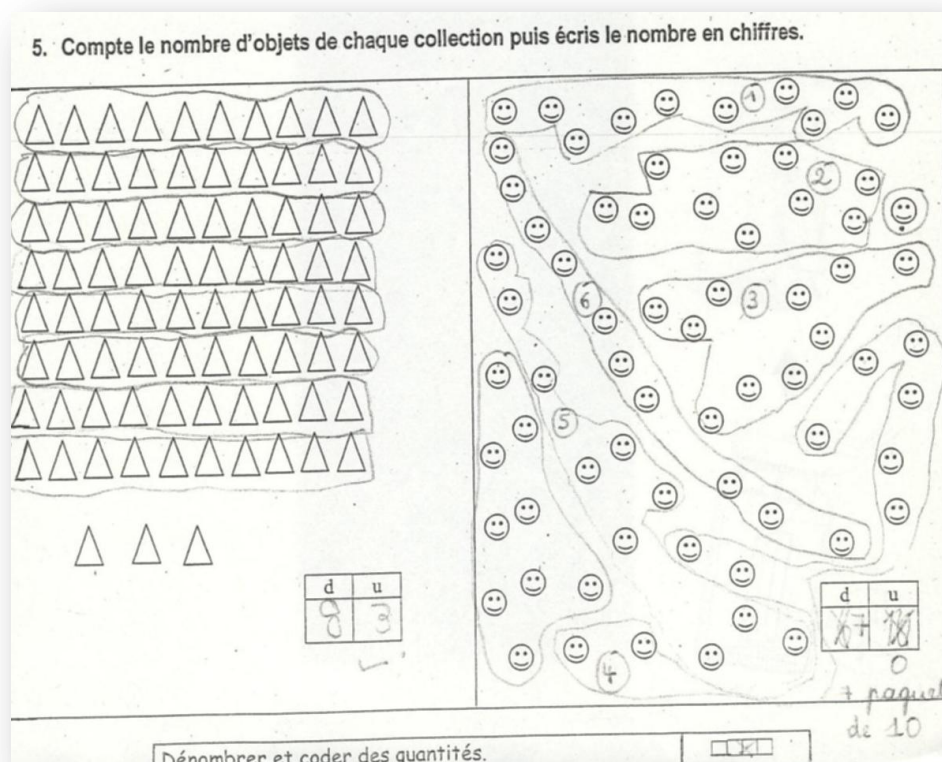


Figure 61 : Épreuve MT3 (Annexe C, p. 224) avec réponses d'élève

¹⁴⁴ La numérotation des paquets de 10 de 1 à 6 conduit à penser qu'il écrit suite à cela « 6 » dans la colonne « d » et qu'il dénombre ensuite 10 objets dans le groupe restant, valeur qu'il reporte dans la case libre dans la colonne « u ».

2/ L'élève dont la copie est reproduite ci-dessous réalise correctement la consigne de groupement et produit la bonne réponse. Comme dans le cas précédent, il lui est demandé de formuler sa réponse dans un tableau de numération imposé. Nous avons précédemment indiqué que l'élève n'en a pas besoin pour produire la réponse : la constitution des 4 groupes de 10 étoiles auxquels s'ajoutent les 3 étoiles restantes peut le conduire à écrire directement le résultat.

Consigne : Fais des paquets de 10 et écris le nombre d'étoiles.

Dizaines	Unités
4 X	3

Il y a 43 étoiles.

Figure 62 : Épreuve MT3 (Annexe C, p. 220) avec réponse d'élève

Le remplissage du tableau de numération est une tâche surajoutée qui peut de surcroît être réalisée après le dénombrement ou après la production de la réponse. Dans le premier cas l'élève est conduit à écrire 4 groupes de 10 et 3 unités (ou déclinaisons du type : 3 groupes et 13), dans le second cas, il peut s'appuyer sur la numération parlée pour dire puis écrire 40 et 3.

Les données recueillies ne nous permettent pas de savoir comment l'élève a procédé mais nous constatons que la démarche utilisée pour dénombrer et le résultat produit sont corrects alors que la tâche de remplissage du tableau de numération est incorrectement réalisée du fait de connaissances d'un autre ordre (40 dizaines équivalent à 4 centaines) : remplir le tableau de numération est clairement une tâche d'une difficulté supplémentaire non négligeable qui ne semble pas prise en compte par l'enseignant.

Le choix du mode de présentation des éléments des collections à dénombrer influe sur le type de stratégie à adopter. Lorsque les éléments sont présentés par groupes de 10 éléments (par ligne ou par colonne¹⁴⁵), la tâche prescrite à l'élève l'oriente vers l'utilisation de ces groupements pour dénombrer, et les compétences d'énumération (cf. IV-1.1.3b) sont moins sollicitées. Lorsqu'à l'inverse les éléments sont présentés de façon non ordonnée¹⁴⁶,

¹⁴⁵ Cf. épreuves 85-4 utilisée par l'enseignant L (Annexe C, p. 221) et 102-2 utilisée par l'enseignant N (Annexe C, p. 223).

¹⁴⁶ Cf. les épreuves 43-3 utilisée par l'enseignant E (Annexe C, p. 219), 86-1 utilisée par l'enseignant L (Annexe C, p. 221), et 102-1 utilisée par l'enseignant N (Annexe C, p. 223).

les compétences d'énumération sont sollicitées. L'enseignant N utilise l'épreuve ci-dessous dans laquelle l'élève doit comparer deux collections dont les éléments sont représentés.

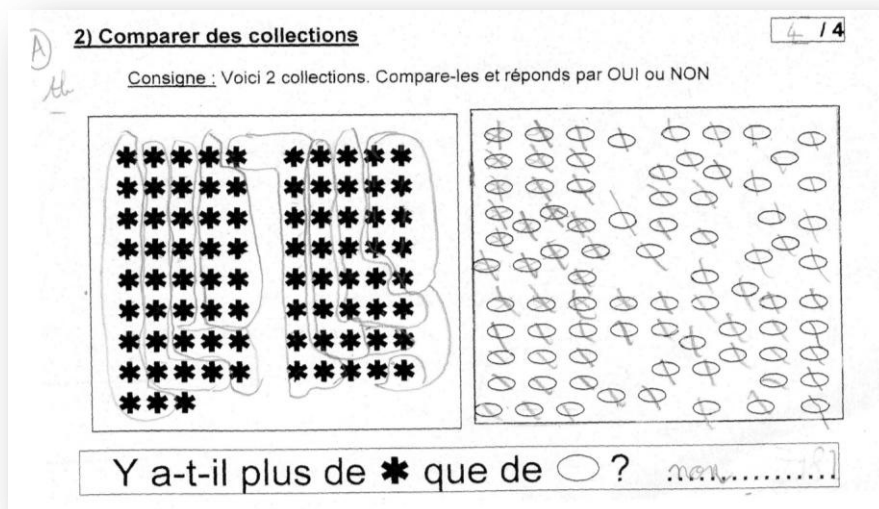


Figure 63 : Épreuve MT3 (Annexe C, p. 223) avec réponse d'élève

Les traces laissées par l'élève dans la partie de gauche permettent de constater le décalage qui s'est installé entre la tâche prescrite à l'élève et la tâche effective. La collection d'étoiles est présentée en 8 lignes de 10 étoiles (sous la forme de deux groupes de 5 séparés par un espace) plus 1 ligne de 3 étoiles. L'élève ne détecte pas cette « logique » et recherche à constituer un groupe vertical de 10 étoiles. Cela le conduit à utiliser des étoiles d'autres colonnes pour former des dizaines. De ce fait, il mobilise des compétences d'énumération alors que l'usage de celles-ci n'était pas prévu.

Les traces laissées par l'élève dans la partie de droite ne montrent pas que l'élève organise la collection en groupes de 10. On peut concevoir qu'il procède au dénombrement par comptage un à un des éléments (chaque élément compté étant barré).

Le cadre de réponse n'exige pas que l'élève indique le cardinal de chaque collection. Sur la seule base de la copie de l'élève, nous ne pouvons pas savoir comment s'effectue la comparaison des cardinaux de chaque collection : l'élève n'a pas écrit ces valeurs. Sur cette même base, l'enseignant n'a pas les moyens de savoir si l'erreur (ou la réussite) de l'élève est fondée sur un dénombrement correct de tous les éléments de chacune des collections puis une comparaison correcte des cardinaux de ces collections (il y a 83 étoiles et 81 ovales, la réponse à la question est donc oui).

En définitive, les épreuves examinées engagent une tâche de dénombrement puis d'écriture des cardinaux des collections dénombrées. Nous avons considéré que l'exigence de l'écriture du cardinal dans un tableau de numération était de nature à invalider l'épreuve utilisée. Ceci a pour conséquence directe la remise en cause la validité du test O qui ne contient qu'une seule épreuve de cette nature.

Nous avons indiqué que les autres épreuves orientaient, plus ou moins explicitement¹⁴⁷, l'activité de l'élève vers la réalisation de groupements par 10. Ces épreuves orientent donc les élèves vers l'utilisation de procédures de calcul fondées sur des décompositions additives et/ou multiplicatives des nombres qui sont ensuite sollicitées dans les épreuves de calcul réfléchi.

-Calcul réfléchi

Au troisième trimestre, la connaissance des faits numériques (tables, doubles et moitiés) est unanimement évaluée dans les tests considérés.

Certaines épreuves sont susceptibles d'engager des démarches propres au calcul en ligne (décompositions / recompositions additives) comme dans l'épreuve reproduite ci-dessous.

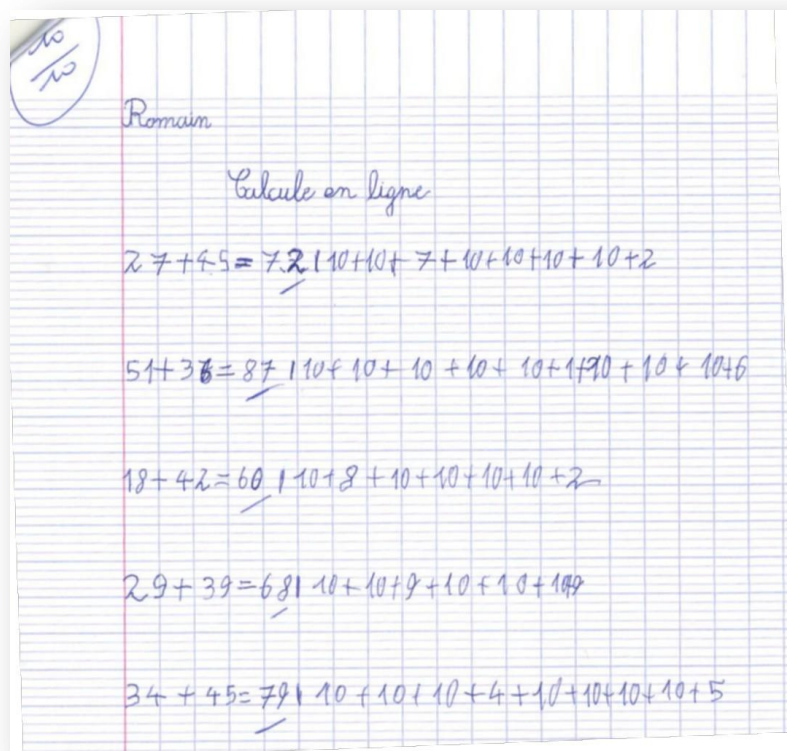


Figure 64 : Épreuve MT3 (Annexe C, p. 232) avec réponses d'élève

Pour répondre l'élève, doit décomposer chaque terme en sommes de 10 et d'unités. Par exemple $29 + 39 = 10 + 10 + 9 + 10 + 10 + 10 + 9$. Les réponses de l'élève ne nous permettent toutefois pas de savoir comment il additionne ensuite 9 et 9 (utilise-t-il un fait numérique mémorisé ? ou décompose-t-il un des deux termes de façon à constituer un nouveau groupe de 10 ?)

Nous relevons peu d'épreuves exigeant de telles décompositions ou recompositions dans les tests étudiés : elles ont pour caractéristiques de dépasser la somme de deux termes.

¹⁴⁷147 L'épreuve 99-1 reproduite p. 222 en Annexe C propose deux tâches dans lesquelles les éléments sont présentés de façon non organisée et deux tâches dans lesquelles des groupes de 10 sont constitués.

Dans l'épreuve ci-dessous, il n'est pas explicitement demandé à l'élève de recomposer le nombre mais les nombres en jeu favorisent l'emploi d'une telle stratégie :

1. Calcule en ligne

$2 + 0 + 5 + 8 + 5 =$ _____	$3 + 5 + 9 + 5 =$ _____	/3
$9 + 7 =$ _____	$6 + 8 =$ _____	
$22 + 45 =$ _____	$37 + 22 =$ _____	

Figure 65 : Épreuve MT3 (Annexe C, p. 228)

Pour trouver le résultat de « $2+0+5+8+5$ » l'élève peut adopter plusieurs stratégies. Il peut additionner successivement chaque terme (en sur-comptant, en combinant résultat mémorisé et surcomptage ...). Il peut aussi rechercher les termes qui, additionnés, équivalent à 10 : il peut ainsi regrouper 2 et 8 puis 5 et 5 et additionner les résultats partiels précédemment calculés pour trouver 20 (notons qu'ici le calcul nécessite la connaissance des différents rôles du 0 : élément neutre de l'addition, marqueur de l'absence d'utilisation d'un groupe de base dans l'écriture d'un nombre). Le nombre de termes à additionner nous invite à penser que la tâche prescrite à l'élève vise l'emploi de la deuxième stratégie. L'épreuve permet à l'élève de la mettre en œuvre bien qu'il lui soit possible de répondre en utilisant une stratégie moins efficace.

La suite de l'épreuve présente des calculs mobilisant deux termes. Soit deux termes inférieurs ou égaux à 10 sont à ajouter, auquel cas ils correspondent à des faits numériques étudiés et mémorisés (tables), soit au contraire ces nombres sont supérieurs ce qui nécessite l'emploi d'autres stratégies en fonction des nombres en jeu (décompositions/recompositions, surcomptage, décomptage (dans le cas d'une soustraction), résultat mémorisé, combinaison de procédures). Dans l'épreuve ci-dessus, les nombres en présence, 37 et 22, sont tels qu'il est coûteux de sur-compter et plus efficace de considérer que $22 + 37 = 20 + 30 + 2 + 7 = 59$. La place laissée pour le calcul agit comme une contrainte : l'élève ne peut pas poser d'addition en colonne.

Cette épreuve provenant du test C nous semble donc particulièrement intéressante : elle engage l'utilisation de diverses stratégies de calcul.

Nous trouvons trace de ce type d'épreuve dans les tests C, D, E, L, M, N, O.

Le test H contient deux épreuves qui ne permettent pas, ou très peu, aux élèves d'engager de procédures de décompositions/recompositions de nombres. La première épreuve (épreuve 72-4 reproduite en Annexe C, p. 235) exige la production de résultats mémorisés (tables). La deuxième épreuve (épreuve 72-5 reproduite en Annexe C, p. 235) exige un calcul (additif ou soustractif) portant sur deux termes. Les calculs additifs demandés pourraient figurer dans la première épreuve : ils font partie des tables. Les calculs soustractifs engagent plutôt les élèves dans une stratégie de décomptage (décompter 4 à partir de 14, 7 à partir de 18 par exemple). Le même type de remarque s'applique au test K.

La validité globale et locale de ces deux tests est remise en cause concernant l'évaluation du sous-composant « calcul réfléchi ».

-Calcul posé

La connaissance de l'algorithme de l'addition (voire de la soustraction) posée est unanimement évaluée dans une proportion qui laisse toutefois une part importante au calcul réfléchi : on trouve en effet dans l'ensemble du corpus, un ratio d'1 épreuve portant sur le calcul posé pour 3 épreuves de calcul réfléchi (« mental » ou « en ligne »).

Selon Roditi (2015), les erreurs susceptibles de se produire ont trois origines possibles :

1. un défaut de connaissance de faits numériques mémorisés (tables),
2. une mauvaise gestion des retenues,
3. un alignement incorrect des nombres.

La connaissance des faits numériques mémorisés a fait l'objet d'épreuves dédiées que nous venons de commenter. Nous laissons donc de côté cet aspect dans la suite de nos analyses.

Tous les tests étudiés dans ce chapitre contiennent des épreuves nécessitant dans un premier temps la pose d'une addition sans retenue puis dans un second temps la pose d'une addition avec retenue. Cela permet de distinguer éventuellement des erreurs ayant pour origine la gestion des retenues. La deuxième origine d'erreur possible est prise en compte par les enseignants.

En revanche, seul le test C (Annexe C, p. 246) propose une épreuve, reproduite ci-dessous, portant sur des nombres de tailles différentes : c'est la seule épreuve qui permet de prendre en compte la troisième source d'erreur repérée relative à l'alignement des nombres.



Figure 66 : Épreuve MT3 (Annexe C, p. 246)

On peut donc considérer que seul le test C est valide du point de vue global et local : il permet d'évaluer, outre la connaissance des faits numériques, l'alignement des chiffres des nombres et (dans une autre épreuve) la gestion des retenues.

1.3. Synthèse

Nous avons synthétisé dans le tableau ci-dessous les constats effectués tout au long du chapitre au sujet de la validité globale et locale des tests et des épreuves concernant l'évaluation des principaux composants du sous-domaine mathématique des nombres et du calcul : « désignations et traductions », « relation d'ordre », « calcul (dénombrement, calcul réfléchi, calcul posé) ». Ils permettent d'affiner l'estimation première de la validité globale des tests que nous formulé précédemment dans le point 1.2.3.

Pour certains tests (tests A, B, F, G, I, J et P) cette étude de validité est sans objet car ils n'évaluent pas tous les composants considérés au moyen d'au moins une épreuve à chacun des deux trimestres considérés¹⁴⁸.

Le tableau ci-dessous permet d'indiquer (par le marquage de la lettre v) selon le composant du domaine Nombre et calcul observé et le trimestre considéré la validité globale et locale des tests couvrant le domaine.

Composant	Trim.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P								
Désignations	T1	Sans objet			v		Sans objet			Sans objet								Sans objet							
	T3			v		v			v			v	v	v											
Relations	T1					v						v			v				v					v	
	T3					v			v			v			v				v	v	v	v			
Dénombrement	T1														v				v		v	v	v		
	T3					v			v			v			v				v	v	v	v			
Calcul réfléchi	T1					v												v							
	T3			v	v	v							v	v	v	v									
Calcul posé	T1			Sans objet (SO)					SO			SO													
	T3			v																					

Tableau 23 : Validité globale et locale des tests pour les composants du domaine du nombre et calcul considérés (composant résolution de problèmes mis à part) au 1er et 3^e trimestres

Nous constatons que selon les critères de couverture que nous adoptés et détaillés, aucun des tests étudiés n'est valide globalement et localement quel que soit le champ et le trimestre considéré.

Aucun test n'est valide sur le plan global et local au premier trimestre.

Au troisième trimestre, seul le test proposé par l'enseignant C contient des épreuves valides permettant d'évaluer les principaux composants examinés : il peut être considéré comme valide sur le plan global et local. Excepté le sous-composant « calcul posé », 4 autres des 9 tests étudiés (tests E, L, M et N) disposeraient au troisième trimestre de la même qualité. Dans ces tests, la couverture didactique du domaine est donc assez satisfaisante et parmi les épreuves qu'ils contiennent, certaines permettent bien d'engager une activité de l'élève correspondant à celle qui est visée. Aucun test du premier trimestre ne s'approche de ce cas de figure : les tests du premier trimestre sont plus lacunaires, ils évaluent moins bien les différents composants du domaine au moyen d'épreuves valides.

¹⁴⁸ Nous n'avons pas collecté de données au deuxième trimestre. Nous ne pouvons donc pas présumer de la validité globale et locale des tests à l'échelle de l'année de CP.

Un défaut de validité au sujet de la désignation des nombres (et des traductions entre les modes de désignations) existe, bien que ces connaissances soient pourtant mobilisées dans toutes les autres épreuves. Cela permet de mettre en évidence une particularité de l'évaluation : s'assurer que des compétences jugées « de base » (dans le sens où elles sont nécessairement mobilisées dans toutes les autres tâches) sont bien évaluées pour elles-mêmes au sein du test au moyen d'épreuves particulières. D'après les résultats obtenus, il semble peu aisé, pour les enseignants, d'évaluer des connaissances « de base » ayant trait aux désignations des nombres dans les différentes numérations et les traductions permettant de passer d'une désignation à une autre au moyen d'épreuves valides. Nous constatons qu'au troisième trimestre, suite à l'apprentissage de l'expression des nombres dans la numération écrite littérale, ces connaissances sont de nouveau évaluées et de meilleure façon (au sens où la validité des tests et des épreuves est assurée dans un plus grand nombre de cas).

Le composant « Relations » est évalué en fin de CP par la majorité des enseignants sans que, pour cela, les élèves n'aient à maîtriser l'usage des signes de comparaison « < », « > » et « = ». C'est le composant le mieux évalué (au sens où l'on constate une validité globale et locale dans le plus grand nombre de cas selon le composant en fonction du trimestre considéré). On peut concevoir que la présence d'un item spécifique rattaché à la comparaison ou la relation d'ordre dans les progressions soit de nature à attirer l'attention des enseignants.

Le sous-composant « dénombrement » donne lieu à une évaluation dont la nature varie : du dénombrement d'objets d'une collection par des stratégies de comptage-numérotage les épreuves évoluent vers des tâches de dénombrement fondées sur des décompositions/recompositions additives, une préoccupation que l'on retrouve prise en compte dans l'évaluation du sous-composant « calcul réfléchi ». Il est intéressant de rappeler que bien que le dénombrement soit absent des progressions sur lesquelles les livrets sont bâtis, les enseignants évaluent ce sous-composant. C'est sans doute sur ce point précis que l'écart est le plus manifeste entre les progressions qui ne mentionnent pas ces compétences explicitement et les programmes qui, à l'inverse, les soulignent. Même si les livrets sont bâtis sur les progressions, sous l'influence des manuels, des programmes ou d'une culture du métier¹⁴⁹, les enseignants estiment important d'évaluer la compétence des élèves à dénombrer les éléments d'une collection. Pour compléter nos conclusions au sujet de l'évaluation du composant « calcul », il est intéressant de souligner que la maîtrise des algorithmes de calcul (addition posée) ne constitue qu'une part de l'évaluation en calcul : en fin de CP, calculer une somme ou une différence ne se fait pas seulement au moyen d'une opération posée.

¹⁴⁹ Nous entendons par cette expression le fait que les enseignants formant la population étudiée ont connu les précédentes réformes des programmes : 1995, 2002. Dans ces derniers la place du dénombrement était explicite : en 1995, le programme mentionne la compétence de « dénombrement des éléments d'une collection et le codage dans le système décimal » (1995, p.48) et en 2002, le programme précise que la compétence « dénombrer et réaliser des quantités en utilisant le comptage un à un ou des groupements et des échanges par dizaines et centaines » est un attendu de fin cycle 2 (2002, p.111).

Chapitre 2. Analyse des tests et des épreuves d'évaluation des compétences de résolution de problème (additif).

Dans ce chapitre, nous analysons les tests et les tâches destinés à évaluer le quatrième composant du domaine des nombres et du calcul précédemment défini (cf. IV. 1.1).

2.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences de résolution de problèmes additifs¹⁵⁰ en CP

Les problèmes additifs constituent le quatrième des composants du domaine des nombres et du calcul précédemment déterminés. Pour Vergnaud, dans la résolution de ces problèmes, les compétences des élèves « ne reposent pas principalement sur leur maîtrise des algorithmes d'addition et de soustraction, mais plutôt sur leur conceptualisation des relations de base » (1981, p. 19). Cet auteur détermine et décrit (1981, p.133) six catégories de relations additives de base pouvant donner lieu à des opérations d'addition ou de soustraction. Elles constituent la base de la typologie de tâches d'évaluation que nous utilisons pour analyser les tests :

1. La relation partie - partie - tout : combinaisons de mesures pour former une troisième mesure.
2. La relation état initial – transformation – état final : transformations opérant sur une mesure pour donner une mesure (les transformations peuvent être positives ou négatives, les états sont des mesures).
3. La relation de comparaison d'états (avoir tant de plus ou de moins que).
4. La composition de transformations.
5. La composition de relations de comparaisons, ou de relations additives quelconques débit/crédit...
6. La transformation d'une relation ou la composition d'états relatifs.

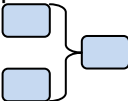
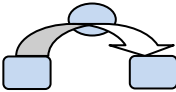
L'auteur souligne au sujet des deux premières catégories de relations que « la réunion de deux parties en un tout constitue un modèle de référence si fort pour l'addition et la soustraction, qu'on a parfois peine à s'en détacher pour analyser d'autres cas d'addition et de soustraction. Or il existe d'autres relations, et les enfants s'y intéressent très tôt au cours de leur développement » (1997, p.11). Chaque type de relation donne lieu à des classes de problèmes suivant ce qui est questionné (par exemple, un problème de relation état initial - transformation - état final peut porter sur la recherche de l'état initial, de la transformation ou de l'état final).

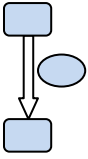
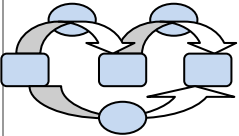
Vergnaud précise que l'enjeu didactique de cette catégorisation est de permettre une classification des raisonnements face à des problèmes additifs, et que « ce sont d'abord ces raisonnements que nous cherchons à développer et évaluer chez les élèves » (1997, p.17) ce qui conduit Maury (2001) à qualifier cette approche de « cognitivo-didactique ».

Nous reproduisons ci-dessous les six catégories de relations additives définies par Vergnaud. Elles donnent lieu à des classes de problèmes dont nous reproduisons chaque

¹⁵⁰ Nous avons collecté trop peu de problèmes multiplicatifs (2 au premier trimestre et 7 au troisième trimestre) pour pouvoir entreprendre une analyse globale et locale des tests et des épreuves. Nous ne détaillerons donc pas le travail de catégorisation entrepris par Vergnaud au sujet des relations multiplicatives (Cf. Vergnaud 1981, p. 177).

fois que possible un exemple provenant du document d'accompagnement des programmes « Le nombre au cycle 2 » (2010, pp. 57-58) et dans lequel ils sont numérotés (n° placé entre parenthèses en fin d'énoncé ou de l'ouvrage de 1981 de Vergnaud.

Catégorie de relations additives	Classe de problème		Exemple de problème
I. Composition de deux mesures pour former une 3^e mesure (Les expressions « en tout », « ensemble » sont souvent employées dans ces problèmes) 	Ia- Connaissant les mesures élémentaires, trouver la composée		Léo a 3 billes. Juliette a 7 billes. Combien de billes ont Léo et Juliette ensemble ? (Pb n°7)
	Ib- Connaissant la composée et l'une des mesures élémentaires, trouver l'autre		Léo et Juliette ont 17 billes ensemble. Juliette a 8 billes. Combien Léo a-t-il de billes ? (Pb n°8)
II. Transformation opérant sur une mesure pour donner une mesure (Une temporalité est marquée (« maintenant », « avant », « hier » ...) et les termes « reste », « gagné », « perdu »... sont souvent employés dans ces problèmes) 	IIa- Recherche de l'état final	1. Transformation positive 2. Transformation négative	Léo avait 3 billes. Puis Juliette lui a donné 5 billes. Combien de billes a maintenant Léo ? (Pb n°1) Léo avait 8 billes. Puis il a donné 5 billes à Juliette. Combien de billes a maintenant Léo ? ((Pb n° 2)
	IIb- Recherche de la transformation	1. Transformation positive 2. Transformation négative	Léo avait 3 billes. Puis Juliette lui a donné des billes. Léo a maintenant 9 billes. Combien de billes Juliette a-t-elle données à Léo ? (Pb n°5) Léo avait 9 billes. Puis il a donné des billes à Juliette. Maintenant Léo a 4 billes. Combien de billes Léo a-t-il données à Juliette ? (Pb n°6)
	IIc- Recherche de l'état initial	1. Transformation positive 2. Transformation négative	Léo avait des billes. Puis Juliette lui a donné 5 billes. Là, Léo a 9 billes. Combien de billes avait Léo ? (Pb n°3) Léo avait des billes. Puis il en a donné 5 à Juliette. Maintenant Léo a 3 billes. Combien avait-il de billes ? (Pb n°4)

Catégorie de relations additives	Classe de problème		Exemple de problème
<p>III. Comparaison d'états On compare deux états (les expressions de plus / de moins sont souvent employées dans ces problèmes)</p> 	<p>IIIa- Recherche de l'état à comparer connaissant l'état comparé et la transformation</p> <p>IIIb- Recherche de l'état comparé connaissant l'état à comparer et la transformation</p> <p>IIIc- Recherche de la comparaison</p>	<p>1. Transformation positive</p> <p>2. Transformation négative</p> <p>1. Transformation positive</p> <p>2. Transformation négative</p> <p>1. Comparaison négative</p> <p>2. Comparaison positive</p>	<p>Léo a 3 billes. Juliette a 5 billes de plus que lui. Combien de billes Juliette a-t-elle ? (Pb n° 9)</p> <p>Léo a 9 billes. Juliette a 5 billes de moins que lui. Combien de billes Juliette a-t-elle ? (Pb n° 10)</p> <p>Léo a 9 billes. Il en a 7 de plus que Juliette. Combien de billes Juliette a-t-elle ? (Pb n° 11)</p> <p>Léo a 9 billes. Il en a 5 de moins que Juliette ? Combien de billes Juliette a-t-elle ? (Pb n° 12)</p> <p>Léo a 8 billes. Juliette en a 6. Combien de billes Juliette a-t-elle de moins que Léo ? (Pb n° 13)</p> <p>Léo a 3 billes. Juliette en a 9. Combien de billes Juliette a-t-elle de plus que Léo ? (Pb n° 14)</p>
<p>IV. Composition de transformations</p> 	<p>Iva- Connaissant les deux transformations élémentaires, trouver la composée</p> <p>IVb- Connaissant la composée et l'une des élémentaires, trouver l'autre</p>	<p>3 cas possibles : transformations positives, négatives ou de signes opposés</p>	<p><i>Paul a gagné 6 billes hier et il en a perdu 9 aujourd'hui. Combien en a-t-il perdu en tout ? (Vergnaud, 1981, p. 136)</i></p> <p><i>Paul a gagné 2 billes en tout. Il avait pourtant gagné 6 billes au départ. Combien en a-t-il perdu ensuite ?</i></p>

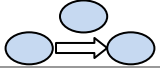
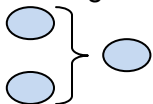
Catégorie de relations additives	Classe de problème	Exemple de problème
<p>V. Composition¹⁵¹ de relations de comparaisons, ou de relations additives Une transformation opère sur un état relatif pour donner un état relatif.</p> 		<p><i>Paul devait 6 billes à Henri. Il lui en rend 4. Combien lui en doit-il encore ? (Vergnaud, 1981, p. 136)</i></p>
<p>VI. Transformation d'une relation Composition d'états relatifs Deux états relatifs¹⁵² se composent pour donner un état relatif (proche de la 4^{ième} catégorie)</p> 		<p><i>Paul doit 6 billes à Henri mais Henri lui en doit 4. Combien de billes Paul doit à Henri ? (Vergnaud, 1981, p. 137)</i></p> <p><i>Paul doit 6 billes à Henri et 4 billes à Antoine. Combien de billes doit-il en tout ? (Vergnaud, 1981, p. 137)</i></p>

Tableau 24 : Catégories de relations additives et classes de problèmes additifs (d'après Vergnaud, 1981)

Bien que portant toutes sur des relations additives, Vergnaud prévient que les six catégories de problèmes ainsi déterminées ne sont pas de difficulté égale¹⁵³. Il faut en outre considérer que « la facilité du calcul numérique nécessaire, l'ordre et la présentation des informations, le type de contenu et de relations envisagés, constituent autant de facteurs susceptibles d'affecter la difficulté des problèmes » (1981, pp. 141-143).

La typologie de relations additives de Vergnaud que nous venons de présenter permet de mettre en évidence un grand nombre de classes de problèmes différentes relatives à six grandes catégories de relations additives.

Toutes les classes de problèmes ne sont pas enseignées au niveau du CP. D'autres travaux portant sur la réussite des élèves selon le type de problème proposé nous permettent d'affiner les critères que nous retenons pour évaluer la validité de couverture du test d'un point de vue didactique.

¹⁵¹ Dans le cas présent, cela correspond à la recherche de ce que l'on doit encore sachant que l'on devait 4 et que l'on a 10. Cette catégorie se différencie de la précédente en ce que la loi de composition ne s'applique pas à deux transformations.

¹⁵² Vergnaud précise que ce qui différencie cette catégorie de la quatrième c'est l'aspect « contemporain » des états relatifs quand on les compose, ce qui n'est pas le cas des transformations en jeu dans la quatrième catégorie.

¹⁵³ La complexité est d'ordre conceptuelle ce qui constitue une différence importante avec l'approche de Rey et coll. (2006). Cf. I-2.1.1

Thevenot, Barrouillet et Fayol (2010, p. 160) se réfèrent à une autre typologie (Riley et coll. 1983) comportant de nombreux points communs avec la typologie de Vergnaud et dans laquelle trois types de problèmes sont distingués. Ils renvoient aux trois premières catégories de Vergnaud (dans le même ordre) : les problèmes de combinaison, de changement ou de comparaison.

Ces auteurs apportent des éléments de nature à caractériser la difficulté des problèmes selon la catégorie et la classe à laquelle ils appartiennent.

Ils indiquent notamment que (id. p. 160) :

-« les problèmes de type changement (catégorie II) sont mieux réussis par des enfants de CE1 et de CE2 (93% de réussite) que par des enfants de maternelle et de CP (64% de réussite) »,

-« la recherche de l'état final (classe IIa) ne pose jamais de problèmes mêmes aux enfants les plus jeunes »,

-« les problèmes de combinaison (catégorie I) sont réussis massivement et de manière précoce lorsqu'ils portent sur le tout (classe Ia) mais leur réussite est plus tardive lorsqu'il s'agit de déterminer le cardinal de l'une des parties (classe Ib) (30% de réussite en maternelle et en CP) »,

-« les problèmes de comparaison (type III) présentent un niveau de difficulté bien supérieur aux deux autres types de problèmes ».

Croisant la catégorisation de Vergnaud et les résultats décrits par Thevenot, Barrouillet et Fayol, on pourrait donc estimer qu'en CP, un test est valide du point de vue de la couverture des contenus didactique s'il contient des épreuves portant sur des problèmes de combinaison (Catégorie I de Vergnaud) et particulièrement des problèmes portant sur la détermination de la composée de mesure (Classe Ia), et s'il contient des problèmes de relation entre un état initial, une transformation et un état final (Catégorie II de Vergnaud) portant a minima sur la recherche de l'état final. De ce point de vue, on pourrait donc estimer qu'une validité de couverture minimale du composant serait assuré si le test contenait au moins une épreuve de chaque classe de problèmes (Ia et IIa).

2.2. Analyse des tests et des épreuves

Nous cherchons à présent à apporter des preuves de validité de contenu didactique des tests et des preuves de validité psycho-didactiques des épreuves destinées à évaluer le composant « Problème additif » du sous-domaine des nombres et du calcul.

2.2.1. Les tests d'évaluation du 1er trimestre

Nous avons recensé 26 problèmes additifs¹⁵⁴ utilisés pour évaluer la compétence de résolution de problème (additif) au premier trimestre. Le tableau ci-dessous indique, pour chaque test, le nombre de problèmes appartenant aux classes et catégories de problèmes précédemment détaillées. Une ligne sans valeurs signifie que l'enseignant n'a pas proposé de problème d'évaluation au cours du trimestre.

Ens.	Composition de mesures (catégorie I)		Transformation opérant sur une mesure pour donner une mesure (catégorie II)					
	Ia.Recherche de la composée	Ib.Recherche d'une mesure	IIa.Recherche état final		IIb.Recherche transformation		IIc.Recherche état initial	
			Transfo. positive	Transfo. négative	Transfo. positive	Transfo. négative	Transfo. positive	Transfo. négative
A			1					
B			2	3				
C								
D	1				1			
E	1	1			1			
F			1	1				
G	1		2			1		
H	1							
I	1			1				
J					3			
K								
L	1	1						
M		1						
N								
O			1					
P								
Total	6	3	7	4	5	1		

Tableau 25 : Répartition des problèmes additifs collectés au 1^{er} trimestre par enseignant et, par catégorie et classe de problèmes (cf. catégorisation présentée dans le tableau 24)

Nous constatons que les enseignants C, K, N et P (i.e. 25% des enseignants de la population étudiée) ne proposent pas d'épreuves de résolution de problème au premier trimestre. La recherche de preuves de validité de ces tests est donc sans objet pour ce trimestre.

Les tests A, H, M et O ne proposent qu'une seule épreuve ce qui rend également sans objet la recherche de preuve de couverture du composant « problèmes ».

Les 26 problèmes se répartissent dans six classes de problèmes à l'exception des deux classes de problèmes nécessitant la recherche de l'état initial connaissant l'état final et la transformation. Nous avons précédemment estimé qu'en raison de leur plus grande réussite par des élèves de CP, la présence d'au moins un problème portant sur la recherche de

¹⁵⁴ Cf. Annexe C pp. 181-188. Cela constitue un corpus supérieur à ce que les enseignants déclaraient affecter à l'item du programme « résoudre des problèmes simples à une opération » : 12 problèmes de type additif au premier trimestre et 20 problèmes au troisième trimestre. Ceci s'explique par l'affectation réalisée par les enseignants de problèmes à d'autres items du programme que l'item « résoudre un problème simple à une opération » du sous-domaine du nombre et du calcul, comme par exemple à l'item « résoudre des problèmes de vie courante » rattaché au sous-domaine des grandeurs et des mesures.

composée de mesures (Ia) et d'un problème de recherche d'un état final après transformation (IIa) constituait un premier indicateur de validité didactique du test. Si ces deux classes de problèmes sont les plus représentées dans le corpus de données recueillies (un problème sur deux appartient à l'une ou l'autre de ces classes), on ne trouve cette caractéristique de l'évaluation que dans les tests G et I. Le test G couvre davantage le composant examiné : il contient un problème de recherche de la transformation d'un état initial connu connaissant également l'état final.

Si le composant examiné est couvert a minima par les tests G et I, en termes de validité de couverture du composant examiné, la validité des tests est nulle au premier trimestre : si deux classes de problèmes sont écartées de tous les tests, aucun test ne couvre les six autres classes de problèmes.

2.2.2. Les tests d'évaluation du troisième trimestre

Nous avons recensé 26 problèmes au troisième trimestre¹⁵⁵ utilisés pour évaluer la compétence de résolution de problème (additif) au troisième trimestre. Le tableau ci-dessous indique, pour chaque test, le nombre de problèmes appartenant aux classes et catégorise précédemment détaillées. Une ligne sans valeurs signifie que l'enseignant n'a pas proposé de problème d'évaluation au cours du trimestre.

Ens.	Composition de mesures (catégorie I)		Transformation opérant sur une mesure pour donner une mesure (catégorie II)					
	Ia. Recherche de la composée	Ib. Recherche d'une mesure	IIa. Recherche état final		IIb. Recherche transformation		IIc. Recherche état initial	
			Transfo. positive	Transfo. négative	Transfo. positive	Transfo. négative	Transfo. positive	Transfo. négative
A	1			1				
B	2			2				
C	2							
D	2			2				
E	2							
F	1		1	1				
G	1			2		1		
H	1			1				
I								
J								
K								
L				1				
M	1							
N								
O	1							
P								
Total	14		1	10		1		

Tableau 26 : Répartition des problèmes additifs collectés au 3^e trimestre par enseignant et, par catégorie et classe de problèmes (cf. catégorisation présentée dans le tableau 24)

Nous constatons que les enseignants K, N et P qui ne proposaient aucune épreuve de résolution de problème additif au premier trimestre n'en proposent pas davantage au

¹⁵⁵ Cf. Annexe C pp. 258-267.

troisième trimestre. Les enseignants I et J n'évaluent pas non plus le composant « problèmes additifs » au troisième trimestre. La recherche de preuves de validité de ces tests est donc sans objet pour ce trimestre.

Les tests L, M et O ne proposent qu'une seule épreuve ce qui rend également sans objet la recherche de preuve de couverture du composant « problèmes ».

Les 26 problèmes se répartissent dans quatre classes de problèmes seulement. Nous avons précédemment estimé qu'en raison de leur plus grande réussite par des élèves de CP, la présence d'au moins un problème portant sur la recherche de composée de mesures (Ia) et d'un problème de recherche d'un état final après transformation (IIa) constituait un premier indicateur de validité didactique du test. Si ces deux classes de problèmes sont les plus représentées dans le corpus de données recueillies (92% des problèmes additifs appartiennent à l'une ou l'autre de ces classes), on ne trouve de tels problèmes que dans les tests A, B, D, F, G et H.

Les tests F et G couvrent davantage le composant examiné : ils contiennent des problèmes appartenant à trois classes de problèmes différentes.

En termes de validité de couverture du composant examiné, la validité des tests est nulle au troisième trimestre : quatre classes de problèmes sont écartées de tous les tests (ce qui est davantage qu'au premier trimestre).

2.2.3. Validité de couverture didactique des tests au premier et au troisième trimestre

Aucun test ne couvre les différentes classes de problèmes relatives aux deux catégories de relations additives prises pour référence. La recherche de validité à la fois globale et locale des tests et des épreuves qui les constituent est de ce fait sans objet.

Parmi les tests examinés, seul le test G couvre a minima le composant examiné au premier trimestre et au troisième trimestre : il comporte un problème de recherche de composée de mesures (Ia) et un problème de recherche d'un état final après transformation (IIa) à chacun des trimestres considérés.

Nous constatons que les tests du troisième trimestre comportent moins de problèmes variés (en termes de classes de problèmes auxquels ils appartiennent) qu'au premier trimestre.

Cette conclusion nous conduit à une discussion relative à la formulation de l'item « résoudre un problème simple à une opération » figurant dans les progressions. Nous avons souligné le rôle important de guidage de l'évaluation occupé par les items des progressions dans la partie II de la thèse. Ni ces progressions ni les programmes ne définissent clairement ce qu'est un tel problème. Nous nous posons donc la question de savoir comment on peut caractériser un problème « simple à une opération » en fin de CP à travers l'analyse des problèmes donnés en évaluation en fin de CP (ce sera le moyen d'y répondre).

2.3. Discussion autour du « problème simple à une opération »

2.3.1. Le problème « simple à une opération » dans les programmes

Les programmes de 2008 mentionnent les problèmes dans deux sous-domaines distincts : dans le sous-domaine « Nombre et calcul » au moyen de l’item « résoudre des problèmes simples à une opération »¹⁵⁶ et, dans le sous-domaine « Grandeur et mesure » au moyen de l’item « résoudre des problèmes de la vie courante ».

Ces programmes présentent par ailleurs un sous-domaine particulier dénommé « Organisation et gestion de données » dont les termes sont assez proches du sous-domaine de rattachement des problèmes dans les précédents programmes datant de 2002 intitulé « Exploitation de données numériques ».

Ce préambule permet de dire que la structuration des programmes de 2008, rompant avec la logique des programmes 2002 risque d’induire une certaine confusion chez les enseignants quant au fait de savoir à quel sous-domaine des mathématiques rattacher un item spécifique à la résolution de problèmes (numériques mais aussi géométriques ou logiques).

8 enseignants¹⁵⁷ sur 16 réfèrent la résolution de problèmes au sous-domaine des nombres et du calcul des programmes¹⁵⁸. Exception faite de l’enseignant I qui ne présente aucun item relatif à la résolution de problème dans le livret qu’il utilise, les autres enseignants réfèrent à d’autres sous-domaines un item propre à la résolution de problèmes. Parmi eux, l’enseignant C réfère cet item au sous-domaine « Résolution de problèmes » : ce sous-domaine existait dans les programmes 2002 mais n’existe plus dans les programmes de 2008. Ceci montre bien l’influence passée des précédents programmes pour cet enseignant. L’enseignant G réfère cet item au sous-domaine « Calcul », l’enseignant H le réfère à un sous-domaine intitulé « Nombres, calcul, résolution de problèmes ». L’enseignant O ne réfère cet item à aucun sous-domaine particulier. Les enseignants J et P le réfèrent au sous-domaine « Organisation et gestion de données » et l’enseignant K et à un agrégat de plusieurs sous-domaines : « Organisation et gestion de données / Mesure ». Une certaine confusion règne ainsi quant au sous-domaine de rattachement des problèmes. Elle nous semble entretenue par le choix de termes proches à ceux employés dans les programmes 2002 et désignant autre chose dans les programmes de 2008.

Elle nous semble également imputable, pour partie, au fait de placer dans le domaine des grandeurs et des mesures des problèmes dits de vie courante. Disons ici, avec Margolinas et Wozniak (2012, p.44), qu’il n’existe pas de définition mathématique standardisée de la grandeur mais que l’on peut considérer, qu’en CP tout au moins, les grandeurs abordées sont toutes mesurables¹⁵⁹ (Brousseau, 2001) et renvoient à des variables comparables entre elles (ce qui correspond à l’un des quatre sens attribués par Brousseau (2001) au terme « grandeur »). Cet aparté permet de considérer que la distinction fine des sous-domaines d’appartenance des problèmes dits de vie courante n’est pas simple : ils risquent bien d’être également des problèmes « simples à une opération » dont le résultat s’écrit dans une unité correspondant à un usage social (ex : l’euro).

¹⁵⁶ Comme cela était le cas dans les programmes de 1995 : l’objectif « Problèmes simples relevant de l’addition, de la soustraction, de la multiplication. » était placé dans le sous-domaine du nombre et du calcul »

¹⁵⁷ Enseignants A, B, D, E, F, L M, N

¹⁵⁸ On relève sans en explorer les conséquences que les problèmes géométriques ou logiques n’ont plus de place dans ces programmes.

¹⁵⁹ D’après Brousseau (2001, p.340), « une grandeur est un ensemble de classes d’objets, stables par rapport aux manipulations ou aux transformations compatibles avec la résolution d’une situation d’ordination ».

2.3.2. Les problèmes « simple à une opération » contenus dans les épreuves du troisième trimestre

Sur les 26 problèmes additifs collectés au troisième nous dénombrons (cf. tableau 26) :

- 14¹⁶⁰ problèmes engagent des raisonnements portant sur des relations partie-tout (composition de mesures, première catégorie de relations additives de Vergnaud).
- 12¹⁶¹ problèmes engagent des raisonnements portant sur des transformations d'état (deuxième catégorie).

L'utilisation de la catégorisation de Vergnaud permet de faire apparaître qu'en fin de CP, dans les épreuves d'évaluation de l'item examiné, le problème « simple » est en première approche plutôt un problème de type additif (mais pas exclusivement) appartenant à l'une ou l'autre de deux principales catégories de problèmes et, à l'intérieur de celles-ci à l'une ou l'autre d'une seule classe de problèmes. Nous pouvons préciser que le problème « simple » est un problème de recherche de la composée de deux mesures relevant de la première catégorie de relations additives explicitée par Vergnaud (combinaison de mesures), ou est, à l'exception d'un problème portant sur la recherche de la transformation, un problème de recherche de l'état final connaissant l'état initial et la transformation (plutôt négative) relevant de la deuxième catégorie de relations additives (transformation opérant sur une mesure pour donner une mesure).

D'après les travaux de Thevenot, Barrouillet et Fayol (2010) précités, ces problèmes sont parmi les problèmes les mieux résolus par de jeunes élèves. De ce point de vue, les tests proposant un problème de composition de mesure et de recherche d'état final suite à une transformation d'état offrirait aux élèves de pouvoir résoudre un problème adapté à leurs compétences ce qui renvoie à une qualité de l'évaluation importante dont seuls les tests A, B, D, F G et H (i.e. un peu plus du tiers des tests) disposent.

Nous nous intéressons à présent à l'affinage des caractéristiques du problème simple du point de vue mathématique mais aussi du point de vue sémantique car, selon Ehrlich (1990, p.8), « il apparaît que les difficultés majeures rencontrées par les élèves dans la résolution des problèmes arithmétiques simples ne sont pas d'ordre logico-mathématique ... elles sont d'ordre sémantique ».

Cet auteur prend pour point de départ le fait que pour résoudre un problème arithmétique simple « l'élève doit opérer une sorte de traduction qui comporte deux aspects :

- un aspect formel [caractérisé par le fait de] passer du langage courant au langage mathématique,
- un aspect sémantique, conceptuel et thématique [caractérisé par le fait de] passer de ce dont on parle dans l'énoncé à une structure mathématique (additive pour ce qui nous occupe dans ce travail) » (1990, p.20).

