

Processus de référentialisation de l'évaluation et validité : analyse de livrets scolaires et d'épreuves d'évaluation de la compréhension en lecture en cours préparatoire

Assessment Referentialization Process and Validity: Analysis of School Reports Books and Reading Comprehension Assessment tests in 1st Grade

Christophe Blanc – christophe.blanc@ac-paris.fr – <https://orcid.org/0009-0005-7737-6772>
Université Paris Cité - France

Pour citer cet article : Blanc, C. (2024). Processus de référentialisation de l'évaluation et validité : analyse de livrets scolaires et d'épreuves d'évaluation de la compréhension en lecture en cours préparatoire. *Évaluer. Journal international de recherche en éducation et formation*, 10(1), 59-83. <https://doi.org/10.48782/e-jiref-10-1-59>

Résumé

Cette étude s'intéresse à l'effet des processus de référentialisation de l'évaluation propres à seize enseignants expérimentés en cours préparatoire (CP- 6ans) dans le cadre de l'évaluation interne instituée. Elle porte particulièrement sur l'analyse de deux produits de leur activité : le choix des tâches pour évaluer et le livret scolaire abritant les jugements évaluatifs résultant de l'évaluation et transmis aux familles. Elle sera conduite sous l'angle de la validité envisagée dans différentes dimensions (didactique, curriculaire, pédagogique).

L'étude permet de mettre en exergue des conceptions hétérogènes des évaluateurs et de prouver la faible validité de l'évaluation. Par-delà ces résultats, elle invite à en interroger les fondements et à esquisser quelques pistes dont celle de la formation des enseignants à la cinquième des compétences commune à tous les professeurs du référentiel métier : « Evaluer ».

Mots-clés

Évaluation, référentialisation, validité didactique, évaluation interne, cours préparatoire, compréhension en lecture

Abstract

This study focuses on the effect of assessment referentialization processes specific to sixteen experienced teachers in cours préparatoires (CP- 1st Grade -6 years) within the framework of the established internal evaluation. It particularly focuses on the analysis of two products of their activity: the choice of tasks to evaluate and the school report containing the evaluative judgments resulting from the evaluation and transmitted to families. It will be conducted from the angle of validity considered from different dimensions (didactic, curricular, pedagogical).

The study highlights heterogeneous conceptions of evaluators and proves the low validity of the evaluation. Beyond these results, it invites us to question their foundations and to outline some avenues, including that of training teachers in the fifth skill common to all teachers in the professional framework: "Assess".

Keywords

Assessment, referentialization, didactic validity, curriculum validity, internal assessment, 1st Grade, reading comprehension

1. Introduction

Si, dans le cas de l'évaluation interne instituée¹ (Barbier, 1985, p.34), évaluer relève du « défi » pour un enseignant (Rey, 2012), c'est qu'il a la charge d'opérationnaliser une double prescription institutionnelle relevant du contrat professionnel (Rogalski, 2003) : *évaluer* et *informer*². Pour ce faire, et à moins qu'un ensemble de tâches utilisées à des fins d'évaluation, ou épreuves lui soit fourni, il lui revient la responsabilité de les sélectionner ou de les élaborer, et de formaliser le résultat de l'évaluation, le jugement évaluatif, pour le communiquer au moyen d'un livret scolaire.

Citant l'enquête Cnesco-Crédoc 2022, les rédacteurs du dossier de synthèse de la conférence de consensus sur l'évaluation en classe soulignent que « 3 parents sur 4 considèrent les résultats scolaires comme l'élément le plus important de la scolarité de leur enfant » (Florin *et al.*, 2023, p.24). Or, indiquent-ils, « l'évaluation en classe (est) faiblement encadrée », constituant une « source de malentendus et de tensions » (2023, p.13), d'où la nécessité de s'intéresser, entre autres, à sa « qualité ».

Pour cela, nous nous intéresserons aux épreuves d'évaluation utilisées par des enseignants de cours préparatoire (CP – 6 ans) et aux livrets scolaires leur permettant de communiquer le résultat de l'évaluation. La qualité que nous cherchons à observer s'apparente à la validité, et nous conduit à poser une série de questions à ce sujet : ce qui est évalué correspond-il à ce qui a été annoncé être évalué ? Cela est-il conforme à ce qui a été enseigné ou ce qui a été prescrit ? L'activité de réponse des évalués correspond-elle à celle que l'évaluateur souhaite apprécier ? Les épreuves couvrent-elles conceptuellement le champ disciplinaire étudié ? Le jugement évaluatif est-il fondé ? Si l'on se place dans une perspective formative de l'évaluation ou plus largement de soutien d'apprentissages (ESA), alors il est nécessaire de s'assurer que celle-ci est bien valide et renseigne précisément sur les échecs et les réussites des élèves.