Pour cet auteur, la difficulté d'un problème « se trouve certainement du côté évènementiel » (1990, p.22), ce qui le conduit à distinguer trois grands groupes de variables pour analyser les résultats d'expérimentations conduites auprès d'élèves de fin de primaire (idem p. 39).

¹⁶⁰ Problèmes 13-1 (test A), 23-1 et 28-2 (test B), 37-1 et 37-2 (test C), 40-5 et 41-1a (test D), 45-5 et 48-2 (test E), 61-3 (test F), 67-2 (test G), 69-3 (test H), 100-3 (test M) et 112-1 (test O) reproduits pp. 258-268 de l'Annexe C.

¹⁶¹ Problèmes 13-2 (test A), 23-2 et 28-1 (test B), 40-4 et 41-1b (test D), 57-5 et 58-1 (test F), 67-1, 67-3 et 67-5 (test G), 69-2 (test H) et 89-3 (test L) reproduits pp. 258-268 de l'Annexe C.

Le premier groupe de variables porte sur la nature des informations événementielles fournies dans l'énoncé. Cela renvoie dans notre travail à interroger le problème du point de vue du type de relations additives mobilisées ce que le référencement des épreuves à la catégorisation de Vergnaud nous a permis de faire.

Le deuxième groupe de variables a trait au signe des informations événementielles et à leur rapport avec le signe de l'opération à poser.

Enfin, le troisième groupe de variables porte sur l'opération à effectuer. Dans notre travail, nous ne nous intéressons qu'à l'addition dont la maîtrise de la technique et du sens sont évaluées dans d'autres épreuves. Nous ne nous attacherons donc pas à développer d'analyses particulières sur ce groupe de variables.

Nous nous intéressons plus particulièrement au deuxième groupe de variables distingué par Ehrlich autour du signe des informations événementielles et à leur rapport avec le signe de l'opération à poser. Ehrlich indique (idem p. 96) que « ce qui importe c'est la conformité de l'ordre énonciatif et de l'ordre événementiel : il est utile pour le jeune enfant de conserver, dans l'énonciation, l'organisation temporelle de l'évènement dont on parle ». Nous cherchons à évaluer si le « problème simple » respecte une telle organisation. Pour conduire nos analyses, nous séparons dans un premier temps les problèmes suivant la catégorie à laquelle ils appartiennent.

2.3.2a Analyse des problèmes de la 2^e catégorie de relations additives de Vergnaud

Nous présentons les 12 problèmes appartenant à la deuxième catégorie de relation additive dans le tableau ci-dessous.

Pour chaque problème, nous présentons quatre variables.

Premièrement, l'organisation événementielle i.e., selon Ehrlich (1990, p.92), « l'organisation de ce dont on parle dans l'énoncé ». Cela correspond pour cette catégorie de problèmes à l'ordre de présentation de l'état initial, de la transformation de cet état, puis de l'état final.

L'organisation énonciative du problème constitue la deuxième variable. Elle est, selon Ehrlich (idem, p. 93), « l'organisation de ce qui est dit dans l'énoncé » : que dit-on en premier, en deuxième puis en troisième ?

L'opérateur sémantique contenu dans l'énoncé est une troisième variable qu'Ehrlich (1990, p.31) définit comme étant « une unité sémantique réunissant un concept et une expression verbale pouvant être limitée à un seul mot. Cette unité sémantique exerce une fonction spécifique dans l'énoncé en marquant selon les cas, un processus d'accumulation ou de comparaison » : « acheter » est un opérateur sémantique d'accumulation. Cet auteur précise (idem) que l'opérateur sémantique « réfère à l'évènement dont on parle dans l'énoncé et non à l'opération terminale conduisant à la solution attendue » et se définit « par sa fonction dans l'énoncé et non pas par ses propriétés intrinsèques » (idem).

Enfin, l'organisation opérative de l'énoncé est la quatrième variable définie par Ehrlich (1990, p.93) comme « l'organisation des informations critiques dans l'opération à poser ». L'opérateur mathématique est l'opération à poser.

N°	Enoncé du problème	Organisation événementielle	Organisation énonciative	Opérateur sémantique	Opérateur mathématique et Organisation opérative
1	Marie ramasse 14 carottes dans son jardin (A). Jeannot le lapin mange 6 carottes (B). Combien de carottes reste-t-il ? (C)	A - B - C	A - B - C	MANGER indique un processus d'accumulation négative	$14 (A) - 6 (B) = 9 (C)$
2	Paul avait un paquet de 24 biscuits (A). Pendant le goûter, il en a mangé 5 (B). Combien reste-t-il de biscuits dans le paquet ? (C)	A - B - C	A - B - C	MANGER indique un processus d'accumulation négative	$24 (A) - 5 (B) = 19 (C)$
3	Martin a un billet de 50€ (A). Il achète un jeu à 20€ (B). Quelle somme lui reste-t-il ? (C)	A - B - C	A - B - C	ACHERTER indique un processus d'accumulation positive	$50 (A) - 20 (B) = 30 (C)$
4	Dans la boîte à aimants, il y a 16 aimants (A). J'en enlève 5 (B). Combien en reste-t-il ? (C)	A - B - C	A - B - C	ENLEVER indique un processus d'accumulation négative	$16 (A) - 5 (B) = 11 (C)$
5	Dans la trousse de Denis, il y a 16 feutres le matin (A), mais il en perd 8 (B), combien lui en reste-t-il ? (C)	A - B - C	A - B - C	PERDRE indique un processus d'accumulation négative	$16 (A) - 8 (B) = 8 (C)$
6	Lucas est venu ce matin à l'école avec des billes dans un sac. Il en avait 14 en arrivant à l'école (A) et il en a gagné 6 pendant la récréation (B). Combien de billes Lucas a-t-il maintenant ? (C)	A - B - C	A - B - C	GAGNER indique un processus d'accumulation positive	$14 (A) + 6 (B) = 20 (C)$
7	Zoé a reçu 25€ pour son anniversaire (A). Elle s'achète un livre qui coûte 2€ (B). Combien lui reste-t-il d'euros ? (C)	A - B - C	A - B - C	ACHERTER indique un processus d'accumulation positive	$25 (A) - 2 (B) = 22 (C)$
8	Yves avait 10 billes ce matin en arrivant à l'école (A). A la récréation, il en a perdu 3 (B). Combien de billes lui reste-t-il ? (C)	A - B - C	A - B - C	PERDRE indique un processus d'accumulation négative	$10 (A) - 3 (B) = 7 (C)$
9	Maman a acheté une douzaine d'œufs (A). En les portant, elle en casse 2 (B). Combien d'œufs lui reste-t-il (C) ?	A - B - C	A - B - C	CASSER indique un processus d'accumulation négative	$12 (A) - 2 (B) = 10 (C)$

N°	Enoncé du problème	Organisation événementielle	Organisation énonciative	Opérateur sémantique	Opérateur mathématique et Organisation opérative
10	Lundi, maman a acheté 12 bouteilles d'eau minérale (A) ; dimanche il en reste 3 (B). Quel est le nombre de bouteilles qui ont été bues ? (C)	A - C - B	A - B- C	BOIRE indique un processus d'accumulation négative	12 (A) - 3 (B) = 9 (C)
11	Gribouille a mangé 8 chocolats (A). Combien reste-t-il de chocolats ? (B) (Suit un dessin comportant la mention 20 chocolats (C))	C - A - B	A - B- C	MANGER indique un processus d'accumulation négative	20 (A) - 8 (B) = 12 (C)
12	Combien le marchand a-t-il vendu de ballons (A) ? (Suit la bulle « Ce matin j'avais 25 ballons » (B) et un dessin représentant le marchand avec 9 ballons (C))	C - A - B	A - B- C	VENDRE indique un processus d'accumulation négative	25 (A) - 9 (B) = 16 (C)

Tableau 27 : Organisations événementielle, énonciative et opérative des énoncés et opérateurs sémantiques et mathématiques des 12 problèmes de catégorie II (transformation opérant sur une mesure pour donner une mesure) recueillis au 3^e trimestre

Dans la majorité des problèmes examinés (qui sont des problèmes d'accumulation plutôt négative), nous constatons que les organisations événementielles et énonciatives de l'énoncé coïncident.

Nous constatons par ailleurs que, dans la plupart des cas, l'opérateur sémantique est de même signe que l'opérateur mathématique ce qui constitue un facteur de réussite. En effet, selon Ehrlich, « lorsque les signes de l'opérateur sémantique (S) et de l'opérateur mathématique (M) sont identique, les performances sont élevées » (1990, p. 80).

Le cas de l'opérateur sémantique « Acheter » se démarque toutefois. Dans les problèmes n° 3 et n°7, ce verbe a valeur d'opérateur sémantique indiquant un processus d'accumulation négative : il est synonyme de « Dépenser ». Or, « Acheter » peut, à l'inverse, indiquer un processus d'accumulation positive comme dans l'exemple suivant où l'on pourrait ajouter « ... *de plus* »: « Martin a 5 pommes. Il achète 2 pommes *de plus* ». Pour résoudre les problèmes n°3 et n°7, les élèves doivent décider du signe de l'opérateur « Acheter » : ils sont de ce fait moins simples que les autres du point de vue du traitement sémantique. Le terme « reste » de la question doit permettre à l'élève, dans chacun des problèmes, de conforter le choix opéré.

Coïncidence des organisations événementielles et énonciatives, coïncidence des signes des opérateurs sémantiques et mathématiques constituent des éléments de convergence entre les problèmes de cette classe de nature à permettre de caractériser leur « simplicité ».

3 problèmes sur les 12 proposés échappent toutefois à cette double caractéristique au sein des problèmes de la classe étudiée :

- Les problèmes n°11 et n°12 du tableau précédent positionnent la question en en-tête. Selon Devidal, Fayol et Barrouillet (1997) un tel placement de la question permettrait

l'activation précoce d'un schéma de problème. Thevenot, Barrouillet et Fayol reprennent et valident en 2004 l'hypothèse précédente selon laquelle « une fois le schéma activé, les informations pourraient y être intégrées en temps réel. Les opérations arithmétiques pourraient ainsi être réalisées au fur et à mesure que les données numériques sont rencontrées, et la charge en mémoire de travail s'en trouverait réduite, d'où l'amélioration des performances » (cf. aussi Thevenot et Perret, 2009). Cela nous conduit à affiner le marqueur de simplicité relatif à la coïncidence des organisations événementielles et énonciatives et à ne pas considérer comme nécessaire la coïncidence des organisations événementielles et énonciatives

Devidal, Fayol et Barrouillet se réfèrent à un modèle de la compréhension issu des modèles de la compréhension de textes de Kintsch et coll. (1978, 1983) qui n'est pas celui utilisé par Vergnaud. Dans ces modèles psycho cognitifs, les connaissances antérieures relatives au mode d'organisation des textes (ou des problèmes) préalablement étudiés peuvent être mémorisées en tant que connaissances et intégrées. Julo (1995, p. 49) utilisent le terme de schéma (de problème) pour désigner « globalement toutes les formes sous lesquelles cette mémoire des problèmes peut intervenir dans la construction d'une représentation particularisée ».

Sans qu'il n'y ait de lien, dans ces mêmes problèmes, une partie de l'information numérique à prendre en compte est dessinée. L'ordre de lecture des informations de l'énoncé peut donc différer (dans le cas où l'élève prendrait en compte d'abord les éléments dessinés) ce qui signe une situation de non-congruence sémantique au sens de Duval (1988) : cela revient à confirmer que l'hypothèse que le dessin peut induire une modification de la tâche de l'élève (modification de l'organisation énonciative du problème),

- Le problème n°10 reproduit ci-dessous :

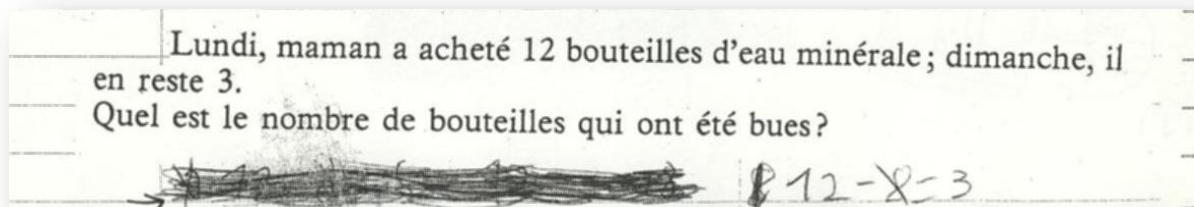


Figure 67 : Épreuve MT3 (Annexe B p. 307) avec réponses d'élève

Résoudre le problème engage, sur le plan mathématique, de trouver la transformation (négative) appliquée à un état initial connu pour donner un état final connu. C'est le seul problème additif du 3^e trimestre qui appartient à une autre classe de problèmes que les classes Ia et IIa (cf. tableau n°26). Il est sans aucun doute plus « difficile » que les autres.

La question que pose ce problème présupposé « simple » est celle du moyen dont dispose l'élève pour répondre. Une ligne placée sous le problème limite fortement la possibilité de dessiner ou schématiser les données. L'écriture de la réponse est extrêmement difficile pour un élève de CP qui n'a pas les moyens conceptuels d'écrire une équation à une inconnue du type $12-x=3$ et d'en déduire que la réponse est la valeur de x . La réponse de l'élève « $12-8=3$ » ne conduit pas immédiatement à la réponse. Lorsque l'élève recherche la situation finale (classe IIa), la valeur située après le signe « = » est la réponse. Dans ce problème, ce

n'est pas le cas et manifestement l'élève ne parvient pas à donner d'autre réponse qu'un calcul (par ailleurs erroné).

L'étude des problèmes de la 2^{ème} catégorie de relations additives nous permet d'indiquer que 9 des 12 problèmes appartenant à cette catégorie se caractérisent par l'égalité de signe des opérateurs sémantiques et mathématiques, et par le traitement linéaire des événements de l'énoncé i.e. il y a identité de l'organisation événementielle et de l'organisation opérative. Cette égalité des organisations ne se retrouve pas lorsque la question est placée en-tête du problème. Cela constitue pourtant un puissant organisateur de l'activité de l'élève : elle contient une unité sémantique marquant l'accumulation (« vendre » « manger » dans les problèmes considérés) susceptible de permettre l'activation d'un schéma de problème.

Nous avons souligné que l'emploi du verbe « Acheter » (problèmes n°3 et n°7) induit une difficulté sémantique : ce verbe indique en effet de façon concomitante une accumulation positive (accordée par défaut par les élèves ?) et négative. « Acheter » indique tout à la fois une acquisition (une accumulation positive fait suite à l'action d'acheter : « acheter » = « acquérir ») et un échange (ce que l'on achète est échangé contre sa valeur en devise et le résultat de l'action d'acheter est une diminution de la somme de devise disponible, « acheter »= « dépenser »). Dans les problèmes de la classe étudiée, i.e. de recherche de l'état final après transformation, « Acheter » est toujours synonyme de « dépenser »

Nous avons également souligné le fait que les problèmes portant sur la transformation (problème n°10) ne peuvent être qualifiés de simples en CP.

Ces trois problèmes (n°3, n°7 et n°10) ne constituent donc pas des problèmes simples sur le plan sémantique et mathématique.

2.3.2b Analyse des problèmes de la 1^{ère} catégorie de relations additives

14 problèmes de relation « partie-tout » constituent l'autre type de problème le plus utilisé pour évaluer l'item au troisième trimestre. Dans les problèmes de composition de mesures que nous avons collectés, la question ne porte que sur la composée de mesures données. L'organisation événementielle des énoncés est donc proche de celle définie précédemment i.e. présentation de la mesure n°1 (A), présentation de la mesure n°2 (B), cumul des mesures (C).

Nous avons placé un astérisque à côté du numéro de certains problèmes : ils font l'objet d'un commentaire particulier.

N°	Enoncé du problème	Organisation événementielle	Organisation énonciative	Opérateur sémantique	Opérateur mathématique et Organisation opérative
1*	Pierre a cueilli 4 pommes vertes (A), 5 pommes rouges (B) et 6 poires (X). Combien de pommes Pierre a-t-il cueilli ? (C)	A - B - C	A - B - X - C	CUEILLIR indique un processus d'accumulation positive	$4 (A) + 5 (B) = 9 (C)$
2	Lundi, Marc a acheté 19 billes. Mercredi, il en a acheté 23. Combien de billes a-t-il en tout ?	A - B - C	A - B - C	ACHETER indique un processus d'accumulation positive	$19 (A) + 23 (B) = 42 (C)$
3*	Camille a 6 euro (A). Elle a acheté deux jouets (B). Quels jouets a-t-elle acheté ? (C) (entourer dans une sélection de 3 jouets présentés avec prix) (E)	E - B - A - C	A - B - C - D	Neutre. D'autres informations événementielles (« en tout », à inférer) indiquent un processus d'accumulation positive puis de comparaison	$2 (E1) + 4 (E2) = 6 (B)$ $6 \leq 6 (A)$ La balle et la peluche (C) Ou $2(E1) + 3 (E2) = 5 (B)$ $5 < 6 (A)$ La balle et le biberon (C)
4*	Dans l'école, il y a deux classes de CP (A). Dans celle de Mme Le Corr il y a 25 enfants (B). Dans celle de Mme Mayer, il y a 12 enfants (C). Combien y a-t-il d'élèves de CP dans l'école ? (D)	B - C - D	A - B - C - D	Neutre. D'autres informations événementielles (« en tout ») indiquent un processus d'accumulation positive	$25 (B) + 12 (C) = 37 (D)$

N°	Enoncé du problème	Organisation événementielle	Organisation énonciative	Opérateur sémantique	Opérateur mathématique et Organisation opérative
5*	Sophie va à la boulangerie (A). Elle achète 3 croissants (B). 2 pains au chocolat (C) et 2 baguettes (D). Un croissant coûte 1€ (E), 1 pain au chocolat 2€ (F), une baguette 1€ (G) (1 croissant = 1€, pain au chocolat = 2€, 1 baguette = 1€). Combien va-t-elle payer ? (H)	A-(B-E)-(C-F)-(D-G)-H	A-B-C-D-E-F-G-H	Neutre. D'autres informations événementielles (« en tout », à inférer) indiquent un processus d'accumulation positive	3×1 (B-E) + 2×2 (C-F) + 2×1 (D-G) = 9 (H)
6	Lola a utilisé 6 cahiers du jour, (A) 2 cahiers du soir (B) et 5 autres cahiers cette année (C), combien a-t-elle utilisée de cahiers ? (D)	A - B - C - D	A - B - C - D	Neutre. D'autres informations événementielles (« en tout », à inférer) indiquent un processus d'accumulation positive	$6(A) + 2(B) + 5(C) = 13(D)$
7	Dans notre classe, il y a 26 élèves de CPb (A) et le maître (B), 4 parents (C) et les 3 élèves de CLIN (D), combien sommes-nous en tout ? (E)	A - B - C - D - E	A - B - C - D - E	Neutre. D'autres informations événementielles (« en tout ») indiquent un processus d'accumulation positive	$26(A) + 1(B) + 4(C) + 3(D) = 34(E)$
8	Pierre offre un bouquet de fleurs à sa maman. Il y a 6 roses rouges (A) et 8 roses jaunes (B). Combien de fleurs y a-t-il dans ce bouquet ?	A - B - C	A - B - C	Neutre. D'autres informations événementielles (« en tout ») indiquent un processus d'accumulation positive	$6(A) + 8(B) = 14(C)$
9	Rémi achète un bateau (prix et dessin accolé 12€ (A)) et un ballon (5€ (B)). Calcule sa dépense.	A - B - C	A - B - C	Neutre. D'autres informations événementielles (« en tout ») indiquent un processus d'accumulation positive	$12(A) + 5(B) = 17(C)$

N°	Enoncé du problème	Organisation événementielle	Organisation énonciative	Opérateur sémantique	Opérateur mathématique et Organisation opérative
10*	Compte ce que Tania et Léo avaient dans leur porte-monnaie. Puis entoure le porte-monnaie qui contenait le plus d'argent. <i>(Les données sont présentées sous forme de dessins de pièces et de billets)</i>			LE PLUS DE indique un processus de comparaison	
11	Gérald possède 34 livres de bibliothèque (A) et 12 albums de bandes dessinées (B). Combien de livres possède-t-il en tout ? (C)	A - B - C	A - B - C	Non explicite. D'autres informations événementielles (« en tout ») indiquent un processus d'accumulation positive	34 (A) + 12 (B) = 36 (C)
12	Arthur a 6 bonbons (A), Zoé en a 8 (B) et Gribouille en a 4 (C). Ils mettent tous les bonbons dans une boîte (D). Combien y a-t-il de bonbons dans la boîte ? (E)	A - B - C - D - E	A - B - C - D - E	Non explicite. D'autres informations événementielles (« tous ») indiquent un processus d'accumulation positive	6 (A) + 8 (B) + 4 (C) = 18 (E)
13*	Au marché, maman a acheté 5 roses (A), 10 pommes (B), 4 bananes (C) et 3 poires (D). Combien maman a-t-elle acheté de fruits ? (E)	B - C - D - E	A - B - C - D - E	Neutre. D'autres informations événementielles (« en tout ») indiquent un processus d'accumulation positive	10 (B) + 4 (C) + 3 (D) = 17 (E)
14*	Tania va dans un magasin de jouets. (Présentation de jouets et prix A, B, C, D) Tania choisit la poupée (43€ - A) et les billes (4€ - C). Combien devra-t-elle payer ? (E)	A - C - E	A - B - C - D - E	Neutre. D'autres informations événementielles (« en tout ») indiquent un processus d'accumulation positive	43 (A) + 4 (C) = 47 (E)

Tableau 28 : Organisations événementielles, énonciatives et opératives des énoncés et opérateurs sémantiques et mathématiques des 14 problèmes de catégorie I (composition de mesures) recueillis au 3^e trimestre

Certains problèmes appellent un commentaire : on ne peut leur attribuer le qualificatif de « simple » pour plusieurs raisons :

-les problèmes n°3 et n°10 du tableau se rangent bien dans la catégorie I (composé de mesures) mais sont aussi des problèmes de comparaison : 3 combinaisons sont possibles pour répondre à la question posée du problème n°3 dont seulement 2 sont correctes. Après avoir trouvé la composée des valeurs monétaires données (i.e. la somme dont dispose Tania et la somme dont dispose Léo), le problème n°10 exige une tâche de comparaison.

-les problèmes n°1, n°4, n°13 et n°14 se rangent aussi dans la catégorie I (composé de mesures) mais nécessitent de mettre à l'écart une mesure (chiffrée ou littérale) constituant une donnée inutile.

-enfin, le problème n°5 qui se range également dans la catégorie I nécessite de résoudre trois problèmes multiplicatifs quaternaires préalables (trouver la valeur de 3 pains au chocolat connaissant le prix unitaire d'un pain au chocolat etc.) qui, s'ils peuvent être résolus par addition répétée, constituent des étapes obligatoires.

Nous soulignons le rôle particulier d'une donnée numérique utile fournie dans la numération écrite littérale dans le problème n°7 : elle ajoute une difficulté supplémentaire indépendante de la compétence de résolution d'un problème additif « simple ».

Ces problèmes « non-simples » étant écartés, pour résoudre chacun des autres problèmes proposés, l'élève doit rechercher la composée positive de mesures contenues dans l'énoncé : des billes, des effectifs d'élèves, des variétés de cahiers, des fleurs, des livres, chocolats ou des euros. Dans chaque problème, la question se situe systématiquement après l'énoncé des données, et l'énoncé comporte soit une unité sémantique explicite (Cueillir...) soit l'expression « en tout », soit une reprise lexicale qui suggère un ensemble formé d'éléments dont les quantités ont été données (effectif d'élèves pour former un effectif d'école, nombre de fleurs pour former un bouquet, nombre de bonbons additionnés dans une boîte, somme des dépenses partielles). Ces expressions suggèrent de sommer les mesures.

Contrairement aux problèmes de la deuxième catégorie de relations additives précédemment analysés, la plupart des énoncés des problèmes de première catégorie analysés ici contiennent un opérateur sémantique neutre : l'élève doit inférer d'autres éléments que la situation décrit une accumulation positive. Dans les problèmes de recherche d'état final après transformation d'un état initial précédemment étudiés, « Acheter » était synonyme de « Dépenser » ou « Payer ». Dans les problèmes de composition de mesure étudiés, « Acheter » est, au contraire, synonyme de « Acquérir » ou « Obtenir ». Cela permet de constater que la polysémie entourant ce verbe est source de difficulté sémantique et mathématique.

Dans ces problèmes « simples » examinés (qui sont des problèmes d'accumulation plutôt positive), nous constatons que l'organisation événementielle et énonciative de l'énoncé coïncide. En outre, dans tous les cas, l'opérateur sémantique est soit neutre, auquel cas d'autres informations événementielles prennent un signe de même signe que l'opérateur mathématique, soit de même signe que l'opérateur mathématique.

2.3.2c Caractérisation mathématique et sémantique du problème simple à une opération donné en évaluation en fin de CP par les enseignants de la population étudiée.

Synthétisant les deux points précédents, nous pouvons indiquer que le problème « simple à une opération » donné en évaluation en fin de CP apparaît être un problème réaliste (i.e. ancré dans des situations, sinon réelles du moins plausibles), relevant exclusivement de deux catégories de relations additives (partie-tout ; transformation d'état). A l'intérieur de celles-ci, les problèmes de recherche d'une mesure finale suite à la transformation négative d'une mesure initiale ou, de combinaison positive de mesures sont les plus simples.

Ces problèmes se caractérisent en outre par la coïncidence des organisations événementielles énonciatives et opératives : ce sont des énoncés sémantiquement congruents au sens de Duval (1988). Nous avons relevé que la question placée en en-tête, constitue un puissant organisateur de l'activité de l'élève. Elle est susceptible de permettre l'activation d'un schéma de problème.

Nous avons souligné que le verbe « Acheter » sous-tend deux processus d'accumulation positive ou négative selon que l'énoncé renvoie à une situation de dépense (plutôt rencontrée dans les problèmes de recherche d'état final après transformation d'état initial) ou de gain (plutôt rencontrée dans les problèmes de combinaison de mesure). L'élève doit rechercher d'autres informations dans le texte de l'énoncé ou dans la question pour inférer le signe de l'accumulation (*a priori* positive si l'on renvoie à l'usage du mot dans la vie quotidienne de l'élève) que ce mot supporte. « Acheter » n'agit donc pas toujours comme un opérateur sémantique. Lorsque c'est le cas, on peut estimer que cela facilite le traitement sémantique de l'énoncé.

Parmi les 26 problèmes utilisés pour évaluer la compétence de résolution de problèmes « simples », nous avons écarté 3 des 12 problèmes appartenant à la deuxième catégorie (problème n°3, n°7 et n°10, tableau n°27) et 7 des 14 problèmes de la première catégorie (problèmes n°1, 3, 4, 5, 10, 13 et 14).

Environ 39% des problèmes utilisée en évaluation en fin de CP ne peuvent ainsi être considérés comme des problèmes simples quand bien même ils sont « à une opération ».

Concernant le traitement opératoire des données, nous pouvons indiquer que les calculs numériques nécessités sont de difficulté variable selon la catégorie de problème considérée (parmi les problèmes collectés). Les problèmes de catégorie I (composition de mesures) occasionnent des calculs plus difficiles que ceux de la catégorie II (recherche d'état final après transformation de l'état initial). Dans les premiers problèmes, l'élève doit sommer plusieurs termes, gérer une retenue ou effectuer un calcul réfléchi (en ligne) nécessitant de procéder à des décompositions/recompositions de nombres. Dans les autres problèmes le calcul ne porte que sur deux termes et, dans la plupart des cas, le résultat s'obtient sans avoir à gérer de retenue.

Nous poursuivons notre investigation du problème simple à une opération avec un questionnement d'une autre nature, relevant de l'écrit. Le problème simple est-il un support de lecture simple pour un élève de fin CP ? Implique-t-il une réponse simple à écrire ?

2.3.2d Le problème simple est-il un support de lecture simple ?

Nous venons de caractériser le « simple » d'un point de vue mathématique et sémantique. Nous considérons à présent les problèmes comme des textes comportant un questionnaire. A l'image de Maury (1984 ; 1985), Julo (1995, p. 153) a indiqué que « l'on peut montrer facilement la responsabilité de l'énoncé dans certaines formes de dysfonctionnement : une simple variation de formulation peut faire que l'on réussit ou que l'on échoue et ceci indépendamment de toutes les autres variables ».

Nous revenons sur le problème n°10 (cf. tableau n°27) que nous avons précédemment écarté de la liste des problèmes « simples » :

Épreuve n°33 : « Les bouteilles »

a- l'item :

Cet enseignant souhaite évaluer l'item « Résoudre des problèmes simples à une opération »

b- l'épreuve :

Pour évaluer cet item, il utilise l'épreuve ci-dessous :

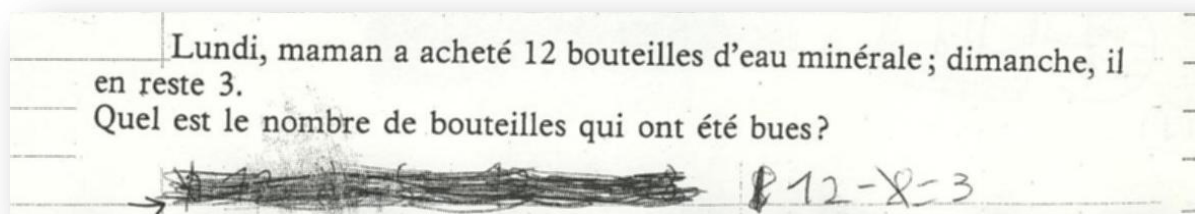


Figure 68 : Épreuve MT3 (Annexe B p. 307) avec réponses d'élève

Outre la difficulté mathématique du problème, nous pouvons isoler trois autres sources de difficultés relatives à la lecture.

-L'identification du mot « bouteilles », présent deux fois dans l'énoncé, est une source de difficulté non négligeable pour un certain nombre d'élèves en fin de CP (longueur du mot, graphème complexe en fin de mot,(cf. III-4.2.1 pour une comparaison avec le texte « Caroline »)).

-La formulation de la question constitue une autre source de difficulté de lecture importante. « Quel est le nombre de bouteilles qui ont été bues ? » repose sur l'emploi de l'adjectif interrogatif « quel » et du pronom relatif « qui » et sur l'usage de deux temps et formes du verbe qui indiquent l'antériorité d'un composant de la phrase par rapport à l'autre (présent du verbe être à la voix active, passé composé du verbe boire à la voix passive). L'étude de la voix passive n'est évidemment pas du programme du cycle 2 pas plus que l'étude du passé composé ou la valeur comparée des temps verbaux en CP.

-La compréhension de l'énoncé qui fournit l'état d'un stock de bouteilles à deux moments de la semaine oblige la réalisation d'une inférence de cause à effet. Faisant appel à ses connaissances sur le monde, l'élève doit inférer de l'élément de la question « des bouteilles ont été bues », que le stock de bouteilles d'eau minérale a diminué parce qu'elles ont été bues. Une inférence obligatoire pour pouvoir traiter mathématiquement le problème est ainsi laissée à la charge de l'élève.

La réponse de l'élève peut être difficilement analysée du fait des nombreuses ratures contenues dans la copie. Il est toutefois patent de constater qu'elle a fait l'objet de nombreux essais et que si l'opérateur mathématique a été reconnu, l'élève ne parvient pas à analyser le résultat du calcul effectué. Cela souligne la difficulté de la tâche de production d'une réponse : rien ne permet d'être sûr que l'élève associe le résultat du calcul « 3 » au « nombre de bouteilles bues », cette inférence est entièrement laissée à la charge de l'évaluateur.

c-discussion de la validité psycho-didactique de l'épreuve :

Le problème n'est pas un problème simple du point de vue mathématique, fût-il à une opération. Il engage une tâche complexe du point de vue des relations entre les données.

Le problème n'est pas un problème simple non plus du point de vue de la lecture. Il nécessite une certaine expertise de lecture sur les plans de l'identification des mots, de la syntaxe et de la réalisation d'inférences.

Ces deux constats nous conduisent à remettre en cause la validité de cette épreuve.

Cette épreuve permet de constater que la tâche prescrite à l'élève imaginée par l'enseignant (résoudre un problème simple) occasionne une tâche de lecture d'une difficulté importante non évaluée pour elle-même, une tâche mathématique d'une difficulté non négligeable elle-aussi de par la nature des relations entre les données, une tâche d'écriture de la réponse que l'on peut considérer elle-aussi comme triviale.

L'épreuve suivante permet de donner un autre exemple des difficultés d'identification des mots de l'énoncé et des inférences laissées à la charge de l'élève pour comprendre la question :

Épreuve n°34 : « Les livres »

a- l'item :

L'enseignant G souhaite évaluer l'item « Résoudre des problèmes simples à une opération »

b- l'épreuve :

Pour évaluer cet item, il utilise l'épreuve reproduite ci-dessous :

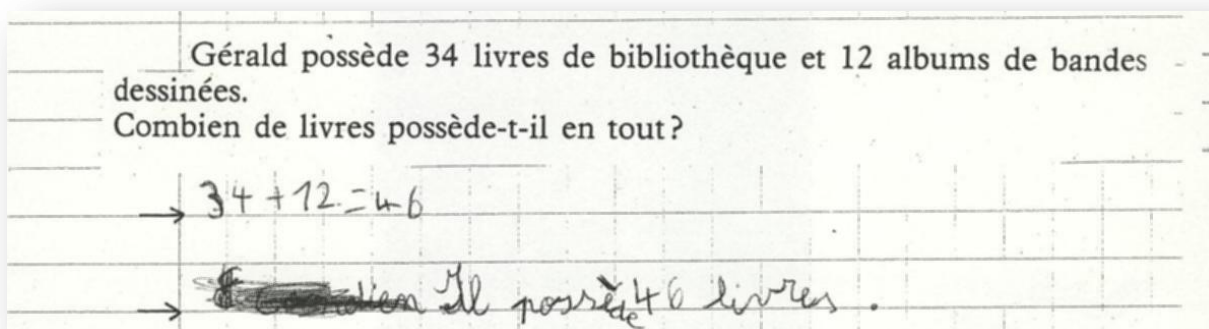


Figure 69 : Épreuve MT3 (Annexe B p. 307) avec réponses d'élève

Pour résoudre ce problème, du point de vue mathématique, l'élève doit composer positivement deux mesures données « 34 » (livres de bibliothèque) et « 12 » (albums de

bandes dessinées). Le calcul de la somme des deux termes ne pose aucun problème particulier.

En termes de lecture, l'identification des mots « bibliothèque » et « albums » pose un certain nombre de difficultés déjà abordées (cf. III 4.2.1). Le mot « album » ajoute une difficulté liée au non-respect des règles du code graphophonologique. Si ces règles sont strictement appliquées, le mot « album » ne peut pas être identifié : autrement dit pour identifier ce mot il faut déjà le connaître ou avoir une connaissance (souvent implicite) de la façon dont on lit les terminaisons des mots d'origine latine (comme aquarium). On souligne que l'expression « albums de bandes dessinées » est rare, les élèves de CP sont sans doute plus familiers du mot « BD ».

La question porte sur la détermination d'un total de livres. Seule la mesure « 34 » s'exprime en livres (de bibliothèque). L'inférence « l'album de bande dessinée est un livre » est obligatoire pour donner la bonne réponse, sans quoi on peut répondre « 34 » à la question posée. Cette inférence est soutenue par l'expression « en tout » qui oriente fortement l'activité de l'élève vers l'addition des deux seules mesures contenues dans l'énoncé.

c-discussion de la validité psycho-didactique de l'épreuve :

La validité de cette épreuve est remise en cause. Elle comporte, comme dans l'exemple précédent, des difficultés de lecture qui ne peuvent être tenues pour non négligeables (identification de certains mots, compréhension de la question).

Indépendamment du contenu mathématique des problèmes étudiés et des raisonnements que l'élève doit suivre pour les résoudre, les difficultés de lecture (et d'écriture des réponses) sont susceptibles d'interdire l'accès au problème aux élèves les moins bons lecteurs.

On pourrait concevoir qu'un élève réponde correctement à la tâche précédente en se fondant sur une connaissance fine des attentes de l'enseignant : par effet de contrat si deux données sont fournies dans l'énoncé, il y a de grandes chances qu'il faille les additionner ou les soustraire, et l'expression « en tout » suffit à comprendre qu'il faut additionner.

On peut donc concevoir que, pour limiter cet effet de contrat, certains enseignants choisissent des problèmes dans lesquels toutes les données ne sont pas utiles pour répondre. L'épreuve suivante propose un tel problème.

Épreuve n°35 : « Les pommes »

a- l'item :

L'enseignant A souhaite évaluer l'item « Résoudre un problème simple » au troisième trimestre ce qui est une reformulation proche de l'item des programmes « Résoudre des problèmes simples à une opération »

b- l'épreuve :

Pour évaluer cet item, l'enseignant utilise l'épreuve reproduite ci-dessous :

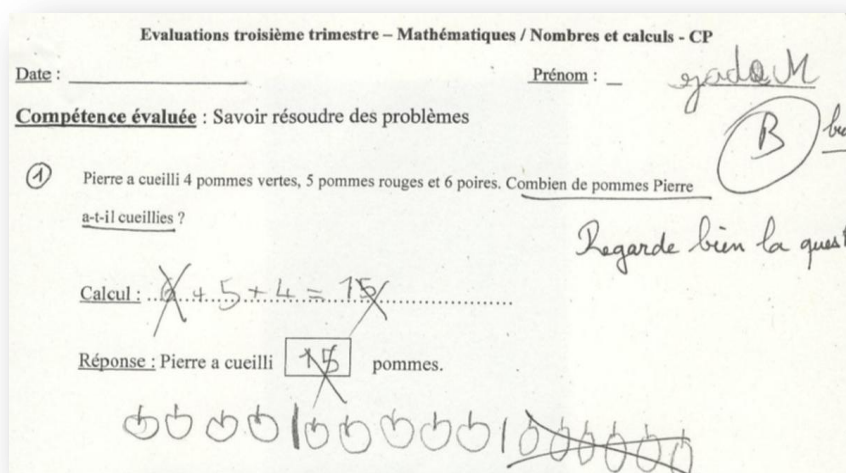


Figure 70 : Épreuve MT3 (Annexe B p. 306) avec réponses d'élève

Le problème proposé appartient à la 1^{ère} catégorie de relations additives (relation partie-partie-tout), et à l'intérieur de celle-ci, à la classe des problèmes de composition de mesures toutes connues.

Préalablement au calcul de la solution ($4 + 5 = 9$), l'élève doit écarter une donnée inutile (nombre de poires). C'est la raison pour laquelle nous l'avons écarté des problèmes que nous avons déterminé comme étant des problèmes simples.

c- la copie de l'élève:

L'élève a additionné l'ensemble des données numériques pour former sa réponse. Une première analyse conduit à voir que la réponse (15) est bien la somme des trois données numériques. L'élève a su reconnaître une situation additive, il a également correctement su calculer une somme. L'élève a su matérialiser dans la numération chiffrée le résultat d'un calcul additif en ligne. Cela constitue un ensemble de compétences non négligeables, mais la réponse est erronée : l'élève n'a pas écarté la donnée inutile.

Plusieurs hypothèses peuvent être formulées pour l'expliquer :

-l'effet de contrat « toutes les données fournies par le maître servent sinon elles ne seraient pas données » peut fonctionner à l'inverse de ce que souhaite l'enseignant. Cela pourrait expliquer pourquoi l'élève additionne toutes les données. On peut concevoir que l'enseignant évalue dans quelle mesure l'élève se distancie du contrat habituel. On peut alors souligner que cela constitue un nouveau contrat, lui aussi implicite.

-on peut observer, sur la copie, la figuration des éléments de l'énoncé. Les 6 « poires » sont barrées ainsi que dans l'écriture de la décomposition de la somme. Ceci est probablement le fait d'une reprise du travail suite à la remarque de l'enseignant « regarde bien la question ». L'enseignant estime que l'erreur provient du traitement de la question « combien de pommes Pierre a-t-il cueillies ? ». Laissons de côté, sans l'occulter, la réelle difficulté linguistique de la consigne en termes de syntaxe ou de déchiffrement particulièrement au niveau du mot « cueilli(es) » rencontré 3 fois sur 19 mots (ce mot est absent de la base Manulex en CP) car l'élève a manifestement traité la consigne. Pour répondre correctement, l'élève doit accepter la première donnée (ce sont des pommes). Il doit ensuite accepter la deuxième donnée (ce sont des pommes). Il peut, ou ne pas, additionner ces deux mesures, puis il doit laisser de côté la dernière donnée (ce sont des poires). Autrement dit, il peut traiter numériquement le

problème en effectuant l'addition $4+5$. Mais ce n'est pas ce qu'il fait. En proposant « $6+5+4$ » on peut estimer qu'il traite les données après les avoir toutes parcourues. Soit il a perdu de vue l'unité dans laquelle exprimer les données (la grandeur) parce qu'il est occupé à calculer, soit il lit vite (vu la difficulté de lecture c'est une hypothèse plausible) et lit « pomme » pour « poire » n'ayant eu aucune raison de se méfier, soit, la question intervenant longtemps après le mot « poire », il ne se souvient plus que 6 s'exprime en poires.

Ceci permet de mettre en avant que répondre correctement oblige l'élève à effectuer un rétrocontrôle sur sa lecture pour se rendre compte qu'il doit inhiber une donnée : c'est donc surtout une compétence de lecture, non triviale, qui est mobilisée ici.

Nous relevons enfin que l'élève inscrit « 15 » dans le cadre de réponse imposé par l'épreuve. Autrement dit, la prise en charge de l'expression du résultat (en « pommes ») est réalisée par le cadre de réponse. De sorte que nous ne pouvons pas savoir si l'élève a bien effectué un lien entre le résultat du calcul numérique (« 15 ») et la grandeur à laquelle il est rattaché (des pommes).

d-discussion de la validité psycho-didactique de l'épreuve :

La validité de l'épreuve est remise en cause. Elle se caractérise par le caractère imbriqué des compétences mobilisées par les élèves (que l'enseignant recherche peut-être à évaluer) en lecture, en résolution de problèmes, en gestion et organisation des données utiles pour résoudre le problème.

2.4. Synthèse

Dans ce chapitre, nous avons montré qu'aucun test destiné à évaluer la capacité des élèves à résoudre un problème additif proposé au troisième trimestre n'était valide sur le plan global et sur le plan local, si l'on considère les critères de couverture du domaine considérés.

De ce fait, notre étude de validité globale et locale se révélait sans objet.

Nous référant à des travaux de psychologie et de didactique, nous pouvons conclure, de l'analyse des problèmes additifs donnés en évaluation pour évaluer la compétence de résolution de problèmes (additifs) « simples à une opération », que les enseignants de la population étudiée emploient des problèmes d'une difficulté non négligeable du point de vue sémantique et cognitivo-didactique. A côté des caractéristiques sémantiques et mathématiques précédemment mises en lumière, il est de surcroît nécessaire de prendre en compte les caractéristiques linguistiques des énoncés et les compétences de lecteur mobilisées pour les traiter (en termes d'identification des mots, de réalisation d'inférences, de connaissances syntaxiques, de stratégie de lecture, de rédaction des phrases réponse). L'effet de ces caractéristiques, comme l'effet de contrat, ne peuvent, en CP surtout, être considérés comme négligeables sous peine de défavoriser très nettement les élèves les moins experts en lecture/écriture indépendamment de leurs compétences mathématiques de résolution de problèmes.

Chapitre 3. Analyse des tests et des épreuves d'évaluation des compétences en lecture

L'identification des mots et le développement de la compréhension constituent deux des principaux axes du dossier de synthèse de la conférence de consensus « Lire, comprendre, apprendre » rédigé par le CNSECO (2016). Ce rapport prolonge les travaux de la conférence de consensus de 2003 au terme de laquelle le jury soulignait que « tous les experts insistent sur la nécessité de mener de front plusieurs apprentissages complémentaires, de combiner le travail sur le code et le travail sur le sens ».

Considérant l'enjeu fondamental pour l'apprentissage de la lecture que représente la « maîtrise du code » par le lecteur débutant, nous nous intéresserons particulièrement à l'évaluation qui en est conduite en tout début d'apprentissage. Puis, nous analyserons les tests d'évaluation relativement à « l'identification des mots » et à la « compréhension autonome de l'écrit » en fin d'année. Comme dans les chapitres précédents, nous recherchons des preuves de validité des tests (en termes de couverture des composants du domaine) et des épreuves (VPD).

Préalablement à cela, afin de déterminer un référentiel de nature didactique pour analyser la validité de contenu des tests, nous nous appuyerons sur une typologie de tâches d'enseignement de la lecture-écriture utilisée par le collectif de chercheurs coordonné par Goigoux (Recherche Lire-écrire au CP ¹⁶², Ifé, 2013) s'intéressant aux pratiques d'enseignement de la lecture et de l'écriture au CP. Les instruments de l'enquête et de l'observation des classes (notamment une catégorisation des tâches proposées par l'enseignant à ses élèves) ont été élaborés à partir d'une série d'hypothèses sur ce qui pouvait conduire à une plus ou moins grande efficacité et à une plus ou moins grande équité des pratiques des enseignants dans le domaine (cf. Goigoux, Jarlégan et Piquée, 2015) pour une présentation détaillée de la méthodologie utilisée). Ces hypothèses, didactiques (en ce sens qu'elles intègrent les résultats des différentes disciplines qui contribuent à la didactique: psychologie, linguistique, sociologie...), ont été construites à partir de l'état des recherches sur l'enseignement et l'apprentissage de la lecture-écriture à ce niveau de la scolarité. Elles tiennent compte de ce que les programmes prescrivent. Il serait sans doute excessif de qualifier cette typologie de consensuelle du fait de débats parfois vifs entre chercheurs. Dans la mesure où cette typologie sert à ce collectif pour produire un certain nombre de résultats, il nous a semblé opportun de l'utiliser à des fins d'analyse de tâches d'évaluation de la lecture.

¹⁶² Ce programme est placé sous la responsabilité scientifique de Roland Goigoux (ACTé (EA 4281), Clermont 2) et mobilise 13 équipes de recherche : ACTé (EA 4281), Clermont 2 ; CREM (EA 3474) + (LISEC et INTERPSY), Lorraine ; CIRCEFT-ESCOL (EA 4384) Paris Est- Créteil ; CIREL-Théodile (EA 4354) Lille ; CREAD (EA 3875) Rennes 2 & UBO ; CREN (EA 2661) U. Nantes ; CTRF (EA 1392) U. Cergy Pontoise ; EFTS (UMR MA 2011.09.122), U. Toulouse-Le Mirail ; IFÉ – ENS Lyon ; LACES (EA 4140) U. Bordeaux 4 et 3 ; LIDILEM (EA 609), Grenoble 3 ; LIRDEF alpha (EA 3749) : Montpellier 2 et 3 ; Université de Rouen (IUFM).

3.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences de lecture en CP

Un certain consensus prévaut (cf. Giasson, 1990, pp.6-7; Remond, 2007, p.1) autour d'une définition didactique¹⁶³, générale, de la lecture telle que formulée par Gombert, Colé, Valdois et coll. (2000/2002, p.11) :

« La lecture est une construction de significations réalisée par une personne à partir d'un texte écrit, dans un contexte donné. Cette construction est le fruit d'une interaction entre les données propres au texte et les connaissances du lecteur (linguistiques, conceptuelles) en fonction des buts qu'il poursuit à travers cette lecture ».

Ce modèle comporte trois variables en interaction : le lecteur, le texte et le contexte. Pour analyser quelques épreuves d'évaluation de la compréhension, nous nous sommes appuyé dans la partie III de la thèse sur deux des trois variables de ce modèle : le lecteur et le texte.

La variable « lecteur » du modèle comprend selon Giasson (1990, p.16) « les structures du sujet et les processus de lecture qu'il met en œuvre ». Pour cet auteur, cela renvoie aux connaissances et attitudes du lecteur et aux habiletés mises en œuvre pendant qu'il lit.

La variable « texte » concerne « le matériel à lire, et peut être considérée sous trois aspects principaux : l'intention de l'auteur, la structure du texte et le contenu ».

La variable « contexte » comprend les éléments qui ne font pas partie du texte et ne concernent pas directement les structures ou les processus de lecture, mais qui influent sur la compréhension du texte. Giasson (1990) distingue particulièrement trois contextes : le contexte psychologique (motivation, intention de lecture), le contexte social (interventions des pairs, de l'enseignant..), et le contexte physique (temps disponible, bruit...).

Selon ce modèle, la compréhension est donc fonction de ces trois variables, si bien que, selon Giasson (1990, p.9), on devrait être conduit à particulariser la difficulté en compréhension éprouvée par un élève en indiquant le type de texte et le contexte. Nous précisons à ce stade que nous n'avons pas collecté de données permettant d'analyser la troisième variable du modèle (variable contexte). Nous ne développerons donc pas cet aspect dans ce qui suit. Nous ne nous intéresserons donc seulement aux variables « Texte » et « Lecteur » pour construire notre référentiel.

3.1.1. La variable « Lecteur »

Au niveau du lecteur, Goigoux et Cèbe (2006, p. 13) mettent en avant quatre composants indissociables, lesquels sont également communément admis au niveau de l'enseignement de la lecture et de l'écriture (cf. ONL 1998, pp. 66-71 ; Gombert et coll., 2000/2002, pp.12-14) : la culture écrite, la lecture de mots, la compréhension d'écrit (de phrases, de mots) et la production d'écrit (rédaction de texte, écriture de mots). Pour élaborer notre référentiel, nous sommes conduit à écarter le composant « culture écrite » et « production d'écrit ». Nous nous en expliquons.

¹⁶³ Cette définition peut être complétée par la prise en compte d'éléments provenant de la vie psychique des enfants soulignés par Diatkine (1983, p.13) : « les processus d'acquisition du langage écrit, comme du langage oral, nécessitent un désir actif, voire violent d'appropriation ». Ils permettent d'interpréter autrement, dans d'autres cadres, les mêmes difficultés d'entrée dans la lecture (référence à la démarche d'individuation mise en exergue par Diatkine, idem p.7, ou au statut de la forme incorrecte en lecture Diatkine, idem p. 9 ; Bettelheim et Zelan, 1983 p. 144-175).

Nous écartons le composant « culture écrite » en raison d'une première raison liée à la collecte des données. Même si Tiré et coll. (2016) concluent leurs premières analyses de l'approche des pratiques d'acculturation réalisée dans l'enquête Lire-écrire au CP (IFé, 2013) en estimant que « les classes très acculturantes favorisent la progression des élèves de niveau faible et intermédiaire dans tous les domaines d'apprentissage du lire-écrire », nous ne pouvons analyser ce composant car nos données (les évaluations auxquelles les élèves sont soumis) ne nous permettent pas de l'apprécier. De plus, si les connaissances antérieures ainsi que le rapport réflexif à la langue interviennent dans le savoir-lire, Martinet et Rieben (2010, p.198) soulignent qu'à ce jour « la littérature scientifique ne fournit pas encore, et particulièrement en français, de modèles suffisamment étayés permettant de synthétiser les relations entre la familiarisation à ces différentes notions et les autres composantes de l'apprentissage de la lecture ».

Nous écartons également le composant « production d'écrit ». Une lecture stricte de la définition de la lecture précédemment mise en exergue -« une construction de significations réalisée par une personne à partir d'un texte écrit, dans un contexte donné »- sous-tend ce choix qui est également celui des progressions. Elles présentent des progressions séparées pour la lecture et pour l'écriture. Les dimensions graphiques, orthographiques et rédactionnelles relèvent typiquement de l'écrit et de modèles ou de cadres particuliers (cf. Zerbato-Poudou (2015) en ce qui concerne la dimension graphique, Pothier (2002/2004) pour l'orthographe, Harris et Graham (1996) ; Hayes et Flower (1980) pour la modélisation de l'écriture « rédactionnelle ») que nous ne détaillerons pas dans le cadre de ce travail.

En définitive, concernant la variable « Lecteur », nous nous intéressons aux deux composants « lecture de mots » et « compréhension (autonome) de l'écrit ».

La typologie de tâches établie pour les besoins de l'enquête « Lire-écrire au CP » précitée, intègre ces deux composants et les réorganise en trois types de tâches (les deux premiers types de tâches relevant de fait du même composant « lecture de mots ») :

- celles qui mobilisent des connaissances portant sur les liens entre des phonèmes et des graphèmes,
- celles qui mobilisent des connaissances permettant l'identification des mots, et
- celles qui mobilisent des connaissances (ou des processus) permettant de comprendre un texte écrit (lu seul et/ou lu par un tiers).

Pour les besoins de notre recherche, nous retiendrons ces trois grands types de tâches lesquels sont déclinés dans la typologie utilisée pour l'enquête « Lire-écrire au CP »¹⁶⁴ comme suit. Ils correspondent pour nous aux trois principaux composants de la lecture.

Les tâches relevant des relations phonographiques¹⁶⁵ (abrégé PG) qui permettent d'enseigner :

- PG 1. les phonèmes (sans écrit)
- PG 2. les syllabes orales ou d'autres unités de taille supérieure au phonème (rime) (sans écrit)
- PG 3. les lettres (noms, différentes écritures, sans valeur sonore)
- PG 4. les correspondances entre phonèmes et graphèmes
- PG 5. la combinatoire ou les syllabes

Les tâches relevant de la lecture (abrégé L) qui permettent d'enseigner :

- L1. la lecture silencieuse
- L2. la reconnaissance de mot entier
- L3. le déchiffrement de mot
- L4. la lecture à haute voix
- L5. l'écoute de l'enseignant lire à haute voix

Les tâches relevant de l'enseignement de la compréhension (abrégé C) déclinées au niveau du mot, de la phrase et du texte qui permettent de :

- C1. Définir, expliciter une intention de lecture
- C2. Anticiper, formuler, vérifier des hypothèses
- C3. Décrire une illustration
- C4. Expliquer, reformuler le sens ou évoquer une représentation mentale
- C5. Produire un rappel de récit
- C6. Rendre explicite une information implicite
- C7. Proposer, débattre ou négocier une interprétation/ des interprétations
- C8. Réaliser une tâche écrite impliquant la compréhension (explicite et/ou implicite)
- C9. Corriger une tâche écrite portant sur la compréhension

Nous avons mis en lumière le guidage de l'évaluation interne réalisé par les items des progressions. Nous proposons une réaffectation de ces items selon la typologie présentée ci-dessus : elle nous permet de réorganiser les données collectées (cf. Annexe C).

¹⁶⁴ Extraits du cahier des charges de l'enquêteur (9 pages), Goigoux dir. (2013), p. 4.

¹⁶⁵ Goigoux et coll. en distinguent deux de plus relatives à la valeur des phonèmes (sans écrit) et aux syllabes orales (sans écrit). Ces tâches ne sont qualifiées par aucun item des programmes.

Item du programme	Classification typologique (Goigoux et coll., 2013)
- Connaître le nom des lettres et l'ordre alphabétique.	PG3
- Distinguer entre la lettre et le son qu'elle transcrit ; connaître les correspondances entre les lettres et les sons dans les graphies simples (ex. <i>f</i> ; <i>o</i>) et complexes (ex. <i>ph</i> ; <i>au</i> , <i>eau</i>).	PG4
- Savoir qu'une syllabe est composée d'une ou plusieurs graphies, qu'un mot est composé d'une ou plusieurs syllabes ; être capable de repérer ces éléments (graphies, syllabes) dans un mot.	PG5
- Connaître les correspondances entre minuscules et majuscules d'imprimerie, minuscules et majuscules cursives.	PG3
- Lire aisément les mots étudiés.	L3
- Déchiffrer des mots réguliers inconnus.	L2
- Lire aisément les mots les plus fréquemment rencontrés (dits mots-outils).	L3
- Lire à haute voix un texte court dont les mots ont été étudiés, en articulant correctement et en respectant la ponctuation.	L4
- Connaître et utiliser le vocabulaire spécifique de la lecture d'un texte : le livre, la couverture, la page, la ligne ; l'auteur, le titre ; le texte, la phrase, le mot ; le début, la fin, le personnage, l'histoire.	*166
- Dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens.	C8
-Écouter lire des œuvres intégrales, notamment de littérature jeunesse.	? (L5 ?)

Tableau 29 : Classification des items des progressions suivant les catégories de la typologie de tâches établie par Goigoux et coll. (2013)

A la lecture du tableau, un premier constat peut être dressé. Il concerne l'absence d'items portant spécifiquement et explicitement sur la conscience phonémique¹⁶⁷ (capacité à percevoir et manipuler les phonèmes) dans les programmes (correspondant aux types de tâches PG1 et PG2), et la faible décomposition dont la compréhension de l'écrit (en général) fait l'objet. Si l'hypothèse du guidage de l'évaluation par les items des progressions (cf. Partie II) se confirme, ces (sous-)composants de la lecture risquent d'être peu évalués ce qui remettrait en cause la validité de contenu (didactique) des tests examinés.

Nous revenons sur la typologie de Goigoux et coll. (2016) afin d'en justifier l'emploi pour analyser le corpus d'épreuves recueillie dans le cadre de notre travail.

Tous les modèles d'apprentissage de la lecture soulignent le rôle central joué par la reconnaissance visuelle rapide et automatique de mots dans la compétence en lecture. On se reportera par exemple au cadre d'évaluation des composants des compétences en

¹⁶⁶ Il s'agit d'un item très rarement évalué (2 épreuves au 1^{er} trimestre, 1 seule au troisième) et qui n'entre dans aucune des catégories de la typologie adoptée. Cet item ne renvoie pas stricto sensu à la lecture mais plutôt au vocabulaire du livre, il ne nécessite sans doute pas véritablement d'épreuve d'évaluation.

¹⁶⁷ La conscience phonologique est un terme général qui correspond à l'habileté à identifier et à manipuler les mots dans une phrase et les parties d'un mot (les syllabes, les rimes et les phonèmes). La conscience phonémique, qui est le plus haut niveau de la conscience phonologique, correspond à l'habileté à identifier et à manipuler les phonèmes.

lecture du Literacy Assessment and Monitoring Programme de l'UNESCO, lequel programme a été validé pour 10 langues¹⁶⁸ (Sabatini, 2004).

Selon un large consensus ainsi formulé par Goigoux (2000, p. 235), « la compréhension de l'écrit dépend étroitement du degré d'automatisation des processus d'identification des mots écrits » (cf. notamment Fayol, 1996 ; Lecocq et coll., 1996). Ces processus d'automatisation des mots, ajoutent Lecocq et coll. (1996, p.6), « sont liés à l'existence et au développement de la compétence métaphonologique ainsi qu'à l'efficacité d'une mémoire de travail que l'âge et la lecture contribuent à développer ». Celles-ci n'émergent pas spontanément, justifiant un enseignement explicite (Ziegler et Goswami, 2005). Ce résultat est un résultat majeur de ces dernières années.

Dans les modèles développementaux, dits modèles à « doubles voies » (Frith, 1985), l'accès au lexique s'effectue selon une voie directe (dite « d'adressage » ou « orthographique »), ou une voie indirecte (dite « d'assemblage »). Pour cette dernière, la reconnaissance du mot s'effectue par utilisation des règles de correspondances graphèmes-phonèmes.

Ces modèles ont évolué depuis, mais ont pour eux une certaine organisation que les enseignants ont pu s'approprier (cf. Ecalle et Magnan 2002, pp.25-63 pour une synthèse, ou Martinet et Rieben, 2010 pp. 205-211). Ils ont été remis en cause du fait notamment de leur conception « étapiste » de la lecture (nous y revenons un peu plus loin.)

Ecalle et Magnan (2002, p. 59) proposent un cadre général du développement des processus de reconnaissance de mots écrits dans lequel le traitement d'un mot écrit emprunte prioritairement (par recherche du moindre coût cognitif) la voie directe (appelée aussi voie orthographique ou voie d'adressage). Pour ce faire, le lecteur s'appuie sur un ensemble de connaissances phonologiques (syllabes, unités infra-syllabiques, phonèmes) et visuelles (lettres, graphèmes, configuration orthographiques) connectées et en cours de constitution. Si l'adressage n'est pas réalisé indiquent ces auteurs, i.e. si aucune représentation lexicale dans ses différents aspects sémantique, phonologique, orthographique, syntaxique, imagé ou conceptuel n'est activée, le recours à la voie d'assemblage s'impose. Au cours du traitement du mot écrit, il est par ailleurs possible que le lecteur bascule vers une procédure d'adressage dès lors qu'une représentation lexicale est activée grâce à l'apport d'indices contextuels. Cela constitue une troisième voie, mixte, permettant à l'élève d'identifier un mot. Le stockage des représentations orthographiques s'appuie sur des connaissances phonologiques et visuelles des mots développées par apprentissage explicite. Il est à noter que le modèle décrit insiste bien sur la récursivité que les connaissances entretiennent avec les représentations qui sont regroupées sous l'appellation générique « lexique mental¹⁶⁹ ». Apprendre à lire ne revient pas, dès lors, pas à élaborer un module spécialisé dans la reconnaissance des mots écrits, inexistant chez le prélecteur ou chez l'analphabète, ce qui remet en cause le principe « étapiste » des modèles développementaux précédemment évoqués. Au contraire, un même système, initialement incapable de traiter l'information écrite, devient progressivement capable de le faire sous le double effet de l'enseignement et de l'expérience comme le soulignent Crinon et Legros (2003). On ne peut donc conserver l'idée d'un stock de mots situés en mémoire à long terme : « reconnaître un mot n'est pas retrouver ce mot quelque part en mémoire, mais

¹⁶⁸ Cf. <http://www.uis.unesco.org/literacy/pages/lamp-literacy-assessment.aspx> (consulté en septembre 2014)

¹⁶⁹ « On appelle lexique mental l'organisation en mémoire des informations que tout individu stocke à propos des mots de la langue. Celles-ci sont au moins de trois ordres pour une personne sachant lire et parler : sémantique, phonologique et orthographique. » Gombert et coll. (2000/2002, p. 21).

recouvrer un certain état d'activation des unités qui, dans le système cognitif, sont concernées par le traitement de l'information lexicale. Chaque configuration différente d'activation correspond alors à la reconnaissance d'un mot différent. » (Gombert, in Ecalle et Magnan, 2002, p. 8). Nous reviendrons plus loin (cf. 3.1.2) sur la différence entre « reconnaître un mot » et « identifier un mot ».

Ce modèle met en avant deux bases de connaissances en interaction : des connaissances visuelles et des connaissances phonologiques. L'expérience avec le langage écrit changerait la nature des représentations phonologiques selon Martinet et Rieben (2010, p. 200) : « en particulier, elle favoriserait la conscience phonémique, qui en retour deviendrait un fort prédicteur de la réussite de l'apprentissage de la lecture. »

3.1.1a. Les relations entre les phonèmes et les graphèmes (critères PG)

- Percevoir et manipuler les phonèmes (tâches PG1) et des syllabes orales (tâches PG2)

L'importance de l'entraînement de la conscience phonémique ne fait plus débat : le principe alphabétique ne peut s'acquérir sans l'intervention explicite de l'enseignement (Gombert et coll., 2000/2002, p. 37). Les travaux relatifs à ce sujet font l'objet d'une large diffusion à destination des enseignants à l'instar de « Phono » (Goigoux, Cèbe et Paour, 2004) ou des publications dédiées à l'entraînement phonologique des éditions La Cigale (Jacquier-Roux et Zorman, 1998). Les programmes de maternelle ont largement repris ces avancées, et une étude de la DEPP¹⁷⁰ de 2013 souligne un accroissement des compétences phonémiques des élèves à l'entrée en CP entre 1997 et 2011.

Ecalle, Magnan et Bouchafa (2002) soulignent que « conformément aux nombreux travaux montrant les liens entre la conscience phonémique et l'apprentissage de la lecture, l'entraînement de type métaphonologique et plus précisément lorsqu'il porte sur les phonèmes contribue au développement des processus de reconnaissance de mots écrits, centraux dans l'acquisition de la lecture ».

Les tâches relatives aux sous-composants PG1 et PG2 renvoient toutes deux à des tâches orales de développement de la **conscience phonémique** : l'élève s'exerce au cours de ce type de tâches à manipuler des phonèmes isolés ou placés dans une syllabe orale.

- Connaissances visuelles et connaissance des lettres (tâches PG3)

Les variables « oculomotrice », « mémoire visuelle et/ou spatiale » et « perception visuelle » prises en compte par les auteurs du National Early Literacy Panel (NIFL, 2008) n'ont, semble-t-il, qu'une faible valeur prédictive concernant l'identification des mots (Martinet et Rieben, 2010, p. 203). La reconnaissance visuelle du mot passe par l'identification de quelques-unes de ses lettres à l'ensemble des lettres. Puis, s'impose la nécessité que le mot soit traité comme une suite de lettres. La prise en compte d'un ordre strict des lettres du mot est apparue comme une acquisition importante pour l'apprentissage des correspondances graphophonologique (Bastien, 1993). Ecalle et Magnan (2002, p. 73 et p. 268) font état de la maîtrise de l'ordre (des lettres dans le mot) comme un outil cognitif extrêmement important en se référant aux travaux de Bastien et de ses collaborateurs. Ils indiquent que le traitement de l'écrit ne peut se réduire au développement de capacités spécifiques de traitement de percepts linguistiques visuels, mais suppose aussi le développement de connaissances

¹⁷⁰ Note d'information de la DEPP du 13-19/09/2013. Repéré à http://cache.media.education.gouv.fr/file/2013/11/2/DEPP_NI_2013_19_forte_augmentation_niveau_acquis_elev es_entree_CP_entre_1997_2011_269112.pdf

morphologiques, orthographiques et phonologiques ce qui peut se résumer dans la formule « du visuel au phonologique ».

En CP, les tâches de reconnaissance visuelle de lettres, hors de tout traitement phonologique, visent principalement à assurer la connaissance des **équivalences** des écritures dans les systèmes d'écriture cursive et scripte, minuscule et majuscule.

- Connaissance des correspondances entre phonèmes et graphèmes (tâches PG4), combinatoire (tâches PG5)

Les connaissances relatives aux lettres engagent des connaissances relatives aux phonèmes et leur représentation au moyen de signes graphiques.

Une lettre est « un signe graphique, qui employé seul ou combiné avec d'autres, représente dans la langue écrite un phonème ou un groupe de phonèmes » (définition du dictionnaire « le petit Robert »). Grévisse et Goosse distinguent cinq aspects dans le système de signes écrits du français (1995, pp.30-50) :

1. L'écriture : on a recours à 26 lettres pour noter les phonèmes du français. Ces signes existent sous deux formes du point de vue du dessin : la minuscule, la majuscule. 6 des 26 lettres du français sont dites « lettres-voyelles » parce qu'elles servent à noter les voyelles. Les autres lettres étant nommées « lettres-consonnes ».
2. L'orthographe : « l'ensemble des fonctions que l'on donne aux lettres et aux autres signes écrits ». Elle ne reproduit pas exactement la prononciation : l'alphabet possède moins de lettres que le français n'a de phonèmes (26 lettres pour 36 phonèmes). Certains phonèmes sont représentés par une lettre, d'autres par une combinaison de lettres (un digramme, un trigramme). Une même lettre peut en outre représenter des phonèmes différents et un même phonème peut, selon les mots, être représenté par différentes lettres. Enfin, beaucoup de mots contiennent des lettres muettes. Il existe 130 graphèmes.
3. Un ensemble de signes auxiliaires tels que : accents, tréma, cédille, apostrophe, trait d'union.
4. Des abréviations et des symboles (dont les chiffres arabes et romains).
5. La ponctuation.

Les programmes n'utilisent pas le terme de graphème, concept bien distinct de celui de lettre qu'ils lui préfèrent. La lettre est un signe qui renvoie à un concept phonographique, or, l'orthographe française, fondamentalement phonétique, n'est pas un système d'écriture qui note seulement les sons. Dans les années 1970, Catach (1973a, 1973b) a proposé une étude du système orthographique français, débouchant sur la notion de plurisystème, classant tous les signes graphiques du français à partir de l'hypothèse selon laquelle les signes écrits transcrivent majoritairement les sons de la langue parlée. Le système orthographique du français est un système à double articulation : phonographique d'une part, supra ou extraphonographique d'autre part (Catach, 1973b, p.30). Le français écrit est une écriture de type alphabétique qui s'appuie sur une correspondance plus ou moins complexe entre des phonèmes et des graphèmes. Le phonème est la « plus petite unité distinctive de la chaîne orale » (Catach, Gruaz & Duprez, 1995, p.16) tandis que le graphème est la « plus petite unité distinctive et/ou significative de la chaîne écrite composée d'une lettre, d'un groupe de lettres (digramme, trigramme), d'une lettre accentuée ou pourvue d'un signe auxiliaire, ayant une référence phonique et/ou sémique dans la chaîne parlée » (idem). Certains graphèmes ont une autre fonction que celle de coder du son (c'est le cas du « -s »

dans un pluriel par exemple) : cette dimension morphologique (dénommée plus haut articulation supra ou extra phonologique) est selon Tauveron (2014, p.11) insuffisamment prise en compte dans l'enseignement de la lecture.

A la suite de Catach, Gruaz & Duprez (1995), on distingue trois types de graphèmes qui se complètent et s'opposent, et qu'il est bien difficile de séparer :

- les phonogrammes, qui sont fidèles au principe de symbole visuel codant un son. Ils représentent plus de 80% du système et ont pour fonction de transcrire un phonème : comme par exemple [u] dans le mot « loup ». 32 à 33 phonogrammes suffisent à transcrire la quasi-totalité (80 à 90%) de notre communication orale (Catach, 1973a, p.5). Ce sont les signes qui transcrivent un son.
- les morphogrammes lexicaux (-h de « inhabité») ou grammaticaux (-x de « choux »). Ils sont porteurs d'une information d'ordre lexical ou d'ordre grammatical (genre, nombre). Ce sont les signes qui ajoutent des éléments de significations (comme les désinences verbales).
- les logogrammes lexicaux (balais/ballet) ou grammaticaux (ou/où). Ils permettent de distinguer les homophones. Ces signes ne traduisent pas les sons du mot, mais le mot tout entier (lettres étymologiques...).

Ces graphèmes permettent la transcription écrite des phonèmes et des morphèmes (« les plus petites unités de sens d'une langue : qu'elle soit de nature lexicale ou grammaticale » selon Catach (idem)).

Ainsi le graphème est-il un signe graphique multiface (considérant les graphies scriptes et cursives, minuscules et majuscules). Il symbolise surtout une unité orale (le phonème) de façon non univoque. Tantôt le graphème code un son ou son absence, tantôt il participe du codage d'un ou plusieurs sons, tantôt il code un autre type d'information, non phonétique : une information lexicale ou grammaticale, voire même encore une information d'une autre nature, extra linguistique : repérage spatial dans un texte (paragraphe a), dans un exposé (A, B, C), dans un espace (le CP A par exemple), voire encore symbolisation d'une appréciation (A dans une évaluation).

Ecalte (2010, p.107) rapporte les travaux de Riley (1996) et Bloodgood (1999) selon lesquels les connaissances relatives au prénom écrit constituent un prédicteur de la réussite en lecture-écriture. Des résultats confirmés récemment par Labat, Farhat, Andreu et coll. (2013, p.45) sur une population de plus de 3 000 élèves : « outre les connaissances phonologiques (phonèmes), l'enfant doit aussi acquérir des connaissances orthographiques (graphèmes) afin de connecter l'écrit et l'oral. En ce sens, la capacité à dénommer les lettres chez les enfants pré-lecteurs constitue l'un des prédicteurs les plus puissants de la lecture ». Puolakanaho, Ahonen, Aro et coll. (2007) soulignent, à l'issue d'une méta-analyse, que la dénomination rapide de lettres, les habiletés phonologiques et la connaissance de lettres sont trois (parmi 6) puissants prédicteurs de la réussite ultérieure en identification de mots écrit (la mémoire de travail, la lecture répétée de pseudo-mots, le vocabulaire expressif sont les trois autres). L'écriture du prénom en maternelle est un indicateur de l'état des connaissances du jeune enfant dans le domaine de la langue écrite et plus précisément sur le code alphabétique. L'écriture correcte du prénom dès 5 ans est liée à un ensemble de connaissances telles que, notamment, la connaissance de l'alphabet. Autrement dit, la connaissance des lettres (et de leur valeur phonologique) dès 5 ans, en fin d'école

maternelle, constitue un indicateur fondé de nature à renseigner sur l'état des connaissances des élèves, connaissances qui interviennent directement dans le savoir-lire de l'enfant.

Recherchant sur quels types de connaissances visuelles et phonologiques s'appuient la maîtrise du code alphabétique, Ecalle (op.cit., p. 105) souligne que la compétence métaphonologique en CP (prise de conscience de la structure segmentale des mots en syllabes (consciencs syllabique) et en phonèmes (conscience phonémique) et possibilité de manipuler ces segments) semble être clairement prédite par une compétence épiphonologique (capacités utilisées automatiquement par l'enfant lui permettant d'identifier une rime commune à un ensemble de mots par exemple) présente en Grande Section de maternelle.

Ecale souligne par ailleurs que la capacité précoce à établir des relations oral-écrit (percevoir le lien entre la quantité d'oral et la quantité d'écrit) explique pour une large part les performances en identification et en production de mots réalisées en fin de CP. D'autres connaissances relatives aux lettres (non assemblées) sont tout aussi importants à pouvoir repérer surtout si l'on considère que certaines connaissances implicites (par exemple règles sous-tendant l'organisation de séquences de lettres) ont un impact sur les performances en lecture (Ecale, op.cit. p.103). Le problème soulevé étant celui du passage de connaissances implicites à des connaissances explicites.

Les tâches de correspondance phonographiques mobilisent des connaissances sur le code alphabétique permettant aux élèves de proposer une transcription écrite d'un phonème, d'une syllabe ou d'un mot à l'opposé des tâches de correspondance graphophonologiques qui nécessitent de proposer une traduction orale d'un graphème, d'une syllabe ou d'un mot écrit.

Ces traductions reposent sur la maîtrise de la **combinatoire** qui, selon Charmeux¹⁷¹, est « l'ensemble des règles, spécifiques à chaque langue, et par voie de conséquence, arbitraires mais organisées en système, qui permettent que des unités linguistiques abstraites puissent devenir productrices de sens par leur présence (ou leur absence) et leur ordre, dans des productions langagières ». Pour cet auteur, la maîtrise de la combinatoire constitue un indicateur majeur de la réussite du savoir lire/écrire. Elle implique la maîtrise d'un système complexe de combinaisons entre des lettres pour former des syllabes et des mots à l'écrit et, entre les lettres, les syllabes et les mots écrits ainsi obtenus et leur traduction orale. Cela signifie que l'élève maîtrise l'arbitraire des signes linguistiques : le signifié du graphème renvoie en effet à la prononciation (d'un phonème dont le signifié renvoie à des éléments de l'expérience, des objets, des idées, des faits...) et la valeur phonologique d'un assemblage de lettres dépend du contexte linguistique (la lettre « a » ne possède pas de prononciation à laquelle on pourrait se reporter de façon mécanique).

Au total, les cinq types de tâches phonographiques listées par Goigoux et coll. (2013) renvoient à plusieurs sous-composants essentiels du composant « Relations phonographiques » : la conscience phonémique (qui résulte d'un enseignement explicite) au niveau de phonèmes isolés ou de syllabes orales, la connaissance des traductions entre systèmes d'écriture, la maîtrise des correspondances phonographiques et la combinatoire. On s'intéressera à observer comment ces différents sous-composants des relations phonographiques sont pris en charge dans les tests d'évaluation en particulier en début de

¹⁷¹ Repéré à <http://www.charmeux.fr/combinatoire.html>

CP. En début de CP et de façon prioritaire, le repérage de l'état de disponibilité de ces connaissances devrait être articulé avec un repérage de l'état des compétences métaphonologiques. En revanche, l'évaluation de ces connaissances ne devrait pas avoir lieu en fin de CP, sauf cas particuliers : il semblerait inutile de demander à un élève de lire un texte et de répondre à des questions de compréhension au sujet de ce texte si ces connaissances ne sont pas acquises.

3.1.1b. L'identification des mots (critères L)

Grévisse et Goosse (1995, p. 52) définissent le mot comme « une suite de sons (ou de lettres si l'on envisage la langue écrite) qui a une fonction dans une phrase donnée, et qui ne peut se diviser en unités plus petites répondant à la même définition ». Ensemble, « les mots d'une langue forment le lexique » (idem.)

Cette définition du mot appelle trois commentaires :

1. La première partie de la définition du mot insiste sur la constitution du mot : une suite de sons ou de lettres. Les combinaisons de ses constituants constituent un premier axe d'apprentissage contribuant à identifier un mot.
2. La deuxième porte sur sa fonction, ce qui amène à considérer le mot dans un ensemble plus vaste : la phrase.
3. La troisième partie renvoie au lexique. Identifier un mot implique de pouvoir référer ce mot au lexique et lui attribuer un sens (c'est aussi la définition¹⁷² retenue par Gombert et coll. 2000/2002, p.24). Le lexique (ensemble des mots) du français emprunte au latin l'essentiel de son lexique (le fonds primitif). Il emprunte à d'autres langues classiques ou vivantes, suivant les périodes de l'histoire, un autre fonds lexical. Deux procédés (dérivation et composition) permettent d'obtenir de nouveaux mots : soit un élément non autonome est ajouté (suffixe, préfixe), soit plusieurs mots sont composés. Le sens d'un mot (ou sa signification) « n'est pas la réalité qu'il désigne, mais l'idée que l'on se fait de cette réalité » (idem p.67).

Lire un mot n'est donc pas qu'affaire d'association de graphèmes, mais mobilise également des connaissances linguistiques précises liées à la grammaire et au lexique. La définition précitée mêle trois aspects que les programmes d'enseignement distinguent : « Vocabulaire », « Orthographe » et « Grammaire » sont constitués en trois sous-domaines du français au même titre que « Langage oral », « Lecture » et « Ecriture ».

On peut penser que cela aurait pour conséquence une évaluation dissociée de la capacité à accéder à une modalité de recodage phonologique d'un mot écrit (quelle que soit la voie utilisée) et de la capacité à l'identifier, une capacité qui ne peut être dissociée d'une représentation lexicale.

La place du lexique n'est pas explicite dans la typologie de Goigoux et coll. (2013), nous la rendons explicite dans la définition du composant « Lecture » qu'il pourrait être judicieux de renommer « Identification ».

Pour déchiffrer (terme synonyme de décoder) le mot « écolier », plusieurs stratégies peuvent coexister qui rendent difficile le fait de dissocier véritablement les critères L1 (lecture

¹⁷² Nous adopterons cette définition du terme dans notre travail. Nous signalons simplement qu'elle est fluctuante suivant les auteurs. Ainsi, pour Giasson (1990, p. 39) le terme « reconnaissance » renvoie à « identification » et vice-versa.

silencieuse), L2 (reconnaissance de mots), L3 (déchiffrage) et L4 (lecture à voix haute) de la typologie de Goigoux et coll. (2013) :

- Soit un déchiffrage intégral du mot par la voie d'assemblage si aucune représentation lexicale (dans ses aspects sémantique, phonologique, orthographique, syntaxique, imagée ou conceptuelle) n'est activée. C'est le traitement cognitif le plus coûteux. Il être également entrepris par attachement procédural (Ecalte et Magnan 2002 p.58).
- Soit une reconnaissance directe du mot (dans le cas où le mot a déjà été rencontré). A l'inverse de ce qui précède, c'est le traitement cognitif le plus économique.
- Soit un traitement mixte dans lequel l'apport d'indices contextuels intervient : reconnaissance du radical « écol- », puis reconnaissance du suffixe « -ier » ou déchiffrage du groupe de lettres « -ier », et enfin assemblage.

Nous discuterons plus loin d'une formule classique établie par Gough et Tunmer (1986) liant décodage et compréhension orale dans la lecture. Par anticipation, nous disons ici que Gough précisera en 1996 que, par décodage, il faut entendre « lire des pseudo-mots » -i.e. une combinaison régulière de graphèmes dénuée de sens- ce qui constitue une activité de lecture très particulière dans laquelle le sens n'est pas convoqué et qui ne renvoie donc non pas à l'identification de mots. On peut concevoir l'intérêt d'un entraînement à la maîtrise du code au cours de tâches d'apprentissages. Ceci ne permet toutefois pas de lire une phrase comportant le verbe « être » conjugué à la troisième personne du présent. **L'identification** d'un mot ne trouve sa pleine et entière réalisation que si elle permet au lecteur d'attribuer une valeur sémantique acceptable au déchiffrage réalisé.

Les programmes demandent d'enseigner aux élèves à reconnaître rapidement des mots outils, mots hyper fréquents. Nous renvoyons le lecteur à l'analyse de la tâche « Niche » (cf. III-5-3) au cours de laquelle nous avons présenté cette liste.

- Que permet d'évaluer la lecture à haute voix ?

Le rapport du PIREF recommandait en 2003 que la lecture à haute voix « ne soit pas, prioritairement, une opération de contrôle du déchiffrage ». Les conclusions du CNETCO en 2016 diffèrent sur ce point : « la lecture à haute voix par les élèves a un effet bénéfique sur la maîtrise du code ». En d'autres termes, il est nécessaire de préciser la finalité que l'on accorde à la lecture à haute voix : exercer (ou évaluer) la maîtrise des relations graphophonologiques ou évaluer la compréhension ?

E. Fijalkow (2003) établit que la lecture à haute voix est le moyen d'évaluation le plus fréquemment utilisé surtout en début d'année de CP. Les inspecteurs généraux Bouysse et Pétreault soulignent dans leur évaluation des dispositifs PARLER et ROLL (2012, p.9) le risque de lier amélioration de la fluence et la compréhension : l'un ne garantit pas l'autre. Lecoq et coll. affirment (1996, p.54) qu'« au moins dans les premières années d'apprentissage de la lecture, il n'existe pas de différences significatives dans les performances en compréhension lorsque les sujets procèdent à une lecture à voix haute ou à une lecture silencieuse ». Reste à savoir si la lecture à haute voix permet véritablement d'évaluer la compréhension. Cèbe, Goigoux et Thomazet indiquent pour leur part que « le plus souvent, la lecture à haute voix ne laisse pas apparaître de problèmes majeurs » (2004, p.2).

La **fluence** de lecture est définie comme la capacité à lire avec aisance, rapidement, sans erreurs et avec une intonation adaptée. Zorman et coll. ont élaboré un protocole d'évaluation de la lecture en Fluence (2008). 40 représente la valeur chiffrée de la variable MLCM

(nombre de mots lus correctement par minute¹⁷³) en juin en CP en deçà de laquelle les « élèves doivent participer aux ateliers de fluence », ce qui contribue à établir une norme¹⁷⁴. Goigoux et Cèbe (2013, p. 8) font état d'une fluence moyenne de 55 mots (MLCM) en CE1 qui corrobore les résultats précédemment indiqués. Ces auteurs prennent toutefois la précaution de souligner que les écarts entre élèves sont importants. Le test E.L.FE¹⁷⁵ (Zorman et coll.) établit le MLCM moyen à 68 en CE1. La prise en compte du lien entre vitesse et compréhension reste significative dans deux cas minoritaires et opposés, dont celui du CP, affirme Mesnager¹⁷⁶. « Pour nombre d'élèves de fin de CP, ou encore faibles en CE2, la lenteur est corrélée à une mauvaise compréhension parce que, ne possédant que peu de mots identifiables « directement », ils déchiffrent la plupart de ceux qu'ils rencontrent et ne peuvent, trop enlisés sur chaque mot, construire de façon efficace des ensembles de sens ». Des questions se posent quant aux situations de mesure de la fluence : s'il est établi que les élèves lisent à vue (sans préparation préalable) et à voix haute (d'après les travaux de Lecocq et coll. précités, la compréhension est identique à une situation de lecture silencieuse), on peut en revanche douter que lire un texte ou une liste de mots isolés soit une tâche comparable. Dans la situation de lecture de texte (telle que celle du test ELFE), d'autres contraintes sont convoquées que la seule reconnaissance de mots. Aussi conviendrait-il sans doute mieux de préciser qu'il s'agit d'une fluence textuelle. Ceci renvoie nécessairement à l'observation fine du texte soumis à l'élève en termes de difficultés (de décodage graphophonologique, en termes de prédicats et d'inférences, en termes de fréquence lexicale des formes de mots contenus dans le texte par exemple).

- La prise en compte des représentations lexicales

Cette dimension doit être évaluée conjointement avec les aspects liés au recodage phonologique. Le cadre d'évaluation des composants des compétences en lecture du LAMP (Sabatini, 2004) met en avant l'évaluation de mots « quotidiens » dans le but de pouvoir mesurer si une personne est capable de lire des textes qu'elle comprendrait à l'écoute. On peut penser qu'il serait opportun d'évaluer prioritairement les mots les plus fréquemment rencontrés dans les manuels ou les textes que les élèves lisent quotidiennement (la base Manulex résultant des travaux de Lété, Sprenger-Charolles & Colé (2004) permet de quantifier la fréquence d'apparition des mots des manuels de lecture).

De façon synthétique, il semble difficile de prendre en compte séparément les critères L1, L2, L3 et L4 sauf à considérer à part des autres les mots dits mots outils. Pour réaliser des tâches de type phonographiques (PG) ou de type lecture (L), l'élève met en œuvre des processus de lecture que Giasson (1990, p.16) qualifie de microprocessus. Ces processus servent selon cet auteur (idem) à « comprendre l'information contenue dans une phrase ».

3.1.1c. La compréhension de l'écrit (critères C)

Nous nous intéressons à présent au sous-composant « Compréhension de l'écrit ». Une note du directeur général de l'enseignement scolaire (Blanquer, 2011) indique qu'en fin de CM2, en lecture, « les résultats font apparaître une bonne maîtrise de l'identification des mots chez la plupart des élèves et leur capacité à prélever des informations ponctuelles dans

¹⁷³ Correspond au nombre de mots correctement lus en une minute auquel on retranche les erreurs et les omissions.

¹⁷⁴ Diaporama consulté le 11 juillet 2013 à l'adresse <http://circ41-contres.ac-orleans-tours.fr/php5/animations/AP%202011-2012/fluence/cycle2/fluencecycle2.pdf> (diapositive 21)

¹⁷⁵ Cognisciences

¹⁷⁶ Cf. Mesnager (non daté) article Vitesse de lecture sur le site TFL

un texte, comme à en connaître le thème général. En revanche, de nombreux élèves (près de la moitié) éprouvent des difficultés à inférer une information nouvelle de leur lecture. Enfin, très peu d'entre eux parviennent à une compréhension fine des textes et à exprimer et justifier une interprétation ». Cette note conclut sur la nécessité de conduire un enseignement de la compréhension des textes annoncé en 2003 dans le rapport du jury du PIREF : « l'obstacle majeur que rencontrent les élèves dans l'activité de compréhension d'un texte est la mise en relation des informations nouvelles avec celles qu'ils avaient déjà mémorisées et traitées ».

A ce jour, cet enseignement n'est pas très développé, et pourtant, « identifier clairement les problèmes de compréhension que pose le récit littéraire aux élèves est une première étape indispensable si l'on se donne pour objectif d'apprendre aux élèves à les surmonter » (Tauveron, 2002, p.23). Cette remarque a également été établie par les inspecteurs généraux (MEN, 2010, p.12) : « En lecture, les observations témoignent d'un travail très nourri sur le code au cycle 2. La découverte des textes reste indigente au Cours Préparatoire et les pratiques courantes au CE1 ne constituent pas un réel enseignement de la compréhension. Ces pratiques sont essentiellement fondées sur des exercices de questions / réponses corrigés sans être réellement expliqués, que ce soit du point de vue de la langue (formation des mots, ...), du texte (identification des protagonistes et des événements référés aux éléments du texte, mise en relations d'éléments éloignés...) ou de la culture (explications de points significatifs qui peuvent échapper aux élèves) ».

Se posent ainsi et particulièrement deux questions : la première est celle que l'enseignement de la compréhension entretient avec celui de l'identification des mots écrits, la deuxième est celle des connaissances spécifiques de la compréhension qui ne sont qu'évoquées par le rapport du PIREF (2003).

- Les liens entre l'enseignement de la compréhension et l'identification de mots

Martinet et Rieben (2010, p. 190) soulignent l'acquis majeur de la recherche en lecture selon lequel on distingue clairement les deux facettes de la lecture que sont, d'une part l'identification de mots et d'autre part, la compréhension écrite (de textes lus par le lecteur ou par autrui). Cependant l'articulation entre ces deux facettes est discutée. En 1986, Gough et Tunmer dénommaient sous le nom de « Simple View of reading » (Gough, 1996, p.4) l'idée que lire (comprendre un texte écrit de manière autonome) résultait de deux activités : une activité de reconnaissance des mots du texte et une activité non spécifique de la lecture : la compréhension.

Sans lui définir de domaine de définition (par exemple celui de la nature des textes), cette « simple vue » est traduite en équation : $L = D * C$ où, « L » est « extraire l'information de ce qui est écrit » et est considérée égale au produit de « D » « la reconnaissance de mots » (dont nous avons dit précédemment que Gough précisera en 1996 que par cela il faut entendre lire des pseudo-mots ce qui correspond bien à la notion de reconnaissance de mots par opposition à celle d'identification de mots) et « C » « la compréhension orale, sémantique et syntaxique » (selon la traduction proposée par Zorman, 2010 par exemple). La variable « L », produit des deux autres, désigne l'extraction d'information, et non le traitement de l'information extraite, deux opérations cognitives qui ne sont pas typiquement équivalentes. « Les progrès de la recherche sur les processus en jeu dans la lecture ont mis en évidence la nécessité de conjuguer l'identification des mots par leur décodage, et (le) travail de compréhension.... Ces conclusions sous-tendent les instructions officielles pour

l'école primaire arrêtées en 2002. (Rapport 2005-123 ONL - IGEN) ». Le même rapport souligne que « le problème auquel se trouve confronté l'enfant qui entre à l'école élémentaire est d'apprendre à identifier les mots écrits et de mettre en œuvre l'activité de compréhension à partir de cette identification ». La compréhension n'étant pas du seul apanage de l'écrit, la question a donc prioritairement porté sur l'aspect du décodage, ce qui est très explicite dans lesdits programmes (cf. p.71 des programmes 2002). La difficulté de compréhension est réputée supérieure pour des « textes longs ou complexes » (IO 2002, p.72).

Dans les programmes de 2008, l'accentuation sur les aspects d'automatisation du décodage et de compréhension du langage oral est palpable.

La formule $L=D*C$ a ainsi été re-présentée lors du séminaire national des conseillers pédagogiques tenu en le 11 avril 2012 (DGESCO – MENJVA) à Paris dans un document intitulé « Ateliers pour la classe et l'aide personnalisée au cycle 2 ». Dehaene, Huron et Sprenger-Charolles introduisent ce document par la « présentation de quelques grands principes de l'apprentissage de la lecture » :

- « le principe d'enseignement explicite du code alphabétique,
- le principe d'une progression rationnelle (« prochainement proposée » mais restée à l'état d'annonce),
- le principe d'apprentissage actif associant lecture et écriture,
- le principe de transfert de l'explicite vers l'implicite,
- le principe de choix rationnel des exemples et des exercices,
- le principe d'effort et de récompense ».

L'objectif poursuivi par ces auteurs est de « permettre à l'enfant d'automatiser rapidement les procédures d'identification et de reconnaissance des mots écrits, afin de libérer des ressources cognitives qu'il pourra utiliser pour bien comprendre ce qu'il lit, et devenir ainsi un lecteur autonome, qui lit pour apprendre et pour son plaisir. » La référence à la méthode de Borel-Maisonny (une méthodologie de rééducation utilisée par l'enseignant C de notre population par exemple) est explicite, elle sert à l'énonciation d'une proposition de progression pédagogique en adéquation avec les principes précédemment développés. Outre l'absence de tout principe relevant de la morphologie ou de la compréhension, le recours à ce type d'outil de rééducation, pose la question de l'apprentissage : quels mots peut-on raisonnablement lire avec les seules voyelles a, é, u et i au bout de 5 semaines d'apprentissage de la lecture (les combinaisons « au » « ai » étant évitées de surcroît) ? Le rapport du CNECSCO préconise que « pour ne pas enfermer les élèves dans des tâches entièrement disjointes (des activités de mise en correspondances des graphèmes/phonèmes sans prendre en compte de la signification des mots d'un côté, et de l'autre des activités de compréhension de textes entendus), il est important de leur donner le plus vite possible les moyens d'une autonomie dans la reconnaissance des mots écrits » (2016, p. 7). Cela passe par l'introduction » d'au moins une dizaine de correspondances graphèmes/phonèmes, dès le début du CP, afin de permettre aux élèves de décoder des mots de façon autonome » (idem).

La formule $L=D*C$ et surtout les conceptions qu'elle sous-tend, bien que reprise par l'ONL (1998, p.157), est critiquée. Elle est critiquée par J. Fijalkow (2000, p. 133) qui réfute un point de vue psycholinguistique et fonctionnaliste : « si l'on accepte cette équation, et si l'on admet par ailleurs que la compréhension est la même à l'écrit qu'à l'oral, alors, le seul facteur spécifique de la lecture est le décodage. Ainsi, lire c'est décoder ». « Que fait alors

l'enfant durant la période où il n'a pas automatisé le déchiffrage, il ne comprend rien de ce qu'il déchiffre avec peine ? » s'interrogent aussi Lecocq et coll. (1996, p. 109).

Selon Cèbe et Goigoux (2007, p.194), Goigoux (1998a) la conception sous-jacente du modèle de la simple vue « masque les spécificités de la compréhension du texte écrit : les différences entre langue orale et langue écrite (lexique ou syntaxe), les particularités des organisations textuelles et, bien sûr, les régulations de l'activité rendues possibles par la permanence de la trace écrite, à l'opposé de l'éphémère oral. » Avec pour conséquence, poursuivent Cèbe et Goigoux (idem), le fait que les enseignants les moins expérimentés négligent les difficultés de compréhension des élèves dès lors que l'identification des mots ne pose plus de problèmes majeurs. La comparaison des propriétés de l'oral et de l'écrit montre que ces registres se distinguent sur la quasi-totalité des dimensions d'analyse du langage : outre celles déjà soulignées par Cèbe et Goigoux, Bianco (2010) relève les thèmes abordés, le mode d'interlocution. La plateforme « M@gistère » désormais utilisée pour la formation à distance des enseignants propose en 2014 un parcours concernant la compréhension en cycle 3 (version 2014) dans lequel un diaporama présente cette formule réactualisée : $L = D * CO * Cé$. La conception de la variable C est modifiée : elle est à présent subdivisée en Co (compréhension orale) et Cé (compréhension de l'écrit).

La compréhension n'est pas spécifique du langage écrit, « puisqu'elle s'exerce dès le plus jeune âge, sur le langage oral ou l'image » (cf. Simard ; Dufays ; Dolz et Garcia-Debanco 2010, p.240). Peut-on justifier d'un enseignement de la compréhension écrite dès lors, surtout si l'on s'accorde à penser avec Martinet et Rieben (2010, p.212) que « fluence et compréhension sont intimement liés », et que l'on postule avec ces auteurs que « les lecteurs non fluents ne peuvent pas comprendre facilement ce qu'ils ont tant de mal à décoder puisqu'une grande part des ressources attentionnelles sont allouées aux mécanismes dévolus à l'identification des mots » (Rapport du National Reading Panel, NICHD, 2002 cité par Martinet, idem) ? En d'autres termes, est-il légitime d'évaluer la compréhension d'un texte que l'élève doit lire seul s'il n'est qu'un « faible » décodeur ?

Les connaissances spécifiques de la compréhension.

La compréhension en lecture telle que l'envisagent les rapporteurs des premières analyses de l'enquête Lire-écrire au CP, correspond selon Bishop, Cèbe et Picquée (2016) à une conception largement partagée dans le domaine de la recherche :

« Il s'agit d'une activité complexe consistant à élaborer de manière dynamique et progressive un modèle mental¹⁷⁷ de la situation évoquée (selon le modèle présenté par Van Dijk et Kintsch en 1983¹⁷⁸). (...) Il s'agit d'un objet à enseigner encore peu élaboré, et il semble exister un manque d'informations et de formation sur l'enseignement de la compréhension, sa didactique (par rapport à la quantité de travaux didactiques et pédagogiques sur l'écriture et sur le décodage) étant encore en voie de constitution (Bishop, 2009). Cette particularité explique que la place accordée à un tel enseignement au cycle 2 ainsi que les pratiques mises en œuvre dans les classes sont encore mal connues ».

¹⁷⁷ Modèle mental et modèle de situation sont généralement utilisés comme des synonymes (Gyselinck, 1996) ; nous retiendrons dans la suite le terme de modèle de situation.