Nous nous inscrivons pleinement dans la lignée de travaux menés au sein du laboratoire EDA (Education, Discours, Apprentissages) de l'Université Paris Cité et du réseau EVADIDA de l'ADMEE autour de la validité, et plus particulièrement de la validité didactique des épreuves d'évaluation (voir Vantourout, 2020, pour une synthèse). Délaissant toute approche psychométrique, l'approche qualitative adoptée nous est utile pour analyser finement les épreuves collectées, en prenant en compte l'activité de réponse des évalués et en interrogeant les objets de savoir mobilisés. Nous avons précédemment montré que la présence d'illustrations était susceptible d'avoir une influence sur l'activité de réponse des évalués, de nature à engendrer un écart entre la tâche attendue et les tâches redéfinies et réalisées (Blanc *et al.*, 2018). Dans ce cas précis, la validité de ces tâches serait dès lors remise en cause pour une partie des élèves, ce qui constitue une problématique eu égard à la production de jugements évaluatifs faisant suite à la passation de ces épreuves, lesquels seraient alors susceptibles d'être mal fondés. L'analyse développée dans cet article nous permet de prolonger cette réflexion autour de la validité didactique d'épreuves de compréhension en lecture. En ce qui concerne les livrets, le format étant libre au moment de

¹ Cette forme d'évaluation régulière est dite **interne** en ce qu'elle est conduite au cœur de la classe. Elle s'oppose en cela à l'évaluation externe standardisée. Elle est en outre **instituée** dans la mesure où elle constitue la partie explicite de l'évaluation interne sur laquelle les enseignants se fondent pour rédiger le livret scolaire de l'élève.

² « 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 » (Code de l'éducation, D321-6).

L'étude, nous interrogerons leur contenu en ce qu'ils constituent un référentiel pour évaluer et nous analyserons l'effet des choix de ces référentiels sur l'évaluation.

Après avoir délimité le cadre théorique et méthodologique, didactique et qualitatif, dans lequel nous inscrivons notre travail, nous présenterons le résultat de nos analyses. Celles-ci seront conduites sous l'angle de la validité et porteront sur deux produits de l'activité évaluative de l'enseignant : le livret scolaire de l'élève et les épreuves d'évaluation. Nous nous limiterons aux épreuves du sous-domaine disciplinaire du Français que constitue la lecture. Nous concluons notre texte en recontextualisant les résultats obtenus et en soulignant l'apport de cette recherche à une meilleure compréhension de l'activité évaluative des enseignants, susceptible de contribuer à une professionnalisation des enseignants à l'évaluation que Jorro (2009) appelle de ses vœux.

2. Cadrage théorique

2.1. *Construction du référent de l'évaluateur*

Si évaluer est un processus de « transformation de représentations » (Barbier, 1985, p.63-64), il est nécessaire de penser le référé - « modèle réduit » de « l'ensemble de l'information qu'il a été possible de produire pour l'évaluation » selon Hadji (2018, p.89) - et le référent de l'évaluation - « système temporaire de critères », « modèle réduit de la réalité » (p.89) - comme des représentations.

Selon Hadji, la production d'un jugement évaluatif par confrontation entre un système d'attentes exprimées sous forme de critères et une réalité saisie à travers des signes ou éléments concrets jouant le rôle d'indicateurs constitue un « élément structurel constant de l'activité d'évaluation » (2018, p.85-86), l'évaluation disant « si le réel est acceptable par rapport à l'attendu » (p.88). Cela exigerait une double articulation : entre des éléments factuels (indicateurs) et des attentes (critères), d'une part, et entre les éléments retenus pour juger (qui forment un système complexe d'attentes possibles à l'égard de la réalité évaluée) avec une réalité qui les englobe et les dépasse, d'autre part.

La question de la rigueur de la spécification des attentes occupe ainsi une place centrale dans l'évaluation dont le produit sera transmis sous forme interprétative via le livret scolaire. Rapportée à notre étude, la spécification des attentes précédemment évoquée est double. Elle comprend la spécification du système d'attentes en lui-même et la spécification des tâches choisies pour y correspondre le plus fidèlement possible. Or souligne Figari (2006, p.103), « le référent fait l'objet d'un choix de l'évaluateur, (...) (et) représente une certaine conception de l'évaluation ». Ce choix s'opère dans un contexte cadrant de référentiels renvoyant eux-mêmes à un ensemble de référents.

Dans le cadre de notre étude, plusieurs référentiels cohabitent parmi lesquels :

- Les programmes³. Rédigés pour indiquer ce qu'il convient d'enseigner, ils orientent assez logiquement l'activité évaluative de l'enseignant autour des objectifs d'apprentissage prescrits.

³ Consultables pour le CP à https://www.education.gouv.fr/bo/2008/hs3/programme_CP_CE1.htm

- Les progressions annexées aux programmes⁴. Présentées comme de simples repères, elles constituent en fait une liste d'objectifs plus opérationnels que les programmes, différant parfois dans leur formulation, et susceptibles de constituer un référentiel pour évaluer à partir duquel chaque enseignant élabore son propre référentiel.

L'enseignant, pour décider de ce qu'il va évaluer, dispose de fait d'une pluralité de référentiels, « organisateurs du travail » (Amigues, 2009), qui, s'ils sont tous rattachés aux programmes, sont plus ou moins opérationnels et articulés. Il doit les traduire dans des proportions qui peuvent être variables afin de matérialiser son propre système d'attentes par deux éléments concrets : le livret scolaire et les épreuves d'évaluation.