¹⁷⁸ Pour une synthèse, voir N. Blanc et Brouillet, 2003

La typologie des tâches d'enseignement de la lecture/écriture utilisée par Goigoux et coll. (2013) dans le cadre de l'enquête IFé correspond selon ces auteurs (ibidem) « aux attendus institutionnels tels qu'ils sont définis dans le programme de 2008, aux données issues de la recherche sur la description des processus de compréhension (Van Dijk et Kintsch, 1983 ; N. Blanc et Brouillet 2003) et à ce que les chercheurs, enseignants et formateurs connaissent des pratiques habituelles des maîtres de CP ».

Nous détaillons ci-dessous la référence faite par ces auteurs aux travaux de Kintsch en psychologie cognitive.

N. Blanc indique que « les processus de compréhension sont des processus généraux de l'activité cognitive » (2010, p. 8). Un large consensus se dégage pour considérer que comprendre un texte « c'est construire une représentation mentale intégrée et cohérente de la situation décrite par ce texte - ce qu'on appelle souvent un modèle mental » (N. Blanc, 2010 ; Fayol, 2003). Cette représentation signifiante de la situation exposée finalise la compréhension par l'établissement d'une représentation cohérente qui suppose un rétrocontrôle du lecteur. Pour cela l'activation (ou instanciation) de connaissances antérieures est l'opération cognitive la plus largement évoquée (Baudet et Denhière, 1990, p.7). Les recherches issues de la psychologie en matière de compréhension de texte ont été largement influencées par la théorie élaborée par Kintsch et van Dijk (Kintsch, 1974, Kintsch et van Dijk, 1978 ; van Dijk et Kintsch, 1983 ; Kintsch 1998). Cette théorie « propositionnelle » de la compréhension a été largement diffusée dans le monde de la recherche (cf. Denhière et Baudet, 1992). Elle se fonde sur une méthode de représentation des textes sous la forme d'une hiérarchie de propositions sémantiques.

Dans cette théorie, la compréhension d'un texte se déroule sous la forme de cycles de traitement mobilisant des microprocessus qui consistent à identifier, connecter les propositions et sélectionner celles à garder en mémoire, et des macro processus qui visent à réduire et organiser l'information détaillée qui constitue la microstructure du texte (Rouet, 2000, p.7).

Notons que le terme microprocessus employé par Rouet ne renvoie pas au même sens que lui accorde Giasson (1990, p.16) (et que nous utilisons dans notre travail) pour qui les microprocessus « renvoient aux processus qui servent à comprendre l'information contenue dans une phrase (identification des mots, lecture par groupe de mots) ». Cet auteur distingue d'autres types processus de lecture :

- des processus d'intégration (« qui ont pour fonction d'effectuer des liens entre les propositions ou les phrases » (idem)) ;
- des macro processus (« orientés vers la compréhension globale du texte vers les liens qui permettent de faire du texte un tout cohérent » (idem)),
- des processus d'élaboration (« ils permettent au lecteur de dépasser le texte, d'effectuer des inférences non prévues par l'auteur » (idem)),
- des processus métacognitifs (« ils gèrent la compréhension et permettent au lecteur de s'ajuster au texte et à la situation » (idem)).

Denhière et Baudet se sont intéressés aux processus de lecture mis en œuvre par le lecteur au cours de l'activité de lecture. Ces auteurs distinguent deux classes principales de modèles de la compréhension : les modèles cognitivistes à instanciation de schémas et les modèles connexionnistes d'activation (1992, p.147). Ils ont en commun le rôle essentiel accordé à l'intervention de connaissances qui ne sont plus uniquement ou principalement

linguistiques (Baudet et Denhière, 1990, p.8) dans la compréhension. Ces modèles diffèrent principalement quant à leurs représentations des connaissances en mémoire : pour les uns, elles sont des unités mémorielles disponibles, pour les autres, les connaissances en tant qu'unités pré-stockées n'existent pas. L'émergence par activation d'une connaissance étant dans ce cas le résultat d'un parcours spécifique du réseau à partir d'une position donnée. Ce parcours est dépendant du contexte.

Modèles de la compréhension de classe 1 :

Les premiers modèles postulent l'intervention de schémas (Anderson, 1977) contraignant la construction de la représentation. Les schémas sont des « structures de données préconstruites et disponibles en mémoire qui, une fois activées, vont contraindre la construction de la représentation à tous les niveaux ». Le modèle de compréhension élaboré par Kintsch et van Dijk en 1978, revu en 1983, est représentatif de cette première classe de modèles. Pour ces auteurs, comprendre un texte c'est traduire une représentation propositionnelle élaborée à partir de l'identification de mots en modèle mental de situation décrivant l'état des choses grâce à la mise en œuvre d'inférences fondées sur les connaissances générales ou spécifiques du sujet. Il s'agit de procéder à l'élaboration de représentations se situant à trois niveaux :

1. la structure de surface constituée par la liste des mots,
2. le contenu sémantique *local* (la microstructure) et *global* (la macrostructure) et,
3. le modèle de situation.

L'articulation des trois niveaux de représentation permet de saisir la mise en œuvre des processus de compréhension au cours de la lecture (Ecalte et Magnan 2002, p. 62 ; Bianco, 2010, p.231).

Comprendre un texte consiste ainsi, dans ce modèle, à reconnaître dans un premier temps les mots et à accéder à leur signification, dans l'ordre syntaxique, ce qui permet d'établir une représentation de surface du texte. Le traitement de chaque mot active une représentation lexicale à différents niveaux selon le modèle développé par Ecalte et présenté précédemment (2002, p.61). Des représentations conceptuelles liées à la représentation lexicale mais inappropriées par rapport au domaine de référence seront inhibées et rejetées.

Dans un second temps, La base de texte est construite. Le lecteur établit la cohérence locale d'un énoncé à l'autre (les idées énoncées séquentiellement dans chaque phrase entretiennent des relations que l'analyse syntaxique ne permet pas de saisir, ce qui fait privilégier, pour l'analyse, la proposition plutôt que la phrase). Il s'agit pour l'essentiel des relations de référence, temporelles et causales (cela renvoie aux processus d'intégration chez Giasson). Parallèlement, l'essentiel des informations est extrait (le thème), et le lecteur accède à une représentation globalement cohérente (ou macrostructure) qui peut être considérée comme une forme d'interprétation partagée. Les macro processus sont gouvernés par des règles permettant de réduire et d'organiser l'information détaillée qui constitue la microstructure du texte. Trois règles sont définies : suppression, généralisation, construction (Rouet, 2000, p.7).

Au-delà des connaissances impliquées jusqu'ici, les connaissances antérieures sont intégrées. Cela a pour effet de renforcer la cohérence des informations lues, mais aussi d'enrichir la représentation avec des données plus ou moins personnelles, conduisant à des modèles de situation variables suivant les individus. Ce niveau est lui-même déterminé par les connaissances initiales du lecteur et les buts qu'il poursuit. Cette activation, produite par

la forme du texte et les structures linguistiques, s'accompagne de l'inhibition d'autres modèles de situation » (cf. Ecalle, 2002, p. 62). Dans un mouvement cyclique, le modèle activé influe les deux premiers niveaux. Ce qui fait dire à Ecalle et coll. (idem) que le modèle est dépendant de la cohésion du texte (assurée par les anaphores et les connecteurs notamment) et de la cohérence du texte (fonction des croyances et connaissances du lecteur). Au cours de la construction de sa compréhension, le lecteur doit, de plus, exercer un contrôle de sa compréhension souligne Bianco (idem) : une dimension importante dont la mise en œuvre peut varier en fonction de différentes variables dont le contenu des textes.

Modèles de la compréhension de classe 2 :

Dans les seconds modèles, la compréhension n'est plus décrite comme sur une série d'instanciation de schémas. Elle consiste à activer des connaissances (envisagées comme des portions d'un réseau d'associations), et à établir de nouvelles associations entre les nœuds du réseau, à former un réseau représentationnel (N. Blanc, 2010, p.9). Le modèle de Construction-Intégration de Kintsch (1988) illustre cette classe. Le système de construction engendre une base de texte mêlant information provenant du texte et connaissances du lecteur. Sa principale caractéristique est d'être non cohérente (Bianco, 2010, p.236 ; Denhière, 1992, p.148 et p.273) du fait qu'elle contient l'ensemble des éléments activés et comporte notamment de ce fait les différentes significations d'un même mot. Le système d'intégration renforce les éléments appropriés et inhibe les autres, intégrant ainsi la base de texte en un tout cohérent. Si la base obtenue n'est pas cohérente, le lecteur, dans une activité délibérée, recherche à produire les inférences nécessaires, rapprochant alors l'activité d'une résolution de problème. N. Blanc (2010, p.43) insiste particulièrement sur les inférences de type causal « au cœur de l'habileté à structurer un récit ». De cette capacité (à identifier les relations existant entre les différents éléments) dépend « la construction d'une représentation cohérente en mémoire, représentation qui notamment facilite le rappel de texte ».

Aussi Denhière affirme-t-il la nécessité de « prendre en compte simultanément les structures linguistiques du texte, les représentations langagières du lecteur et sa représentation du monde auquel le texte réfère » (1992, p.289).

Des limites sont toutefois formulées à ces modèles par Rouet (2000, pp.16- 19) qui souligne particulièrement trois problèmes théoriques.

Le premier concerne le rapport entre texte et situations. A ce sujet, Rouet indique que l'analyse propositionnelle ne permet pas de représenter la source de l'information ce qui permettrait de prendre en compte les aspects illocutoires (relatifs à l'intention de l'auteur) et perlocutoire (effet produit) de la production d'un énoncé. La source de l'information joue pourtant un rôle dans la construction de la cohérence textuelle (Tauveron, 2014) l'évoque sous le terme de « coups narratifs ») qui suggère la mobilisation de connaissances sur les propriétés rhétoriques des textes.

Le deuxième problème souligné par Rouet est celui de l'intégration entre textes : lire deux textes ne se réduirait pas à lire deux fois plus. Nous retiendrons de ce problème soulevé par Rouet une interrogation fondamentale sur la conduite de l'évaluation de la compréhension d'un texte A par un questionnaire, que nous avons considéré comme un (méta-)texte B (cf. III. 4).

Le troisième problème que relève cet auteur (2010, p. 19) est celui des objectifs de lecture : « dans la plupart des expériences de laboratoire, les tâches de compréhension n'ont aucune finalité précise, si ce n'est de se préparer à des tests dont la nature précise n'est pas

toujours connue du sujet ». Tauveron (2014) souligne la non-prise en compte des stratégies narratives (rétablir la valeur illocutoire du texte) dans ces modélisations. Particulièrement la compréhension débute, selon Rouet, par l'évaluation des besoins d'informations. Cette dernière dimension est absente des épreuves collectées. Le besoin d'information est entièrement pris en charge par l'épreuve. Il est donné aux élèves et correspond au fait de pouvoir répondre à des questions (ce terme est employé dans la définition de l'item des progressions relatif à la compréhension de texte).

Nous nous intéressons à présent à la deuxième variable du modèle présenté au début de ce chapitre : la variable « Texte ».

3.1.2. La variable « Texte »

Les propos précédemment cités de Denhière convergent avec la nécessité affirmée en début de chapitre d'articuler la variable Lecteur à la variable Texte. On ne peut faire l'économie d'une description fine des textes (et des questionnaires) lorsque l'on s'intéresse à l'évaluation de la lecture. Mais il n'existe pas de familles de situations clairement définies permettant d'effectuer des catégorisations fines comme l'a fait Vergnaud pour les relations additives en mathématiques (cf. partie IV, chapitre 2.1). Il existe des typologies de texte mais leur enseignement n'est pas au programme du cycle 2 en 2008 (cf. texte des programmes, MEN BO 2008, p.17) ce qui rend ce travail complexe. Les programmes précisent simplement qu'au CE1 « des textes plus longs, plus variés, comportant des phrases plus complexes, sont progressivement proposées aux élèves » (idem). Tout comme en mathématiques pour l'item « résoudre un problème simple à une opération » (cf. discussion conduite en IV, 2.3), la simplicité (du texte) n'est pas caractérisée. A la lecture des programmes précédemment cités, la simplicité du texte évoquée pourrait se résumer à sa faible longueur, l'unicité du type de texte donné et la simplicité des phrases qu'il contient. Nous renvoyons le lecteur aux analyses faites dans le chapitre 4.2 (partie III) et aux communications des résultats d'une recherche conduite en parallèle de cette étude concernant la conception d'une épreuve de lecture en fin de CP, valide du point de vue psycho-didactique (cf. C. Blanc, 2013, 2014 ; C. Blanc et Vantourout, 2015 ; Vantourout et C. Blanc, soumis ; Vantourout et Maury, accepté).

Deux dimensions de la variable « Texte » du modèle général doivent être distinguées :

- La première est en réalité liée au contexte (et au co-texte) du texte et de la situation dans laquelle il est présenté au lecteur. Dans le contexte de l'évaluation, et cela peut être élargi à l'enseignement dans bien des cas, le texte est accompagné d'un questionnaire : à l'école, le lecteur doit rendre compte de ce qu'il a « compris » de sa lecture. La lecture est à la fois outil et objet d'enseignement : elle est aussi outil et objet de l'évaluation. Elle s'emploie pour son propre enseignement (cf. Reuter (1994, p.105) et Simard, Dufays, Dolz et Gracia-Debanco (2010, p.19)) mais aussi pour sa propre évaluation.

Les tâches utilisées par les psychologues sont fort éloignées de celles utilisées par les enseignants. Le rappel de texte, intégral, en situation individuelle, ou l'entretien tel que décrit par Besse et coll. ne sont pas usuellement employés à l'école en situation individuelle. Comme souligné par E. Fijalkow (2003, op. citée ou encore plus récemment Goigoux et Cèbe (2014, p.9), les enseignants ont massivement recours à des questionnaires (un résultat mis en évidence par E. Fijalkow 2003). Le rapport du PIREF (2003) précisait que « les questionnaires et la lecture à voix haute sont aujourd'hui des moyens imparfaits d'évaluation, qui n'ont pas d'efficacité avérée dans la maîtrise de la lecture » sans toutefois proposer d'alternatives suffisamment opératoires, et Giasson souligne (1990, p. 224) que le problème n'est pas tant de savoir si les enseignants doivent ou non utiliser les questions « mais plutôt de savoir comment, quand et où ils doivent le faire ? ». La question de l'évaluation de la compréhension semble surtout poser celle de l'enseignement de la compréhension dont une certaine unanimité des chercheurs (PIREF 2003) déplore la faible qualité. Particulièrement, avertissent Cèbe, Goigoux et Thomazet (2004, pp. 9-10), il est important de « distinguer les compétences qui permettent de comprendre les textes de celles qui sont sollicitées par le traitement des questionnaires ».

Nous renvoyons le lecteur aux analyses développées dans le chapitre 4 de la Partie III. Elles convergent vers l'importance de s'intéresser au rapport que chaque question entretient avec la réponse (cf. modèle RQR de Pearson & Johnson, 1978, 1984), aux processus impliqués et au lien qui peut être établi avec ce qui lie question et réponse (cf. typologie d'Irwin, 1991), mais aussi au fonctionnement et à l'économie interne du questionnaire dont nous avons montré qu'il pouvait permettre de répondre sans même avoir lu le texte dans deux épreuves standardisées données en fin de CM2 (cf. C. Blanc et Goasdoué, 2016).

Des connaissances sur le fonctionnement des questionnaires sont pour l'élève des connaissances de nature à contribuer efficacement à leur activité de réponse.

- La seconde dimension renvoie au texte et à ses propriétés linguistiques, indépendamment du lecteur. Selon J. Fijalkow « savoir lire c'est savoir tout lire » (2000, p.107-108) i.e. le savoir lire doit pouvoir s'exercer quel que soit le texte à lire. À des fins de caractérisation de la nature des textes présentés à des élèves de CP, Goigoux et Cèbe (2006, pp.50-51) distinguent grossièrement trois types de textes selon l'objectif poursuivi :

- Les textes destinés à initier les enfants au monde de l'écrit (souvent issus de la littérature jeunesse),
- Les textes facilitant un travail systématique sur la compréhension (souvent écrits pour des manuels),
- Les textes constitués de mots qu'il faut apprendre à déchiffrer (la notion même de texte pouvant être questionnée quand l'accumulation de phrases donne un récit incohérent).

Cette catégorisation permet de mettre en évidence que, suivant le type de texte choisi, l'activité de lecture de l'élève n'est pas de même nature. Typiquement en début CP, les élèves rencontrent des textes de type 2 voire 3 dans les tâches de lecture quotidienne. Cette typologie est toutefois assez générale et centrée sur une problématique « littéraire » : elle ne permet pas de ranger les questionnaires sur les textes ou les énoncés de problèmes, deux textes que les élèves lisent fréquemment.

Cette typologie peut être affinée selon que le texte est un extrait d'un texte plus long ou un texte intégral. La lecture opère sur des écrits que l'on qualifie d'énoncés ou de textes. Un énoncé est défini comme « une suite finie de mots émises par un locuteur » (Dubois, et coll. 1994/1999). Or les écrits auxquels sont confrontés les lecteurs débutants ne sont pas nécessairement des énoncés : particulièrement lorsqu'ils ont à traiter des mots (présentés avec ou sans déterminants) ou des textes de type 3 dans la classification générale précédemment mentionnée. Si l'on suit la définition proposée par Hjelmslev dans le dictionnaire de linguistique des sciences du langage, le mot texte désigne « un énoncé quel qu'il soit, parlé ou écrit, long ou bref ». « Stop » constitue de ce point de vue un texte. A la suite de ce linguiste nous ne distinguerons plus ensuite « énoncé » de « texte » qu'il s'agisse d'oral ou d'écrit. En effet, nous retiendrons que c'est une unité signifiante qui caractérise le texte : il résulte de l'expression finie d'un locuteur. A cet égard, l'extrait de texte ne constitue donc pas un texte : la délimitation n'est pas du fait du locuteur, mais de celui qui propose le texte au lecteur. La nature même des textes que les lecteurs débutants ont à lire est immédiatement questionnable de ce point de vue et permet sans doute d'ajouter un quatrième type à la typologie de Goigoux et Cèbe : celui de pseudo-texte (qui inclut texte tronqué ou extrait de texte, suite de phrases sans lien particulier...).

Par « texte » en CP, il faut bien souvent comprendre « texte narratif » à l'exclusion de tout autre type de texte. Ce type de texte appartient au genre littéraire narratif. Les textes d'apprentissage relèvent dans une très large majorité de ce type. Les textes d'évaluation sont, pour une part importante, de même nature : l'évaluation conduite par l'enseignant se fonde sur le repérage d'indices de compréhension demandés par des questionnaires, textes d'un type différent. L'évaluation de la lecture de textes narratifs se fonde, pour partie, sur une évaluation implicite de la lecture de texte injonctif.

Lors de la lecture d'un texte narratif, le lecteur mobilise des schémas (relevant de connaissances sur le monde (« tous les ogres mangent les enfants »), relevant des connaissances sur le fonctionnement des textes (« les récits commencent toujours par il était une fois »). Les textes narratifs manifestent une organisation générale commune (dite schéma quinaire (Larivaille, 1974) : situation initiale, élément déclencheur, péripéties, élément de résolution et situation finale), et la connaissance des schémas narratifs intervient dans la compréhension de texte. Ce sont ces schémas, construits puis stockés en mémoire à long terme qui peuvent être activés et servir en quelque sorte de guide pour la tâche de lecture comme d'écriture (Besse, Luis, Paire et coll., 2004, p.21). Les travaux de Propp (1970) ont permis de déterminer une certaine morphologie des contes qui permet de les organiser en « tableaux¹⁷⁹ » (1970, pp. 146-154). La connaissance du fonctionnement du récit constitue une connaissance importante intervenant dans la compréhension.

¹⁷⁹ Situation initiale, partie préparatoire, nœud de l'intrigue, donateurs, entrée en scène de l'auxiliaire et fin de la première séquence, deuxième séquence (qui contient l'accomplissement de la tâche).

Tauveron (2002, pp. 47-80 et 2014) souligne par ailleurs un ensemble de connaissances textuelles à développer chez les élèves telles que des connaissances sur le fonctionnement éditorial, sur l'acte d'écrire et le processus de fonctionnalisation, sur la « préhistoire » de l'histoire en cours, sur les stéréotypes culturels, sur les mythes et les symboles, sur les techniques narratives, sur le déroulement et les caractéristiques textuelles du genre auquel appartient le texte, sur l'auteur, sur le référent textuel, sur l'intertexte, sur les stratégies textuelles. Elle précise (2014, p. 11) que les élèves doivent développer une culture des « coups narratifs » :

« Apprendre à lire, c'est donc aussi apprendre (très tôt) que, sans marques de surface évidentes, le texte peut recourir à des stratégies de « perte » (le non-dit, le dit à demi, le dit et son contraire, le dit pour son contraire, le sarcasme, la mauvaise foi, la construction concertée de fausses pistes), ou, sous couvert du dit, peut vouloir dire métaphoriquement ou allégoriquement autre chose. Il me semble important de donner aux enfants une culture de ces « coups » narratifs, comme il est important d'avoir une culture des « coups » si l'on veut devenir un bon joueur d'échec, ce qui suppose que l'on choisisse dès l'entrée dans l'écrit des textes qui exhibent quelques-uns de ces « coups » et singulièrement posent la question de la vérité ou de fiabilité de ce qui est dit et écrit (naïveté vs vantardise ; sincérité vs mensonge, affabulation, fabulation, manipulation...). »

Le texte narratif implique également des choix d'écriture qui engendrent un résultat particulier quant à la représentation verbale de l'histoire. On peut ainsi distinguer l'histoire du récit (cf. Genette, 1972) : un récit peut débiter par la fin de l'histoire. La narration joue avec l'ordre des faits, mais également avec la durée : certains faits ou arguments sont plus ou moins développés. Le texte narratif mobilise en outre une capacité de décentration du lecteur : auteur, narrateur et personnage sont autant de points de vue à prendre en compte pour construire le positionnement du lecteur. Différents voix narratives sont susceptibles d'intervenir dans le récit. Comprendre un texte engage la compréhension de ces phénomènes narratifs. .

3.1.3. Synthèse

Pour conclure, pour construire notre référentiel pour analyser les épreuves d'évaluation nous sommes référés à un modèle général dans lequel les variables « Lecteur », « Texte » et « Contexte » sont en interaction. La prise en compte de cadres provenant de la psychologie cognitive et de la didactique de la lecture (modélisation de l'activité du lecteur, modèle d'apprentissage de la lecture) mais aussi de cadres linguistiques (typologie de textes, narratologie) nous conduisent à un déterminer un référentiel dans lequel la compétence, la tâche et l'activité interagissent (cf. Leplat et de Montmolin, 2001).

Notre référentiel vise l'analyse d'épreuves d'évaluation de la lecture collectées auprès d'enseignant du point de vue de l'activité de réponse engagée, des compétences (ou habiletés) mobilisées en tenant compte de ce que l'institution scolaire prescrit d'enseigner. Il délaisse ainsi les composants « culture écrite » et « écriture » de la variable « Lecteur » ainsi que les composants relatifs à la variable « Contexte » lesquels ne peuvent être analysés au moyen des données collectées dans cette étude.

Goigoux (2003, pp. 51-52) isole trois sources majeures de difficultés de compréhension des textes écrits qui sembleraient, de ce fait, devoir faire l'objet d'épreuves d'évaluation :

- « les déficits des traitements de «bas niveau» (notamment la faiblesse de l'automatisation de l'identification des mots) (tâches PG et L)
- les déficits généraux des capacités de compréhension (déficits non spécifiques à la lecture et qui affectent également la compréhension du langage oral) et
- les déficits spécifiques au traitement du texte écrit, liés principalement à une mauvaise régulation de l'activité de lecture par l'élève » (tâches C).

S'intéresser à rechercher la couverture des trois composants (PG, L et C) réalisée par les tests d'évaluation est ainsi cohérent.

L'ensemble des aspects phonographiques (nom des lettres, correspondances graphophonologique, conscience phonémique et syllabique, combinatoire) mobilisés dans l'identification des mots et leur pendant graphophonologique ; l'automatisation des procédures d'identification et de reconnaissance des mots ; la compréhension autonome de l'écrit constituent les principaux composants de notre référentiel.

La classe de CP constitue un point d'articulation essentielle entre eux.

Giasson (2012, p. 168) propose d'affiner ces aspects en CP et de considérer dans l'évaluation :

- « La lecture orale (analyse des méprises) (i.e. type de tâche L en référence à la typologie de Goigoux et coll. (2013), ndlr),
- La compréhension à l'écrit (rappel de texte), (i.e. type de tâche C5, idem, ndlr)
- Le lexique orthographique (mots reconnus immédiatement) (i.e. type de tâche L2, ndlr)
- Les habiletés de décodage (lecture de mots) (i.e. type de tâche L3, ndlr)
- La connaissance des lettres (type de tâche PG3, ndlr). »

Cet auteur précise en outre (idem) que l'évaluation doit évoluer au cours de l'année. Ainsi l'évaluation de la connaissance des lettres devient moins utile en fin d'année pour la plupart des élèves dans un mouvement contraire à l'évaluation de la lecture de texte qui prend de plus en plus de place au fur et à mesure du déroulement de l'année. De surcroît, elle souligne que « toutes les évaluations ne sont pas nécessaires pour tous les enfants et qu'il n'est pas nécessaire d'évaluer en détail les habiletés en identification des mots d'un enfant qui lit couramment » (idem).

L'évaluation des principaux composants de la lecture pourrait donc être observée de façon différente suivant le moment de l'année : l'évaluation devrait porter prioritairement sur les « traitements de bas niveau » (i.e. des tâches portant sur les relations phonographiques et tâches d'identification de mots) en début d'année puis inclure des tâches de compréhension écrite (les unes n'excluant pas les autres). La validité de contenu (didactique) des tests pourrait être analysée selon cette logique.

Ces propositions semblent compatibles avec les pratiques (déclarées) des enseignants collectés par E. Fijalkow (2003). Cet auteur a coordonné la seule enquête d'ampleur réalisée en France sur l'enseignement de la lecture-écriture au CP dont nous disposons en 2013 avant celle en cours de l'Ifé coordonnée par Goigoux. Elle porte sur l'enseignement/apprentissage de la lecture en CP. La méthodologie adoptée est celle d'un questionnaire envoyé à 2501 enseignants et retourné par 1253 d'entre eux. Le questionnaire se présente sous la forme d'un livret de 35 pages comportant 7 questions auxquelles

correspondent 399 variables. Cette enquête ne porte pas sur les pratiques effectives mais sur les pratiques déclarées d'enseignement/apprentissage de la lecture. Ce premier point est essentiel : les enseignants ont dû répondre à chaque énoncé en indiquant « Jamais, Quelques fois, Souvent, Très souvent, En début d'année ou en Fin d'année », de sorte qu'une grille de tâches d'enseignement préexistait et que l'on ne peut donc qualifier l'étude d'étude portant sur les pratiques effectives. Les résultats de l'enquête conduisent E. Fijalkow à indiquer que :

- En début d'année, la lecture à haute voix est le moyen le plus plébiscité (83 % des enseignants affirment y avoir recours) suivi par des tâches de remise en ordre de mots pour faire des phrases (81%), ou la reconnaissance de mots isolés (74%). Cela conduit E. Fijalkow à qualifier cette évaluation de traditionnelle¹⁸⁰, reposant surtout sur la perception visuelle et la mémoire.
- En fin d'année : la pratique d'évaluation choisie de façon quasi unanime consiste à répondre à des questions après lecture d'un texte (95%). Lire à haute voix et suivre une consigne (environ 80%) constituent les deux autres modalités d'évaluation les plus plébiscitées. E. Fijalkow ((id., pp. 92-93) estime que les modalités d'évaluation indiquent que les enseignants n'adoptent pas un comportement identique tout au long de l'année, que les tâches qui font appel à la signification sont surtout présentes en fin d'année.

L'aspect séquentiel¹⁸¹ de la vision du développement des habiletés de compréhension apparaît à la lecture des résultats de l'enquête d'E. Fijalkow. A des habiletés langagières dites « de base », suit le décodage auquel succèdent les habiletés de compréhension puis la compréhension en situation de lecture. L'évaluation suit la même linéarité, nos analyses se dérouleront selon la même chronologie.

¹⁸⁰ Fijalkow précise (2003, p.17) que le sens donné à ce mot se limite à son étymologie « issu de la tradition », et n'est ainsi pas l'expression d'un jugement de valeur.

¹⁸¹ Un autre modèle existe développé par van den Broeck et coll. (2005) qui propose une vision simultanée de ces mêmes habiletés : habiletés langagières de base et habiletés de compréhension se développent simultanément. A la première succède le décodage. Jointe à celle-ci, la deuxième habileté précitée permet de forger la compréhension en situation de lecture. (N. Blanc, 2010, p.13).

3.2. Analyse de la dimension phonographique des tests et des épreuves du 1er trimestre

3.2.1. Les tests

80 épreuves¹⁸² au sein des 16 tests du premier trimestre sont utilisées pour évaluer les connaissances phonographiques des élèves.

Le tableau ci-dessous indique, à l'intérieur de chaque test, le nombre d'épreuves destinées à évaluer chaque sous-composant du composant « Relations phonographiques » (cf. typologie de tâches d'enseignement établie par Goigoux et coll. (2013), présentée plus haut). Cela permet d'apprécier la couverture du domaine de la lecture occasionnée par les tests du premier trimestre.

T1	Sous composants du composant PG (Relations phonographiques)				
	PG1. Phon.	PG2. Syll. O	PG3. Lett.	PG4. CPG	PG5. Combin.
A	1	1	1		3
B			1		2
C		2	1		1
D			2	1	1
E	1	2	4	3	
F		2	1		
G	1				2
H			2		5
I			1	1	5
J			4	1	4
K			3	1	2
L		1	3		2
M		2	1		2
N					1
O					1
P		1	2		2
Tot.	3	11	26	7	33

Tableau 30 : Répartition des épreuves d'évaluation recueillies au 1^{er} trimestre par enseignant et par sous-composant du composant « Relations phonographiques »

Près d'une case du tableau sur deux ne contient pas de valeur numérique. D'un point de vue général, cela traduit l'aspect lacunaire de l'évaluation instituée du composant « Relations phonographiques ». L'évaluation réalisée ne permet pas d'apporter d'informations précises sur la nature des obstacles rencontrés par les élèves, en particulier les élèves les moins avancés en lecture en fin de premier trimestre.

De telles absences de valeurs sont également constatées dans chacune des lignes, i.e. pour chaque test : au premier trimestre, aucun test n'évalue l'ensemble des différents sous-composants des relations phonographiques, bien que la place centrale et première de ces relations dans l'apprentissage de la lecture soit établie.

¹⁸² Elles sont classées par type de tâches et par enseignant dans l'Annexe C, pp.2-28

Caractériser la validité de contenu des tests en termes de couverture des principaux composants effectuée par les tests est ainsi, selon le référentiel utilisé, sans objet : aucun test ne propose d'épreuves portant sur tous les aspects des relations phonographiques.

Une analyse plus fine du tableau permet toutefois de considérer la présence de certains « consensus » chez les évaluateurs. Si certains sous-composants sont très peu évalués (PG1, PG2 et PG4), d'autres à l'inverse, sont évalués par la plupart des enseignants (PG3 et PG5). Cela nous conduit à formuler une remarque importante : même si elles sont peu nombreuses et peu présentes dans chacun des corpus d'épreuves, la présence d'épreuves portant sur la conscience phonémique (PG1 et PG2) peut être interprétée comme la trace d'une réelle assimilation des travaux de recherches portant sur ces sujets dans les pratiques enseignantes. Nous avons dit qu'aucun item des programmes ne porte sur cet aspect dans la partie II. Etant donné le guidage de l'évaluation par les items des progressions souligné dans le début de notre travail, nous aurions pu penser que ces sous-composants ne seraient pas évalués. Ce n'est pas le cas et cela constitue un résultat important.

Les sous-composants PG1 et PG2 de la typologie de Goigoux et coll. (2013) ne sont décrits par aucun item spécifique des programmes (cf. tableau n°29 Partie IV, 3.1) au contraire des sous-composants PG3 ; PG4 ; PG5 qui correspondant aux items n°8 ; n°9 ; n°10 et n°11 des programmes. Seuls les tests D ; I ; J et K (i.e. un quart des tests examinés seulement) contiennent pourtant des épreuves permettant d'évaluer chacun de ces trois derniers sous-composants. Le guidage de l'évaluation par les items des progressions ne conduit pas à une évaluation qui en assure la couverture. Les tests D, I, J et K couvrent le mieux l'ensemble de ces trois sous-composants. C'est la raison pour laquelle nous nous intéressons maintenant aux épreuves qu'ils contiennent afin de poursuivre notre analyse de validité des tests et des épreuves.

3.2.2. Les épreuves des tests proposés par les enseignants D, I, J et K

3.2.2a. Analyse des épreuves permettant d'évaluer les connaissances des équivalences entre les différents systèmes d'écriture des lettres (tâches PG3)

10 épreuves¹⁸³ permettent aux enseignants D ; I ; J et K d'évaluer ces connaissances.

Il est important de souligner que la connaissance du nom des lettres est toujours associée à une autre connaissance dans les items des progressions :

- « Connaître le nom des lettres et l'ordre alphabétique »
- « Connaître les correspondances entre minuscules et majuscules d'imprimerie, minuscules et majuscules cursives »

Cette conjonction de connaissances au sein d'un même item (connaissances relatives au nom des lettres, à l'ordre alphabétique, à l'équivalence de leurs modes de désignation graphique) conduit les enseignants à évaluer différentes connaissances en même temps au moyen de mêmes épreuves. Cela fait naître le risque d'une évaluation conforme aux programmes mais imprécise sur le plan didactique.

Observons les épreuves en indiquant en préambule que la tâche consistant à épeler les lettres d'un mot (en tout premier lieu les lettres du prénom) nous semble constituer un exemple de tâche appropriée pour évaluer la connaissance du nom des lettres.

Les épreuves utilisées par l'enseignant D

Cet enseignant utilise deux épreuves pour évaluer les deux items du programme précédemment mentionnés.

La première épreuve¹⁸⁴ évalue en même temps la connaissance du nom des lettres et la connaissance de l'ordre dans lequel les lettres sont rangées dans l'alphabet. On peut donc imaginer qu'un élève capable de dire le nom des lettres mais ne maîtrisant pas l'ordre alphabétique échouerait. Par ailleurs, la connaissance de l'ordre alphabétique est hautement liée à la connaissance du nom des lettres (comment la manifester autrement qu'en disant la suite des lettres ?). Mais, restituer oralement la suite des mots-lettres ne signifie pas pour autant que l'élève dispose de connaissances relevant d'une prise de conscience explicite des unités traitées que Gombert et Colé (2000) nomme « capacité métaphonologique ». Indépendamment de tout traitement de l'écriture des lettres, réciter l'alphabet relève d'un processus de traitement moins conscient que ces mêmes auteurs qualifient d'épiphonologique (les unités ne sont pas directement disponibles et manipulables) (cf. Ecalle, Magnan et Bouchafa, 2002). Un élève pourrait réussir cette épreuve sans que l'on soit assuré qu'il connaît le nom de toutes les lettres si celles-ci ne sont justement pas présentées dans l'ordre (alphabétique). Cet élément est de nature à invalider l'épreuve du point de vue psycho-didactique.

¹⁸³ Les épreuves se situent toutes dans l'Annexe C : p. 9 pour celles du test D ; p.11 pour celles du test I ; p. 12 pour celles du test J et p. 13 pour celles du test K.

¹⁸⁴ Epreuve 42-1, p. 9 Annexe C

La seconde épreuve¹⁸⁵ porte sur les équivalences des modes de représentation des lettres (graphies scriptes et cursives, minuscule et majuscules). Pour répondre, l'élève doit « colorier d'une même couleur les lettres identiques ». Cette consigne contient plusieurs sources de difficultés. « Colorier d'une même couleur » ne doit pas être interprété par l'élève comme « tout colorier de la même couleur ». L'élève doit implicitement déterminer le nombre de couleurs à utiliser (donc de crayons à choisir) ce qui dépend du nombre de triplets de lettres équivalentes (selon la relation d'équivalence « à la même valeur phonologique que »). L'élève doit donc conjointement traiter une deuxième source de difficulté provenant du mot « identique ». Ce mot laisse à l'élève la responsabilité d'interpréter correctement le fait que le caractère identique des lettres dont il est question s'exprime relativement à la valeur phonologique des lettres et non à leur forme graphique. Enfin, il est implicite que l'élève doit colorier « toutes » les lettres « identiques ». La tâche d'association contenue dans l'épreuve ne pose pas de problèmes de validité en tant que telle : pour répondre, l'élève doit bien s'appuyer sur les connaissances visées. En revanche, la façon dont la tâche est prescrite au moyen de la consigne conduit à invalider l'épreuve : la reconnaissance du critère d'identité des lettres est implicite et laissé à la charge de l'élève.

L'épreuve utilisée par l'enseignant I

Cet enseignant utilise une seule épreuve¹⁸⁶ proche de la seconde épreuve utilisée par l'enseignant D que nous venons de décrire. La consigne étant strictement identique, nous ne reviendrons pas sur les raisons qui nous conduisent à remettre la validité de l'épreuve en cause.

Les épreuves utilisées par l'enseignant J

L'enseignant J utilise 4 épreuves.

L'épreuve ci-après porte sur la connaissance du nom des lettres.

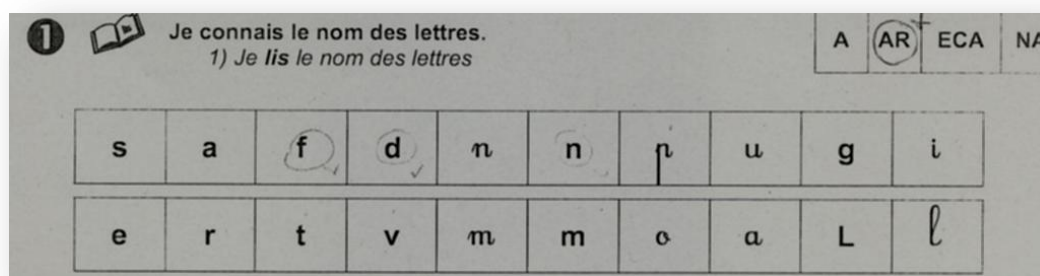


Figure 71 : Épreuve FT1 (Annexe C, p. 12) avec réponses d'élève

La consigne est donnée oralement (« lire les lettres »). Pour répondre, l'élève doit lire la série de lettres présentées dans des graphies variées (minuscules cursives, minuscules et majuscules scriptes). L'activité de réponse de l'élève se fonde sur une mobilisation de

¹⁸⁵ Epreuve 42-4, page 9 Annexe C

¹⁸⁶ Epreuve 76-2, page 11 Annexe C

connaissances métaphonologiques et de connaissances relatives à l'équivalence des modes de présentation des lettres. Cette épreuve évite les sources de difficultés précédemment isolées : elle peut être considérée comme valide et suffirait pour évaluer les connaissances visées.

Toutefois, cet enseignant utilise trois autres épreuves reproduites ci-après et dont la validité peut être discutée pour des raisons différentes.

-Une première raison est directement dépendante du choix des distracteurs dans le cas d'une tâche de détection d'intrus.

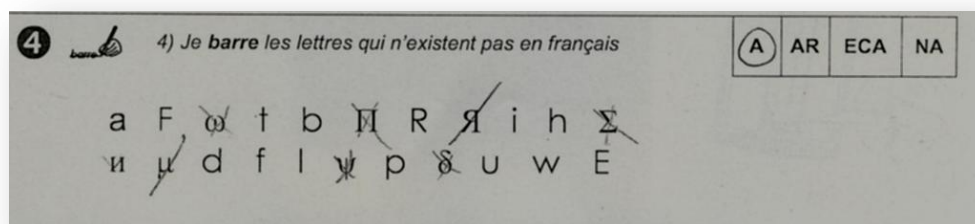


Figure 72 : Épreuve MT1 (Annexe C p. 12) avec réponses d'élève

Pour répondre, l'élève doit barrer (toutes, ce qui est implicite) les lettres qui n'appartiennent pas au système graphique du français. L'élève dont la copie est reproduite n'a pas barré la lettre « И » : cela pourrait s'expliquer par un fait déjà commenté (cf. tâche 8, Partie III 3-2) concernant la symétrie des lettres. Cette lettre ressemble à un H majuscule ou un N majuscule. L'élève a bien barré la lettre « μ » qui présente aussi de grandes similitudes graphiques avec le « u ». L'échec à cette tâche intéressante ne suffit pas pour conclure à un défaut de la capacité des élèves à connaître les lettres (et leur valeur phonologique).

-Une deuxième raison est liée au cadre de réponse imposé par la tâche d'association de propositions de réponses. Il peut induire un effet de contrat : dans le cas de l'épreuve reproduite ci-dessous celui-ci serait « relier tous les éléments de réponse »

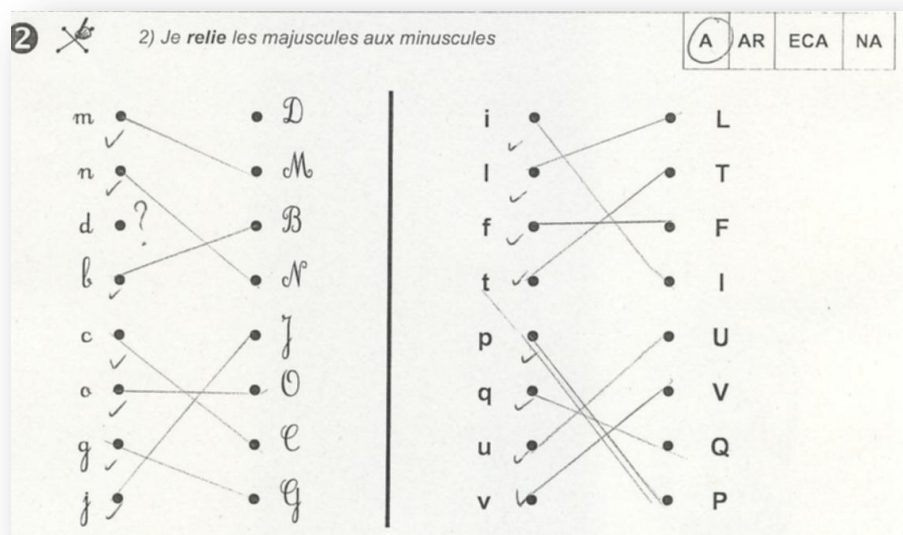


Figure 73 : Épreuve MT1 (Annexe C p. 12) avec réponses d'élève

L'élève dont la copie est reproduite ci-dessus n'a pas associé le « d » script minuscule au « d » cursif majuscule. Associer ces deux éléments de réponse pourrait pourtant être motivé par le simple fait que ce soit la seule possibilité restante, indépendamment de leur valeur. Cette épreuve permet également de considérer le cas où l'élève aurait associé deux éléments de réponse de façon erronée. L'erreur commise pourrait traduire un défaut de connaissance. Il est alors envisageable qu'une seconde erreur soit commise (par effet de contrat, l'élève pourrait associer les deux derniers éléments restants). Cette seconde erreur n'aurait pas le même statut que la première et serait alors directement liée au cadre de réponse imposé par la tâche.

-Une troisième raison peut être considérée, elle est également relative au cadre de réponse. Dans le cas de l'épreuve reproduite ci-dessous, il repose sur un système de codage de la réponse inutilement compliqué.

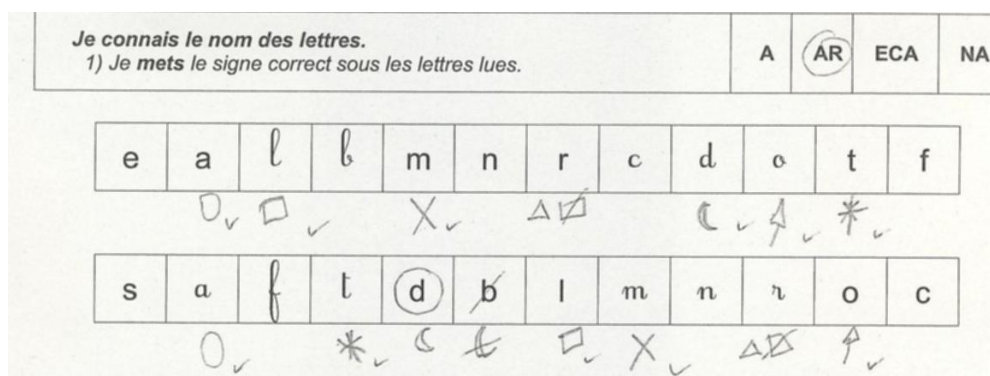


Figure 74 : Épreuve MT1 (Annexe C p. 12) avec réponses d'élève

L'élève doit matérialiser au moyen d'un même symbole les deux graphies différentes d'un certain nombre de lettres dont le nom est dit par l'enseignant. Par exemple « placer une

étoile sous les lettres qui se lisent [t] ». L'élève dont la copie est reproduite ci-dessus reconnaît les graphies minuscules scriptes et cursives de A, L, M, O et T. Une petite encoche en forme de « v » marquée au pied des symboles par l'enseignant indique que l'association réalisée est correcte. Deux paires ne sont pas validées :

- la paire de D : l'élève hésite entre les graphies minuscules scriptes « b » et « d » (cf. commentaire dans Partie III-3.2).
- la paire de R : on ne peut pas savoir si l'élève a utilisé le bon double symbole (triangle - carré barré) ou s'il n'a pas su associer les graphies de la lettre R.

Ce système de codage des réponses occasionne un « parasitage » visuel de nature à augmenter la difficulté de la tâche initiale.

Les épreuves utilisées par l'enseignant K

L'enseignant K utilise 3 épreuves.

La première épreuve¹⁸⁷ est semblable à la première épreuve utilisée par l'enseignant J. Nous ne revenons pas sur les raisons qui nous ont conduit à la considérer comme étant valide.

Nous écarterons la deuxième épreuve¹⁸⁸ pour des raisons précédemment développées au sujet de l'épreuve d'association de paires utilisée par l'enseignant J.

La troisième épreuve¹⁸⁹ se distingue des autres épreuves examinées. Cette épreuve reproduite ci-dessous engage une activité d'écriture sous la dictée. L'élève doit écrire les 10 lettres dictées par l'enseignant.

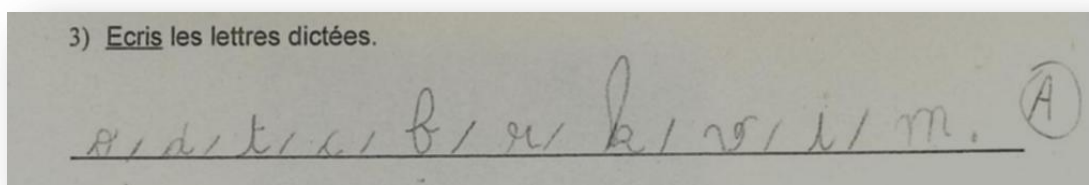


Figure 75 : Épreuve FT1 (Annexe C, p. 13) avec réponse d'élève

La ligne support de l'écriture des lettres occasionne une gêne pour cet élève qui ne parvient pas à écrire correctement la lettre « f ». Toutefois, l'appréciation portée par l'enseignant conduit à estimer que le tracé des lettres n'est pas strictement évalué en lui-même. L'activité de réponse nécessite bien la mobilisation d'une traduction graphique d'une information phonologique, ce qui est la connaissance visée. L'épreuve est à ce titre valide.

¹⁸⁷ Epreuve 99-1 p.13 Annexe C

¹⁸⁸ Epreuve 99-2, p. 13 Annexe C

¹⁸⁹ Epreuve 100-1, p. 13 Annexe C

3.2.2b. Analyse des épreuves permettant d'évaluer les correspondances entre phonèmes et graphèmes (tâches PG4)

Les enseignants D ; I ; J et K évaluent, au moyen d'une épreuve chacun¹⁹⁰, ces connaissances renvoyant à l'item des programmes :

« Distinguer entre la lettre et le son qu'elle transcrit ; connaître les correspondances entre les lettres et les sons dans les graphies simples (ex. *f* ; *o*) et complexes (ex. *ph* ; *au*, *eau*). »

L'enseignant D utilise une épreuve dans laquelle l'élève doit indiquer au moyen d'un numéro « la lettre qui produit le son produit par le maître » i.e. associer un graphème à un phonème. Les graphèmes proposés sont tous des graphèmes simples¹⁹¹ à une seule lettre, présentés dans des contextes non ambigus (pour le s et le a), et les phonèmes donnés par l'enseignant sont des phonèmes consonantiques ([p], [r], [s]), ou vocaliques ([a]). Cette épreuve complète l'épreuve précédemment utilisée par cet enseignant dans laquelle l'élève devait effectuer une conversion inverse : dire le nom (plus que le phonème correspondant) de graphèmes. Cette épreuve, bien que très simple, permet bien d'engager une activité de correspondance phonographique pour l'élève.

L'enseignant I utilise à peu près la même épreuve que les enseignants J et K. Dans cette épreuve reproduite ci-dessous, l'élève doit associer la lettre qui, parmi quatre consonnes proposées, transcrit l'attaque du mot.

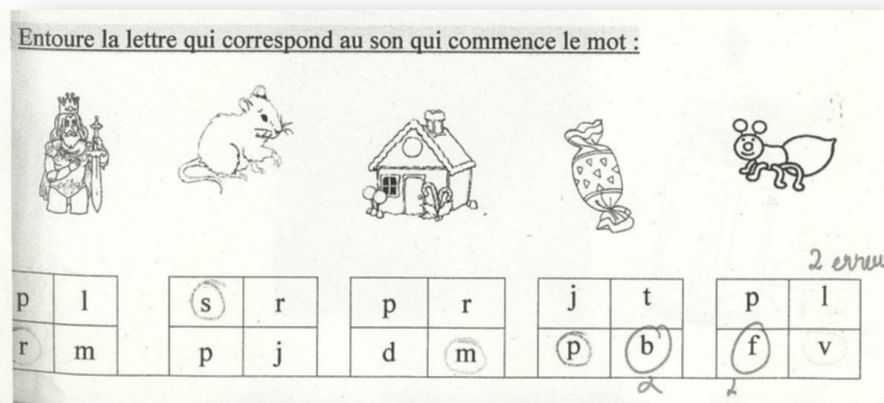


Figure 76 : Épreuve FT1 (cf. Annexe C, p. 17) avec réponses d'élève.

Ce type d'épreuve est plus riche que celle précédemment étudiée en raison de la proximité de certains phonèmes susceptibles de poser problème aux élèves. Ils sont source d'une difficulté de distinction phonologique pour les élèves. Les phonèmes [b] et [p] par exemple diffèrent au niveau de la tension : le premier est moins tendu que le second.

Pour répondre, l'élève doit tout d'abord établir l'image acoustique du mot représenté par un dessin ([bōbō]). Nous avons discuté dans la partie III, 3.2 de l'obstacle que le dessin

¹⁹⁰ Les épreuves se situent toutes dans l'Annexe C : p. 16 pour celles du test D ; p.17 pour celles des tests I ; J et K.

¹⁹¹ Voir Perez (2014) pour sa proposition de hiérarchisation des 45 graphèmes de base de l'orthographe du français.

représente : le deuxième dessin proposé est-il celui d'une souris (bonne réponse produite en entourant la lettre s) ou celui d'un rat (auquel cas la bonne réponse est produite en entourant la lettre r) ? Il doit ensuite se donner l'image acoustique de chacune des lettres proposées [ʒ], [t], [p], [b] à comparer à l'attaque du mot représenté par le dessin qui doit avoir été conservé en mémoire : b correspond à la lettre qui transcrit le phonème [b] du mot [bōbō]. Ce type d'épreuve est très intéressant : il permet à l'enseignant de collecter une information de nature phonologique lui permettant d'envisager un travail de correction verbo-tonale fondé sur le rapprochement de phonèmes possédant deux traits phonologiques communs sur trois (parmi l'acuité, la labialité et la tension) placés dans des contextes facilitants (cf. Kaneman-Pougatch & Pedoya-Guimbretière, 1991). Toutefois, la validité de ces épreuves est remise en cause du fait de l'usage du dessin : même si l'enseignant a pris le soin de nommer les dessins (auquel cas on peut concevoir que ces dessins sont inutiles pour la tâche demandée), il est possible que des élèves élaborent un autre sens au cours de la redéfinition ou de la réalisation de la tâche (soit parce qu'ils ont oublié ce qu'a dit l'enseignant, soit parce qu'un détail attire leur attention...). Sans indication supplémentaire, nous ne pouvons par exemple pas répondre à l'épreuve pour le quatrième dessin (est-ce un ver, une luciole, une fourmi ? ce qui conduit à trois bonnes réponses possibles).

3.2.2c. Analyse des épreuves permettant d'évaluer la combinatoire (tâches PG5)

12 épreuves¹⁹² permettent aux enseignants D ; I ; J et K d'évaluer ces connaissances renvoyant à l'item des programmes :

« Savoir qu'une syllabe est composée d'une ou plusieurs graphies, qu'un mot est composé d'une ou plusieurs syllabes ; être capable de repérer ces éléments (graphies, syllabes) dans un mot. »

Les tâches attendues *a priori* seraient des tâches permettant d'évaluer la capacité des élèves à assembler des graphèmes pour former des syllabes, à assembler des syllabes pour former un mot. L'ensemble des assemblages de syllabes et/ou de lettres demandés aux élèves devrait faire apparaître leurs connaissances de l'arbitraire des signes linguistiques : le signifié du graphème renvoie en effet à la prononciation (d'un phonème dont le signifié renvoie à des éléments de l'expérience, des objets, des idées, des faits...) et la valeur phonologique d'un assemblage de lettres dépend du contexte linguistique (la lettre « a » ne possède pas de prononciation à laquelle on pourrait se reporter de façon mécanique).

L'épreuve utilisée par l'enseignant D

L'enseignant D utilise une seule épreuve, reproduite ci-dessous, pour évaluer ces connaissances. Pour répondre, l'élève doit entourer la syllabe dictée par l'enseignant située à l'intérieur d'un mot écrit dans une série de 5 mots. Par exemple entourer le « mar » de « mardi ».

¹⁹² Les épreuves se situent toutes dans l'Annexe C : p. 20 pour celles du test D ; p. 22 à 25 pour celles des tests I ; J et K.

3. Repérer les syllabes dans un mot.
J'entoure la syllabe dictée par le maître.

Compétence 11 : 10/10

1. mardi 2. balançait 3. dimanche 4. dévorait 5. maison

Figure 77 : Épreuve FT1 (Annexe C, p. 20) avec réponses d'élève

Les syllabes ciblées se situent à une place variable dans le mot ce qui constitue une variable intéressante. L'activité de l'élève s'appuie sur une mobilisation de connaissances des correspondances graphophonologiques et d'assemblage (il lui faut déchiffrer le mot), et de connaissances de l'ordre de la conscience syllabique (il lui faut découper le mot oral en syllabes). Cette tâche est donc une tâche qui s'appuie sur une activité de décodage d'un mot entier et/ou sur la reconnaissance directe d'une syllabe graphique (« ma » de « mardi »).

Dans le cas où l'activité de l'élève se fonde sur ses seules capacités à reconnaître un certain groupe de lettres (« m » et « a » pour former la syllabe « ma ») et non pas sur sa capacité à décoder le mot, la validité de l'épreuve est discutable : la prononciation du groupe de lettres « ma » dépend en effet de l'environnement linguistique où elles se trouvent (« ma » se prononce différemment dans les mots « mardi », « dimanche » et « maison » et l'ajout d'une lettre ne modifie pas nécessairement le son produit (mais)). Dans l'épreuve proposée, la syllabe que l'élève doit entourer n'est jamais la même ce qui constituerait pourtant une variable intéressante (cf. discussion autour de la tâche 10, Partie III, 3.2).

L'épreuve proposée engage une activité de l'élève supérieure à celle qui est annoncée : le repérage des syllabes n'est possible que parce que l'élève peut décoder le mot entier puis le décomposer en syllabes. Dans le cas où l'activité de l'élève se fonde sur sa reconnaissance d'un groupe de lettres, l'épreuve ne permet pas de savoir si l'élève sait distinguer la valeur des lettres en fonction de celles qui l'entoure. La validité de cette épreuve est ainsi remise en cause.

L'épreuve utilisée par l'enseignant I

Contrairement à l'enseignant D, l'enseignant I utilise un nombre d'épreuves assez élevé (5 épreuves¹⁹³) ce qui traduit l'importance qu'il accorde à ce point essentiel de l'apprentissage de la lecture.

Dans la première de ces épreuves, l'élève doit « entourer la syllabe qui correspond au son qui commence le mot ». Nous ne revenons pas sur le motif d'invalidation de l'épreuve qui correspond au fait de présenter un mot au moyen d'une image du mot. Contrairement à l'épreuve discutée ci-avant, la syllabe orale est donnée par l'enseignant et il revient à l'élève de l'associer à sa traduction graphique. La demande de l'enseignant est toutefois ambiguë : elle porte sur la détection d'une syllabe mais « qui correspond au son qui commence le mot ». L'élève peut donc entourer « ma » ou « mi » dans les propositions correspondantes

¹⁹³ Cf. Annexe C, pp. 22-24.

au dessin d'un miroir, la lettre « m » produit le phonème [m] qui débute le mot s'il accorde implicitement au mot « son » de la consigne la valeur de « phonème ». Pour entourer la première syllabe du mot « poisson » (dont la représentation imagée est seulement fournie) parmi trois possibilités « poi », « moi », « mou », le phonème initial étant [p], il est possible de répondre à la consigne portant sur un repérage de syllabe par le simple repérage du graphème correspondant au premier phonème entendu (« le son qui commence le mot »). Ce type de tâche est toutefois très intéressant et proche de l'épreuve utilisée par cet enseignant et précédemment examinée dans le paragraphe 3.2.2b, à condition que les distracteurs proposés soient choisis en raison de difficultés connues comme, par exemple, la proximité phonologique de [j] et [s] dans le mot « sapin ».

La deuxième épreuve, reproduite ci-après¹⁹⁴, demande aux élèves de compléter par « ma » ou « mi » des mots dont la représentation imagée est fournie et dont une partie du mot est donnée : ex. « un rteau ».

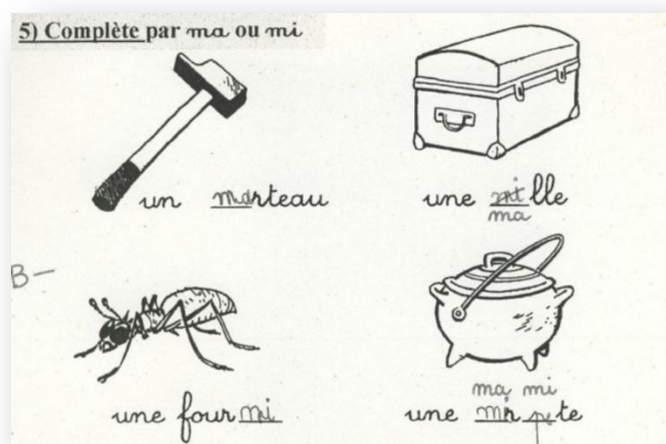


Figure 78 : Épreuve MT1 (Annexe C p. 22) avec réponses d'élève

Pour répondre, l'élève doit dire le nom du mot représenté ([marto]), détecter laquelle des deux modalités de réponses proposées ([ma] et [mi]) s'entend dans le mot [marto] puis écrire le groupe de lettres permettant de transcrire cette syllabe (« ma »). Le travail demandé à l'élève porte sur l'écriture d'un groupe de lettres fonctionnant tantôt comme une syllabe (pour « fourmi »), tantôt comme une portion de syllabe (« marteau ») ce qui implique de prendre en compte l'environnement de ces lettres dans le mot.

L'élève commet une première erreur pour le mot « malle ». Elle peut être comprise comme un défaut dans la chaîne procédurale que nous avons détaillée ci-dessus pour le mot « marteau ». Elle peut aussi être comprise comme résultant d'un effet de contrat : la syllabe « ma » ayant été utilisée pour répondre dans le premier mot, elle est, par contrat, écartée pour le deuxième mot pour lequel ne reste donc que « mi ».

Compléter le mot « ...r... te » correspondant au mot « marmite » est une tâche difficile qui requière un traitement fin du cadre de réponse imposé par la tâche. L'élève doit

¹⁹⁴ Nous reproduisons la copie d'un élève de la classe de l'enseignant H (cf. Annexe C, p.22). L'épreuve est identique mais la copie collectée contient des erreurs intéressantes à analyser.

successivement écrire « ma » et « mi » en tenant compte du graphème isolé (« r ») faisant partie de la première syllabe (« mar ») et dont il ne doit pas tenir compte dans l'écriture de la réponse qui est à sa charge. On peut penser qu'il serait sans doute plus simple, et plus cohérent du point de vue du sens des apprentissages, de demander à l'élève d'écrire le mot « marmite » sous la dictée. Sans entretien avec l'élève il est difficile de comprendre la production erronée de la deuxième partie de syllabe « pi »

La copie reproduite ici montre bien qu'une partie de la tâche est parfaitement exécutée, que la (portion de) syllabe « ma » ou « mi » soit à écrire au début ou à la fin du mot. A l'inverse, dans les deux autres mots, « malle » et « marmite », les erreurs contrebalancent la réussite sans que cela apporte des informations intéressantes sur la maîtrise de la combinatoire : on ne sait pas quel sens fait pour l'élève la réponse produite.

Ces éléments nous conduisent à écarter cette épreuve

La troisième épreuve¹⁹⁵ propose une série de syllabes (ou de portions de syllabes) présentées dans un tableau de 9 colonnes sur 3 lignes. L'enseignant dicte 6 syllabes « cibles » que l'élève a pour tâche de retrouver et d'entourer : « ru », « li », « ma », « mur », « ar », « mil ». Pour répondre, l'élève doit parcourir l'ensemble des cases, et extraire de chacune la valeur phonologique de ce qui écrit pour le comparer au modèle phonologique donnée par l'enseignant. Le parcours des cases du tableau peut être restreint lorsque l'élève s'appuie sur un phonème saillant (attaque ou rime) : par exemple si [ma] est donné on peut écarter toutes les syllabes commençant par « r ». L'activité de l'élève s'appuie donc sur sa maîtrise des relations graphophonologiques. Si l'épreuve porte sur des syllabes ou groupes de syllabes, elle ne les articule toutefois à la question du sens. Or, si l'on s'attache à la définition de la combinatoire soulignée par Charmeux¹⁹⁶, c'est un aspect essentiel qui est négligé. Cet auteur rappelle que « la combinatoire est l'ensemble des règles, spécifiques à chaque langue, et par voie de conséquence, arbitraires mais organisées en système, qui permettent que des unités linguistiques abstraites puissent devenir productrices de sens par leur présence (ou leur absence) et leur ordre, dans des productions langagières » (idem).

La quatrième épreuve demande à l'élève de reconstituer un mot découpé en syllabes (graphiques). Cette épreuve reproduite ci-dessous prend en compte la question du sens et constitue une épreuve valide.

¹⁹⁵ Epreuve 77-2, Annexe C, p. 23

¹⁹⁶ Repéré à <http://www.charmeux.fr/combinatoire.html>

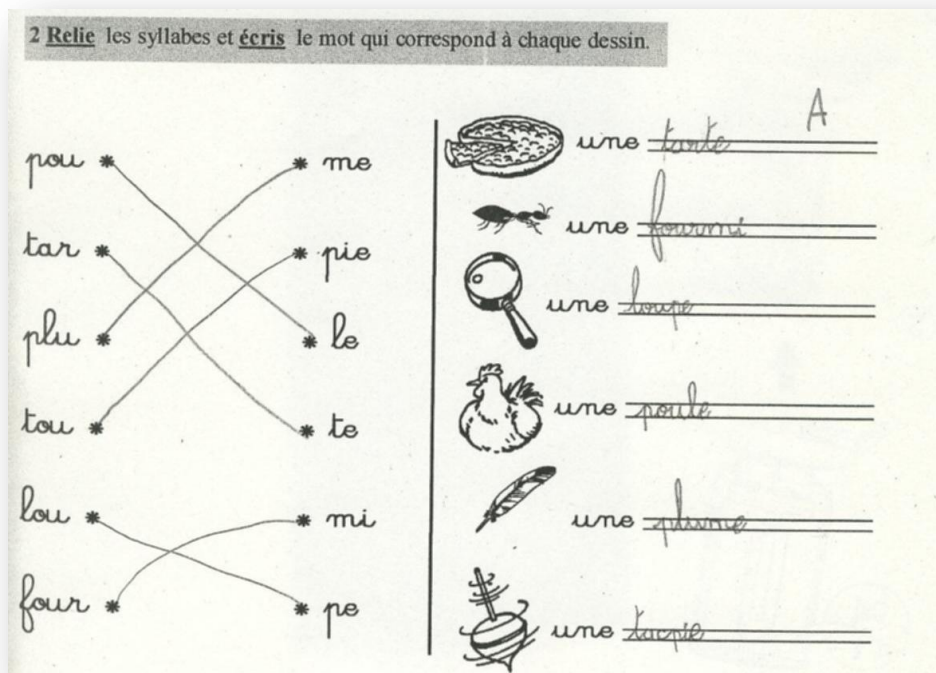


Figure 79 : Épreuve FT1 (Annexe C, p. 23) avec réponses d'élève.

Les assemblages que l'élève doit concevoir (« pou » + « le ») sont à comparer avec une représentation imagée du mot que l'élève doit ensuite réécrire de façon assemblée (« poule »). Le dessin intervient ici comme une modalité d'autocontrôle tout à fait intéressante. Cela limite d'éventuelles erreurs liées au cadre de réponse (une mauvaise association en entraîne nécessairement une autre). Pour répondre, l'élève mobilise bien des connaissances d'ordre graphophonologique, d'assemblage et de combinaison de syllabes dans le but de produire une suite de lettres ayant un sens.

La dernière épreuve étant semblable à la troisième, nous ne reviendrons pas sur les raisons qui nous conduisent à l'écartier.

Les épreuves utilisées par l'enseignant J

L'enseignant J utilise également plusieurs épreuves pour évaluer ces connaissances.

Deux épreuves¹⁹⁷ portent sur la lecture oralisée de syllabes ou le repérage de syllabes lues par l'enseignant dans une liste. Dans les deux cas, la question du sens n'est pas prise en charge : l'élève engage une activité de déchiffrage sur des syllabes isolées. Ces épreuves n'évaluent donc pas véritablement la maîtrise de la combinatoire.

La troisième épreuve utilisée¹⁹⁸ propose 6 listes de mots dans chacune desquelles l'élève doit repérer des syllabes cibles. Nous la reproduisons :

¹⁹⁷ Épreuves 91-4 p. 24 et épreuve 93-3, p. 25 Annexe C

¹⁹⁸ Épreuve 91-2, Annexe C, p. 24

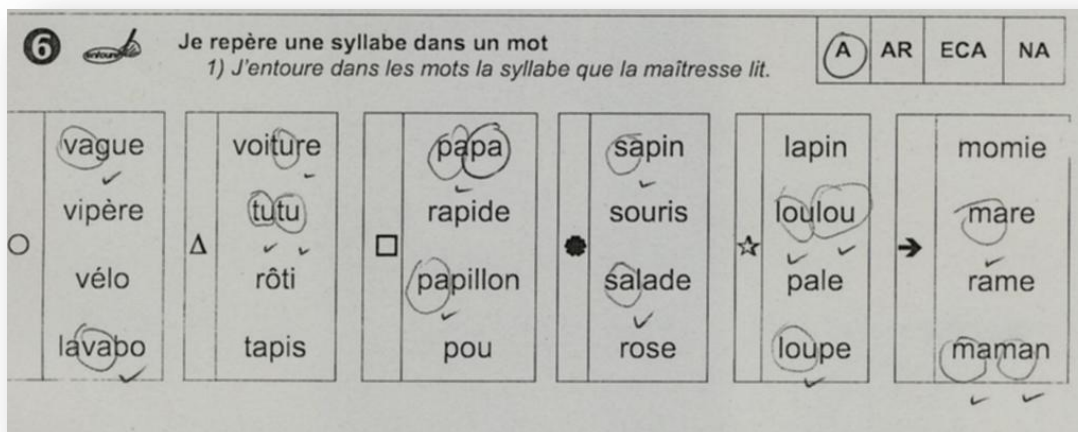


Figure 80 : Épreuve FT1 (Annexe C, p. 24) avec réponses d'élève

Pour répondre à la cinquième tâche (localiser [ma] dans « momie », « mare », « rame », « maman »), l'élève doit déchiffrer chaque mot : i.e. transformer la suite de graphèmes en mot phonologique. L'erreur commise dans la cinquième liste (pourtant validée par l'enseignant) est importante et montre la confusion entretenue entre une suite de graphèmes et la syllabe. L'élève ne prend pas suffisamment en compte l'environnement linguistique de la deuxième syllabe du mot « maman » et entoure le deuxième groupe « ma ». Or, dans cette syllabe précisément, la suite de lettres « ma » ne se lit pas [ma] en raison du « n » suivant qui nasalise la voyelle « a ». Cette épreuve correspond donc parfaitement à ce qui est visé (à savoir la maîtrise de la combinatoire) : sa validité psycho-didactique est établie. Elle est en revanche mal utilisée par l'enseignant J. Cela pose un autre problème, de correction.

La quatrième épreuve utilisée par cet enseignant ressemble à la première épreuve utilisée par l'enseignant I. Nous ne revenons pas sur les motifs qui nous conduisent à l'invalidation.

Les épreuves utilisées par l'enseignant K

L'enseignant K évalue les compétences visées au moyen de deux épreuves.

La première¹⁹⁹ épreuve demande à l'élève de localiser des syllabes lues par l'enseignant dans une liste de 10 mots. Cette épreuve se rapproche de la deuxième épreuve proposée par l'enseignant J. Nous la reproduisons ci-après :

¹⁹⁹ Épreuve 100-3, Annexe C, p. 25

5) Pour chaque mot, entoure la syllabe lue par l'enseignante.

1	maman	6	rame
2	lune	7	musique
3	vélo	8	avenir
4	souris	9	renard
5	auto	10	gâteau

A

Figure 81 : Épreuve FT1 (Annexe C, p. 25) avec réponses d'élève

La tâche de l'élève est plus difficile que ne laisse supposer la consigne : il doit localiser une syllabe orale dans un autre contexte que celui dans lequel elle est donnée. Pour répondre correctement à la demande de l'enseignant, l'élève doit entourer « ne » dans le mot « lune ». Or, le mot « lune » ne contient qu'une seule syllabe phonologique [lyn]. La consigne est en réalité d'entourer, non pas la syllabe lue par l'enseignant mais, le groupe de lettres qui, dans le mot n^o., se dit [...]. Dans le mot « renard », entourer la syllabe [nar] conduit l'élève à entourer « nar » laissant le « d » final seul. Cela laisse sous-entendre que deux autres syllabes existent dans le mot : « re » et « d ». Cette épreuve intéressante comporte plusieurs imprécisions qui conduisent à ne pas la considérer comme étant valide : elle ne conduit pas les élèves à isoler les syllabes d'un mot mais des ensembles de lettres formant une syllabe orale donnée par le maître de façon indépendante du contexte dans lequel elle est employée dans le mot, ce qui est justement l'aspect essentiel de ce sous-composant.

La deuxième épreuve²⁰⁰ engage une activité de déchiffrage oral. Nous ne revenons pas sur les raisons qui nous ont précédemment amené à écarter ce type d'épreuve.

3.2.3. Synthèse

Si l'on considère que l'ensemble des cinq types de relations phonographiques (cf. typologie de Goigoux et coll. 2013 présentée en IV 2.1) doit être évalué, alors aucun test étudié n'est valide au sens de la validité de couverture du domaine.

Si l'on considère que les enseignants se réfèrent aux progressions pour évaluer (lesquelles n'explicitent pas d'items portant sur les deux premiers types de relations phonographiques de la typologie de Goigoux et coll. (2013), cf. tableau n^o29), alors seuls les tests des enseignants D, I, J et K apparaissent en première approche globalement valides du point de vue de la couverture didactique du domaine. Toutefois, l'analyse des 27 épreuves proposées dans ces 4 tests au 1^{er} trimestre pour évaluer le composant « Relations phonographiques » de la lecture montre que seulement 5 d'entre elles (environ 18,5%) peuvent être considérées comme valides. Aucun test ne propose d'épreuves valides pour chacun des trois sous-composants examinés.

La validité de l'évaluation des aspects les plus significatifs du début d'apprentissage de la lecture au cours du premier trimestre apparaît ainsi rarement établie au vu des épreuves collectées et du référentiel utilisé.

²⁰⁰ Epreuve 100-4, Annexe C, p.25

3.3. Analyse de la dimension lecture des tests et des épreuves du troisième trimestre

Nous rappelons que ce composant renvoie à l'identification de mots, un objectif central de l'apprentissage de la lecture en CP. Il renvoie à la capacité des élèves à reconnaître immédiatement des mots fréquents (par voie orthographique ou voie d'adressage ou voie directe d'accès au mot), à décoder des mots inconnus en prenant appui sur les régularités du code graphophonologique (voie d'assemblage ou voie indirecte d'accès au mot) puis la capacité à associer au produit du décodage du sens (ce qui fait appel à des connaissances lexicales qui ne sont pas nécessairement à disposition au moment où l'élève lit)²⁰¹.

3.3.1. Les tests

37 épreuves²⁰² au sein des 16 tests du premier trimestre sont utilisées pour évaluer ce composant.

T3	Sous composants du composant L (Lecture)		
Ens.	L2 Reconnaissance	L3 Décodage/Identification	L4 Lecture voix haute
A	1	4	2
B	1	2	2
C			
D	1	1	1
E		1	1
F	1	1	1
G			
H			1
I			1
J	1	1	
K			
L	1	3	1
M		1	
N		1	1
O			1
P		3	1
Tot.	6	18	13

Tableau 31 : Répartition des épreuves d'évaluation recueillies au 3^e trimestre par enseignant et par sous-composant du composant de la lecture.

Bien que la maîtrise de ces habiletés (en tout premier lieu le décodage) soit un objectif prioritaire du CP, on constate à la lecture du tableau que certains sous-composants ne sont pas évalués dans de nombreux cas, bien que chacun d'eux fasse l'objet d'un item dans les progressions. Ce premier résultat important conduit à souligner que le guidage de l'évaluation par les items ne produit pas l'effet escompté en termes de production d'épreuves spécifiques malgré l'insistance de la prescription.

²⁰¹ Décoder le mot « soliflore » est accessible à un élève de CP. En revanche, sauf travail particulier, il y a peu de chance qu'il parvienne à l'identifier.

²⁰² Elles sont classées par type de tâches et par enseignant dans l'Annexe C, pp.85-101

Seuls les enseignants A, B, D, F et L (environ un tiers de la population étudiée) évaluent au moyen d'épreuves spécifiques le lexique orthographique, les habiletés de décodage, et la lecture à haute voix au terme de l'année de CP : 3 des 5 compétences clés de l'évaluation de la lecture en CP d'après Giasson (2012) (cf. paragraphe 3.1.4 de cette partie). Cette proportion est extrêmement faible en regard des enjeux décisifs qui sont associés à la maîtrise de ces compétences de lecture. Cela constitue un autre résultat important de l'analyse du tableau.

Un peu plus des deux-tiers des enseignants de la population étudiée (11 sur 16) évalue ces compétences de lecture au moyen d'une épreuve au moins de lecture à haute voix. On peut sans doute estimer que les épreuves de lecture à haute voix permettent d'évaluer en même temps les deux autres sous-composants, ce que les enseignants H, I, O et P ont sans doute estimé. Nous invalidons ce choix parce que ces enseignants n'utilisent qu'une épreuve de lecture, chronométrée, d'un texte. Or, dans la lecture d'un texte (voire même d'un groupe de mots), des processus de compréhension sont nécessairement convoqués ce qui augmente la difficulté de la tâche et peut ralentir le temps de lecture. Par ailleurs, dans une étude portant sur la compréhension en fin de CP nous avons constaté que la lecture à haute voix d'un texte conduit de bons lecteurs (i.e. obtenant un bon score en fluence et un bon score en compréhension) à commettre des erreurs de lecture de mots fréquents (cf. C. Blanc et Vantourout, 2015). Nous renvoyons le lecteur au paragraphe 5.5 de la Partie III pour les analyses que nous avons conduites des épreuves de fluence en lecture : toutes les épreuves de lecture à haute voix collectées sont des épreuves qui portent sur des textes et qui, de ce fait, sollicitent des processus de compréhension au cours de la lecture. Leur validité psychodidactique est remise en cause.

3.3.2. Les épreuves des tests A ; B ; D ; F et L

1/ Épreuves de reconnaissance

Chacun de ces enseignants utilise une épreuve²⁰³ pour évaluer le lexique orthographique des élèves (i.e. reconnaissance de mots par voie d'adressage). Dans chacune de ces épreuves, pour répondre, l'élève doit lire à haute voix une série de mots fréquents.

A l'exception d'un seul mot dans chaque liste, les enseignants A et D ne proposent que des mots monosyllabiques fréquents. Ils peuvent être immédiatement reconnus, de façon plus rapide que certains mots contenus dans les listes des épreuves proposées par les enseignants B ; F et L qui proposent au contraire davantage de mots contenant plusieurs syllabes (« avec, comment, parce que, jamais, dessus, encore, pourquoi ... »). Ces mots sont susceptibles d'être reconnus par voie d'assemblage ce qui illustre la difficulté de concevoir des épreuves permettant de n'évaluer que l'une ou l'autre voie d'accès au lexique.

En fin de CP, l'usage de ces épreuves est pleinement justifié. Elles engagent l'activité de lecture de l'élève visée laquelle est fondée sur la reconnaissance rapide de mots fréquents. Leur validité nous semble avérée.

²⁰³ Épreuves 18-1 (Enseignant A), 28-4 (Enseignant B), 52-5 (Enseignant D) situées en Annexe C p.85 et épreuves 68-1 (Enseignant F) et 101-4 (Enseignant L) situées en Annexe C p. 86.

2/ Épreuves de décodage

Les enseignants utilisent tous au moins une épreuve²⁰⁴ demandant à l'élève de lire une série de mots « inconnus » : i.e. dont l'enseignant suppose *a priori* qu'ils ne sont pas immédiatement reconnaissables par l'élève. L'épreuve reproduite ci-dessous utilisée par l'enseignant F nous permet de discuter ce point précis :

Consigne : Lis les mots suivants :

la main	le pigeon	le poing	un poisson	un garçon
un éléphant	La fille	du feu	les oiseaux	une cigogne
un chameau	la marelle	le pied	un lapin	le loup

Figure 82: Épreuve FT3 (Annexe C, p. 90) avec réponses d'élève

Pour répondre et lire le mot « cigogne », l'élève mobilise probablement ses connaissances sur le code graphophonologique pour proposer [si.gɔŋ] en deux syllabes (le mot « une » étant reconnu directement). Il mobilise donc probablement plutôt la voie d'assemblage.

En revanche, il est hautement probable qu'il reconnaisse et identifie le mot « fille » directement : ce mot n'est pas un mot « outil » mais c'est un mot très fréquenté, auquel cas l'élève mobilise plus sûrement la voie d'adressage. Il est probable qu'il en soit ainsi pour d'autres mots de la liste : « main », « poisson », « garçon », feu », « oiseaux », « lapin » et « loup ». Le mot « loup » reconnu directement sera lu [lu] tandis que, si l'on se fie aux règles de fonctionnement du code, il y a un risque de prononcer le « p » final : [lup]

Toutes les épreuves utilisées par les enseignants pour évaluer les compétences de décodage ou de reconnaissance des mots sont valides considérant qu'il n'est pas aisé de savoir à quel moment les élèves utilisent l'une ou l'autre voie d'accès à la reconnaissance d'un mot.

Trois épreuves seulement, contenues toutes dans le test A, proposent une tâche supplémentaire d'identification, i.e. qui engage une activité d'attribution de sens au déchiffrement effectué. Dans l'épreuve ci-après, l'élève doit déchiffrer les cinq mots puis les associer au dessin qui les représentent le mieux pour répondre. L'univers de référence de chacun des mots est supposé connu des élèves.

²⁰⁴ Épreuves 10-2, 18-2, 18-3 et 20-3 (Enseignant A), 28-3 et 29-1 (Enseignant B), 53-1 (Enseignant D) situées en Annexe C pp.88-90 et épreuves 68-2, 68-3 (Enseignant F), 101-2, 101-3 et 105-4 (Enseignant L) situées en Annexe C pp. 91-93

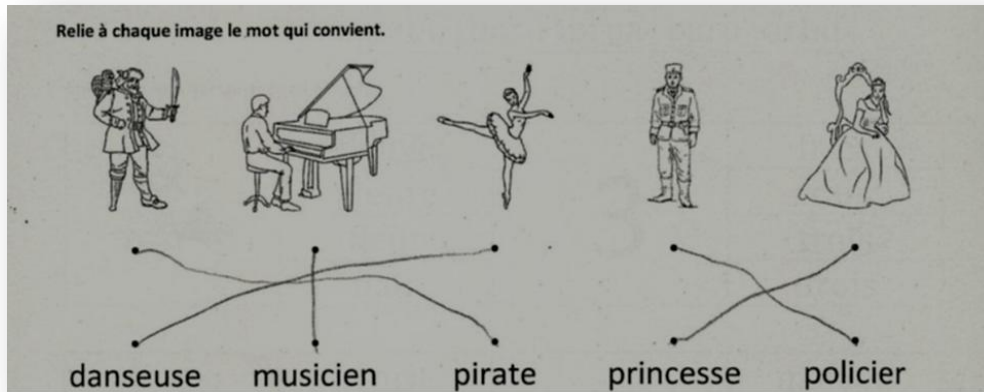


Figure 83 : Épreuve FT3 (Annexe C, p. 88) avec réponses d'élève.

Ce type d'épreuve mobilise bien une activité de reconnaissance et/ou de décodage (il se peut que les mots « pirate » ; « princesse » soient immédiatement reconnus) et d'identification de mots.

3/ Épreuves de lecture à haute voix

Nous avons invalidé l'usage des épreuves évaluant de façon explicite la lecture à haute voix au motif, brièvement rappelé ici, qu'elles évaluent également, et de façon implicite, la compréhension de texte.

Toutefois, dans les épreuves de reconnaissance ou de décodage de mots examinées dans ce paragraphe, les élèves sont invités à lire à haute voix pour répondre. Nous avons conclu à la validité des épreuves. Nous pouvons donc souligner que les épreuves de lecture à haute voix sont des épreuves valides pour évaluer les habiletés de décodage et/ou de reconnaissance ou d'identification de mots isolés, ce qui rejoint les préconisations du CNECSO (2016).

3.3.3. Synthèse

Au total, les tests examinés (les tests A ; B ; D ; F et L) contiennent tous au moins une épreuve valide permettant d'évaluer les habiletés de décodage et le lexique orthographique des élèves. Seul le test A contient en plus des épreuves permettant d'évaluer la capacité d'identification des mots des élèves : à ce titre, il nous apparaît être le test disposant de la meilleure validité globale et locale du point de la couverture du composant « Identification des mots » en lecture.

3.4. Analyse de la dimension compréhension des tests et des épreuves du troisième trimestre

Nous nous intéressons à présent à l'évaluation réalisée au cours du 3^e trimestre de CP au sujet du composant Compréhension : le composant le plus expert de la lecture.

Si « la lecture est une construction de significations réalisées par une personne à partir d'un texte écrit, dans un contexte donné » (cf. définition posée en début de chapitre, 3.1) comment les épreuves permettent-elles d'évaluer cette construction de significations propres à chaque lecteur ?

3.4.1. Les tests

38 épreuves²⁰⁵ au sein des 16 tests du premier trimestre sont utilisées pour évaluer la compréhension d'un écrit (lu par l'élève). Elles impliquent toutes la réalisation de tâches écrites impliquant la compréhension d'unités infra-textuelles (des phrases ou groupes de 2 phrases) ou de textes.

T3	Sous- composant du composant C (Compréhension)	
Ens.	C8 Unités infra-textuelles	C8 Unités textuelles
A	3	2
B		2
C	2	1
D		1
E	2	1
F		1
G		1
H		1
I	3	2
J	3	2
K		
L	1	1
M		2
N		2
O		1
P	1	3
Tot.	15	23

Tableau 32 : Répartition des épreuves d'évaluation recueillies au 3^e trimestre par enseignant et par sous-composant du composant « Compréhension »

L'enseignant K se démarque des autres : il ne propose aucune épreuve d'évaluation de la compréhension au motif de projets de classe que l'échéance de la fin de l'année scolaire a conduit à privilégier au détriment de l'évaluation. L'enseignant K s'est donc servi d'évaluations informelles, réalisées au fil de la classe, pour rédiger un commentaire de synthèse dans le livret scolaire.

Cela démontre que l'évaluation que nous étudions dans notre travail n'est qu'une part de l'évaluation réalisée par les enseignants, et que la part de celle-ci, bien que non-instituée, est loin d'être négligeable dans certains cas.

²⁰⁵ Elles sont classées par types de tâches et par enseignants dans l'Annexe C, pp.2-28.

Dans le chapitre 4 de la partie III, nous avons détaillé les raisons pour lesquelles la validité des épreuves d'évaluation de la compréhension uniquement basées sur des questions écrites à réponses écrites était discutable. Nous y renvoyons le lecteur. Deux études complémentaires (C. Blanc, 2013, 2014 ; C. Blanc & Vantourout, 2015 ; C. Blanc & Vantourout, soumis ; Vantourout & Maury, accepté, pour la première recherche et C. Blanc et Goasdoué, 2016 pour la seconde) nous ont permis d'affiner ce point de vue : les résultats obtenus par des élèves à une épreuve de compréhension d'un texte lu seul dépend du texte, de la conception du questionnaire et du mode de questionnement. Répondre à un questionnaire sur un texte est en tout premier lieu une évaluation, implicite, de la compréhension du questionnaire.

C'est la raison pour laquelle, nous recherchons d'autres preuves de validité globale dans les tests examinés. Parmi les 16 tests examinés, 7 d'entre eux contiennent des épreuves portant sur la compréhension de phrases qui ciblent plus particulièrement les microprocessus de lecture voire les processus d'intégration d'après le cadre proposé par Giasson (1990) lorsque les élèves ont à lire au moins deux phrases. Pour poursuivre nos analyses de validité des tests et des épreuves de compréhension en lecture en fin de CP, nous nous intéressons donc aux épreuves contenues dans ces 7 tests (A ; C ; E ; I ; J ; L et P) portant spécifiquement sur la compréhension de phrases et sur la compréhension de textes. Nous recherchons à observer en quoi elles se complètent.

3.4.2. Les épreuves des tests A ; C ; E ; I ; J ; L et P

Les épreuves utilisées par l'enseignant A

1/ Pour évaluer la compréhension de phrases, l'enseignant utilise trois épreuves. Nous reproduisons ci-dessous la première d'entre elles :

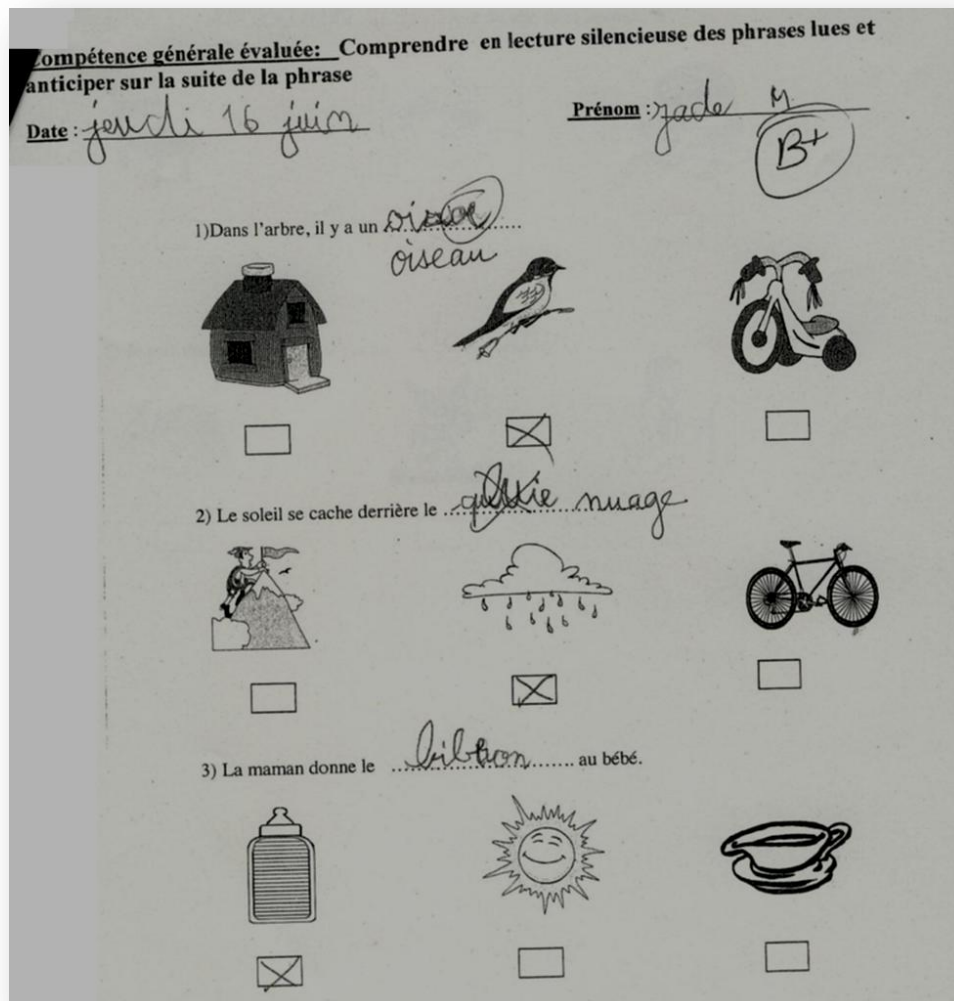


Figure 84 : Extrait d'épreuve FT3 (Annexe C, pp.103-104) avec réponses d'élève

Pour chaque phrase de l'épreuve, on peut distinguer trois phases dans l'activité de réponse de l'élève :

- Compléter la phrase avec un mot, ce qui nécessite la réalisation d'une inférence. Plusieurs possibilités existent pour compléter « le soleil se cache derrière le ... ». Par exemple « bois », « bateau » ... C'est la tâche principale évaluée par cette épreuve.
- Choisir un mot correspondant à un dessin proposé par le cadre de réponse. Le cadre de réponse soutient, mais limite, l'activité de l'élève en proposant trois dessins : le dessin d'un alpiniste parvenant à un sommet, le dessin d'un nuage se transformant en pluie, et le dessin d'un vélo. Chacun des éléments pourrait convenir (le soleil pourrait se cacher derrière le sommet, derrière le nuage, derrière la roue du vélo (si l'observateur est allongé ?)). Ces propositions orientent l'activité de réponse.
- Encoder le mot correspondant.

Dans la copie reproduite ici, l'élève écrit « pluie » pour compléter la deuxième phrase. Cette réponse, considérée comme erronée par le correcteur qui estime que « nuage » est la bonne réponse, est imputable au dessin. Du point de vue du sens, la réponse est pourtant acceptable (le soleil peut très bien se cacher derrière la pluie). Mais du point de vue syntaxique, la réponse est en effet erronée du fait d'un accord en genre qui doit être réalisé. Il s'appuie sur la mobilisation d'une connaissance d'un autre ordre (par extension « le rideau de pluie » conviendrait), une connaissance syntaxique.

En somme, cette épreuve intéressante engage bien la mobilisation de processus d'intégration qui interviennent dans la compréhension. Mais l'épreuve nécessite la réalisation de tâches différentes non évaluées pour elles-mêmes : après choix du mot réponse (la tâche principale), il faut vérifier qu'il correspond aux choix proposés (et sinon en choisir un autre), il faut écrire le mot et l'accorder en genre ce qui peut conduire à remettre en cause la réponse trouvée et à tout recommencer. Cela nous conduit à invalider l'usage de cette épreuve.

On pourrait aménager cette épreuve dans le but de la rendre valide en ne donnant pas de dessins et en demandant à l'élève de donner sa réponse soit par écrit, s'il peut et sait le faire, soit par oral. En tout état de cause, le respect de la contrainte orthographique est un critère surajouté de réussite de l'épreuve : il est pourtant pris en compte par l'enseignant au moment de la correction comme la réponse à la première phrase le montre.

La deuxième épreuve²⁰⁶ demande à l'élève d'associer une phrase à un dessin.

F17 *Dire de quoi et de qui parle le texte lu*
 - associer une image à une phrase :
 - trouver des informations explicites :

4 /4 (1 pt par bonne réponse)
 3 /4 (1 pt par bonne réponse)

Résultat **BF**

Relie à chaque image la phrase qui convient.

Bruno pleure. Bruno écoute une histoire. Bruno est endormi. Bruno boit du sirop.

Figure 85 : Épreuve FT3 (Annexe C, p. 104) avec réponses d'élève

Pour répondre, l'élève doit lire chaque phrase et trouver parmi les 4 vignettes dessinées celle qui correspond (le mieux). Le mot « Bruno » débute toutes les phrases, un enfant est dessiné dans chaque vignette : l'enfant s'appelle « Bruno ». L'élève doit lire le mot suivant « pleure » et associer la phrase lue au deuxième dessin (le dessin de larmes le permet, il ne faut pas prendre compte le dessin de la mère). L'élève ne peut se contenter de trouver le

²⁰⁶ Epreuve 23-1, Annexe C, p. 104

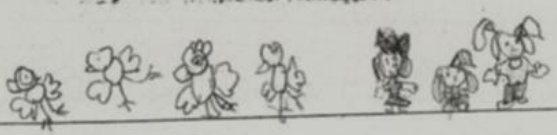
mot « livre » ou « lit » pour associer la deuxième phrase au troisième dessin. La dernière phrase est associée au dernier dessin ce qui limite l'effet des connaissances sur le monde (ici le sirop à boire est synonyme de prendre un médicament). Au total, cette épreuve engage bien une tâche de compréhension fondée sur la mobilisation de microprocessus de lecture (identification des mots) dans le contexte d'une phrase. Elle peut être considérée comme valide.

La troisième épreuve proposée par l'enseignant A est reproduite ci-après.


Compétence générale évaluée: Comprendre en lecture silencieuse des phrases lues et en rendre compte

Date : mardi 14 juin Prénom : grande M A-

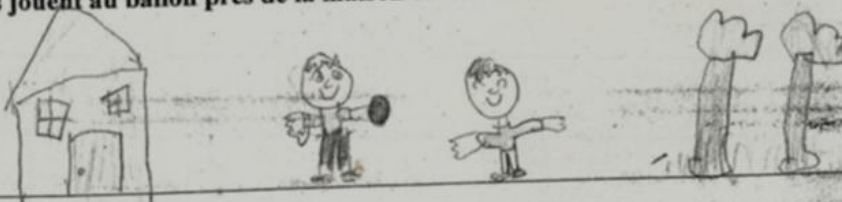
1. Je vois quatre oiseaux et trois dames.



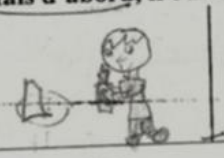
2. Il y a deux chats. Le plus grand est tout rouge et le plus petit est bleu.



3. Ils jouent au ballon près de la maison et de la forêt.



4. Il se sert un verre. Mais d'abord, il ouvre la bouteille de jus.



5. Avec ma fusée, je vais sur la lune.

?

Figure 86 : Épreuve FT3 (Annexe C, p. 105) avec réponses d'élève.

L'élève doit traduire par un dessin la compréhension de chaque phrase lue seul.

La première phrase engage un dessin cumulatif (l'élève peut lire « quatre oiseaux » et les dessiner, puis lire « trois dames » et les dessiner). L'illustration de cette phrase ne mobilise pas véritablement d'autres processus de lecture que les microprocessus (identification de mots et lecture par groupes de mots). Elle devrait être réussie par la majorité des élèves.

La deuxième phrase est en réalité un petit texte : la tâche nécessite la mobilisation de processus d'intégration et d'inférences référentielles (« le plus grand » renvoie à la phrase précédente : « il y a deux chats »). Pour produire sa réponse, l'élève doit avoir lu les deux phrases sans quoi la taille des chats ne sera pas différenciée.

La troisième phrase nécessite la réalisation d'une inférence logique dont le référent est implicite (« ils » renvoient nécessairement à « plusieurs enfants »).

La quatrième phrase fonctionne aussi comme un texte. L'élève doit restituer la chronologie de l'histoire (en deux dessins) : elle est inverse à la présentation des événements du texte ce qui nécessite d'avoir bien compris le sens du groupe de mots « mais d'abord ».

Enfin, la cinquième phrase nécessite la réalisation d'une inférence de moyen (avec ma « fusée », je vais sur la Lune). L'élève écrit le mot correspondant ou le dit à l'enseignant, ce qui évite le défaut souligné au cours de l'analyse de l'épreuve précédente.