Ces éléments (livret, épreuves) peuvent être considérés comme le produit d'un processus complexe de référentialisation, terme qui désigne « 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. Dans ce sens, toute évaluation est tributaire de sa référentialisation. » (Figari, 2006, p. 104). S'intéresser à la référentialisation conduit donc à s'intéresser au système d'attentes de l'évaluateur, aux modélisations de l'évaluation, à la détermination de critères (ce qu'on estime légitime d'attendre) et d'indicateurs (ce qui s'est manifesté dans la réalité à l'égard du critère) qui eux-mêmes ne peuvent faire l'économie d'un référent didactique, comme le souligne Vergnaud (2006, p.167).

Eu égard au retentissement identitaire d'un échec sur les biographies évaluatives (Baudouin et Pita, 2011, p.93), au poids des jugements évaluatifs sur les trajectoires scolaires, il est important de s'intéresser à la qualité du produit de l'activité de référentialisation de l'évaluateur, à son référent, que l'on peut appréhender par deux construits observables : le livret scolaire, les épreuves d'évaluation. Une façon de s'y intéresser consiste à interroger cette qualité sous l'angle de la validité selon deux axes : l'un est celui de la dimension de conformité au prescrit (validité curriculaire), l'autre de la qualité théorique des épreuves choisies (validité didactique).

2.2. Le livret scolaire

Puisqu'évaluer est un processus fondé sur des représentations visant la communication d'un jugement, on peut s'intéresser au produit de l'activité conduisant un enseignant, au terme du processus de référentialisation, à rendre tangible son système d'attentes et sa traduction dans le livret scolaire communiqué aux familles.

Deux options s'offrent concrètement à lui en vertu de la liberté pédagogique qui lui est attribuée. Il peut construire le livret scolaire en fonction de ce qui a été enseigné et seulement cela et y reporter les résultats de l'évaluation. Dans ce cas, le référentiel pour évaluer de l'enseignant est ce qui a été enseigné. Il peut aussi choisir un autre type de livret scolaire davantage articulé à la prescription et qui présente l'ensemble des compétences devant être

⁴ Consultables pour le CP à <https://www.education.gouv.fr/bo/2008/hs3/apprentissages.htm> « Les tableaux suivants donnent des repères aux équipes pédagogiques pour organiser la progressivité des apprentissages ». Les programmes 2008 sont en vigueur dans le contexte de cette étude. Ils seront révisés en 2015. Il en subsiste une telle organisation, avec d'un côté des programmes établis pour un cycle d'apprentissage de 3 ans explicitant des attendus de fin de cycle et, d'un autre côté, des points de repères annuels : <https://www.education.gouv.fr/l-evaluation-des-acquis-des-eleves-en-cp-des-reperes-pour-la-reussite-5318> . Pour simplifier la lecture, nous les désignerons dans le texte par les Progressions.

maîtrisées par les élèves. Dans ce cas, le référentiel pour évaluer de l'enseignant est ce qui doit être enseigné par le professeur et maîtrisé par l'élève.

Dans le premier cas, l'analyse conduirait à interroger la validité pédagogique de l'évaluation. Dans le second cas, elle mène à observer sa validité curriculaire. Sans entrer ici dans une discussion sur les curriculums, nous retiendrons à la suite de Crahay (2011, p11) qu'« en définitive, toutes les définitions renvoient à l'idée de plan et d'organisation. Lorsqu'on parle d'un curriculum, on parle (...) d'un *artefact culturel* qui a le projet d'influer sur les processus d'enseignement apprentissage qui se déroulent concrètement dans les classes ». La question de savoir ce qu'il faut évaluer et comment le faire restant, dans le cadre de l'évaluation interne de notre étude, à charge de l'enseignant. Rechercher la validité curriculaire de l'évaluation renvoie pour Scallon (2004, p. 241) au besoin « d'établir que le contenu d'un examen, ou d'un procédé de vérification, correspond bien à ce qui a été enseigné ». Or entre ce qui a été enseigné, ce qui doit l'être et ce qui est enseigné mais n'est pas nécessairement inscrit dans un curriculum (le curriculum caché), des écarts sont constatés. De Landsheere (1988) qualifie cette forme de validité de *validité pédagogique* c'est-à-dire 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, Crahay, 1996, p. 39). La *validité curriculaire* serait alors à rechercher du côté d'une conformité au prescrit plutôt qu'au réalisé. Elle correspondrait au fait de savoir si ce qui est évalué correspond à ce qui est prescrit d'évaluer, lequel est formulé par le curriculum.

Ces différences lexicales font apparaître un nouveau problème concernant le prescrit. L'institution prescrit ce qu'il faut enseigner mais pas ce qu'il convient d'évaluer. Le référentiel le plus opérationnel pour évaluer se situe en Annexe des programmes, sous la forme de Progressions (listes de compétences), comme nous le verrons plus loin.

Les enseignants sont libres de bâtir le livret scolaire comme ils le souhaitent, ce qui donne lieu, avant l'instauration d'un Livret Scolaire Unique (LSU) en 2016, à une grande variété de formats de livrets, y compris entre deux enseignants dans le même niveau au sein d'une même école. Le format même du livret scolaire et son contenu sont donc une première piste d'appréhension du processus de référentialisation de l'évaluateur : sur quoi porte l'évaluation ?