Nous estimons que cette épreuve est pertinente pour évaluer la compréhension en lecture de phrases isolées (surtout fondée sur la mobilisation de microprocessus de lecture), mais aussi de phénomènes apparaissant dans un texte et nécessitant la mobilisation de processus d'intégration.

2/ L'enseignant A utilise ensuite deux épreuves pour évaluer la compréhension d'un texte.

La première épreuve²⁰⁷ reproduite ci-après porte sur un texte informatif. Quatre questions écrites sont posées et, pour répondre, l'élève doit souligner dans le texte « la réponse ». Ce système de réponse évite l'écriture des réponses mais de ce fait, la relation questions-réponses est exclusivement de type textuelle-explicite (la question porte sur un élément explicitement contenu dans le texte²⁰⁸) et mobilise des microprocessus de lecture.

²⁰⁷ Epreuve 23-2, Annexe C, p. 113.

²⁰⁸ Cf. Partie III, chapitre 4.2

2 Lis le texte suivant.

Le lapin sauvage

La lapine peut avoir 4 à 10 petits à la fois.
 Les petits naissent sans poils, les yeux fermés.
 Ils tètent leur mère.
 Le lapin mange de l'herbe et des légumes.
 Pour échapper aux renards, aux chasseurs et aux chiens, il se cache dans son terrier.

Combien la lapine peut-elle avoir de petits ? Souligne en vert la réponse dans le texte. 0,5

Que mange le lapin ? Souligne en noir la réponse dans le texte. 1

Où le lapin se cache-t-il ? Souligne en rouge la réponse dans le texte. 1

Comment sont les petits quand ils naissent ? Souligne en bleu la réponse dans le texte. 0,5

Figure 87 : Épreuve FT3 (Annexe C, p. 113) avec réponses d'élève

L'élève souligne « 10 » dans le texte pour répondre à la première question « combien la lapine peut-elle avoir de petits ». Cette réponse reçoit un crédit partiel : la bonne réponse attendue est « 4 à 10 ». Cette épreuve nécessite la lecture d'un texte sur lequel les questions posées n'engagent que des microprocessus de lecture : elle ne permet pas d'évaluer la signification construite par le lecteur. Elle est de ce fait moins riche que les épreuves portant sur des phrases ou des groupes de deux phrases, et le modèle de situation élaboré par l'élève n'est pas interrogé. Comprendre le texte est synonyme de répondre correctement à des questions de repérage d'éléments explicitement contenus dans les questions et dans le texte.

L'autre épreuve²⁰⁹ de compréhension de texte engage la production de réponses écrites à des questions écrites portant sur un texte lu par l'enseignant. Nous avons invalidé l'usage de telles épreuves : les compétences rédactionnelles sont fortement mobilisées et l'évaluation de la compréhension passe par l'évaluation implicite de celles-ci.

Les épreuves utilisées par l'enseignant C

1/ L'enseignant C utilise deux épreuves pour évaluer la compréhension en lecture au niveau de la phrase.

La première épreuve²¹⁰ demande à l'élève d'associer une phrase (parmi 4 modalités de réponse) à un dessin. Elle engage la lecture des 12 phrases contenues dans l'épreuve pour n'en retenir que 3 à associer à un dessin. L'épreuve engage donc 12 tâches de compréhension (dans 9 cas il s'agit après lecture de conclure qu'aucune phrase ne

²⁰⁹ Épreuve 26, Annexe C, p. 114.

²¹⁰ Épreuve 37, Annexe C, p. 106

correspond au dessin cible, dans 3 cas il s'agit après lecture de conclure que la phrase correspond au dessin cible). L'organisation des mots dans les différentes phrases oblige l'élève à lire toute la phrase, il ne peut se contenter d'un repérage de surface pour répondre. L'activité de lecture de l'élève mobilise essentiellement des microprocessus de lecture. Cette épreuve est donc valide pour évaluer la compréhension.

La deuxième épreuve porte sur la compréhension de phrases particulières : des consignes. La consigne de l'épreuve est de « lire attentivement » et de faire ce qui est demandé (par chacune des 4 consignes suivantes). La quatrième consigne demande à l'élève de colorier le mot (parmi 4) qui correspond au dessin. Pour répondre à la consigne et colorier le mot « guitare » correspondant au dessin d'une guitare, l'élève doit lire « gitare », « guitare », « guidare » et « quitare ». Il peut échouer en colorier le mot « guidare » (la bonne réponse nécessite de repérer le mot « guitare ») ou en entourant le mot « guitare » (la bonne réponse nécessite de colorier et non d'entourer) : des erreurs de nature différente. Répondre nécessite donc de réaliser la tâche sur laquelle porte l'une des consignes de l'épreuve. La troisième consigne semble à cet égard inutilement compliquée (si l'on cherche à voir si l'élève sait exécuter une consigne demandant de colorier quelque chose) : il faut colorier le dessin du chat en bleu (on peut donc échouer si on entoure au lieu de colorier, si on colorie en rouge, si on colorie le mot « chat » au lieu du dessin) et souligner le mot souris. Réussir nécessite donc de respecter cinq critères : colorier ET dessin de chat ET en bleu ET souligne ET le mot souris. Une seule erreur dans cette chaîne d'action entraîne un non succès.

Cette épreuve intéressante évalue bien des compétences de compréhension mais elles relèvent plutôt de l'ordre du méta- ou du contrat (il est plus fréquent d'entourer un mot que de le colorier ce qui peut entraîner des erreurs). Cela nous conduit à invalider l'usage de cette épreuve.

2/ L'enseignant C utilise ensuite une épreuve destinée à évaluer la compréhension d'un texte²¹¹. Nous avons amplement détaillé dans le chapitre 6.1 de la Partie III les raisons qui nous conduisaient à invalider l'usage de cette épreuve pour évaluer la compréhension.

Les épreuves utilisées par l'enseignant E

1/ L'enseignant E utilise 2 épreuves pour évaluer la compréhension de phrases

La première²¹² d'entre elles demande à l'élève de numéroter, de 1 à 5, 5 phrases dans l'ordre donné (par l'enseignant). L'activité de réponse de l'élève est fondée sur l'identification des mots et ne fait, *a priori*, pas appel à la compréhension du sens des différentes phrases. Pour placer le n°1, l'élève doit différencier « ce matin, il a plu » de « ce matin, il a fait froid » et de « ce matin, il neige à gros flocons ». La validité de l'épreuve pour évaluer la compréhension est remise en cause.

La deuxième épreuve²¹³ demande à l'élève de construire 4 paires composées chacune d'une phrase (parmi 8) et d'un dessin (parmi 4). L'élève doit par exemple associer le dessin d'un garçon offrant une fleur à une fille, à la phrase « il lui donne une fleur ». Pour répondre, l'élève doit réaliser des inférences référentielles (« il » = un personnage masculin ; « lui » =

²¹¹ Epreuve 36, Annexe C, p. 117.

²¹² Epreuve 57-2, Annexe C, p. 108.

²¹³ Epreuve 58-2, Annexe C, p. 108

« à elle » = un personnage féminin). L'épreuve contient plus de phrases que de dessins de sorte que la moitié des phrases n'est pas associée à un dessin : cela limite le risque qu'une mauvaise réponse entraîne nécessairement une autre.

Au total, cette épreuve nous semble valide pour évaluer la compréhension.

2/ L'enseignant E utilise une épreuve pour évaluer la compréhension d'un ensemble de consignes. Cette épreuve est reproduite ci-dessous :

10 Dessine en respectant les consignes.

- Dessine un grand sapin.
- Ajoute un oiseau tout en haut de l'arbre.
- Dessine des nuages gris dans le ciel.
- Dessine un enfant qui joue au ballon à côté du sapin.




Figure 88 : Épreuve FT3 (Annexe C, p. 118) avec réponses d'élève

Cette épreuve n'évalue pas la compréhension de consignes isolées mais la compréhension d'un petit texte : les différentes consignes sont liées (l'élève doit dessiner un oiseau en haut d'un sapin précédemment dessiné, puis un enfant à côté de ce même sapin) et, pour répondre, l'élève doit comprendre la substitution lexicale de « un grand sapin » par « l'arbre ». Il doit également inférer qu'« ajouter » signifie « dessine sur le même dessin ». Cela nécessite la mobilisation de processus d'intégration. Nous retrouvons trace de cette épreuve à deux autres reprises (dans les tests I et J). Nous avons ainsi accès à d'autres copies reproduites ci-dessous :

3 Dessine en respectant les consignes.

- Dessine un grand sapin.
- Ajoute un oiseau tout en haut de l'arbre.
- Dessine des nuages gris dans le ciel.
- Dessine un enfant qui joue au ballon à côté du sapin.




Figure 89 : Épreuve FT3 (Annexe C, p. 123) avec réponses d'élève

6/ Je comprends les informations explicites de phrases.				A	AR	ECA	NA
---	--	--	--	---	----	-----	----

♦ Dessine dans le cadre en respectant les consignes.

- ♦ Dessine un grand sapin. ✓
- ♦ Ajoute un oiseau tout en haut de l'arbre. ✓
- ♦ Dessine des nuages gris dans le ciel. ✓
- ♦ Dessine un enfant qui joue au ballon à côté du sapin. ✓

Figure 90 : Épreuve FT3 (Annexe C, p. 124) avec réponses d'élève

Ces différentes réponses d'élèves permettent de constater que les trois élèves ont chacun construit une représentation différente du texte : le premier élève dessine une fille tenant un ballon dans les mains²¹⁴, le deuxième élève dessine un garçon jouant au football, et le troisième élève schématise davantage qu'il ne dessine un enfant ballé au pied.

Même si le texte ne comporte pas de difficultés de compréhension particulière, l'épreuve n'évalue ainsi pas seulement la compréhension de phrases isolées.

Nous soulignons que l'élève ayant produit la troisième copie, bien qu'il ait correctement exécuté les consignes successives, n'obtient qu'un crédit partiel (« AR » pour « Acquis à renforcer ») : « dessiner » semble donc pris au pied de la lettre par le correcteur. Dans ce cas, on n'évalue plus la compréhension en lecture mais la capacité à dessiner. Les deux premières copies d'élèves présentent des dessins coloriés sans que cela ne soit nécessaire pour indiquer qu'ils ont compris ce que les consignes demandaient de faire.

Il est possible aussi que l'enseignant considère que l'arbre est mal dessiné. Ce qui serait très intéressant car, le dessin produit par ce troisième élève peut être aussi analysé comme lié à la nature de la consigne générale (i.e. exécuter une série de consignes). Le sapin dessiné par l'élève montre deux parties distinctes. On peut penser que l'élève a répondu à la première consigne et qu'il laisse de la place pour l'exécution des autres consignes (puisque la consigne générale demande de « dessiner »). N'ayant pas véritablement la place d'« ajouter un oiseau tout en haut de l'arbre », il le dessine donc légèrement décalé (une trace de dessin effacé au sommet de l'arbre nous conduit à ce raisonnement). Il dessine ensuite les nuages conformément à la troisième demande puis, afin de faire reposer son personnage sur un sol, il prolonge le dessin du premier arbre de façon que celui-ci s'aligne également sur le bord inférieur du cadre qui devient alors le sol. Cette suite d'actions correspond à la demande du concepteur : l'élève traite la tâche comme une succession de tâches ce qui l'empêche d'élaborer une représentation globale de la tâche.

En somme, au cours de la réalisation de cette tâche, l'activité de l'élève s'appuie sur la mobilisation de microprocessus et de processus d'intégration. En ce sens, son usage pour évaluer la compréhension en lecture au niveau de la phrase n'est pas remis en cause bien

²¹⁴ Par extension elle joue avec car il n'est pas facile de dessiner l'action de jouer au ballon

que les phrases soient pour partie liées et forment en réalité un texte. Son usage pour évaluer la compréhension d'un texte est davantage discuté : le lien entre les phrases est faible et un traitement pas à pas de la tâche peut suffire pour répondre. Un tel traitement de la tâche est susceptible de compliquer la réponse de l'élève lorsqu'il doit harmoniser le dessin des différents éléments de réponse : il n'est pas facile de modifier un dessin comme le montre l'analyse de la troisième copie d'élève.

Le système de réponse imposé par la tâche est susceptible d'invalider l'usage de cette épreuve pour évaluer la compréhension lorsque la tâche de dessin est évaluée du point de sa qualité graphique.

Les épreuves utilisées par l'enseignant I

1/ L'enseignant I utilise 3 épreuves pour évaluer la compréhension de phrases

La première épreuve²¹⁵ demande à l'élève d'entourer les phrases « vraies ». L'épreuve vise le repérage de la cohérence des phrases du point de vue du sens. Les trois premières phrases sont à écarter (le canard n'a pas quatre pattes, la voiture ne doit pas être remplie d'eau pour rouler, les enfants ne mangent pas à la cantine le soir) et les deux autres à entourer (le mécanicien répare les voitures en panne, quand on est malade on va chez le docteur). Les propositions « vraies » sont sujettes à discussion : on peut être malade sans aller chez le docteur (ce qui conduit à répondre que la cinquième phrase est fausse), on peut être malade et aller chez le médecin (ce qui conduit à répondre que la cinquième phrase est fausse parce que l'équivalence lexicale « médecin »-« docteur » ne serait pas connue de l'élève). De la même façon, on peut réparer soi-même sa voiture en panne (ce qui conduit à répondre que la quatrième phrase est fausse), et un garagiste peut réparer les voitures en panne (ce qui conduit à répondre que la quatrième phrase est fausse parce que l'équivalence lexicale « mécanicien »-« garagiste » n'est pas connue).

Cette épreuve évalue des compétences lexicales ainsi que des connaissances sur le monde implicitement convoquées liées à l'univers du médecin ou du mécanicien automobile. L'élève peut fournir des réponses erronées qui ne seraient pas à liées à un défaut de mobilisation des processus de lecture que l'on veut évaluer. A ce titre, sa validité est remise en cause.

La deuxième épreuve²¹⁶ demande à l'élève d'exécuter une série de 8 consignes partiellement illustrées (pour probablement éviter des difficultés d'identification des mots) lesquelles nécessitent le respect d'un certain nombre de critères de réalisation. Par exemple :

-« Entoure 2 arbres en vert » en comprend ainsi 4 : « entourer » ET « deux » ET « arbres » ET « en vert »

-« Dessine un oiseau jaune sur le toit, une souris grise sur le banc et deux chats noirs dans l'herbe » en comprend 10 : « dessine » ET « un oiseau » ET « sur le toit » ET « une souris » ET « grise » ET « sur le banc » ET « deux » ET « chats » ET « noirs » ET « dans l'herbe ». Pour cette consigne, le critère « jaune » ne peut être réellement vérifié car la première consigne demande de colorier le toit en orange.

Réussir l'épreuve nécessite ainsi le respect de plus de 30 critères. L'épreuve évalue surtout la réalisation d'une série d'actions. En effet, mis à part des noms des couleurs, tous les mots

²¹⁵ Epreuve 87-2, Annexe C, p. 109

²¹⁶ Epreuve 89, Annexe C, p. 109

(« toit », « arbre », « fleur », « oiseau », « souris », « banc », « chat », « herbe », « murs », « nuage ») sont illustrés. La validité de cette épreuve est de ce fait remise en cause.

La troisième épreuve²¹⁷ est identique à la deuxième épreuve utilisée par l'enseignant E pour évaluer la compréhension du sens de phrases. Nous renvoyons le lecteur à l'analyse que nous en avons fait et qui nous conduit à valider l'usage de cette épreuve.

2/ Ce même enseignant utilise ensuite 3 autres épreuves pour évaluer la compréhension de texte.

La première²¹⁸ est constituée d'un texte et d'un questionnaire sur le texte. Nous ne revenons pas sur les raisons qui nous conduisent à invalider son usage.

La deuxième épreuve²¹⁹ contient un texte et un questionnaire en deux parties. La première partie amène l'élève à répondre à 3 questions, texte masqué. Elles portent sur la situation initiale du récit. Puis, l'élève démasque le texte pour répondre aux 3 questions suivantes qui portent sur l'ensemble du texte. Cette épreuve nécessite des réponses écrites dont une partie est prise en charge par le système de réponse de l'épreuve, ce qui limite la tâche d'écriture des réponses. Les questions sont posées sous la forme de phrases à compléter.

La dernière question « ils se dépêchent ... » montre bien un phénomène de dérive méta textuelle que nous avons évoqué dans la partie III au chapitre 4. L'élève répond «... parce que le chien est à leur poursuite », une réponse validée par le correcteur. Or, le texte n'établit pas ce lien de causalité. Il indique simplement que « les enfants se mettent à courir car ils ont peur du chien qui aboie derrière la grille ». La réponse de l'élève est décalée sans que l'on puisse dire que le décalage est totalement imputable à l'élève. Nous pouvons distinguer une erreur provenant du modèle de situation que l'élève a construit pendant la lecture : le chien n'est pas à la poursuite des enfants, l'élève a sans doute extrapolé du fait que le chien aboie et que les enfants courent, le fait que le chien poursuive les enfants. Mais, nous pouvons aussi souligner que la réponse « parce que le chien est à leur poursuite » s'accorde mal avec « ils se dépêchent » : l'amorce de réponse est un élément apporté par le questionnaire.

Le modèle de situation construit par l'élève est insuffisamment questionné et la surinterprétation du texte qu'il évoque est considérée comme valide. Par ailleurs, les contraintes de rédaction de la réponse sont importantes comme on peut le voir sur l'extrait de la copie de l'élève reproduite ci-dessous :

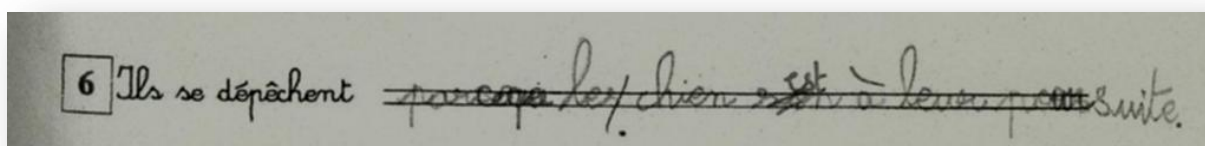


Figure 91 : Épreuve FT3 (Annexe C, p. 123) avec réponses d'élève

Ces différentes raisons nous conduisent à remettre en cause la validité de l'épreuve pour évaluer la compréhension d'un texte.

²¹⁷ Epreuve 92-2, Annexe C, p. 110.

²¹⁸ Epreuve 84, Annexe C, p. 122.

²¹⁹ Epreuve 86, Annexe C, p. 123.

L'enseignant I utilise l'épreuve d'exécution d'une série de consigne formant un petit texte commentée plus haut (cf. épreuve de compréhension de texte utilisée par l'enseignant E). Nous ne revenons pas sur les raisons qui nous conduisent à invalider l'usage de cette épreuve pour évaluer la compréhension d'un texte.

Les épreuves utilisées par l'enseignant J

1/ L'enseignant J utilise 3 épreuves pour évaluer la compréhension de phrases

La première²²⁰ demande à l'élève de numéroter une série de 6 phrases dans l'ordre dans lequel les lit l'enseignant. Pour répondre l'élève doit vérifier si les mots qu'il lit correspondent aux mots lus par l'enseignant. Cette épreuve ne fait pas véritablement appel à la compréhension de ce qui est lu.

La deuxième²²¹ utilisée est identique à la troisième épreuve utilisée par l'enseignant I. Nous ne revenons pas sur les raisons qui nous conduisent à la considérer comme étant valide.

La troisième épreuve²²² demande à l'élève former une phrase avec un ensemble de 8 mots présentés de façon aléatoire. Pour répondre, l'élève doit s'appuyer sur des connaissances syntaxiques (relatives à la majuscule, au point) pour extraire de la suite de mots le premier mot « L'oiseau » et le dernier mot « nid. ». Il doit ensuite articuler un verbe au premier mot : « cherche » doit être préféré à « faire » qui n'est pas conjugué. La phrase « L'oiseau cherche (son) nid. » peut convenir mais il est implicite que la phrase doit contenir « tous » les mots. L'élève doit donc poursuivre pour obtenir « L'oiseau cherche une branche pour faire son nid. » La phrase reconstituée doit être réécrite en écriture cursive dans un lignage de type Seyès. La compréhension de la phrase est évaluée au moyen de l'écriture de la phrase.

Ce type d'épreuve pose plusieurs questions : cette tâche de lecture est tout d'abord entièrement fabriquée pour l'exercice. Elle évalue surtout des connaissances syntaxiques implicites (trouver quel verbe convient une fois le premier mot (qui devient alors le sujet) est choisi). Deuxièmement, cette tâche évalue la compréhension au moyen de l'écriture (normée) de la phrase réponse. Des difficultés de copie ou d'écriture sont susceptibles d'être défavorables à l'évaluation de la compréhension.

Ces raisons nous conduisent à invalider l'usage de cette épreuve.

2/ Il utilise parallèlement 2 épreuves pour évaluer la compréhension de texte.

La première épreuve est identique à la troisième épreuve utilisée par l'enseignant I et examinée plus haut (cf. tâche d'exécution d'une suite de consignes format un petit texte). Nous ne revenons pas sur les raisons qui nous conduisent à écarter cette épreuve.

La deuxième épreuve²²³ est constituée d'un texte suivi de deux modalités de questionnement : un vrai/faux et une série de 6 « why questions »²²⁴. Nous la reproduisons ci-dessous :

²²⁰ Epreuve 93-4, Annexe C, p. 110.

²²¹ Epreuve 94-3, Annexe C, p. 111.

²²² Epreuve 95-1, Annexe C, p. 111.

²²³ Epreuve 95-2, Annexe C, p. 124.

²²⁴ Cf. Note 77 dans le chapitre 4.3 Partie III

/ Je comprends les informations explicites d'un texte. A AR ECA NA

❖ Lis ce texte tout seul silencieusement puis réponds aux questions.

PIPO, LE POISSON ROUGE

De son bocal, Pipo, le petit poisson rouge, regarde les moineaux noirs voler dans le ciel.
 « Moi aussi, j'aimerais voler » dit-il et il pleure.
 Heureusement, la nuit, il rêve qu'il est un poisson volant. Il s'élanche au-dessus des vagues et tout le monde l'admire. Pipo préfère la nuit que le jour.

❖ Réponds aux questions en entourant la bonne case.

Pipo n'est jamais triste.	<input checked="" type="checkbox"/> vrai	<input type="checkbox"/> faux
Pipo pleure parce qu'il ne sait pas voler.	<input checked="" type="checkbox"/> vrai	<input type="checkbox"/> faux
<u>Le jour</u> , Pipo rêve qu'il est un poisson volant.	<input checked="" type="checkbox"/> vrai	<input type="checkbox"/> faux
La nuit, Pipo rêve qu'il s'élanche au-dessus des vagues.	<input checked="" type="checkbox"/> vrai	<input type="checkbox"/> faux
Dans son rêve, personne n'admire Pipo.	<input checked="" type="checkbox"/> vrai	<input type="checkbox"/> faux
Pipo aime mieux le jour que la nuit.	<input checked="" type="checkbox"/> vrai	<input type="checkbox"/> faux
Pipo vole pour de vrai.	<input checked="" type="checkbox"/> vrai	<input type="checkbox"/> faux

❖ Réponds aux questions en écrivant sur la ligne.

Qui est Pipo ? Un petit poisson rouge

Où vit Pipo ? Dans une salle maison et un bocal

Comment est Pipo ? Il est joli

Que voudrait Pipo ? voler

Que fait Pipo ? Il nage

Figure 92 : Épreuve FT3 (Annexe C, p. 124) avec réponses d'élève

Nous avons par ailleurs explicité les raisons qui nous conduisent à invalider l'usage de questionnaires pour évaluer la compréhension d'un texte.

La 3^e question du second questionnaire « Comment est Pipo ? » conduit l'enseignant à valider la réponse de l'élève « Il est joli ». Aucun élément du texte ne permet de donner raison, ou tort, à l'élève. Ce type de question permet de souligner que le recours systématique aux « why questions » n'est pas toujours pertinent (on peut questionner également de ce point de vue la dernière question « Que fait Pipo ? ». Plusieurs réponses sont valides : certains reprennent les verbes d'action du texte (« il regarde les moineaux », « il se parle », « il pleure », « il rêve »), d'autres sont fondées sur le modèle de situation « il s'ennuie ». La réponse de l'élève « il nage » est partiellement valide : elle est fondée ses connaissances générales sur le monde (les poissons nagent) indépendamment du texte (cf. « passage independent item » Keenan & Betjemann, 2006).

L'épreuve utilisée par l'enseignant L

1/ Cet enseignant utilise l'épreuve reproduite ci-dessous pour évaluer la compréhension de phrases.

Lire des phrases dont les sons ont été étudiés et répondre à des questions de compréhension lues par le maître.	A	AR	EA	NA
---	---	----	----	----

Gonzague a soigné des enfants à Djibouti pendant sept ans.

Quand il sera grand, Damien veut voyager pour étudier les insectes.

Matthieu rêve de nager avec des dauphins dans une mer turquoise.

1- Qu'a fait Gonzague à Djibouti pendant sept ans ? /

2- Que veut faire Damien quand il sera grand ? /

3- Quel est le rêve de Matthieu ? /

Figure 93 : Épreuve FT3 (Annexe C, p. 112) avec réponses d'élève

Pour répondre aux questions lues par l'enseignant, l'élève doit lire des phrases isolées (sans liens entre elles). La relation entre questions et réponses est de type textuelle-explicite : les informations sont explicitement présentes dans le texte. Les processus de lecture mobilisés sont donc plutôt des microprocessus. Cette épreuve mobilise donc l'activité de lecture attendue.

2/ L'enseignant L utilise une épreuve pour évaluer la compréhension de texte. Nous renvoyons le lecteur au paragraphe III-3.1 Tâche n°6 pour l'exposé des raisons qui nous conduisent à invalider l'usage de cette épreuve.

L'épreuve utilisée par l'enseignant P

1/ Cet enseignant utilise une épreuve pour évaluer la compréhension de phrases (consignes) L'épreuve utilisée par cet enseignant pour évaluer la compréhension a été analysée en détail dans le chapitre 2 de la Partie III (cf. tâche n°1). Nous y renvoyons le lecteur et ne revenons pas sur les raisons qui nous conduisent à invalider son usage pour évaluer la compréhension en lecture.

2/ Il utilise trois autres épreuves pour évaluer la compréhension de texte.

La première²²⁵ est composée d'un texte documentaire sur la tortue suivi de deux questionnaires (un QCM et un vrai/faux). Nous reproduisons cette épreuve pour analyser l'effet de la formulation négative d'une affirmation à laquelle l'élève doit répondre Vrai ou Faux :

²²⁵ Epreuve 125 (Annexe C, p.132)

8--Lis ce texte.. *lisez*

Dès le matin, la tortue cherche sa nourriture. Elle mange de la salade, des carottes et des radis. Elle n'arrête pas de manger toute la journée.

La tortue est protégée par une carapace dure. Quand elle a peur, elle rentre sa tête et ses quatre pattes.

La femelle pond ses œufs dans un trou, puis elle les recouvre avec de la terre. Les bébés tortues sortent des œufs au bout de trois ou quatre mois.

9 --Entoure la bonne réponse.

- Quand elle a peur, la tortue rentre
 - ses bébés.
 - sa nourriture.
 - ses pattes. -
- La tortue cherche sa nourriture
 - dès le matin.
 - dès le soir.
 - toute la nuit. -
- ★ La tortue mange
 - des œufs.
 - de la terre.
 - des carottes. -
- La tortue pond ses œufs
 - dans un nid.
 - dans un trou.
 - dans l'eau X

10- Ecris vrai ou faux

La tortue dort toute la journée. ... *faux* -

La tortue ne mange pas de viande. ... *faux* X

La tortue ne pond qu'un œuf. *faux* -

Figure 94 : Épreuve FT3 (Annexe C, p. 132) avec réponses d'élève

L'élève répond « Faux » à l'affirmation « La tortue ne mange pas de viande » ce qui ne constitue pas la réponse attendue. Cette réponse n'est pas cohérente avec la réponse fournie à la question 3 dans le QCM (la réponse de l'élève indique que « la tortue mange des carottes »). Cette absence de cohérence peut conduire à s'interroger sur la fiabilité des réponses produites (celles considérées comme correctes, celles considérées comme incorrectes). Elle peut aussi conduire à s'interroger sur les raisons qui conduisent éventuellement l'élève à produire des réponses apparemment contradictoires. L'élève lit « la tortue ne mange pas de viande » et peut indiquer qu'« il est VRAI de dire que la tortue ne mange pas de viande » ou qu'« il est FAUX de dire que la tortue mange de la viande » sans prendre en compte la négation de l'énoncé.

La validité de cette épreuve est ainsi remise en cause.

L'enseignant P utilise un autre questionnaire composé de « why questions » portant sur un texte court²²⁶ similaire à la deuxième épreuve utilisée par l'enseignant I pour évaluer la compréhension de texte précédemment examinée. Nous ne revenons pas sur les raisons qui nous ont conduit à invalider son usage pour évaluer la compréhension d'un texte.

²²⁶ Epreuve 126-1, Annexe C, p. 133.

L'enseignant P utilise une dernière épreuve dans laquelle pour répondre l'élève doit dessiner « ce que raconte le texte ». Nous la reproduisons ci-dessous :

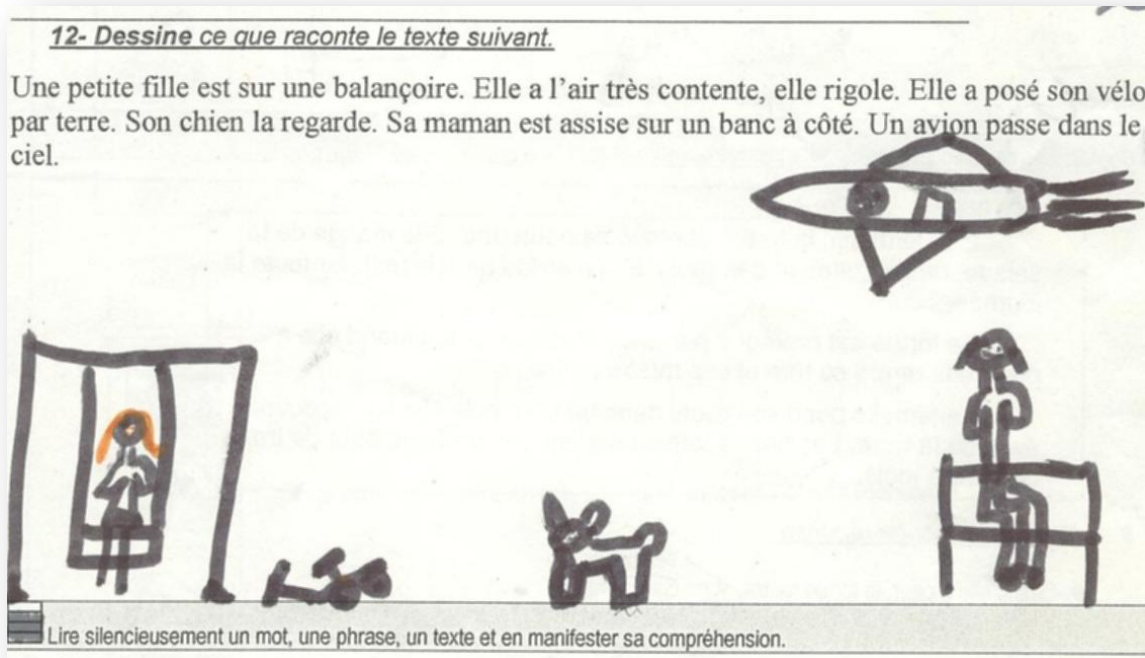


Figure 95 : Épreuve FT3 (Annexe C, p. 133) avec réponses d'élève

Pour répondre, l'élève doit illustrer « ce que raconte le texte » ce qui s'appuie sur la figuration successive des éléments, ou couples d'éléments, suivants : « fille »/ « balançoire » ; « vélo »/ « par terre » ; « chien » ; « maman »/ « assise sur un banc » ; « avion ». Ce texte décrit, plus qu'il ne raconte, une scène de la vie courante d'un enfant. En mobilisant des connaissances générales, on peut inférer l'histoire qui aurait pour cadre ce décor : la petite fille est venue avec sa maman et son chien au parc, elle se précipite à la balançoire etc.

La tâche prescrite ne permet pas à l'élève de faire état d'une représentation mentale élaborée au fur et à mesure de la lecture du texte. Le système de réponse occulte certaines parties du texte. La deuxième phrase du texte renvoie ainsi à un état émotionnel difficile à dessiner. On peut aussi considérer que les personnages dessinés par les jeunes élèves sourient le plus souvent ce qui rend difficile d'évaluer le dessin. Les trois dernières phrases contiennent des informations spatiales difficiles à dessiner : la maman est, de par le cadre de réponse, forcément « à côté » ; l'avion est naturellement « dans le ciel ».

Cette épreuve intéressante ne peut se suffire à elle-même pour évaluer la compréhension : il est nécessaire de recourir à l'interview de l'élève au sujet de ce qu'il a dessiné. Nous avons également souligné le fait que l'épreuve évalue davantage la compréhension d'une suite de phrases que d'un texte.

3.4.3. Synthèse

Nos analyses nous conduisent à souligner l'intérêt de l'usage d'épreuves évaluant la compréhension de phrases en fin de CP. Au cours de la réalisation des tâches de lecture qu'elles comportent, l'activité des élèves se fonde sur l'emploi de microprocessus de lecture dont nous rappelons que Giasson les définit comme « des processus de base en lecture, responsables de la compréhension de la phrase » (1990, p. 50). Ces tâches mobilisent également l'usage de processus d'intégration lorsque la tâche nécessite la compréhension d'indices de relation (inférences, anaphores) entre les éléments de deux phrases (ou d'une phrase et d'une illustration). Ces processus permettent d'effectuer des liens entre les propositions ou entre les phrases. L'ensemble de ces processus sont donc essentiels pour comprendre un texte et, à ce titre, l'utilisation d'épreuves portant explicitement sur leur mobilisation est bien-fondée d'un point de vue didactique. Les tests A, C, E, I, J et L contiennent chacun au moins une épreuve valide pour évaluer la compréhension au niveau de la phrase.

Toutefois, d'autres processus interviennent dans la compréhension d'un texte. Ils sont probablement sollicités au cours de la réalisation de la tâche de lecture d'un texte, mais les épreuves utilisées pour évaluer la compréhension (d'un texte) que nous avons examinées ne sont pas valides sur le plan psycho-didactique pour les évaluer. Le principal motif de remise en cause de leur validité tient à l'usage généralisé de questionnaires dont nous avons montré certaines limites dans la partie III (cf. chapitre 4) et dont nous avons détaillé d'autres limites dans ce chapitre : la représentation que l'élève élabore au cours de la lecture (le modèle de situation) n'est pas véritablement questionnée, le traitement du questionnaire engage une autre activité de lecture, la modalité de réponse mobilise d'autres compétences (rédaction, dessin...) implicitement évaluées. Selon le référentiel que nous avons utilisé, aucun test ne contient donc d'épreuves valides pour évaluer la compréhension d'un texte.

L'usage d'épreuves de lecture de consignes dont on mesure la compréhension par leur exécution, principalement sous forme de dessins, constitue un fait évaluatif intéressant à souligner. Il renvoie à la prise en compte par les enseignants du caractère autocentré de la lecture à la fois objet et outil de son propre enseignement, dès le début de l'apprentissage. La lecture de consignes est en effet au cœur de l'activité quotidienne de l'élève, et elle est implicitement évaluée lors de la réalisation de toutes les tâches scolaires. Nous avons souligné le rôle de guide des items des progressions : la lecture de consignes ne fait l'objet d'aucun item. Toutefois, la compréhension autonome de consignes est un composant de la compétence 1 (Maîtrise de la langue française) du Socle commun dont la maîtrise est attendue en fin de CE1 au palier 1²²⁷.

²²⁷ « Lire seul et comprendre un énoncé, une consigne simples » (BO 2008, p. 20)

Synthèse de la partie IV

A l'issue de l'analyse globale et locale des tests et des épreuves d'évaluation conduite dans cette partie, et selon les référentiels utilisés, nous constatons que, bien que contenant des épreuves valides du point de vue psycho-didactique, aucun test du 1^{er} trimestre n'est valide du point de vue de la couverture didactique concernant l'évaluation des relations phonographiques en lecture et concernant les modes de désignation des nombres, les relations entre les nombres, le calcul et la résolution de problèmes additifs dans le domaine nombres et du calcul. Au 3^e trimestre, le même constat peut être dressé concernant l'évaluation de composants occasionnant des tâches souvent dénommées « complexes » (compréhension en lecture, résolution de problèmes additifs en mathématiques).

Ce défaut de validité a plusieurs origines : soit le test n'évalue pas l'un des principaux composants du référentiel didactique que nous avons explicité au moyen d'au moins une épreuve, soit les épreuves présentent un défaut de validité d'ordre psycho-didactique.

Nous revenons dans la conclusion de notre travail sur ces résultats et sur la ou les méthodologie(s) utilisée(s). Comment interpréter l'absence de validité de l'évaluation globale et locale des tests et des épreuves ?

Conclusion

Nous avons inscrit notre travail dans une conceptualisation générale de l'évaluation dans laquelle l'évaluation est fondamentalement une opération de comparaison de deux construits : un référé et un référent. Il en découle la place centrale accordée dans la thèse à la détermination de ces construits, et plus particulièrement à l'explicitation d'une partie du référent, le référentiel (cf. I.2), qui en constitue la part la plus explicite.

L'examen de la qualité de la comparaison référentiel/référé constituait l'objet de notre étude. La qualité de l'évaluation à laquelle nous nous sommes intéressé est assimilable à la validité. Notre travail poursuivait le but principal de fournir un éclairage de nature didactique sur la validité de l'évaluation conduite de façon régulière par une population de 16 enseignants de CP dans deux sous-domaines à fort enjeu socio-scolaire : la lecture en Français, et les nombres et le calcul en Mathématiques. Trois dimensions de la validité ont été dégagées (cf. I.3). Elles correspondent chacune à des finalités d'analyses différentes, non exclusives. La conformité de l'évaluation réalisée aux programmes, la qualité psychodidactique de chaque épreuve et la qualité de couverture didactique des tests constituent ainsi trois manières de considérer la validité de l'évaluation. Selon la dimension de la validité étudiée, le (ou les) référé(s) et le (ou les) référentiel(s) construits doivent être précisés. Les trois dimensions de la validité déterminées ont orienté notre étude, et ont constitué les trois axes de la problématique de la thèse.

Le premier axe de notre problématique concernait la validité curriculaire de l'évaluation réalisée par chaque enseignant, c'est-à-dire l'appréciation du degré de conformité de l'évaluation aux programmes (cf. II). Les progressions annexées aux programmes constituent le référentiel formel auquel tous les enseignants se rapportent pour élaborer leur propre référentiel pour évaluer. Ces progressions constituent notre propre référentiel pour analyser les données collectées. Pour construire les référés, nous nous sommes intéressé aux référentiels pour évaluer élaborés par les enseignants. Le référentiel pour évaluer de l'enseignant est élaboré au cours d'un processus de référentialisation, propre à chacun. L'aboutissement de ce processus se matérialise par la production, ou le choix, d'un livret scolaire. Nous avons mis en évidence le rôle central du livret scolaire dans l'évaluation : rédigé sous la forme d'une liste de compétences établie en référence aux progressions annexées aux programmes, le livret constitue le référentiel formel de chaque enseignant pour évaluer. Une autre partie du référé est constitué des épreuves affectées par chaque enseignant aux items du livret qu'il utilise. Dès lors, nous avons pu comparer le référentiel (i.e. les progressions) et les référés construits (i.e. les livrets et les épreuves utilisés par chaque enseignant), et ainsi étudier la couverture des programmes réalisée par l'évaluation mise en œuvre par chaque enseignant. Les résultats des analyses renseignent sur la conformité de l'évaluation aux programmes. Ils soulignent certaines tendances qui montrent

l'effet des choix effectués par les enseignants au niveau de la mise en œuvre de l'évaluation. Ils montrent surtout l'importance de prendre en compte cette partie de l'activité évaluative des enseignants, complexe et entièrement déléguée aux enseignants par l'institution scolaire.

Le deuxième axe constituait l'élément central de notre travail. Le point de vue adopté sur l'évaluation réfère aux approches psycho-didactiques de l'évaluation (APDE). Dans le cadre des APDE, on s'intéresse au potentiel de chaque épreuve à engager les élèves dans l'activité de réponse souhaitée par l'évaluateur, c'est-à-dire à mobiliser les compétences évaluées. Nous nous sommes interrogé sur la validité psycho-didactique des épreuves (cf. III). Ce type d'analyse nécessite d'explicitier des référentiels d'ordre didactique en fonction des domaines étudiés, et de prendre en considération des composants de l'épreuve qui auraient pu échapper à l'analyse proprement conceptuelle de la tâche. Ils pourraient en effet s'avérer être des déterminants majeurs du comportement de réponse des élèves. Pour cela, nous avons élaboré une méthodologie d'analyse qualifiée de « psycho-didactique » : elle prend en compte les aspects conceptuels de la tâche, mais aussi d'autres composants des épreuves que nous avons été conduit à isoler et qui sont susceptibles de conduire à l'invalidation de l'usage des épreuves.

Le troisième axe de la thèse intégrait l'étude d'épreuves du point de vue psycho-didactique, un point de vue local sur l'évaluation, à une étude globale du test en termes de couverture des principaux composants des domaines disciplinaires étudiés (cf. IV). Un travail de référentialisation a été nécessaire pour expliciter les référentiels didactiques permettant d'analyser les tests collectés dans les domaines de la lecture, des nombres et du calcul. Pour chacun de ces domaines, nous avons été amené à en déterminer les « principaux composants ».

Le croisement de ces trois points de vue sur la validité de l'évaluation permet de considérer l'ampleur de la tâche dévolue à l'enseignant lorsqu'il doit évaluer les acquis scolaires de ses élèves et en rendre compte, ce qu'O. Rey (2012) considère, à juste titre, comme un véritable « défi ».

Nous revenons sur le rôle central de la construction des référentiels dans notre travail, sur les résultats produits concernant les différentes formes de validité étudiées. Nous présentons les limites de notre démarche en lien avec les aspects pouvant être développés et les perspectives de recherche et de formation des enseignants qui en découlent.

1. La construction de référentiels : une étape clé

Quelle que soit la forme examinée, la validité de l'évaluation est une qualité dépendante des référentiels construits pour l'apprécier. L'explicitation de ces derniers constitue donc un enjeu central pour l'enseignant, pour évaluer, et, pour le chercheur, pour analyser l'évaluation réalisée par l'enseignant.

Nous avons souligné (cf. II.1) que les différentes prescriptions adressées aux enseignants s'accordent de façon imparfaite. En conséquence, les référentiels pour évaluer élaborés par les enseignants (les livrets) sont hétérogènes : certains livrets consignent ce qui a été enseigné par l'enseignant sur la période considérée sans qu'il soit aisé de pouvoir les comparer aux programmes, alors que d'autres contiennent l'ensemble des items des

progressions à chaque trimestre. Nous avons constaté que les référentiels construits par les enseignants ont néanmoins en commun une grande proximité avec les progressions annexées aux programmes. Ces progressions, présentées par l'institution scolaire comme de simples « repères pour organiser la progressivité des apprentissages », sont l'élément le plus opérationnel des programmes pour évaluer. Nous considérons que cela constitue précisément l'origine de la diversité des référentiels pour évaluer des enseignants.

Les progressions constituent en réalité le référentiel formel « alpha » à partir duquel les enseignants élaborent leur propre référentiel pour évaluer. L'analyse des livrets, et de certains propos recueillis pendant les entretiens post-collecte, nous a conduit à souligner la difficulté que représente l'opérationnalisation d'une commande institutionnelle très générale en matière d'évaluation régulière des compétences. Nos observations nous conduisent toutefois à indiquer que les enseignants se conforment tous au cadre d'une évaluation des compétences des élèves telle qu'elle est souhaitée par l'institution scolaire : le savoir-lire, par exemple, est décliné en plusieurs compétences, ou items, dans les livrets.

L'un des effets bénéfique du processus de référentialisation entrepris par chaque enseignant à partir du référentiel formel des progressions est l'ajout de compétences non explicites dans les progressions. Elles sont considérées comme étant importantes par les enseignants, et sont ajoutées à leur propre référentiel pour évaluer : il en est ainsi du dénombrement en mathématiques et de la conscience phonémique en lecture. Ce résultat nous semble très important : il dénote la diffusion de certains résultats provenant de la recherche auprès des enseignants. Précisément, la recherche en didactique de la lecture a souligné le rôle essentiel de la conscience phonémique dans l'apprentissage de la lecture. Elle a plaidé pour un enseignement explicite de celle-ci en fin de grande section de maternelle et en début de CP. L'importance des compositions-décompositions des nombres et du dénombrement d'une collection d'objets sont des aspects mis en lumière par la recherche en didactique des mathématiques. Cet ensemble de résultats et/ou de préconisations apparaît pris en compte par les enseignants de la population étudiée.

Pour analyser les épreuves et les tests (dans notre travail, nous rappelons qu'un test est assimilé à l'ensemble des épreuves proposées par un enseignant pour évaluer les compétences d'un domaine disciplinaire), nous avons été conduit à construire une méthodologie reposant sur une conjonction de cadres issus de champs de recherches scientifiques en psychologie (ergonomique, cognitive) et en didactique (des mathématiques, du français).

L'élaboration de la méthodologie d'analyse des épreuves a nécessité de clarifier différents niveaux de la tâche (cf. III.1 et 2) afin de pouvoir permettre l'analyse d'éventuels écarts entre ce que l'enseignant prescrit à l'élève de faire et la tâche que réalise effectivement l'élève. Elle permet ainsi de prendre en compte ce que l'enseignant dit évaluer (et qui est formulé par la compétence dans l'item du livret) au moyen de l'épreuve choisie. L'application de cette méthodologie d'analyse des épreuves suppose que l'on fasse des hypothèses sur le fonctionnement cognitif des élèves au cours de la réalisation de la tâche : c'est ainsi qu'elle permettra d'apprécier finement l'écart potentiel entre la tâche prescrite (à l'élève) et la tâche réalisée (par l'élève). Cela nous a conduit à souligner que l'échec, ou la réussite, à une épreuve peuvent s'expliquer par l'effet de certains facteurs insuffisamment pris en compte par l'évaluateur et susceptibles d'intervenir de façon importante dans l'activité de l'élève.

Pour analyser les tests, nous avons combiné la méthodologie d'analyse des épreuves précédemment décrite à l'utilisation de référentiels d'ordre didactique relativement aux

domaines disciplinaires considérés. L'élaboration de ces référentiels visait la détermination des « principaux composants » des domaines de la lecture, des nombres et du calcul, au niveau du CP. La comparaison référentiel/référentiel a ainsi porté sur la comparaison des tests et des épreuves qui les constituent, aux référentiels élaborés. Elle visait à établir la validité globale, et locale, des tests, et des épreuves. La qualité recherchée renvoie à la couverture de l'ensemble des principaux composants des domaines, telle qu'elle est réalisée par les tests.

2. Différentes formes de validité de l'évaluation

Nous revenons à présent sur les principaux résultats obtenus suite à l'analyse de chacune des formes de validité de l'évaluation auxquelles nous nous sommes intéressés.

Validité curriculaire

Les référentiels pour évaluer des enseignants résultent de la mise en œuvre de processus individuels de référentialisation conduits à partir des progressions. Au niveau de chaque enseignant, cela a pour effet l'adoption d'une liste personnelle importante d'items, non hiérarchisés. La plupart des enseignants (10 enseignants sur 16) utilisent un livret très proche des progressions (cf. II.2) dans lequel tous les items sont listés à chacun des trimestres considérés. Il est alors possible d'étudier la validité curriculaire de l'évaluation réalisée par ces enseignants. Dans ce cas de figure, nous avons constaté que la couverture des programmes n'est généralement pas assurée, quel que soit le trimestre et le domaine considéré. La validité curriculaire de l'évaluation n'est ainsi pas avérée.

Parmi les principaux résultats obtenus, nous avons montré que certains sous-domaines des domaines disciplinaires du Français et des Mathématiques sont sous-représentés (l'écriture par exemple) quand d'autres, concernant des connaissances formelles, occupent très tôt une place importante (grammaire, orthographe, conjugaison). Ce résultat n'est pas nouveau, il a déjà été décrit par Grisay (1989, 1992). Son actualité est toutefois intéressante à discuter : il peut résulter d'une forme de pratique issue de la tradition qu'E. Fijalkow qualifie ainsi de « traditionnelle » (2003). Il peut s'illustrer par la structure même des progressions qui accordent une grande importance aux sous-domaines « vocabulaire, grammaire, orthographe », lesquels contiennent presque autant de compétences que les sous-domaines « langage oral, lecture, écriture » (18 compétences contre 22). Les progressions ne se posent ainsi pas en rupture avec l'approche traditionnelle, et, pour reprendre la distinction entre habileté et compétence faite par Perrenoud (cf. I. 2.1.1), il est plus aisé d'évaluer des habiletés (par exemple en orthographe « commencer à utiliser correctement la majuscule ») que des compétences « complexes » (par exemple « dire de qui ou de quoi parle le texte lu »). On peut ainsi estimer que le référentiel formel constitué par les progressions encourage une évaluation d'habiletés plutôt que de compétences.

Nous avons également montré que, quel que soit le trimestre considéré, aucun des tests recueillis ne couvre l'ensemble des domaines étudiés (le domaine du Français et le domaine des Mathématiques). L'évaluation est donc lacunaire. Deux sous-domaines sont toutefois particulièrement évalués : la lecture ainsi que les nombres et le calcul. Mais, à l'intérieur de ces sous-domaines, l'évaluation des différents items des progressions n'est pas homogène. L'évaluation des compétences les plus expertes n'est pas homogène parce que celles-ci sont davantage évaluées en français (comprendre un texte) qu'en mathématiques (résoudre

un problème) en fin d'année. L'évaluation des habiletés de base n'est pas homogène non plus parce que ces dernières ne sont pas évaluées par tous les enseignants en début d'année. La résolution de problèmes en mathématiques est, en un sens, une compétence sous-évaluée. Cela peut s'expliquer par une diminution, entre les programmes de 2002 et ceux de 2008, de l'importance accordée à cette compétence dans les programmes au profit du calcul. La formulation des priorités du cycle 2 ne mentionne d'ailleurs pas explicitement en 2008 la résolution de problèmes : « l'apprentissage de la lecture, de l'écriture et de la langue française, la connaissance et la compréhension des nombres, de leur écriture chiffrée et le calcul sur de petites quantités constituent les objectifs prioritaire du CP et du CE1 » (MEN - BO, 2008, p.18). Les progressions ne mentionnent pas de compétences faisant explicitement référence à la « conscience phonologique », ni au « dénombrement ». Certains enseignants n'évacuent pourtant pas ces composants des domaines comme nous l'avons souligné infra, mais on peut estimer que l'absence de visibilité de ces compétences masque leur importance.

Nous avons relevé que le pilotage de l'évaluation réalisé par les items des progressions conduisait certains enseignants à évaluer chacun des items du livret qu'ils utilisent au moyen d'une, et parfois une seule, épreuve, sans les hiérarchiser. Nous interprétons ce fait comme un effet de la fonction de « réponse à la commande de l'institution » attribuée par certains enseignants à l'évaluation. L'enseignant D évalue ainsi les compétences de ses élèves dans les domaines de la lecture, des nombres et du calcul (cf. tableaux n°12 et n°14) selon une modalité typique : chaque item est évalué par une (et une seule) épreuve. Notons que, quand bien même chaque épreuve serait valide, l'évaluation d'une compétence au moyen d'une seule épreuve pose problème. Prenons pour exemple la compréhension en lecture qui n'est évaluée que sur la base d'un seul texte par plusieurs enseignants : des questions se posent relativement aux caractéristiques du savoir-lire engagé par le texte (savoir lire un récit et savoir lire un documentaire n'ont par exemple pas tout à fait le même statut) ainsi qu'à la nature du questionnement sur le texte permettant d'en évaluer la compréhension. Comme il n'existe pas de « familles » de textes, ni de « familles » de questionnement de textes, le savoir-lire évalué est dépendant du degré d'exigence de chaque enseignant, ce dont nous avons de multiples exemples dans les tests collectés. Ainsi, au niveau du choix des items à évaluer, mais aussi du nombre d'épreuves se rapportant à un item, l'évaluation interne se caractérise par un rapport de dépendance étroit à l'enseignant qui la met en œuvre.

Validité psycho-didactique des épreuves

Nous nous sommes ensuite intéressé aux épreuves choisies ou conçues par chaque enseignant pour évaluer les compétences du livret qu'il utilise.

Comme nous l'avons précédemment rappelé, l'analyse *a priori* des épreuves, menée à partir de référentiels didactiques, joue un rôle central sans pour autant suffire pour étudier leur validité au sens de l'approche psycho-didactique de l'évaluation que nous avons adoptée. D'autres composants de la tâche sont pris en considération du point de vue de leur rôle potentiel dans le processus de réponse de l'élève (la prise en compte d'éléments provenant des réponses d'élèves que nous avons pu recueillir nous a parfois permis d'affiner les hypothèses produites). Cet ensemble d'éléments est alors examiné en référence à la formulation de la compétence (l'item du livret) que l'enseignant souhaite évaluer au moyen de l'épreuve.

Nous avons pu montrer qu'une analyse fine des tâches impliquées dans les épreuves renvoyait à des dimensions manifestement peu prises en compte par les enseignants, et qui avaient pour effet potentiel la complexification des tâches sur lesquelles l'élève exerce son activité. Ce phénomène peut être analysé en termes de défaut de validité de l'évaluation, mais aussi sous l'angle de l'accroissement des inégalités scolaires (cf. Rochex et Crinon, 2011, p.13). Dans la majorité des cas, les tests et les épreuves d'évaluation sont susceptibles d'évaluer d'autres composants que ceux qu'ils prétendent explicitement évaluer. Nous avons souligné à plusieurs reprises que, de ce fait, dès le CP, l'évaluation se déroule en défaveur des élèves les moins experts en lecture/écriture, ce qui va dans le même sens que les résultats de Mercier, Bodin et Quilio (2015) : ces auteurs relèvent que « certaines enquêtes internationales montrent la faible réussite de ces (des) élèves sur l'ensemble des tâches demandant un travail écrit » (p.33). Dans le cas de notre travail, ces différences pourraient être liées à la modalité principale de passation de l'évaluation : collective, simultanée et homogène (tous les élèves sont confrontés à la même épreuve en même temps) et, surtout, à réponse écrite dans la très grande majorité des cas. Cela permet de nuancer la portée des propos de Vergnaud (2006, p.167) pour qui « ce sont les contenus de connaissances qui déterminent le plus directement la réussite ou l'échec ».

Dans une épreuve d'évaluation, d'autres facteurs sont susceptibles d'intervenir, non directement liés aux contenus visés. Notre travail a permis d'en isoler certains, ce qui pourrait permettre de limiter certaines causes d'échec imputables à la conception et à la passation de l'évaluation elle-même :

- Le dessin qui figure dans l'épreuve constitue l'un de ces facteurs. Nous avons relevé qu'il est susceptible d'intervenir au niveau du modèle de situation que l'élève construit au cours de la lecture du texte ou de l'énoncé du problème. Il peut modifier les procédures de réponse en mathématiques : par exemple, en conduisant l'élève à utiliser une stratégie de dénombrement alors qu'une stratégie de calcul est attendue (cf. III, chapitre 3.1).
- Nous avons relevé, dès le CP, un usage quasi-unanime de questionnaires écrits exigeant une réponse écrite pour évaluer la compréhension en lecture : un résultat déjà décrit par E. Fijalkow (2003, p.85). Si l'usage des questionnaires est souvent remis en cause, ils sont unanimement utilisés pour évaluer la compréhension en lecture par les enseignants de la population étudiée. Pourtant, ils n'engagent pas le questionnement du modèle de situation élaboré par l'élève au cours de sa lecture. Nous avons insisté sur le fait que ces questionnaires exigent par ailleurs la compréhension des questions elles-mêmes : le questionnaire est un texte sur le texte, doté de caractéristiques propres (en termes d'identification des mots, de réalisation d'inférences...) qui ne semblent pas suffisamment prises en compte par les enseignants. Ils requièrent en outre des compétences rédactionnelles pour formuler et produire les réponses.
- L'usage des QCM, plus rare, limite certains phénomènes mais pose d'autres questions, notamment au sujet de la nature même de l'acte de lire. Il pourrait être assimilé par ces jeunes élèves à « lire c'est trouver dans le QCM la réponse attendue par l'enseignant ». Cette conception se retrouve chez les plus faibles lecteurs comme l'ont souligné Cèbe, Goigoux & Thomazet (2004) au sujet d'élèves de SEGPA (cf. III.4.2.2.). Elle pourrait générer des stratégies de réponse aux QCM. Ainsi, de manière générale, répondre correctement aux questions d'un questionnaire,

qu'il s'agisse d'un questionnaire écrit à réponse écrite ou d'un QCM, ne peut être *a priori* considéré comme synonyme d'avoir bien compris le texte.

- Le « système de réponse » imposé ou suggéré par la tâche mobilise comme nous l'avons souligné les compétences rédactionnelles des élèves. Pour alléger l'activité de réponse de l'élève, certaines épreuves proposent ainsi à l'élève de compléter une phrase réponse lacunaire. Dans de telles épreuves en mathématiques, l'élève peut ainsi être conduit à reporter le résultat d'un calcul dans une phrase. Dans ce cas, rien dans la réponse produite ne permet de s'assurer que l'élève a correctement interprété le résultat et répond bien à la question posée dans le problème.
- Dans la majeure partie des cas, l'évaluation se déroule de façon massée dans le temps (sur une quinzaine de jours). Cela occasionne une charge de travail importante pour les enseignants qui se trouvent confrontés à la nécessité de concevoir des épreuves. Ils ont ainsi souvent recours à des tâches empruntées à un vaste ensemble de ressources disponibles pour lesquelles se posent aussi le problème de la validité, y compris quand ces épreuves proviennent d'opérateurs externes ou sont abritées sur des sites institutionnels. De telles tâches devraient être rapportées au contexte d'utilisation prévu par les concepteurs : une tâche peut en effet être valide pour un certain usage et invalide pour un autre. La validité d'une épreuve n'est pas une qualité endogène : une épreuve peut être valide pour évaluer un item et ne pas l'être pour évaluer ce même item dans un autre contexte ou un item que l'on pourrait estimer comme étant « proche ».
- A de nombreuses reprises, nous avons pu constater que l'élève pouvait produire une bonne réponse sur la base d'un raisonnement erroné : un résultat décrit par Bastien et Bastien-Tonniazzo (2004) qui signifierait que, pour qui s'en tiendrait à la seule considération de la réponse, la tâche ne permet pas d'évaluer ce pourquoi elle est utilisée.

Le résultat majeur des analyses conduites est la remise en cause d'un nombre important d'épreuves du point de vue de leur validité psycho-didactique du fait de l'action potentielle de différents facteurs au moment de la redéfinition et de l'effectuation de la tâche par l'élève.

Validité de couverture didactique des tests

Suite à la mise en évidence des facteurs précités, susceptibles d'invalider l'usage de certaines épreuves, nous avons conduit une analyse combinant une approche globale des tests et une approche locale de la validité des épreuves. Pour cela, nous nous sommes tout d'abord intéressé à la couverture des « principaux composants » des domaines considérés réalisée par chaque test. Cette analyse ne prenant pas en compte des dimensions de l'activité de l'élève, nous l'avons ensuite complétée par une analyse psycho-didactique de la validité de chacune des épreuves composant les tests dont nous avons précédemment estimé qu'ils couvraient les domaines étudiés. Nous recherchions ainsi à déterminer les tests couvrant les domaines de la lecture, des nombres et du calcul au moyen d'épreuves valides du point de vue psycho-didactique.

La double analyse réalisée - globale des tests, et locale des épreuves - nous a conduit, selon les référentiels utilisés, à remettre en cause la validité de couverture didactique des tests.

Deux raisons distinctes peuvent être avancées pour comprendre ce résultat. La première raison tient au fait que certains tests collectés n'évaluent pas tous les principaux composants des sous-domaines disciplinaires de la lecture, des nombres et du calcul selon les

référentiels élaborés. Sans présumer de la validité des épreuves, ces tests ne peuvent ainsi, par définition, être considérés comme valides du point de vue de la couverture des composants du domaine. La deuxième raison est que certains tests, bien que contenant des épreuves portant sur chacun des différents composants des domaines examinés, comportent des épreuves dont la validité n'est pas avérée du point de vue psychodidactique. De ce fait, certains composants ne sont plus couverts par des épreuves valides et, dans ce cas, le test ne peut être considéré comme valide du point de vue de la couverture didactique des domaines considérés.

3. Limites et ouvertures

L'évaluation interne analysée dans ce travail ne représente que la partie explicite de l'évaluation à laquelle un nombre restreint d'enseignants nous a permis d'avoir accès. Cette recherche laisse donc de côté un vaste pan de l'évaluation réalisée au cours de la conduite quotidienne de la classe. Ce pan d'évaluation, bien souvent implicite, entre pour une partie non négligeable dans l'évaluation. L'enseignant K en fournit l'exemple, en lecture, au troisième trimestre (cf. tableau 32, paragraphe IV-3.4.1) : cet enseignant renseigne ainsi les items du livret qu'il utilise sans avoir pu nous fournir d'épreuves d'évaluation qu'il aurait utilisés. Cette forme d'évaluation, difficilement observable au moyen de données, pourrait être qualifiée d'« intuitive » à l'image de l'indicateur intuitif « LAT » formalisé par Florin (1991) (indicateur du niveau de langage (L), du niveau d'attention (A) et de la capacité à suivre le rythme de travail (T)) qu'Ecalte et Magnan (2002, p.111) soupçonnent de permettre aux enseignants de Maternelle de prédire efficacement la réussite scolaire ultérieure dans les trois domaines relatifs à l'apprentissage de la langue écrite, l'identification et la production de mots et la compréhension en lecture. Cette limite permet d'entrevoir un élargissement de notre travail à l'analyse du référent de l'enseignant. Il nous conduit à formuler de nouvelles questions à explorer : comment s'articulent les représentations de l'enseignant au sujet de l'activité de réponse des élèves, au référentiel formel qu'il a élaboré en amont des passations ? Fait-il évoluer le référentiel formel suite aux passations et, le cas échéant, de quelle manière et pour quel effet ?

Dans certains cas, les enseignants n'ont pas élaboré eux-mêmes les épreuves d'évaluation. Une autre limite de notre étude provient des sources des épreuves utilisées par ces enseignants pour évaluer. Nous avons souligné la multiplicité de leurs provenances (manuels, sites,...) et/ou des sources qui les ont inspirées, en particulier les évaluations externes. Dans ces cas, la cohérence et la validité des épreuves empruntées devrait être analysée, ainsi que leur effet modélisant sur les pratiques des enseignants (par exemple l'emploi de QCM). C'est ce que nous avons fait notamment (cf. C. Blanc et Goasdoué, 2016) dans une recherche relative à l'analyse d'évaluations standardisées en lecture en fin de CM2, dans laquelle nous remettons en cause le rôle des questionnaires de compréhension en lecture.

Dans d'autres cas où les enseignants élaborent eux-mêmes les épreuves, la question de la validité se pose de la même manière. Elle invite à questionner la formation à l'évaluation reçue par les enseignants : pour les enseignants formant la population étudiée, celle-ci est minimale sinon nulle. Parallèlement à notre travail, nous avons construit une ingénierie évaluative courte destinée à concevoir, en formation continue, un dispositif d'évaluation du savoir-lire en fin de CP (cf. C. Blanc, 2013, 2014 ; C. Blanc et Vantourout, 2015 ; Vantourout et C. Blanc, soumis ; Vantourout et Maury, accepté). Les exigences, en terme de niveau

« d'expertise », d'une approche psycho-didactique telle que nous l'avons développée, l'investissement en temps que sa mise en œuvre nécessite, nous conduisent à souligner que cette tâche ne peut être entièrement dévolue aux enseignants. Le manque de validité didactique et psycho-didactique de l'évaluation interne analysée dans ce travail doit en effet être considéré en prenant en compte le haut degré d'exigence que requièrent de telles approches. Les travaux de Grugeon-Allys (2012, 2015, 2016) concernant l'évaluation des compétences des élèves en algèbre permettent de souligner à quel point construire une évaluation valide sur le plan global est complexe, même pour des experts, dans un domaine pour lequel une référence épistémologique est pourtant accessible. Il paraîtrait donc important que les didactiques produisent des référentiels et des outils (par exemple des tâches de référence) permettant de penser l'évaluation en termes d'interaction entre la compétence, la tâche et l'activité et que, de son côté, l'institution scolaire prenne en considération ces résultats et s'assure de leur transfert aux enseignants, en particulier à travers une formation ambitieuse aux didactiques disciplinaires et à l'évaluation.

Sommaire détaillé

Remerciements.....	2
Table des matières	3
Partie I- L'évaluation interne : introduction et question de recherche	7
Chapitre 1. Finalité(s) de l'évaluation	9
Chapitre 2. Le référent et le référé de l'évaluation.....	13
2.1. Référent, référentiel, référentialisation	13
2.1.1. Les programmes : un référentiel de compétences.....	16
2.1.2. Les livrets scolaires : des produits de processus de référentialisation.....	21
2.2. Référé	22
2.3. Référentiel et référé pour le chercheur	23
Chapitre 3. La validité : une qualité plurielle de l'évaluation	24
3.1. Validité de contenu des tests.....	25
3.1.1. Validité curriculaire.....	25
3.1.2. Validité de couverture (didactique) des domaines	26
3.2. Validité psycho-didactique des épreuves.....	28
3.3. Validité globale et locale	29
Chapitre 4. Synthèse et formulation de la problématique	30
Chapitre 5. Présentation du corpus de données recueillies.....	32

Partie II – Étude de la validité curriculaire de l'évaluation.....	36
Chapitre 1. Les progressions annexées aux programmes, référentiel des enseignants	38
1.1. Programmes et progressions, un lien ambiguë	38
1.2. Les items « lecture » des progressions.....	39
1.3. Les items du sous-domaine des nombres et du calcul des progressions.....	41
Chapitre 2. Première approche de la validité curriculaire	43
2.1. Les livrets utilisés par les enseignants.....	43
2.1.1. Le livret SCEREN	43
2.1.2. Autres livrets	49
2.2. Le taux de remplissage des livrets.....	52
2.3. Discussion	54
Chapitre 3. Analyse de la validité curriculaire.....	56
3.1. Couverture des domaines des programmes réalisée par l'évaluation	56
3.1.1. Au premier trimestre	56
3.1.2. Au troisième trimestre	59
3.1.3. Discussion	61
3.2. Couverture du sous-domaine « lecture » du domaine du Français	63
3.2.1. Au premier trimestre	63
3.2.2. Au troisième trimestre	64
3.3. Couverture du sous-domaine des nombres et du calcul du domaine des Mathématiques.....	65
3.3.1. Au premier trimestre	65
3.3.2. Au troisième trimestre	66
Synthèse de la partie II	67
Partie III – Étude de la validité psycho-didactique des épreuves.....	69
Chapitre 1. Cadre conceptuel d'analyse de la tâche	69
1.1. La tâche, première approche.....	69
1.2. A quelle tâche répond l'activité ?	70
1.3. Adaptation du modèle pour notre étude.....	72

1.4. Vers un cadre méthodologique pour analyser les épreuves d'évaluation.....	74
Chapitre 2. Méthodologie d'analyse psycho-didactique d'une épreuve d'évaluation	75
2.1. Méthodologie.....	75
2.2. Application à deux épreuves de référence.....	76
2.3. Discussion :.....	81
Chapitre 3. Le dessin : un simple « habillage » de la tâche ?.....	82
3.1. Le dessin dans des tâches d'évaluation de compétences complexes.....	83
Discussion :	91
3.2. Le dessin dans des tâches visant à évaluer des habiletés « de base » en lecture..	92
Discussion :	99
3.3. Le dessin dans des tâches de dénombrement.....	100
3.3-1. L'item relatif à la désignation des nombres en mathématiques.	100
a- l'item des programmes.....	100
b- l'item dans les livrets	100
3.3-2. Épreuves de dénombrement.....	103
Discussion	108
3.4. Synthèse du chapitre 3.....	110
Chapitre 4. Les questions écrites de compréhension en lecture	111
4.1. L'item relatif à la compréhension en lecture.....	112
a- l'item des programmes relatif à la compréhension.	112
b- l'item dans les livrets.....	114
Formulations de l'item dans les livrets :	114
Analyse des formulations :.....	114
4.2. Les questionnaires pour évaluer la compréhension en lecture :	117
4.2.1. Le questionnaire « Caroline »	117
4.2.2. Processus engagés par les questionnaires.....	123
1. Processus attentionnels :	123
2. Processus de contrôle de l'activité et stratégie de lecture.....	124
3. Processus de lecture.....	125
Résultats :.....	126
Analyse des résultats :	127
4.3. Synthèse du chapitre 4.....	129

Chapitre 5. Le système de réponse	131
5.1. Rédiger une (phrase) réponse complète.....	132
5.2. Trouver la bonne réponse parmi plusieurs propositions de réponse	136
5.2a Travaux de références en matière de QCM	136
5.2b Analyses de QCM en Lecture	138
5.2b-1 Proximité graphique ou graphophonologique des solutions.....	138
5.2b-2 Proximité thématique des solutions	141
5.2b-3. Proximité sémantique des solutions	143
Discussion :	149
5.3. Trouver la bonne réponse et rédiger la réponse :	150
5.4. Dessiner en guise de réponse	153
5.4-1 Le dessin en lecture.....	153
5.4-2. Le dessin en mathématiques	156
Discussion :	161
5.5. Lire à haute voix	161
Discussion	169
5.6. Synthèse du chapitre 5.....	170
Chapitre 6. Le choix des tâches d'évaluation	171
6.1. Tâches empruntées au web	171
6.2. Tâches empruntées à des manuels.....	178
Synthèse de la partie III	183

Partie IV – Étude de la validité globale et locale des tests et des épreuves.....185

Chapitre 1. Analyse des tests et des épreuves d'évaluation des compétences relevant des nombres et du calcul.....	186
1.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences du domaine des nombres et du calcul en CP.	186
1.1.1. Les modes de représentation des nombres (ou de désignations) et les traductions.....	188
1.1.2. Les relations entre les nombres	190
1.1.3. Le calcul	191
1.1.3a. Le nombre cardinal, mesure de quantités discrètes.....	191

1.1.3b. Le nombre comme résultat d'un dénombrement et/ou d'un comptage.....	192
1.1.3c. Calcul réfléchi et calcul posé	194
1.2. Analyse des tests et des épreuves	196
1.2.1. Les tests d'évaluation du 1er trimestre.....	196
1.2.2. Les tests d'évaluation du troisième trimestre.....	198
1.2.3. Validité de couverture didactique des tests au premier et au troisième trimestre	199
1.2.4. Validité psycho-didactique des épreuves composants les tests couvrant les principaux composants du domaine des nombres et du calcul au 1er trimestre.....	199
1.2.4a Les épreuves portant sur le composant Désignations-Représentations.....	199
1.2.4b Les épreuves portant sur le composant Relations entre les nombres	201
1.2.4c Les épreuves portant sur le composant Calcul	204
1.2.5. Validité psycho-didactique des épreuves composants les tests couvrant les principaux composants du domaine des nombres et du calcul au troisième trimestre	211
1.2.5a Les épreuves portant sur le composant Désignations-Représentations.....	211
1.2.5b Les épreuves portant sur le composant Relations entre les nombres	214
1.2.5c Les épreuves portant sur le composant Calcul	216
1.3. Synthèse	223

Chapitre 2. Analyse des tests et des épreuves d'évaluation des compétences de résolution de problème (additif).....	226
2.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences de résolution de problèmes additifs en CP	226
2.2. Analyse des tests et des épreuves	230
2.2.1. Les tests d'évaluation du 1er trimestre.....	231
2.2.2. Les tests d'évaluation du troisième trimestre.....	232
2.2.3. Validité de couverture didactique des tests au premier et au troisième trimestre	233
2.3. Discussion autour du « problème simple à une opération ».....	234
2.3.1. Le problème « simple à une opération » dans les programmes	234
2.3.2. Les problèmes « simple à une opération » contenus dans les épreuves du troisième trimestre	235
2.3.2a Analyse des problèmes de la 2 ^e catégorie de relations additives de Vergnaud	236
2.3.2b Analyse des problèmes de la 1 ^{ère} catégorie de relations additives.....	241

2.3.2c Caractérisation mathématique et sémantique du problème simple à une opération donné en évaluation en fin de CP par les enseignants de la population étudiée.....	245
2.3.2d Le problème simple est-il un support de lecture simple ?.....	246
2.4. Synthèse	250
Chapitre 3. Analyse des tests et des épreuves d'évaluation des compétences en lecture	251
3.1. Construction d'un référentiel pour l'analyse d'épreuves d'évaluation des compétences de lecture en CP.....	252
3.1.1. La variable « Lecteur ».....	252
3.1.1a. Les relations entre les phonèmes et les graphèmes (critères PG).....	257
3.1.1b. L'identification des mots (critères L)	261
3.1.1c. La compréhension de l'écrit (critères C).....	263
3.1.2. La variable « Texte »	270
3.1.3. Synthèse.....	273
3.2. Analyse de la dimension phonographique des tests et des épreuves du 1er trimestre	276
3.2.1. Les tests	276
3.2.2. Les épreuves des tests proposés par les enseignants D, I, J et K.....	278
3.2.2a. Analyse des épreuves permettant d'évaluer les connaissances des équivalences entre les différents systèmes d'écriture des lettres (tâches PG3)....	278
3.2.2b. Analyse des épreuves permettant d'évaluer les correspondances entre phonèmes et graphèmes (tâches PG4)	283
3.2.2c. Analyse des épreuves permettant d'évaluer la combinatoire (tâches PG5)	284
3.2.3. Synthèse.....	290
3.3. Analyse de la dimension lecture des tests et des épreuves du troisième trimestre	291
3.3.1. Les tests	291
3.3.2. Les épreuves des tests A ; B ; D ; F et L	292
3.3.3. Synthèse.....	294
3.4. Analyse de la dimension compréhension des tests et des épreuves du troisième trimestre	295
3.4.1. Les tests	295
3.4.2. Les épreuves des tests A ; C ; E ; I ; J ; L et P	297
3.4.3. Synthèse.....	312

Synthèse de la partie IV	313
Conclusion.....	314
1. La construction de référentiels : une étape clé	315
2. Différentes formes de validité de l'évaluation	317
Validité curriculaire	317
Validité psycho-didactique des épreuves	318
Validité de couverture didactique des tests	320
3. Limites et ouvertures.....	321
Sommaire détaillé.....	323
Table des figures.....	330
Table des tableaux	333
Bibliographie	335
Résumé	354
Abstract	356

Table des figures

Figure 1 : Mode d'emploi du livret SCEREN	44
Figure 2 : Préface du livret SCEREN.....	46
Figure 3 : Extrait du livret utilisé par l'enseignant A (Annexe A, p.32)	53
Figure 4 : Épreuve FT3 (Annexe B, p. 156) avec réponses d'élève	77
Figure 5 : Épreuve MT1 (Annexe B, p. 215 et 223) avec réponses d'élève.....	80
Figure 6 : Épreuve MT1 (Annexe B, p.163) avec réponses d'élève	83
Figure 7 : Épreuve MT1 (Annexe B, p. 223) avec réponses d'élève	85
Figure 8 : Épreuve FT3 (Annexe B, pp. 152-153) avec réponses d'élève	86
Figure 9 : Épreuve FT3 (Annexe B, p. 148) avec réponses d'élève.	90
Figure 10 : Épreuve FT1 (Annexe B, p.13) avec réponses d'élève	93
Figure 11 : Épreuve FT1 (Annexe B, p. 13) avec réponses d'élève	94
Figure 12 : Épreuve FT1 (Annexe B, p. 15) avec réponses d'élève	96
Figure 13 : Épreuve FT1 (Annexe B, p. 11) avec réponses d'élève	97
Figure 14 : Épreuve MT1 (Annexe B, p. 174) avec réponses d'élève	103
Figure 15 : Épreuve MT1 (Annexe B, p. 159)	105
Figure 16 : Épreuve MT1 (Annexe B, p. 166) avec réponses d'élève	106
Figure 17 : Épreuve MT1 (Annexe B, p. 162) avec réponse d'élève	107
Figure 18 : Épreuve MT1 (Annexe B, p. 159)	108
Figure 19 : Épreuve FT3 (Annexe B, p. 139) avec réponses d'élève	118
Figure 20 : Épreuve FT3 (Annexe B, p. 154) avec réponses d'élève	133
Figure 21 : Épreuve MT3 (Annexe B, p. 307) avec réponses d'élève	135
Figure 22 : Épreuve FT1 (Annexe B, p. 77) avec réponses d'élève	139
Figure 23 : Épreuve FT1 (Annexe B, p. 67) avec réponses d'élève	141
Figure 24 : Extrait de l'album "Scratch, scratch, dip, clapote" (Crowther, Ecole des loisirs) .	142
Figure 25 : Épreuves « Petit Gaston » (site de circonscription de Cluses pour la 1 ^{ère} version avec réponses d'élève, et Annexe B, p. 134 pour la 2 ^e version)	145
Figure 26 : Épreuve FT1 (Annexe B, p. 40) avec réponses d'élève	152
Figure 27 : Épreuve FT1 (Annexe B, p. 65) avec réponses d'élève	154
Figure 28 : Épreuve FT1 (Annexe B, p. 69)	155
Figure 29 : Épreuve MT1 (Annexe B, p. 220) avec réponse d'élève	157
Figure 30 : Épreuve MT3 (Annexe B, p. 314) avec réponse d'élève	158
Figure 31 : Épreuve MT3 (Annexe B, p. 310) avec réponses d'élève	160
Figure 32 : Épreuves FT1 (Annexe B, p. 58 (à gauche) et p. 61 (à droite)).....	164
Figure 33 : Épreuve FT1 (Annexe B, p. 63) avec réponses d'élève	166
Figure 34 : Épreuve FT3 (Annexe B, p. 119)	167
Figure 35 : Épreuve FT3 (Annexe B, p. 124)	168
Figure 36 : Capture d'écran de la requête « évaluation lecture petit Gaston »	172

Figure 37 : Épreuves « Petit Gaston » (site Cluses (gauche) et Annexe B, p. 134 (droite)) avec réponses d'élève.....	173
Figure 38 : Résultats détaillés à l'épreuve « Petit Gaston ». Évaluation CE2, DEPP 2000.	175
Figure 39 : Épreuves MT1 utilisées par les enseignants L (à gauche) et D (à droite) (Annexe B, p. 208 et 221) avec réponses d'élève.....	178
Figure 40 : Extrait de manuel (A l'école des albums, CP Retz).....	180
Figure 41 : Épreuve FT1 utilisée par l'enseignant E (Annexe p. 70) avec réponse d'élève .	181
Figure 42 : Épreuve FT1 (Annexe B, p. 72) avec réponses d'élève	181
Figure 43 : Épreuve MT1 (Annexe C, p.135) avec réponse d'élève	200
Figure 44 : Épreuve MT1 (Annexe C, p. 155) avec réponse d'élève	201
Figure 45 : Épreuve MT1 (Annexe C, p. 153) avec réponse d'élève	202
Figure 46 : Épreuve MT1 (Annexe C, p. 148) avec réponse d'élève	203
Figure 47 : Extrait d'Épreuve MT1 (Annexe C, p. 157).....	204
Figure 48 : Épreuve MT1 (Annexe C, p.1 58)	205
Figure 49 : Épreuve MT1 (Annexe C, p. 157)	205
Figure 50 : Épreuve MT1 (Annexe C, p. 160) avec réponse d'élève	206
Figure 51 : Épreuve MT1 (Annexe C, p. 164)	206
Figure 52 : Épreuve MT1 (Annexe C, p. 159) avec réponse d'élève	207
Figure 53 : Épreuve 37-3 (Annexe C, p. 160) avec réponse d'élèves	208
Figure 54 : Épreuve MT1 (Annexe C, p. 176) avec réponse d'élève	209
Figure 55 : Épreuve MT3 (Annexe C, p. 191)	211
Figure 56 : Épreuve MT3 (Annexe C, p. 195) avec réponses d'élève	212
Figure 57 : Épreuve FT3 (Annexe A, p. 242) avec réponses d'élève	213
Figure 58 : Épreuve MT3 (Annexe C, p. 203) avec réponses d'élève	214
Figure 59 : Extrait de l'épreuve 95 (Annexe C, pp. 213-214) avec réponse d'élève	215
Figure 60 : Épreuve MT3 (Annexe C, p. 219) avec réponses d'élève	217
Figure 61 : Épreuve MT3 (Annexe C, p. 224) avec réponses d'élève	218
Figure 62 : Épreuve MT3 (Annexe C, p. 220) avec réponse d'élève	219
Figure 63 : Épreuve MT3 (Annexe C, p. 223) avec réponse d'élève	220
Figure 64 : Épreuve MT3 (Annexe C, p. 232) avec réponses d'élève	221
Figure 65 : Épreuve MT3 (Annexe C, p. 228)	222
Figure 66 : Épreuve MT3 (Annexe C, p. 246)	223
Figure 67 : Épreuve MT3 (Annexe B p. 307) avec réponses d'élève	239
Figure 68 : Épreuve MT3 (Annexe B p. 307) avec réponses d'élève	246
Figure 69 : Épreuve MT3 (Annexe B p. 307) avec réponses d'élève	247
Figure 70 : Épreuve MT3 (Annexe B p. 306) avec réponses d'élève	249
Figure 71 : Épreuve FT1 (Annexe C, p. 12) avec réponses d'élève	279
Figure 72 : Épreuve MT1 (Annexe C p. 12) avec réponses d'élève	280
Figure 73 : Épreuve MT1 (Annexe C p. 12) avec réponses d'élève	281
Figure 74 : Épreuve MT1 (Annexe C p. 12) avec réponses d'élève	281
Figure 75 : Épreuve FT1 (Annexe C, p. 13) avec réponse d'élève.....	282
Figure 76 : Épreuve FT1 (cf. Annexe C, p. 17) avec réponses d'élève.	283
Figure 77 : Épreuve FT1 (Annexe C, p. 20) avec réponses d'élève.....	285
Figure 78 : Épreuve MT1 (Annexe C p. 22) avec réponses d'élève	286
Figure 79 : Épreuve FT1 (Annexe C, p. 23) avec réponses d'élève.....	288
Figure 80 : Épreuve FT1 (Annexe C, p. 24) avec réponses d'élève.....	289
Figure 81 : Épreuve FT1 (Annexe C, p. 25) avec réponses d'élève.....	290
Figure 82 : Épreuve FT3 (Annexe C, p. 90) avec réponses d'élève	293

Figure 83 : Épreuve FT3 (Annexe C, p. 88) avec réponses d'élève	294
Figure 84 : Extrait d'épreuve FT3 (Annexe C, pp.103-104) avec réponses d'élève.....	297
Figure 85 : Épreuve FT3 (Annexe C, p. 104) avec réponses d'élève	298
Figure 86 : Épreuve FT3 (Annexe C, p. 105) avec réponses d'élève	299
Figure 87 : Épreuve FT3 (Annexe C, p. 113) avec réponses d'élève	301
Figure 88 : Épreuve FT3 (Annexe C, p. 118) avec réponses d'élève	303
Figure 89 : Épreuve FT3 (Annexe C, p. 123) avec réponses d'élève	303
Figure 90 : Épreuve FT3 (Annexe C, p. 124) avec réponses d'élève	304
Figure 91 : Épreuve FT3 (Annexe C, p. 123) avec réponses d'élève	306
Figure 92 : Épreuve FT3 (Annexe C, p. 124) avec réponses d'élève	308
Figure 93 : Épreuve FT3 (Annexe C, p. 112) avec réponses d'élève	309
Figure 94 : Épreuve FT3 (Annexe C, p. 132) avec réponses d'élève	310
Figure 95 : Épreuve FT3 (Annexe C, p. 133) avec réponses d'élève	311

Table des tableaux

Tableau 1 : Répartition des enseignants formant la population étudiée, par circonscription et par école	32
Tableau 2 : Connaissances et compétences en Lecture (Progressions pour le CP, MEN BO, 2008, p. 31)	39
Tableau 3 : Connaissances et compétences relatives aux nombres et au calcul (Progressions pour le CP, MEN BO, 2008, p. 33)	41
Tableau 4 : Connaissances et compétences en Grandeurs et mesures (Progressions pour le CP, MEN BO, 2008, p. 33)	42
Tableau 5 : Correspondances des items du livret SCEREN aux items des progressions (MEN, BO 2008 p. 31)	47
Tableau 6 : Formats de livrets utilisés par les enseignants	50
Tableau 7 : Distribution des variables L ; R et du taux R/L par enseignant (en ligne), par domaine disciplinaire et trimestre (en colonne)	52
Tableau 8 : Nombre d'épreuves d'évaluation par sous-domaine du Français et des Mathématiques et par enseignant pour le 1 ^{er} trimestre	57
Tableau 9 : Nombre d'épreuves d'évaluation par sous-domaine du Français et des Mathématiques et par enseignant pour le 3 ^e trimestre	60
Tableau 10 : Comparaison des distributions des épreuves de Français dans les tests collectés pour notre étude et ceux étudiés par Grisay (1992, p. 139)	62
Tableau 11 : Nombre d'épreuves d'évaluation du 1 ^{er} trimestre pour chaque item de Lecture, par enseignant	63
Tableau 12 : Nombre d'épreuves d'évaluation du 3 ^e trimestre pour chaque item de Lecture, par enseignant	64
Tableau 13 : Nombre d'épreuves d'évaluation du 1 ^{er} trimestre pour chaque item du sous-domaine des nombres et du calcul, par enseignant	65
Tableau 14 : Nombre d'épreuves d'évaluation du 3 ^e trimestre pour chaque item du sous-domaine des nombres et du calcul, par enseignant	66
Tableau 15 : Formulations de l'item du programme « Connaître (savoir écrire et nommer) les nombres entiers naturels inférieurs à 100 » dans les livrets utilisés par les enseignants. ...	102
Tableau 16 : Formulations de l'item « dire de qui ou de quoi parle le texte lu ; trouver dans le texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens » dans les livrets utilisés par les enseignants.	114
Tableau 17 : Taxonomie de la relation question-réponse (RQR), d'après Irwin (1991, p.182)	126
Tableau 18 : Répartition des questions selon taxonomie de la RQR présentée dans le tableau 17	127

Tableau 19 : Liste des 45 mots les plus fréquents de la langue française écrite (avec mention de leur catégorie grammaticale et de leur fréquence d'occurrence).....	151
Tableau 20 : Nombre d'épreuves d'évaluation par aspects du nombre évalué et par enseignant au 1 ^{er} trimestre.	196
Tableau 21 : Nombre d'épreuves d'évaluation par aspects du nombre évalué et par enseignant au 3 ^e trimestre.....	198
Tableau 22 : Pourcentages d'épreuves de chaque type de tâche d'évaluation du nombre par trimestre	199
Tableau 23 : Validité globale et locale des tests pour les composants du domaine du nombre et calcul considérés (composant résolution de problèmes mis à part) aux 1 ^{er} et 3 ^e trimestres	224
Tableau 24 : Catégories de relations additives et classes de problèmes additifs (d'après Vergnaud, 1981).....	229
Tableau 25 : Répartition des problèmes additifs collectés au 1 ^{er} trimestre par enseignant et, par catégorie et classe de problèmes (cf. catégorisation présentée dans le tableau 24).....	231
Tableau 26 : Répartition des problèmes additifs collectés au 3 ^e trimestre par enseignant et, par catégorie et classe de problèmes (cf. catégorisation présentée dans le tableau 24).....	232
Tableau 27 : Organisations événementielle, énonciative et opérative des énoncés et opérateurs sémantiques et mathématiques des 12 problèmes de catégorie II (transformation opérant sur une mesure pour donner une mesure) recueillis au 3 ^e trimestre	238
Tableau 28 : Organisations événementielle, énonciative et opérative des énoncés et opérateurs sémantiques et mathématiques des 14 problèmes de catégorie I (composition de mesures) recueillis au 3 ^e trimestre.....	243
Tableau 29 : Classification des items des progressions suivant les catégories de la typologie de tâches établie par Goigoux et coll. (2013).....	255
Tableau 30 : Répartition des épreuves d'évaluation recueillies au 1 ^{er} trimestre par enseignant et par sous-composant du composant « Relations phonographiques ».....	276
Tableau 31 : Répartition des épreuves d'évaluation recueillies au 3 ^e trimestre par enseignant et par sous-composant du composant « Lecture »	291
Tableau 32 : Répartition des épreuves d'évaluation recueillies au 3 ^e trimestre par enseignant et par sous-composant du composant « Compréhension »	295

Bibliographie

- Allal, L. (2012). Les exigences inconciliables des activités évaluatives et leurs paradigmes de référence. Dans L. Mottier Lopez & G. Figari (2012) *Modélisations de l'évaluation en éducation*. Bruxelles : De Boeck.
- Allal, L., Bain, D. & Perrenoud, P. (dir.) (1993). *Évaluation formative et didactique du français*. Neuchâtel et Paris : Delachaux et Niestlé.
- Allal, L. (1979) Stratégies d'évaluation formative : conceptions psycho-pédagogiques et modalités d'application. Dans L. Allal, J. Cardinet & P. Perrenoud (Eds.) (1979). *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.
- Allal, L., Cardinet J. & Perrenoud, P. (Eds.) (1979). *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.
- Amigues, R. (2009). Le travail enseignant : prescriptions et dimensions collectives de l'activité. *Les sciences de l'éducation pour l'ère nouvelle*, 42(2), 11-26.
- Amigues, R. (2004). *Compétences, capacités, savoirs*. Document inédit. Repéré à <http://eppee.ouvaton.org/spip.php?article113>
- Amigues, R. & Zerbato-Poudou, M.-T. (1996). *Les pratiques scolaires d'apprentissage et d'évaluation*. Paris: Dunod.
- Anderson, R. (1977). The Notion of Schemata and the Educational Enterprise: General Discussion of the Conference. In R. Anderson, R. Spiro & W. Montague (1977) *Schooling and the Acquisition of Knowledge*. Hillsdale, NJ : Erlbaum.
- Astolfi, J.-P. (1997). *L'erreur un outil pour enseigner*. Paris : ESF.
- Bain, D. & Schneuwly, B. (1993). Pour une évaluation formative intégrée dans la pédagogie du français: de la nécessité et de l'utilité de modèles de référence. Dans : L. Allal, D. Bain & P. Perrenoud. *Évaluation formative et didactique du français*. Neuchâtel : Delachaux et Niestlé, 1993. 51-79
- Barbier, J.-M. (2011). Evaluation. Dans J.-M. Barbier. *Vocabulaire d'analyse des activités* (66-67). Paris : PUF.

Barbier, J.M. (1985). *L'évaluation en formation*. Paris, PUF.

Bardi, A.-M. & Mégard, M. (2009). L'évaluation des élèves en France, à un moment charnière de leur histoire. *Mesure et évaluation en éducation*, 32(3), 125-152.

Barrouillet, P., Camos, V., Morlaix, S., & Suchaut, B., (2008). Progressions scolaires, mémoire de travail et origine sociale : quels liens à l'école élémentaire ? *Revue française de pédagogie*, 162, 5-14.

Barthes, R, Kayser W., Booth W.-C., Hamon P. (1977). *Poétique du récit*. Paris : Points Seuil.

Bastien, C. & Bastien-Tonniazzo, M. (2004). *Apprendre à l'école*. Paris : Armand Colin.

Bastien, C. (1997). *Les connaissances : de l'enfant à l'adulte*. Paris : Armand Colin.

Bastien, C. (1993). Apprentissage : modèles et représentation. Dans : J.-F. Le Ny (Ed.), *Intelligence naturelle et intelligence artificielle*, 257-268. Paris : PUF.

Baudet, S., Denhière, G. (1990). Le fonctionnement cognitif dans la compréhension de textes. *Les cahiers de l'Unadrio*, 19, 4-12.

[Beck](#), I. L., Mc Keown, M. G., Mc Caslin, E. S., & Burke, A. M (1979). *Instructional dimensions that may affect reading comprehension: Examples from two commercial reading programs*. Pittsburgh : University of Pittsburgh, Learning Research and Development Center.

Besse, J.-M., Luis, M.-H., Paire, K., Petiot-Poirson, K., Petit Charles, E. (2004). *Évaluer les illettrismes*. Paris : Retz.

Bettelheim, B. & Zelan, K. (1983). *La lecture et l'enfant* (traduit par T. Carlier) Paris : Robert Laffont.

Bianco, M. (2014). *Propositions pour une programmation de l'enseignement de la compréhension en lecture*. Repéré sur le site du Conseil Supérieur des Programmes : http://cache.media.education.gouv.fr/file/Le_systeme_educatif/70/7/Bianco_Maryse_-_MCF_-_CSP_Contribution_362707.pdf

Bianco, M. (2010). La compréhension de textes : peut-on l'apprendre et l'enseigner ? Dans M. Crahay & M. Dutrevis (Eds) (2010). *Psychologie des apprentissages scolaires*, Bruxelles : De Boeck.

Bideaud, J., Meljac C. & Fischer, J.-P. (1991). *Les chemins du nombre*. Lille : Presses universitaires de Lille.

Bishop, M.-F., Cèbe, S. & Piquée, C. (2016). L'enseignement de la compréhension dans les classes de CP aujourd'hui : temps consacré et choix didactiques. *Repères*, 52, 15-38.

Blanc, C. & Vantourout, M. (*Soumis*). Une approche psycho-didactique pour évaluer la lecture au Cours Préparatoire.

Blanc, C. & Goasdoué R. (2016, janvier). *Qu'évaluent les questionnaires de compréhension de texte ?* Communication présentée au XVIII^e colloque de l'Admee Europe, Lisbonne.

Blanc, C. & Vantourout, M. (2015, janvier). *Evaluation de la lecture en fin de CP dans le cadre des APDE*. Symposium EVADIDA. Communication présentée au XXVII^e colloque de l'Admee Europe, Liège.

Blanc, C. (2014, juillet). *Une approche psycho-didactique de l'évaluation en classe*. Symposium « Cadres psychologiques et aux modélisations de l'évaluation des apprentissages des élèves ». Communication présentée au 28th International Congress of Applied Psychology, Paris.

Blanc, C. (2013, juin). *Evaluer la compréhension de texte en fin de CP : construction d'une épreuve interne de circonscription*. Communication présentée à la Journée de la section française de l'ADMEE Europe, Paris.

Blanc, N., (2010). *Lecture et habileté de compréhension chez l'enfant*. Paris : Dunod

Blanc, N. & Brouillet, D. (2003). *Mémoire et compréhension – Lire pour comprendre*. Paris : In Press.

[Blanquer](#), J.-M. (2011). Note n° 2011-0033 du 6 avril 2011. DGESCO.

Bonniol, J.-J. & Vial M. (1997/2006). *Les modèles de l'évaluation*. Bruxelles : De Boeck.

Bonniol, J.-J. (1981) *Déterminants et mécanismes des comportements d'évaluation d'épreuves scolaires* (Thèse, Université de Bordeaux II, Bordeaux).

Briand, J., Lacave Luciani, M.-J., Harvouët M., Bedere D., & Goua de Baix V. (2000) Enseigner l'énumération en moyenne section. *Grand N*, 66, 7- 22.

Briand, J. (1993). *L'énumération dans le mesurage des collections* (Thèse, Université Bordeaux I, Bordeaux).

Brissiaud, R. (1989/2003). *Comment les enfants apprennent à calculer ?* Paris : Retz.

Brousseau, G. (2007). Introduction à une étude des situations et des champs conceptuels. Dans M. Merri (coord.) (2007) *Activité humaine et conceptualisation. Questions à Gérard Vergnaud*. Toulouse : PUM.

Brousseau, G. (2004). Tâche, situation, activité. Repéré à <http://www.ssrddm.ch/SSRDM/actualite/materiels/tachebrousseau.pdf>

Brousseau, G. (2001). Les grandeurs dans la scolarité obligatoire. Dans J.-L. Dorier, M. Artaud, M. Artigue, R. Berthelot & R. Floris (coord.). *Actes de la XIe Ecole d'Été de*

Didactique des Mathématiques, Corps (Isère) du 21 au 30 août 2001. Grenoble : La pensée sauvage éditions.

Brousseau, G. (1998). Glossaire de quelques concepts de la théorie des situations didactiques en mathématiques. Repéré à http://guy-brousseau.com/wp-content/uploads/2010/09/Glossaire_V5.pdf

Brousseau, G. (1998). *Théorie des situations didactiques*. Grenoble : La Pensée Sauvage.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18 (1), 32-42.