2.3. Les épreuves

Prescrites aux élèves, les épreuves constituent le siège de leur activité de réponse que l'enseignant analyse dans l'optique de produire le jugement évaluatif transmis aux parents via le livret scolaire. Sans que l'on sache précisément si la détermination de l'épreuve suit la détermination de la compétence à évaluer ou l'inverse, l'enseignant sélectionne les épreuves en fonction d'un modèle plus ou moins précis de l'activité de l'élève au cours de la réalisation de la tâche ou du moins d'une représentation des connaissances mobilisées par l'élève.

Classique en psychométrie dès lors que l'on s'intéresse à un test, un certain nombre de preuves de validité peuvent être recherchées (Laveault et Grégoire, 2014; Leclercq, 2006). Dans le cas de l'évaluation interne instituée, une entrée indépendante de questions relatives à la mesure et portant sur la recherche de preuves de validité des épreuves s'impose, puisque le choix des épreuves est effectué par chaque enseignant et que certaines composantes des épreuves peuvent être minorées, ou passer inaperçues. Selon Bastien et Bastien-Tonniazzo (2004), en effet « 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 » (p.11).

C'est pourquoi nous adoptons une approche didactique qualitative de la validité pour examiner les épreuves collectées auprès d'enseignants (Vantourout, 2020). Elle repose sur des analyses *a priori* des tâches (Grapin et al., 2022). L'expression « validité didactique » (Grapin et Grugeon-Allys, 2018) permet d'intégrer celle de validité psycho-didactique –VPD (Blanc et Vantourout, 2018; Vantourout et Goasdoué, 2014) et celle de validité épistémologique –VED (Grugeon-Allys et Grapin, 2016).

La VPD est une approche qui a pour origine un questionnement «local⁵» sur la validité des tâches qui sont utilisées en classe par les enseignants pour évaluer leurs élèves (Vantourout et Goasdoué, 2014), indépendamment de toute considération psychométrique. Le principe essentiel qui fonde la VPD est que « l'évaluation n'a pas pour but de décrire le niveau d'un individu sur une échelle décrivant à son tour une position sur un trait latent, mais de décrire/qualifier le fonctionnement cognitif à partir de l'activité de réponse qui est pour une part déterminée par le couple sujet-situation » (p. 141).

La VPD conjugue les apports des didactiques disciplinaires et de l'ergonomie cognitive : « on complète l'analyse conceptuelle, propre aux didactiques [et impliquant une analyse *a priori*], par une analyse de la tâche, en s'interrogeant sur ce que la tâche et ses spécificités sont susceptibles d'engendrer quant à l'activité de réponse de l'évalué » (Blanc *et al.*, 2018). Pour cela, nous nous référons au modèle d'analyse psychologique de l'activité de Leplat et Hoc (1983), revisité par Rogalski (2003) qui distingue tâche et activité.

L'analyse au niveau local (c'est-à-dire au niveau de chaque tâche prise isolément) peut être complétée au niveau global (c'est-à-dire au niveau de l'ensemble des épreuves que l'on peut assimiler dans le cas de l'évaluation interne instituée à un test) de sorte à examiner la couverture d'un domaine disciplinaire à partir de l'ensemble des analyses *a priori* de toutes les tâches d'un test relevant d'un même domaine.

2.4. Validité didactique d'épreuves d'évaluation de la compréhension en lecture

En lecture, 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 » (PIREF, 2003 ; Blanc, 2010).

Un autre consensus semble prévaloir au niveau des pratiques évaluatives : celui du recours à des questionnaires (Fijalkow, 2003 ; Goigoux et Cèbe, 2014, p.9). Le rapport du PIREF (2003) précisait qu'(ils) « sont des moyens imparfaits d'évaluation qui n'ont pas d'efficacité avérée dans la maîtrise de la lecture » sans toutefois proposer d'alternatives opératoires. Les élèves passent beaucoup de temps à traiter des tâches de lecture-compréhension de manière individuelle et souvent hors de la présence de l'enseignant, souligne Goigoux (2016, p.400) : l'enquête *Lire Ecrire* rapporte que les tâches de type « réaliser une tâche écrite impliquant la compréhension explicite et/ou implicite » occupent, en moyenne, plus du tiers du temps global alloué à la compréhension (ibid., p.229). Or Cèbe *et al.* (2004, p. 9-10) relèvent l'importance de « distinguer les compétences qui permettent de comprendre les textes de

⁵ « Local » au sens où chaque tâche, exercice, question, item est considéré de manière « isolée » lorsque l'on s'assure de sa validité. La VPD n'apporte pas de réponse quant à la couverture « didactique » du domaine, contrairement à la VED.

celles qui sont sollicitées par le traitement des questionnaires », laissant penser qu'une compétence spécifique telle que « traiter un questionnaire » serait en réalité évaluée dans les épreuves de compréhension en lecture de façon implicite, résultat avec lequel nos travaux convergent (Blanc et Vantourout, 2018 ; Blanc, 2015).

Une fois la compétence identifiée, l'enseignant détermine une épreuve permettant de l'évaluer. Nous avons déjà montré l'impact des illustrations sur les comportements de réponse des élèves et sur la validité des tâches : les illustrations, même lorsqu'elles sont plutôt a priori dotées d'une simple fonction « motivationnelle », sont en fait susceptibles d'intervenir au niveau du processus de redéfinition de la tâche. Elles sont ainsi une source potentielle d'écart entre la tâche attendue par l'enseignant et celles effectivement réalisées par les élèves, ce qui conduit à remettre en cause la validité de ces épreuves, au sens des approches psycho didactiques des évaluations (APDE) (Blanc *et al.*, 2018).