Butlen, D. (2007). *Le calcul mental entre sens et technique*. Besançon : Presses universitaires de Franche-Comté.

Camus, J.-F. (1996). *La psychologie cognitive de l'attention*. Paris : Armand Colin.

Cardinet, J. (1990). *Evaluation interne, externe ou négociée ?* Conférence de 1987. Hommage à Jean Cardinet. (39-158). Fribourg : Delval.

Cardinet, J. (1989). Évaluer sans juger. *Revue française de pédagogie*, 88, 41-52.

Cardinet, J. (1989). Le découpage des objectifs pédagogiques. *Mesure et évaluation en éducation*, 7(2), 5-48.

Carette, V. & Rey, B. (2010). *Savoir enseigner dans le secondaire*. Bruxelles : de Boeck.

Carette, V. (2009). Pédagogie par objectifs et approche par compétences : Rupture ou continuité ? *Bulletin de l'ADMEE-Europe*, 2009(2), 11-26.

Catach, N., Gruaz, C. & Duprez, D. (1995). *L'orthographe française*. (3^e éd. révisée). Paris : Nathan Université.

Catach, N. (1980). *L'orthographe française*. Paris : Nathan.

Catach, N. (1973a) Présentation. *Langue française*, 20, 3-5.

Catach, N. (1973b) Que faut-il entendre par système graphique du français ? *Langue française*, 20, 30-44.

Céard, M.-T., Remond, M., Varier, M. (2003). L'appréciation des compétences des élèves et des jeunes en lecture et écriture et l'évolution de ces compétences dans le temps. *Les rapports du Haut Conseil de l'évaluation de l'école*, 15.

Cèbe, S. & Goigoux, R. (2013). *Lectorino & Lectorinette. Apprendre la compréhension des textes narratifs*. Paris : Retz.

Cèbe, S. & Goigoux, R. (2007). Concevoir un instrument didactique pour améliorer l'enseignement de la compréhension de textes. *Repères*, 35, 185-208.

Cèbe, S., Goigoux, R. & Thomazet, S., (2004). *Enseigner la compréhension : principes didactiques, exemples de tâches et d'activités*. Repéré à <http://eduscol.education.fr/cid48095/sommaire.html> puis https://halshs.archives-ouvertes.fr/file/index/docid/922482/filename/lire_ecrire_enseigner_comprehension_115544.pdf

Charmeux, E. (non daté). Combinatoire. Repéré à <http://www.charmeux.fr/combinatoire.html>

Chartier, P. & Loarer, E., (2008). *Evaluer l'intelligence logique*. Paris : Dunod

Chauveau, G. (2007). *Le savoir-lire aujourd'hui*. Paris : Retz.

Chenu, F., Crahay, M. & Lafontaine, D. (2014). Par-delà l'approche par compétences : quelle place réserver aux savoirs, à leur enseignement et à leur évaluation ? *Education & Formation*, e-302, (17-29).

Chervel, A. (1988) L'histoire des disciplines scolaires. Réflexions sur un domaine de recherche. *Histoire de l'éducation*, 38, 59-119.

Chevallard, Y. (1988, novembre) *Esquisse d'une théorie formelle du didactique*. Communication présentée au Premier colloque franco-allemand de didactique des mathématiques et de l'informatique, CIRM, Marseille. Dans : C. Laborde (éd.), *Actes*. (97-106). Grenoble : La Pensée sauvage.

Chevallard, Y. (1985/1991). *La transposition didactique. Du savoir savant au savoir enseigné*. Grenoble : La pensée sauvage.

Clot, Y. (1995). *Le travail sans l'homme. Pour une psychologie des milieux de travail de vie*. Paris : La Découverte.

CNESCO-IFé/ENS Lyon (2016). Conférence de consensus Lire, Comprendre, Apprendre. Comment soutenir le développement de compétences en lecture ? Dossier de synthèse, remise des recommandations du jury. 16 et 17 mars, 2016. ENS, Lyon. Repéré à http://www.cnesco.fr/wp-content/uploads/2016/09/CCLecture_dossier_synthese.pdf

CNESO (2014). *L'évaluation des élèves par les enseignants dans la classe et les établissements : réglementation et pratiques*. 44 pages. <http://www.cnesco.fr>

Collet, M., Grégoire, J. (2008). Le développement du système en base dix chez les enfants de première et de deuxième primaire. Dans M. Crahay, L. Verschaffel, E. De Corte & J. Grégoire (Eds) *Enseignement et apprentissage des mathématiques - Que disent les recherches psychopédagogiques ?* (80-103). Bruxelles : De Boeck.

Crahay, M. & Dutrévis, M. (Eds.) (2010). *Psychologie des apprentissages scolaires*. Bruxelles : De Boeck.

Crahay, M., Verschaffel, L., De Corte E. & Grégoire, J. (Eds) (2008) *Enseignement et apprentissage des mathématiques - Que disent les recherches psychopédagogiques ?* Bruxelles : De Boeck

Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue française de pédagogie*, 154, 97-110.

Crahay, M. (1999/2010). *Psychologie de l'éducation*. Paris : Puf.

Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck

[Crinon, J. & Legros, D. \(2003\). Identification des mots et traitement phonologique.](http://www.cndp.fr/bienlire/04-media/a-identification-lmp.htm) Repéré à <http://www.cndp.fr/bienlire/04-media/a-identification-lmp.htm>

Daunay, B. (1993) Les questions de compréhension au brevet des collèges. *Recherches*, 19, 101-120.

De Croix, S. & Ledur, D. (2014). Construire des savoirs sur l'activité de lecture afin de développer les compétences des élèves en difficulté : l'inaccessible rêve ? *Education & Formation*, e-302, 43-56.

De Ketele, J.-M. (2012). A la recherche de paradigmes unificateurs ou de paradigmes intégrateurs ? Dans L. Mottier Lopez & Figari, G. *Modélisations de l'évaluation en éducation* (195-210). Bruxelles : De Boeck

De Ketele, J.-M. & Gérard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation*, 28(3), 1-26.

De Ketele, J.-M. & Dufays, J.-L. (2003). Vers de nouveaux modes d'évaluation des compétences. Dans Colles, L., Dufays., J.-L. & Maeder, C. *Enseigner le français, l'espagnol et l'italien, Les langues romanes à l'heure des compétences*, Bruxelles-Paris : De Boeck Duculot, 171-182.

De Ketele, J.-M. (1993). L'évaluation conjugquée en paradigmes. *Revue Française de Pédagogie*, 103, 59-80.

De Ketele, J.-M. (1986). *L'évaluation : approches descriptives ou prescriptives ?* Bruxelles : De Boeck.

De Landsheere, G. (1976). *Evaluation continue et examens. Précis de docimologie*. Paris : Nathan, Labor.

De Landsheere V. (1988). *Faire réussir, faire échouer*. Paris : PUF.

Delcambre, I. (2007/2008). Tâche. Dans Y. Reuter (dir). *Dictionnaire des concepts fondamentaux des didactiques* (217-222). Bruxelles : De Boeck.

Demeuse, M. et Henry, G. (2004). *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation*. Liège : Editions de l'Université de Liège.

Denhière, G., Jehan-Larose, S. (2011). *Le vocabulaire et son enseignement*. Repéré à http://cache.media.eduscol.education.fr/file/Dossier_vocabulaire/14/2/Guy_Denhiere_111202_avec_couv_201142.pdf

Denhière, G. & Baudet, S. (1992) *Lecture, compréhension de texte et sciences cognitives*. Paris : PUF.

Detroz, P., Crahay, M. & Fagnant, A. (dir.) (2017). *L'évaluation à la lumière des contextes et des disciplines*. Bruxelles : De Boeck.

Devanne, B. (1996). *Lire, écrire : des apprentissages culturels au CP*. Tome 3. Paris : Armand Colin.

Devidal, M., Fayol, M. & Barrouillet, P. (1997). Stratégie de lecture et résolution de problèmes arithmétiques. *L'année de psychologie*, 97, 9-31.

Diatkine, R. (1983). L'enfant et l'écrit. *Les textes du Centre Alfred Binet*, 3.

Dierendonck, C. & Fagnant, A. (2014). Approche par compétences et évaluation à large échelle : deux logiques incompatibles ? *Mesure et évaluation en éducation*. 37(1), 43-82.

Dubois, J., Giacomo, M., Guespin L., Marcellesi C. et Mevel J.-P. (1994/1999) *Dictionnaire de linguistique des sciences du langage*. Paris: Larousse.

[Durkin](#), D. (1978). *What classroom observations reveal about reading comprehension instruction*. Technical Report n°506. University of Illinois at Urbana-Champaign.

Duval, R. (1988). Ecarts sémantiques et cohérence mathématique. *Annales de didactique et de sciences*, 1, 7-25.

Ecalte, J. & Magnan, A. (2010). *L'apprentissage de la lecture et ses difficultés*. Paris : Armand Colin.

Ecalte, J. (2010). L'évaluation de la lecture et des compétences associées. *Revue française de linguistique appliquée*, XV(1), 105-120.

Ecalte, J. & Magnan, A. (2002). *L'apprentissage de la lecture – Fonctionnement et développement cognitif*. Paris : Armand Colin.

[Ecalte](#), J., Magnan, A. & Bouchafa J. (2002). Le développement des habiletés phonologiques avant et au cours de l'apprentissage de la lecture : de l'évaluation à la remédiation. *Glossa*, 82, 4-12.

Ehrlich S. (1990). *Sémantique et mathématique*. Paris : Nathan.

Endrizzi, L. & Rey, O. (2008). L'évaluation au cœur des apprentissages. Dossier d'actualité du service de veille scientifique et technologique, 39.

ERMEL (1991). Apprentissages numériques et résolution de problèmes- CP. Paris : Hatier.

Fayol, M. (2003). *La compréhension : évaluation, difficultés et interventions*. Repéré à <http://www.cndp.fr/bienlire/01-actualite/document/fayol.pdf>

Fayol, M. (1990). *L'acquisition du nombre*. Paris : PUF.

Figari, G. & Mottier Lopez, L. (2006) *Recherches sur l'évaluation en Education*. Paris : L'Harmattan.

Figari, G. (2006). Les référentiels. Dans Figari, G. & Mottier Lopez, L. (2006) *Recherches sur l'évaluation en Education*. (101-108). Paris : L'Harmattan.

Figari, G. & Achouche, M. (Eds.) (2001), *L'activité évaluative réinterrogée*, Bruxelles : De Boeck

Figari, G. (1994). *Évaluer: Quel référentiel ?* Bruxelles: De Boeck.

Fijalkow, E. (2003). L'enseignement de la lecture-écriture au cours préparatoire, entre *tradition et innovation*. Paris : L'Harmattan.

Fijalkow, J. (2000). *Sur la lecture*. Paris : ESF.

Fuson, K. (1998). *Children's counting and concepts of number*. New York : Springer Verlag.

Fuson, K & Kwon, Y. (1991). Système de mots-nombres et autres outils culturels : effets sur les premiers calculs de l'enfant. Dans J. Bideaud, C. Meljac & J.-P. Fischer. *Les chemins du nombre*. Lille : Presses universitaires de Lille.

[Garcia-Debanc](#), C. (2008) De la configuration didactique au modèle disciplinaire en acte : trente ans de didactique du français avec Jean-François Halté. *Pratiques*, 137-138, 39-56.

Gelman, R. & Meck, E. (1991). Premiers principes et conceptions du nombre. Dans J. Bideaud, C. Meljac & J.-P. Fischer. *Les chemins du nombre*. Lille : Presses universitaires de Lille

Genette, G., (1972), *Figures III*. Paris : Seuil.

Giasson, J. (2012). *La lecture : apprentissages et difficultés*. Bruxelles : De Boeck.

Giasson, J. (1990). *La compréhension en lecture*. Bruxelles : De Boeck.

Gillet, P. (1991) *Construire la formation : outils pour les enseignants et les formateurs*. Paris : PUF.

Goasdoué, R. & Vantourout, M. (2017). Évaluations scolaires et étude du jugement des enseignants : pour une docimologie cognitive. Dans : P. Detroz, M. Crahay & A. Fagnant (dir.) (2017). *L'évaluation à la lumière des contextes et des disciplines*. (141-168). Bruxelles : De Boeck.

Goigoux, R., Jarlegan, A. & Piquée, C. (2015). Évaluer l'influence des pratiques d'enseignement du lire-écrire sur les apprentissages des élèves : enjeux et choix méthodologiques. *Recherches en didactiques*, 19, 33-52.

Goigoux, R., Cèbe, S. (2014). Enseignement de la lecture : réponses à cinq questions posées par le conseil supérieur des programmes. Repéré à http://cache.media.education.gouv.fr/file/CSP/06/3/GOIGOUX_Roland_PU_et_CEBE_Sylvie_MCF_-_CSP_Contribution_395063.pdf

Goigoux, R., Cèbe, S. (2013). Enseigner la compréhension à l'école élémentaire : des résultats de recherches à la conception d'un outil didactique. *Recherches*, 58, 29-46.

Goigoux, R. (2008). (Sans titre) Repéré à http://www.cafepedagogique.net/lesdossiers/Pages/2008/programmes_goigoux.aspx

Goigoux, R., (2007). Un modèle d'analyse de l'activité des enseignants. *Éducation et didactique*, 1(3), 47-69.

Goigoux, R., Cèbe, S. (2006). *Apprendre à lire à l'école*. Paris : Retz

Goigoux, R., Cèbe, S. & Pour, (2004). *PHONO - Développer les compétences phonologiques - GS-CP*. Paris : Hatier.

Goigoux, R., (2003). Quelques points de repère pour une didactique de la compréhension. *Langage & pratiques*, 31, 51-60.

Goigoux, R., (2001, septembre). *Tâche et activité en didactique du français*. Communication présentée au VIIIème colloque de la DFLM, Neuchâtel.

Goigoux, R., (2000a). Apprendre à lire à l'école. *Psychologie française*, 45(3), 234-242.

Goigoux, R., (2000b), *Les élèves en grande difficulté de lecture et les enseignements adaptés*. Paris : Éditions du CNÉFÉI.

Goigoux, R., (coord.), Gombert, J.-E., Colé, P., Valdois, S., Mousty, P. & Fayol, M. (2000/2002). *Enseigner la lecture au cycle 2*. Paris : Nathan.

Goigoux, R., (1998a). Apprendre à lire : de la pratique à la théorie. *Repères*, 18, 147-162.

Goigoux, R., (1998b) *Sept malentendus capitaux*. Repéré à http://www4.ac-nancy-metz.fr/ien-bar-le-duc/IMG/pdf/les_7_malentendus_capitaux_en_maternelle.pdf

Goigoux, R., (1997, octobre). *La psychologie cognitive ergonomique : un cadre d'étude des compétences professionnelles des enseignants de Français*. Journées d'étude de Didactique du Français Langue Maternelle, Montpellier. Non publié, 7 pages.

Gombert, J.-E (2002). Préface. Dans J. Ecalle & A. Magnan, A. (2002). *L'apprentissage de la lecture – Fonctionnement et développement cognitif*. Paris : Armand Colin.

Gombert, J.-E., Colé, P., Valdois, S., Goigoux, R., Mousty, P. & Fayol, M. (2000/2002). *Enseigner la lecture au cycle 2*. Paris : Nathan.

Gombert, J.-E. & Colé, P. (2000). Activités métalinguistiques, lecture et illettrisme. Dans : M. Kail et M. Fayol (Eds), *L'acquisition du langage. Le langage en développement. A delà de trois ans* (117-150). Paris : PUF.

Gonthier-Maurin, B. (2012). Rapport d'information n°601 fait au nom de la commission de la culture, de l'éducation et de la communication par la mission d'information sur le métier d'enseignant. Sénat.

[Gough](#), P. (1996). How children learn to read and why they fail. *Annals of Dyslexia*, 46, 3-20.

Gough, P. & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6–10.

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

Grisay, A. (1992). *Examens et échecs dans l'enseignement fondamental. Une recherche-action sur les pratiques d'évaluation-bilan de cinquante écoles primaires en Communauté française de Belgique*. Liège : Service de Pédagogie expérimentale

Grisay, A. (1989). Améliorer l'évaluation bilan à l'école primaire. *Mesure et évaluation en éducation*, 11(4), 5-26.

Grugeon-Allys, B. & Grapin, N. (2016, novembre). *Validité des évaluations : une approche épistémo-didactique*. Communication présentée au colloque Evaluation en mathématiques, UPEC Université Paris Est Créteil.

[Grugeon-Allys](#), B. & Grapin, N. (2015). *Validité d'une évaluation externe, complémentarité des approches*. Communication présentée au séminaire national de l'IREM Paris.

Grugeon-Allys B. (2015, janvier). *Evaluation diagnostique en algèbre élémentaire : une approche didactique et épistémologique*. Communication présentée au 27^{ième} colloque international de l'ADMEE. Liège : Belgique.

Grugeon-Allys B., Pilet, J., Chenevotot-Quentin, F., Delozanne, E. (2012) Diagnostic et parcours différenciés d'enseignement en algèbre élémentaire. *Recherches en Didactique des Mathématiques, Hors-série*, 137-162.

- Guimard, P. (2010). *L'évaluation des compétences scolaires*. Rennes : PUR.
- Gyselinck, V. (1996). Illustrations et modèles mentaux dans la compréhension de textes. Dans: *L'année psychologique*, 96(3), 495-516.
- Hadji, C. (2012). *Faut-il avoir peur de l'évaluation ?* Bruxelles : De Boeck.
- Hadji, C. (1992). *L'évaluation des actions éducatives*. Paris : PUF.
- Hadji, C. (1989/1990). *L'évaluation, règles du jeu*. Paris: ESF.
- Harris, K. & Graham, S. (1996). *Making the writing process work: strategies for composition and self-regulation*. Cambridge, MA: Brookline.
- Irwin, J.W. (1991). *Teaching reading comprehension processus*. Needham Heights (MA), Allyn et Bacon.
- Jacquier-Roux, M. & Zorman, M. (2002). *Entraînement visuel et apprentissage de la lecture en CP. Cognitif : documentation – Classeur spécial conseiller pédagogique – Grenoble* : Editions la Cigale.
- Jonnaert, P., Ettayebi, M. et Defise, R., (2009). *Curriculum et compétences, un cadre opérationnel*. Bruxelles : De Boeck.
- Jonnaert, P., & Laveault, D., (1994). Évaluation de la familiarité de la tâche: quelle confiance accorder à la perception de l'élève? *Revue des sciences de l'éducation*, 20 (2), 271-291.
- Joole, P. (2008). *Comprendre les textes écrits*. Paris : Retz / CRDP de l'Académie de Versailles.
- Jorro, A. (1999). *Le lecteur interprète*. Paris : PUF.
- Joshua, S. & Dupin, J.-J. (1993/2003). *Introduction à la didactique des mathématiques*. Paris: PUF.
- Julo, J. (1995). *Représentation des problèmes et réussite en mathématiques*. Rennes : PUR.
- Kaneman-Pougatch, M. & Pedoya-Guimbretière, E. (1991), *Plaisir des sons : enseignement des sons du français*. Paris : Hatier-Didier.
- Keenan, J. M. & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without Reading It : Why Comprehension Tests Should Not Include Passage-Independent Items. *Scientific Studies of Reading*, 10(4), 363-380,
- Labat, H., Farhat, S.-L., Andreu, S., Rocher, T., Cros L., Magnan, A. & Ecalle, J. (2013). Évaluation des connaissances précoces prédictives de l'apprentissage de la lecture en grande section de maternelle. *Revue française de pédagogie*, 184, 41-54.

Lafontaine, B. (2003). *Comment faciliter, développer et évaluer la compréhension des textes aux différentes étapes de la scolarité primaire ?* Document envoyé au PIREF en vue de la conférence de consensus sur l'enseignement de la lecture à l'école primaire les 4 et 5 décembre 2003.

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

Leclerq, D. (2006) L'évolution des QCM. Dans : Figari, G. & Mottier Lopez, L. (138-146) *Recherches sur l'évaluation en Education*. Paris : L'Harmattan.

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

Lecocq, P., Casalis, S., Leuwers, C. & Watteau N. (1996). *L'apprentissage de la lecture et compréhension d'énoncés*. Presses Universitaires du Septentrion : Villeneuve d'Ascq.

Lehman, A. (2011). *Le vocabulaire et son enseignement*. MENJVA/DGESCO

Lemoine, A. & Sartiaux P. (2005). *Des mathématiques aux enfants*. Bruxelles : De Boeck.

Leplat, J. (2004) Éléments pour l'étude des documents prescripteurs. *@activités*, 1(2), 195-216.

Leplat, J. & de Montmolin, M. (2001). *Les compétences en ergonomie*. Toulouse : Octarès

Leplat, J. (2000). *L'analyse psychologique de l'activité en ergonomie*. Toulouse : Octarès.

Leplat, J. (1997). *Regards sur l'activité en situation de travail*. Paris : Puf.

Leplat, J. & Hoc, J.M. (1983). Tâche et activité dans l'analyse psychologique des situations. *Cahiers de psychologie cognitive*, 3(1), 49-63.

Leplat, J. & Cluny, X. (1977, rééd. 1984), *Introduction à la psychologie du travail*. Paris : Puf

Leplat, J. (1977) *Regards sur l'activité en situation de travail*. Paris : PUF.

Lesnes, M. (1984). *Lire les pratiques de formation d'adultes : essai de construction théorique à l'usage des formations*. Paris: Edelig.

Lété, B., Sprenger-Charolles, L. & Colé, P. (2004). Manulex : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, 36, 156-166.

Loye, N. (2012, janvier). *Comment assurer l'évaluation diagnostique des compétences scolaires ?* Communication présentée au Colloque Admee L'évaluation des compétences en milieu scolaire et en milieu professionnel. Luxembourg

Margolinas, C., Wozniak, F. (2012). *Le nombre à l'école maternelle*. Bruxelles : De Boeck.

Martinand, J.-L. (1986). *Connaître et transformer la matière*. Berne : Peter Lang.

Martinet, C. & Rieben, L. (2010). L'apprentissage initial de la lecture et ses difficultés. Dans : Crahay, M. & Dutrévis, M. (189-228). *Psychologie des apprentissages scolaires*. Bruxelles : De Boeck.

Mas, M., Garcia-Debanc, C., Tauveron, C., Romian, H., Gadeau, J. & Finet, C. (1991). *Evaluer les écrits à l'école primaire*. Paris : INRP / Groupe EVA Hachette éducation.

Maury, S. (2001). Didactique des mathématiques et psychologie cognitive : un regard comparatif sur trois approches psychologiques. *Revue française de pédagogie*, 137, 85-93.

Maury, S. (1985). Influence de la question dans une épreuve relative à la notion d'indépendance. *Educational Studies in Mathematics*, 16, 283-301.

Maury, S. (1984). La quantification des probabilités : analyse des arguments utilisés par les élèves de la classe de seconde. *Recherche en didactiques des mathématiques*, 5(2), 187-214.

MEN DEPP Moisan C. (2015). *Evaluation des acquis : principes, méthodologie, résultats*. MEN DEPP, 86/87.

MEN - DEPP (2013). *Forte augmentation du niveau des acquis des élèves à l'entrée au CP entre 1997 et 2001*. Note d'information DEPP n° 13.19 - septembre 2013. Paris : MEN.

MEN-MESR (2012). *Les composantes de l'activité professionnelle des enseignants outre l'enseignement dans les classes*. Paris : MEN MESR.

MENJVA-MESR. (2011). *Note de synthèse sur le suivi de la réforme de l'enseignement primaire : pilotage local et suivi des élèves*. Paris : MEN.

MEN - Claus, P, Petreault, G., Louis F. (2010). *Note de synthèse sur la mise en œuvre de la réforme de l'enseignement primaire*. Note n° 2010-092 juillet 2010. IGEN IGAENR.

MEN (2010). *Note de synthèse sur la mise en œuvre de la réforme de l'enseignement primaire*. Paris : MEN.

[MEN](#) - Durpaire, J.-L. & Mégard, M. (2010). *Le nombre au cycle 2. Document d'accompagnement des programmes*. Paris : SCEREN CNDP.

[MEN](#) (2010). *Lire au CP. Document d'accompagnement des programmes*. Paris : MEN SCEREN CNDP.

[MEN](#) -BO. (2008). *Les nouveaux programmes de l'école primaire*. Hors-série n° 3 du 19 juin 2008.

MEN DEPP Rocher T. (coord.) (2008) *Lire, écrire, compter : les performances des élèves de CM2 à vingt ans d'intervalle 1987-2007*. Note d'information DEPP n° 08.38- décembre 2008

MENESR. (2005). *Les acquis des élèves, pierre de touche de la valeur de l'école ?* Paris : MEN.

MEN DEPP - Robin, I. (coord.) (2001). *Evaluations CE2-6^{ième} Résultats nationaux septembre 2000. Les dossiers d'éducation et formations. Août 2001, n°124.* Paris : MEN

MEN DEPP - Colmant, M., Jentheau, J.-P. (1996) *Evaluations CE2-6^{ième} Résultats nationaux septembre 1995. Les dossiers d'éducation et formations. Mars 1996, n°65.* Paris : MEN

[Mercier](#), A., Bodin, A. & Quilio, S. (2015). *Quelques conditions d'efficacité des professeurs pour le premier enseignement des mathématiques dans les pays francophones.* Rapport d'étude pour la banque mondiale. (En ligne sur le site de la CFEM)

Mercier, A. (2008). Pour une lecture anthropologique du programme didactique. *Éducation et didactique*, 2(1), 7-40.

Merri, M. (coord.) (2007) *Activité humaine et conceptualisation. Questions à Gérard Vergnaud.* Toulouse : PUM.

Mesnager, J., (non daté). *Vitesse de lecture.* Site TFL - Paris V.

Monteil, J.-M. & Huguet P. (2002). *Réussir ou échouer à l'école : une question de contexte ?* Grenoble : PUG.

Monteil, J.-M. & Huguet P. (1991). Insertion sociale, catégorisation sociale et activités cognitives. *Psychologie française*, 36, 35-46.

Mottier Lopez, L. (2015). Tâches pour apprendre, tâches pour évaluer. *Formation et pratiques d'enseignements en questions*, 19, 55-65.

Mottier Lopez, L. & Figari G. (Eds). (2012) *Modélisations de l'évaluation en éducation.* De Boeck : Bruxelles.

Mounier, E. (2012) Les signes de la numération écrite chiffrée et parlée en France : une analyse mathématique à usage didactique. *Annales de Didactique et de Sciences Cognitives*, 17, 27-58.

[Mounier](#), E. (2010). *Une analyse de l'enseignement de la numération au CP.* Thèse de doctorat. Université Paris Diderot.

Noizet, G. & Caverni, J.P. (1978). *Psychologie de l'évaluation scolaire.* Paris : PUF.

Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques*, 280, 47-54.

ONL (Germain, B. dir.) (2011). *Analyse de manuels de lecture au CP.* Futuroscope: SCEREN CNDP-CRDP.

ONL (1998). *Apprendre à lire*. Paris : CNDP / Odile Jacob.

[Pearson](#), D. (1984) *Reading comprehension instruction*. Reading Education Report n°54. University of Illinois at Urbana-Champaign.

Pearson, D. & Gallagher, M.C. (1983). *The instruction of reading comprehension*. Technical Report n°297. Illinois: University of Illinois at Urbana-Champaign.

Pearson P. D. & Johnson D. D. (1978). *Teaching reading comprehension*. New York : Holt Rinehart and Winston.

Perez, M. (2004). *Proposition de hiérarchisation des 45 graphèmes de base de l'orthographe du français*. Repéré à <https://halshs.archives-ouvertes.fr/hal-00961511/document>

Perrenoud, P. (2011). *Quand l'école prétend préparer à la vie...* Paris : ESF.

Perrenoud, P. (2001) Evaluation formative et évaluation certificative : postures contradictoires ou complémentaires. *Formation professionnelle suisse*, 4, 25-28.

Perrenoud, P. (2001) Compétences, langage et communication. Dans : Collès, L., Dufays, J.-L., Fabry, G. et Maeder, C. (dir.) (2001) *Didactique des langues romanes. Le développement de compétences chez l'apprenant*, Bruxelles : De Boeck Duculot.

Perrenoud, P. (1998). La transposition didactique à partir de pratiques : des savoirs aux compétences. *Revue des sciences de l'éducation*, XXIV(3), 487-514.

Perrenoud, P. (1991) Pour une approche pragmatique de l'évaluation formative. *Mesure et évaluation en éducation*, 13(4), 49-81.

Piaget, J. (1926/1947) *La représentation du monde chez l'enfant*. Alcan, 1926, 3e édition identique, PUF 1947.

Picoche, J. & Rolland, J.-C. (2002) *Dictionnaire du français usuel*. Bruxelles : De Boeck-Duculot.

Pini, G. (2012). *A propos de la théorie des réponses aux items (TRI-IRT)*. IRDP. Repéré à http://www.irdp.ch/edumetrie/documents/TRI_DICHO.pdf

PIREF (2003). Recommandations du jury. Conférence de consensus sur l'enseignement de la lecture à l'école primaire. Des premiers apprentissages au lecteur compétent. Paris. Repéré à <http://www.cndp.fr/bienlire/01-actualite/c-en-parle06.asp>

Pons, X. (2010). *Evaluer l'action éducative*. Paris : Puf.

Pothier, B. & Pothier P. (2002/2004). *Echelle d'acquisition en orthographe lexicale*. Paris : Retz.

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppanen, P., Poikkeus A.-M., Tolvanen, A., Torppa, M. & Lyytinen, H. (2007) Very early phonological and language skills: estimating individual risk of reading disability. *J Child PsycholPsychiatry*, 48 (9), 923-31.

Remond, M. (2007, mars). *Quelle formation des enseignants à l'évaluation dans la perspective du socle commun ?* Communication présentée au Colloque international : Français, littérature, socle commun : quelle culture pour les élèves, quelle professionnalité pour les enseignants ? INRP- Lyon.

Reuter Y. (éd.), Cohen-Azria, C., Daunay, B., Delcambre, I., Lahanier-Reuter, D. (2007). *Dictionnaire des concepts fondamentaux des didactiques*. Bruxelles : De Boeck.

Reuter, Y. (1994). Didactiques et savoirs professionnels, le cas du français. *Spirale*, 1, 97-112.

Rey, B., Carette, V., Defrance, A., Kahn, S. (2006). *Les compétences à l'école*. Bruxelles : De Boeck.

Rey, O. & Feyfant, A. (2014). Évaluer pour (mieux) faire apprendre. *Dossier de veille de l'IFÉ*, n°94.

Rey, O. (2012). Le défi de l'évaluation des compétences. *Dossier d'actualité Veille et Analyse*, n°76.

Ricco, G., Menotti, G., Boyer, C., Larere, C., Numa-Bocage, L. & Allenbach, L. (2008) L'hétérogénéité des rapports des élèves au domaine numérique au début et à la fin de la première année de l'école élémentaire. *Recherches en didactique des mathématiques*, 28(3), 279-318.

Richard, J.-F. (2004). Préface. Dans : C. Bastien & M. Bastien-Tonniazzo. *Apprendre à l'école*. Paris : Armand Colin.

Robert, A. & Rogalski, M. (2002). Comment peuvent varier les activités mathématiques des élèves sur des exercices ? Le double travail de l'enseignant sur les énoncés et sur la gestion en classe. *Petit x*, 60, 6-25

Rochex, J.-Y., & Crinon, J. (2011). *La construction des inégalités scolaires. Au cœur des pratiques et des dispositifs d'enseignement*. Rennes : PUR.

Roditi, E. (2015). *Le calcul et l'intelligence du calcul*. Conférence donnée lors la conférence nationale de consensus sur le « Nombres et opérations : premiers apprentissages à l'école primaire » CNESCO - Ifé (Paris nov. 2015). Repéré à <http://www.cnesco.fr/numeration/paroles-dexperts/calcul-et-operations/>

Roditi, E. (2005). L'éducation face aux théories de la construction du nombre chez l'enfant. *Spirale*, 36, 37-52.

- Rogalski, J. (2007, juin). *Approche de psychologie ergonomique de l'activité de l'enseignant*. Communication présentée au séminaire international « La professionnalisation des enseignants de l'éducation de base : les recrutements sans formation initiale »- CIEP- Paris.
- Rogalski, J. (2003). Y-a-t-il un pilote dans la classe ? Une analyse de l'activité de l'enseignant comme gestion d'un environnement dynamique ouvert. *Recherches en didactique des mathématiques*. 23(3), 343-388.
- Rouet, J.-F. (2000). *Les activités documentaires complexes aspects cognitifs et développementaux*. HDR. Poitiers.
- Sabatini, J. (2004). *Cadre d'évaluation des composantes des compétences en lecture. Programme d'évaluation et de suivi de l'alphabétisation (LAMP)*. UNESCO.
- Scallon, G. (2004 rééd. 2009). *L'évaluation des apprentissages dans une approche par compétences*. Bruxelles : De Boeck.
- Scallon, G. (1988). *L'évaluation formative des apprentissages, 1 et 2*. Québec : Presses de l'Université de Laval.
- SCEREN. (2012). *Le livret scolaire de l'école primaire*. SCEREN -CRDP Académie Orléans-Tours.
- Simard C., Dufays J.-L., Dolz J., Garcia-Debanco C. (2010). *Didactique du français langue première*. Bruxelles : De Boeck.
- Stillman, G. (2012, July) Applications and modeling research in secondary classrooms: what have we learnt? 12th International Congress on Mathematical Education. COEX, Seoul.
- Tauveron, C. (2014) Contribution aux travaux des groupes d'élaboration des projets de programmes C2, C3 et C4. MEN CSP.
- Tauveron, C. (2002). *Lire la littérature à l'école*. Paris : Hatier.
- Thevenot, C., Barrouillet, P. & Fayol M. (2010) De l'émergence du savoir calculer à la résolution des problèmes arithmétiques verbaux. Dans : Crahay M., Dutrévis M. (2010) *Psychologie des apprentissages scolaires*. Bruxelles : De Boeck.
- Thevenot C., Perret P. (2009). Le développement du raisonnement dans la résolution de problèmes : l'apport de la théorie des modèles mentaux. *Développements*, 2(2),49-56.
- Thevenot C., Barrouillet P. & Fayol M. (2004). Représentation mentale et procédures de résolution de problèmes arithmétiques : l'effet du placement de la question. *L'année psychologique*, 104(4), 683-699.
- Toczek, M.-C. (2009). *Les situations d'enseignement : un programme de recherches centré sur l'analyse des déterminants et des impacts*. HDR. Clermont-Ferrand.

- Trosseille, B. & Rocher, T. (2015). Les évaluations standardisées des élèves. *Éducation & formations, 86-87*, 15-35.
- Van Dooren, W., Verschaffel, L., Greer, B., De Brock, D. & Crahay M. (2010) La modélisation des problèmes mathématiques. Dans : Crahay M., Dutrévis M. (2010) *Psychologie des apprentissages scolaires*. Bruxelles : De Boeck.
- Vanhulle S. & Dufays J.-L. (2010) La lecture. Dans : Simard C., Dufays J.-L., Dolz J., Garcia-Debanc C. (2010). *Didactique du français langue première*. Bruxelles : De Boeck.
- Van Nieuwenhoven, C. (1996). Le comptage et la cardinalité, deux apprentissages de longue haleine qui évoluent en interaction. *Revue des sciences de l'éducation, 22(2)*, 295-320.
- Vantourout, M & Maury, S. (A paraître). Evaluation de la lecture au CP : mise en œuvre d'une approche multidimensionnelle.
- Vantourout, M., & Goasdoué, R. (2014). Approches et validité psycho-didactiques des évaluations. *Education et Formation, e-302*, 139-156.
- Vantourout, M. & Goasdoué, R. (2010, juillet). *Correction de dissertations*. Communication présentée au colloque de l'AREF.
- Vantourout, M. & Maury, S. (2006). Quelques résultats relatifs aux connaissances disciplinaires de professeurs stagiaires dans des situations simulées d'évaluation de productions d'élèves en mathématiques. *Revue des sciences de l'éducation, 32(3)*, 759-782.
- Vantourout, M., (2004). *Étude de l'activité et des compétences de professeurs des écoles et de professeurs de mathématiques dans des situations « simulées » d'évaluation à visée formative en mathématiques*. Thèse de doctorat. Université Paris 5 – René Descartes.
- Vergnaud, G. (2006). Développement cognitif et évaluation des compétences. Dans : Figari G. & Mottier Lopez M. (Eds.), *Recherches sur l'évaluation en éducation*. (167-175) Paris : L'Harmattan.
- Vergnaud, G. (2001). Forme opératoire et forme prédicative de la connaissance. Conférence publiée dans les Actes du Colloque GDM-2001. Portugais, J. (Ed) (2012) *Investigações em Ensino de Ciências, 17(2)*, 287-304.
- Vergnaud, G. (2001). Psychologie du développement cognitif et évaluations des compétences. Dans : Figari G. & Achouche M. (Eds.), *L'activité évaluative réinterrogée*, Bruxelles : De Boeck.
- Vergnaud, G. (1997). *Le moniteur de mathématiques. Résolution de problèmes cycle 3*. Paris : Nathan.
- Vergnaud, G. (1991). Structuration du domaine numérique. Dans : J. Bideaud, C. Meljac & J.-P. Fischer (1991). *Les chemins du nombre*. Lille : Presses universitaires de Lille.

Vergnaud, G. (1991). La théorie des champs conceptuels. *Recherche en didactique des mathématiques*, 10(23), 133-170.

Vergnaud, G. (1981). *L'enfant, la mathématique et la réalité*. Berne : Peter Lang.

Verret, M. (1975) *Le temps des études*. Paris : Honoré Champion.

Verschaffel L., De Corte E., Crahay M. (2008). Apprendre et enseigner les mathématiques : un cadre conceptuel pour concevoir des environnements d'enseignement-apprentissage stimulants. Dans : Crahay M. et al. (2008). *Enseignement et apprentissage des mathématiques. Que disent les recherches psychopédagogiques ?* Bruxelles : De Boeck.

Verschaffel L., Greer B., De Corte E. (2000). Making sense of word problems. *Educational studies in Mathematics*, 42(2), 211-213.

Viola S. (1998). Dis-moi quelles questions tu poses et je te dirai qui tu es ? *Québec français*, 108, 40-46.

Weiss, J. (1992). L'enseignant au cœur froid ou l'objectivité en évaluation. *Mesure et évaluation en éducation*, 14 (4), 19-31.

Weiss, J. (1986). La subjectivité blanchie ? Dans : De Ketele, J.-M. (1986) *L'évaluation : approches descriptives ou prescriptives ?* Bruxelles : De Boeck.

Ziegler, J.C., Goswami U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages : a psycholinguistic grain size theory. *Psychological bulletin*, 131 (1), 3-29.

Zorman, M. (2010, octobre). *Les dyslexies*. Communication présentée à Nantes. Repéré à http://www.apedys-vendee85.fr/uploaded/nantes%20viso_adaptation%20dys.ppt

Zorman, M. (2008). *Évaluation de la lecture en fluence*. Non publié. 12 pages.

Résumé

Mots clés : *évaluation, approche psycho-didactique, validité, évaluation interne, lecture, nombres et calcul, Cours Préparatoire (CP).*

Bien que située au cœur de l'activité quotidienne de l'enseignant dans sa classe, l'évaluation interne est une forme d'évaluation peu étudiée. La thèse s'intéresse à la partie la plus explicite de cette évaluation constituée par l'ensemble des épreuves soumises par chaque enseignant aux élèves à des fins d'évaluation, et qui servent de support aux jugements évaluatifs portés dans les « livrets scolaires ».

Plus précisément, nous étudions la validité de cette évaluation selon trois dimensions : une dimension de conformité, permettant d'apprécier comment l'évaluation remplit le cahier des charges institutionnel ; une dimension psycho-didactique, permettant d'apprécier en quoi chacune des épreuves est susceptible - ou non - de remplir les objectifs qui lui sont assignés ; une dimension didactique permettant de juger comment l'ensemble des épreuves assure la couverture conceptuelle des champs disciplinaires concernés.

Notre travail est centré sur les évaluations internes conduites aux premier et troisième trimestres par des enseignants de Cours Préparatoire en Français et en Mathématiques, plus particulièrement dans les sous-domaines à fort enjeu socio-scolaire que constituent la lecture, les nombres et le calcul.

Pour étudier la première dimension, nous sommes amené à expliciter les référentiels utilisés par les enseignants pour conduire l'évaluation. Une analyse du lien entre ces référentiels et les livrets scolaires utilisés par les enseignants est entreprise. Elle renseigne sur la conformité de l'évaluation à la demande de l'institution scolaire.

Concernant la dimension psycho-didactique, centrale dans notre travail, l'enjeu est de prendre en compte, pour chaque épreuve, les éléments qui auraient pu échapper à l'analyse proprement conceptuelle de la tâche et qui pourraient cependant s'avérer des déterminants majeurs du comportement de réponse des élèves. Cela suppose l'explicitation d'un cadre de référence - celui de l'approche psycho-didactique - qui permette effectivement d'analyser chaque épreuve et le fonctionnement cognitif potentiel des élèves confrontés à ces épreuves.

La détermination de référentiels de nature didactique (ou épistémologique) qui permettent d'analyser le contenu des tests utilisés par chaque enseignant constitue la base de l'analyse de la validité au sens de la troisième dimension (validité de couverture du test).

La conjugaison des dimensions psycho-didactique et didactique de la validité de l'évaluation permet une double analyse, globale et locale, des tests et des épreuves. La validité d'un test est dépendante de la validité des épreuves qui le constitue.

La conclusion la plus saillante est que la validité des épreuves et la validité des tests, dans bien des cas, ne sont pas assurées. Cela conduit à remettre en cause les jugements évaluatifs prononcés suite aux passations de l'évaluation. Ce résultat plaide en faveur d'une analyse de la validité conjuguant approche didactique du contenu des tests et approche psycho-didactique des épreuves composant le test.

Abstract

Key words : assessment, psycho-didactic approach, validity, internal assessment, literacy, numeracy and arithmetic, 1st Grade

Internal assessment is not often studied as a form of assessment although it is the teacher daily school activity. This thesis focuses on the most explicit part of this assessment i.e. all the tasks given to the pupils by their teacher in order to assess them. These tasks are used for evaluative judgment written in the « school report book».

More precisely, we study this assessment validity depending on three distinct issues : a conformity issue that allows us to understand how assessing may fulfill the institutional specifications; a psycho-didactics issue that allows us to see how each task is likely to fulfill the assigned goals– or not ; a didactics issue that allows us to grasp how the tests may conceptually cover the concerned disciplinary fields.

This thesis explores internal assessments in French and Mathematics conducted by 1st grade teachers during the first and third terms. We particularly focus on reading, numeracy and arithmetic that are subfields with high socio-educational stakes.

In order to examine the first issue, we explain the reference frameworks used by teachers when giving the test. French school programs are central in this area. We investigate how those reference frameworks and school reports are linked. It shows how assessment conforms to the academic institution.

When focusing on our main issue, the psycho-didactic one, we emphasize for each task on the elements that could have been missed by the conceptual analysis of the task but yet could turn out to be key factors in the pupils answering behaviors. It means a new reference framework needs to be accounted for : the psycho-didactic approach brings the analysis tools for each task and for the potential cognitive functioning of the pupils who perform those tasks.

In order to construct the validity analysis of the test from our third issue (test cover validity), we point out the didactic (or epistemo-didactic) frameworks that allows us to analyze the content of the tests given by each teacher.

In combining the psycho-didactic and didactic approaches of assessment validity, we are able to conduct a double analysis, a global one and a local one, of both tasks and tests. The validity of a test depends on the validity of its tasks.

The results mainly show that the tasks validity and the tests validity are often not ensured, which leads to question the value of evaluative judgment following the assessment. This advocates for a validity analysis combining a didactic approach of the test content to a psycho-didactic approach of the tasks.

Résumé

Mots clés : *évaluation, approche psycho-didactique, validité, évaluation interne, lecture, nombres et calcul, Cours Préparatoire (CP).*

Bien que située au cœur de l'activité quotidienne de l'enseignant dans sa classe, l'évaluation interne est une forme d'évaluation peu étudiée. La thèse s'intéresse à la partie la plus explicite de cette évaluation constituée par l'ensemble des épreuves soumises par chaque enseignant aux élèves à des fins d'évaluation, et qui servent de support aux jugements évaluatifs portés dans les « livrets scolaires ». Plus précisément, nous étudions la validité de cette évaluation selon trois dimensions : une dimension de conformité, permettant d'apprécier comment l'évaluation remplit le cahier des charges institutionnel ; une dimension psycho-didactique, permettant d'apprécier en quoi chacune des épreuves est susceptible - ou non - de remplir les objectifs qui lui sont assignés ; une dimension didactique permettant de juger comment l'ensemble des épreuves assure la couverture conceptuelle des champs disciplinaires concernés. Notre travail est centré sur les évaluations internes conduites aux premier et troisième trimestres par des enseignants de Cours Préparatoire en Français et en Mathématiques, plus particulièrement dans les sous-domaines à fort enjeu socio-scolaire que constituent la lecture, les nombres et le calcul.

Pour étudier la première dimension, nous sommes amené à expliciter les référentiels utilisés par les enseignants pour conduire l'évaluation. Une analyse du lien entre ces référentiels et les livrets scolaires utilisés par les enseignants est entreprise. Elle renseigne sur la conformité de l'évaluation à la demande de l'institution scolaire. Concernant la dimension psycho-didactique, centrale dans notre travail, l'enjeu est de prendre en compte, pour chaque épreuve, les éléments qui auraient pu échapper à l'analyse proprement conceptuelle de la tâche et qui pourraient cependant s'avérer des déterminants majeurs du comportement de réponse des élèves. Cela suppose l'explicitation d'un cadre de référence - celui de l'approche psycho-didactique - qui permette effectivement d'analyser chaque épreuve et le fonctionnement cognitif potentiel des élèves confrontés à ces épreuves. La détermination de référentiels de nature didactique (ou épistémologique) qui permettent d'analyser le contenu des tests utilisés par chaque enseignant constitue la base de l'analyse de la validité au sens de la troisième dimension (validité de couverture du test). La conjugaison des dimensions psycho-didactique et didactique de la validité de l'évaluation permet une double analyse, globale et locale, des tests et des épreuves. La validité d'un test est dépendante de la validité des épreuves qui le constitue.

La conclusion la plus saillante est que la validité des épreuves et la validité des tests, dans bien des cas, ne sont pas assurées. Cela conduit à remettre en cause les jugements évaluatifs prononcés suite aux passations de l'évaluation. Ce résultat plaide en faveur d'une analyse de la validité conjuguant approche didactique du contenu des tests et approche psycho-didactique des épreuves composant le test.

Abstract

Key words : *assessment, psycho-didactic approach, validity, internal assessment, literacy, numeracy and arithmetic, 1st Grade*

Internal assessment is not often studied as a form of assessment although it is the teacher daily school activity. This thesis focuses on the most explicit part of this assessment i.e. all the tasks given to the pupils by their teacher in order to assess them. These tasks are used for evaluative judgment written in the « school report book ». More precisely, we study this assessment validity depending on three distinct issues : a conformity issue that allows us to understand how assessing may fulfill the institutional specifications; a psycho-didactics issue that allows us to see how each task is likely to fulfill the assigned goals— or not ; a didactics issue that allows us to grasp how the tests may conceptually cover the concerned disciplinary fields. This thesis explores internal assessments in French and Mathematics conducted by 1st grade teachers during the first and third terms. We particularly focus on reading, numeracy and arithmetic that are subfields with high socio-educational stakes.

In order to examine the first issue, we explain the reference frameworks used by teachers when giving the test. French school programs are central in this area. We investigate how those reference frameworks and school reports are linked. It shows how assessment conforms to the academic institution. When focusing on our main issue, the psycho-didactic one, we emphasize for each task on the elements that could have been missed by the conceptual analysis of the task but yet could turn out to be key factors in the pupils answering behaviors. It means a new reference framework needs to be accounted for : the psycho-didactic approach brings the analysis tools for each task and for the potential cognitive functioning of the pupils who perform those tasks. In order to construct the validity analysis of the test from our third issue (test cover validity), we point out the didactic (or epistemo-didactic) frameworks that allows us to analyze the content of the tests given by each teacher. In combining the psycho-didactic and didactic approaches of assessment validity, we are able to conduct a double analysis, a global one and a local one, of both tasks and tests. The validity of a test depends on the validity of its tasks.

The results mainly show that the tasks validity and the tests validity are often not ensured, which leads to question the value of evaluative judgment following the assessment. This advocates for a validity analysis combining a didactic approach of the test content to a psycho-didactic approach of the tasks.