Nous nous intéresserons ici à la modalité d'évaluation de la compréhension en lecture que représente le questionnaire, modalité largement choisie pour évaluer la compréhension en lecture en CP.

Ainsi l'objet de cet article porte-t-il sur une analyse de certains effets résultant de processus de référentialisation propres à des enseignants de CP dans le cadre de l'évaluation interne instituée qu'ils opérationnalisent auprès de leurs élèves. Pour cela, une méthode qualitative originale d'analyse de la validité sera adoptée selon différentes dimensions complémentaires (validité didactique, validité curriculaire, validité pédagogique). Elle prendra appui sur quelques exemples représentatifs issus d'un corpus de livrets et d'épreuves d'évaluation collectées.

3. Méthodologie

Les éléments étudiés sont extraits d'un travail plus ample (Blanc, 2017) dans lequel on s'intéresse à divers types de validité d'épreuves utilisées par des enseignants de CP dans deux sous-domaines disciplinaires : « lecture » en Français et « nombres et calcul » en Mathématiques. Ils ont été collectés auprès de 16 enseignants expérimentés, plus d'un quart sont formateurs, exerçant en CP dans 13 écoles différentes réparties sur 6 circonscriptions scolaires.

Deux types d'éléments 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.
- Un exemplaire du livret scolaire transmis aux familles concernant l'élève dont les copies ont été collectées.

Les documents ont été collectés à la fin du premier trimestre et à la fin du troisième trimestre.

Des entretiens post-collecte, semi-directifs ont été entrepris auprès des enseignants, visant en particulier à préciser le référencement des 1 184 épreuves selon les 61 items des Progressions Ils ont principalement permis de préciser l'affectation de chaque épreuve utilisée par les enseignants aux items de leurs livrets. Ils ont également permis de collecter des réactions et commentaires librement formulés. Ceux-ci seront utilisés de façon marginale dans le cadre de cet article.

4. Résultats

4.1. Les livrets

Quatre résultats peuvent être mis en exergue.

4.1.1 Quels référentiels pour évaluer ?

Dans le but d'outiller la nécessité de communication des acquis des élèves, un livret scolaire type a été publié en 2010 par le SCEREN⁶ – CRDP de l'Académie d'Orléans-Tours. Préfacé par un inspecteur général, il se situe du côté de la prescription et tente d'établir une liaison opérationnelle entre les programmes et l'évaluation, s'insérant dans une dynamique de guidage de l'action. Il a été largement plébiscité par les enseignants : il est économique dans la mesure où il évite de grandes quantités de photocopies⁷ chaque trimestre, c'est un outil éditorial « clé en main » validé par l'institution, il regroupe en un seul document tous les livrets successifs de l'élève...

Suite à une préconisation faite au niveau de la temporalité de l'évaluation, les auteurs précisent, dans l'introduction, 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 ». Cela constitue une proposition inédite d'opérationnalisation de l'évaluation. Le caractère inédit de la proposition renvoie au fait qu'elle provient de l'institution elle-même et non des enseignants, lesquels procédaient sans doute pour partie déjà ainsi.

Les auteurs ont manifestement eu conscience de la grande quantité d'items⁸ contenus dans les Progressions (40 en français, 21 en mathématiques), et ont insisté sur le fait que cela n'impliquait pas une évaluation reposant sur autant d'épreuves. Toutefois, précisent-ils, « en principe, en fin d'année, tous les items, notamment en mathématiques et en français, doivent avoir été travaillés, donc cochés ». Cela sous-entend que l'évaluateur doit évaluer tous les items mais pas nécessairement de façon individuelle et en temps limité. Un fort risque existe alors d'une évaluation subjective et implicite.

Par le biais de ce livret, ce qui est présenté comme des « repères » devient de fait une norme référentielle pour l'évaluation, formulant ce qu'il faut évaluer. L'enseignant s'en saisit ou non : nous constatons des traductions-reformulations des items des Progressions.

⁶ En 2014, le réseau SCÉRÉN-CNDP, opérateur essentiel à la mise en œuvre de la stratégie pour « Faire entrer l'école dans l'ère du numérique », se transforme et devient Canopé, le réseau de création et d'accompagnement pédagogiques.

⁷ Du fait du principe républicain de gratuité de l'école, les dépenses de fonctionnement et d'équipement des écoles relèvent des municipalités qui leur allouent un crédit leur permettant de commander manuels et autres outils pédagogiques sur catalogue. Celui-ci n'incluant généralement pas de photocopies dans de grandes quantités, celles-ci restent à charge de l'école. Les écoles n'ayant pas le statut d'établissement ne disposent pas de trésorerie de fonctionnement. Le financement de photocopies dans une école constitue un sujet souvent sensible.

⁸ Entendre le mot « item » comme « une formulation de la compétence » à des fins de communication.

4.1.2 Évaluer ce qui a été enseigné ou évaluer parmi ce qui doit être enseigné ?

Jusqu'à l'apparition du LSU en 2016, il n'y pas de modèle de livret imposé et l'on en observe une grande variété : dans notre étude, dans 13 cas sur 16, un livret scolaire différent est utilisé quand bien même les enseignants enseignent dans la même école au même niveau.

Les livrets se présentent tous sous la forme de listes de compétences et sont, dans la plupart des cas (pour 10 enseignants sur 16), très proches des listes fournies dans les Progressions et du livret SCEREN. Les enseignants effectuant un autre choix reportent dans le livret scolaire de l'élève les items travaillés durant la période considérée.

6 enseignants utilisent un livret différent entre le 1er et le 3e trimestre, 3 emploient le livret SCEREN et 7 adoptent un livret très proche du précédent, identique pour l'année, largement fondé sur les items des Progressions.

Les livrets utilisés par les enseignants semblent traduire deux conceptions de l'évaluation :

La première, « pédagogique », consisterait à faire état dans le livret de ce qui a été enseigné sur la période considérée et seulement de cela. Dans ce cas, nous constatons que 100% des items du livret sont renseignés. La validité pédagogique est forte. Mais il est difficile d'apprécier cette qualité de l'évaluation si l'on ne peut la rapporter à ce que les élèves doivent apprendre sur l'année. La formulation des compétences, généralement moins précise que dans les Progressions, n'aide pas à établir cette correspondance.

La seconde, « curriculaire », renverrait à un usage du livret visant à faire état de ce qui a été enseigné parmi tous les items des Progressions (le prescrit ou plus exactement une traduction du prescrit puisque les programmes ne prescrivent pas ce qu'il faut évaluer). Dans ce cas, le taux de remplissage des items du livret est variable et, quels que soient l'évaluateur et le trimestre considéré, n'est jamais égal à 100%. Puisque les enseignants évaluent systématiquement au troisième trimestre des items évalués au premier trimestre, cela signifie que certains items ne sont jamais évalués. L'enseignant P, par exemple, ne renseigne que la moitié des items du livret qu'il a conçu, amenant à une question que l'on ne peut trancher : a-t-il tout de même enseigné les compétences non évaluées ? Le guidage de l'évaluation par les items des Progressions constitue de facto une contrainte importante à respecter : il y a trop d'items à évaluer. La validité curriculaire est faible. Mais pourtant, l'enseignant ajoute aux items des Progressions de nouveaux items pour mieux répondre à ce qu'il estime être une commande institutionnelle : « on a eu plusieurs conférences imposées sur ce sujet, il fallait qu'on le fasse ». L'enseignant N adopte un autre raisonnement et indique « ne (pas) respecte(r) pas les IO⁹ plus que ça. C'est clair comme ça pour les parents et pour nous » pour justifier le fait que son livret tienne sur une seule page.

Ce résultat fait apparaître des représentations de nature différente quant à ce qu'il convient d'évaluer et comment en rendre compte.

4.1.3 Évaluer davantage ce qui semble facile à évaluer ?

Nous reproduisons ci-dessous la répartition des épreuves d'évaluation de Français collectées au 3^e trimestre selon le sous-domaine sur lequel porte l'épreuve. En moyenne, 14,1 épreuves sont utilisées par les enseignants avec un écart type de 7,9 épreuves.

⁹ IO : Instructions Officielles

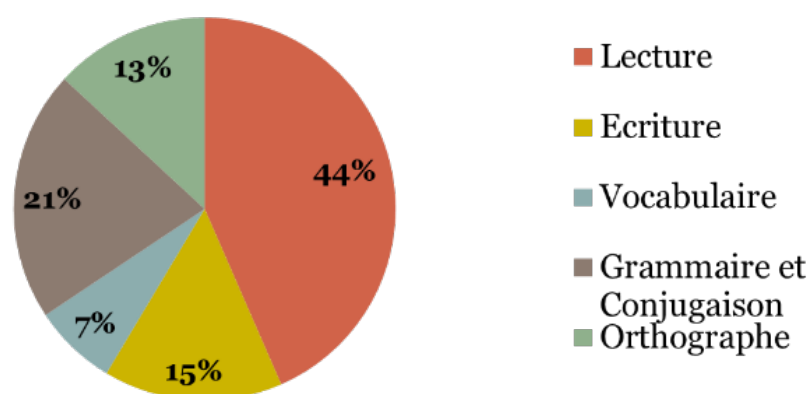


Figure 1 : Répartition en % des 226 épreuves d'évaluation du troisième trimestre selon le sous-domaine du Français évalué

Si l'on se base sur les affectations des épreuves aux items telles que les enseignants les décident, un déséquilibre est constaté entre les différents sous-domaines : si la Lecture concentre la majorité des épreuves (44%), l'Écriture donne lieu, en comparaison, à trois fois moins d'épreuves environ en fin de CP (15%) tandis que les savoirs relevant de la maîtrise de la langue (Vocabulaire V, Grammaire G, Orthographe O, Conjugaison C) concentrent 41% des épreuves.

Ce résultat est à articuler à la prescription qui détaille davantage les repères de progressivité dans les sous-domaines G, C et O qu'en Écriture. Ce résultat témoigne également d'une tendance qui s'affirme dès la fin du CP autour de l'évaluation des compétences formelles de la langue. La facilité de construire des épreuves courtes portant sur des aspects ciblés (une dictée de mots par exemple) peut contribuer à expliquer ce résultat. Ce résultat est à rapprocher de ceux de Grisay, obtenus suite à une étude portant sur les questions de français recueillies dans les épreuves d'examen de 33 classes de première primaire en 1992 :

Tableau 1 : Comparaison des distributions des épreuves de Français dans les tests collectés pour notre étude et ceux étudiés par Grisay (1992, p. 139)

	Lecture	Ecriture	Vocabulaire V	Grammaire G Conjugaison C	Orthographe O
% d'épreuves se rapportant aux sous-domaines dans notre travail au troisième trimestre (N = 16 classes)	43%	15%	7%	21%	13%
% d'épreuves se rapportant aux sous-domaines relevés par Grisay (1992) (N= 33 classes)	27%	12%	18%	18%	24%

A l'exception de la Lecture, une certaine similitude des distributions peut être constatée entre les deux études. Elle met en avant la part des « compétences formelles » (cf. Grisay, 1992, p.128) dans l'évaluation des compétences de Français, déjà importante en fin de CP et qui 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 les évaluer là même, où en 1992, on ne trouvait qu'une description assez générale du savoir lire. Pour comparaison, les Programmes français de 1985 ne contiennent qu'un seul intitulé concernant la lecture en CP : « apprentissage de la lecture » (qui est défini de façon plus détaillée dans la partie Instructions).

4.1.4 Malgré un pilotage par les items des Progressions, une évaluation hétérogène

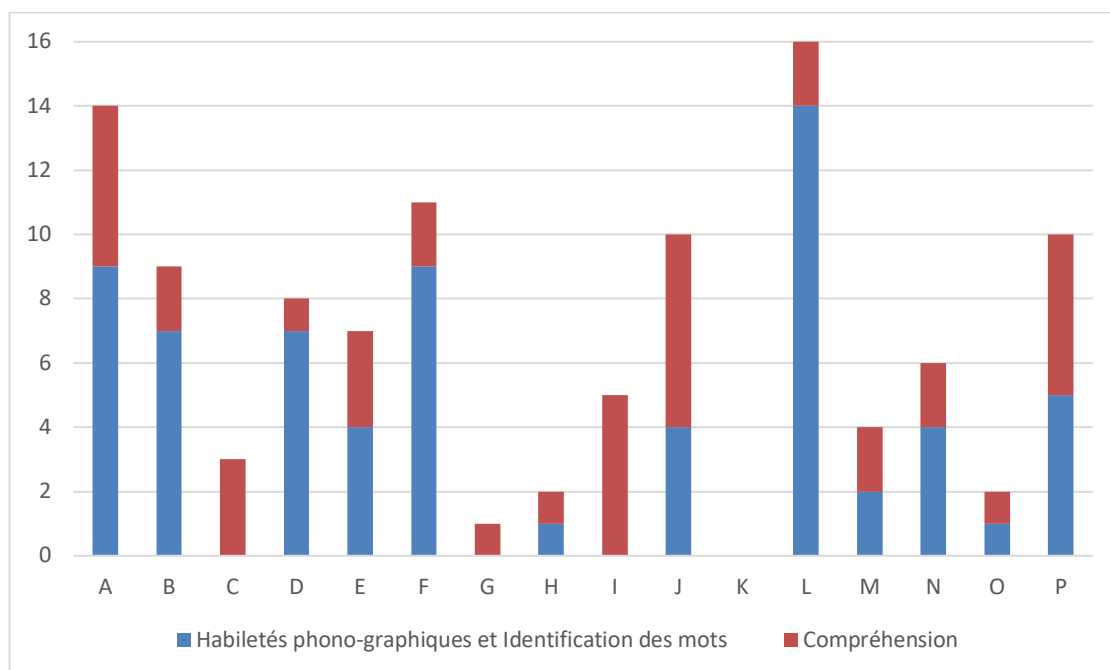


Figure 2 : Nombre d'épreuves utilisées par chaque enseignant au troisième trimestre, par groupe d'items de Lecture

On peut affiner l'analyse en s'intéressant aux répartitions des épreuves utilisées par chaque enseignant par item d'un même sous-domaine. À titre d'exemple, en Lecture, sous-domaine le plus évalué en Français, on peut regrouper les items associés aux compétences de compréhension et les items associés aux habiletés de base (relations phonographiques, identification des mots), et observer comment chacun de ces deux groupes d'items est représenté.

Les résultats montrent d'importantes disparités entre enseignants au sujet du nombre d'épreuves utilisées mais aussi dans la répartition des épreuves selon les compétences de Lecture évaluées :

- L'enseignant L évalue massivement les habiletés de base au moyen de 14 épreuves et n'évalue les compétences de compréhension qu'au moyen de 2 épreuves.
- L'enseignant C n'utilise pas d'épreuve portant sur la maîtrise des habiletés de base bien que les programmes insistent sur ce point. Il renseigne toutefois cet item dans le livret.

- L'enseignant K n'utilise aucune épreuve de lecture bien qu'il ait rempli les items du livret. Cela permet de souligner le recours parfois fait par les enseignants à une évaluation implicite qui échappe à notre observation.

4.2. Validité didactique d'épreuves d'évaluation de la compréhension en lecture

Nous nous intéressons à présent à l'examen de la validité didactique de deux épreuves d'évaluation de la compréhension en lecture en fin de premier trimestre de CP. Dans notre corpus, les enseignants proposent des tâches qui sollicitent soit une activité de réponse fondée sur la production d'un dessin, soit le choix d'une réponse parmi plusieurs (nous écartons les tâches exigeant une réponse écrite à une question écrite). Les épreuves analysées ci-après constituent ainsi un échantillon représentatif des épreuves collectées.

Dans les deux épreuves discutées ci-après, les élèves répondent à des questions de compréhension en dessinant (épreuve 1), ou en produisant une réponse écrite, ou en choisissant une bonne réponse parmi plusieurs (épreuve 2).

4.2.1 Épreuve 1 : évaluer la compréhension d'un texte au moyen du dessin

Pour évaluer la compétence, désignée par la compétence du livret « *Comprendre un texte lu* » en fin de premier trimestre, un enseignant conçoit et utilise l'épreuve suivante :

9- Dessine ce qui est écrit dans le texte.

Léon est sur le ponton. Il a un bâton. Il y a un crocodile dans l'eau.

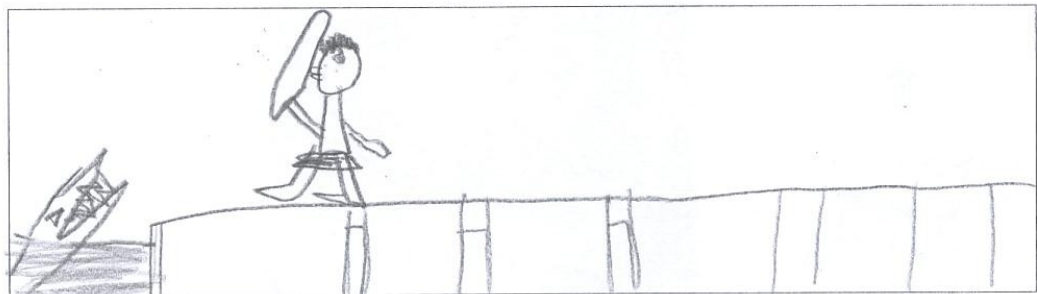


Figure 3 : Épreuve de Lecture du 1er trimestre avec réponse d'élève

Pour réussir l'épreuve, l'élève doit lire un texte de 16 mots, que l'on peut estimer adapté à la capacité de décodage d'élèves de fin de 1^{er} trimestre de CP et sur lequel porte une consigne de 8 mots, et manifester sa compréhension en produisant un dessin.

Analysons successivement trois composantes de la tâche dans leur ordre d'apparition :

La consigne, texte sur le texte dont la compréhension est évaluée, lue par l'enseignant. Elle commande de « *dessiner ce qui est écrit* ». Cette prescription pourrait donner lieu à deux redéfinitions selon le sens attribué à « *ce* », inférence laissée à la charge de l'élève, que l'on pourrait tenter (ce ne sont là que des hypothèses sur l'activité de l'élève) de formuler ainsi :

- l'une conduirait à dessiner *stricto sensu* chaque élément dans l'ordre d'apparition dans le texte : « Léon », « sur + un ponton », « un bâton (dans la main de Léon, ou sur son dos...) », « un crocodile dans l'eau ». Un tel traitement de la consigne pourrait

conduire l'élève à dessiner le crocodile dans l'eau en dernier, donc plutôt à droite du dessin, si l'on estime que les élèves commenceront leur dessin dans le sens de l'écriture. L'activité de lecture sous-tendue serait alors une lecture pas-à-pas, fragmentée, dont chaque étape serait finalisée par un dessin.

- L'autre serait fondée sur une interprétation de la consigne. Il ne s'agit pas tant de « dessiner ce qui est écrit » que d'« illustrer la situation décrite par le texte ». Auquel cas, le dessin intégrerait les différents éléments selon leur relation qu'elle soit explicitement formulée (Léon/bâton ; Léon/ponton) ou inférée (l'interaction Léon/crocodile). L'activité de lecture sous-tendue viserait plutôt l'élaboration d'un modèle dynamique de situation.

Le *texte*, dont la compréhension est évaluée. Considérons une compréhension de la consigne selon la dernière modalité présentée. Pour construire son modèle mental de la situation évoquée par le texte, l'élève doit se représenter « Léon », « sur un ponton ». Il possède un bâton et, sans que l'on sache s'il est en interaction avec lui, « il y a un crocodile dans l'eau ». La connaissance du sens du mot « ponton » est essentielle pour pouvoir dessiner le personnage dans l'espace. Ce mot, non référencé dans la base MANULEX¹⁰ (Lété *et al.*, 2004), a dû être rencontré par les élèves auparavant, sans quoi il est concevable qu'une large partie des élèves ne puisse répondre. L'élève doit comprendre l'anaphore (il = Léon), réaliser plusieurs inférences d'action (que fait Léon avec ce bâton ? comment Léon interagit avec le crocodile ?) et les traduire dans un dessin.

La *réponse*, manifestation tangible de la compréhension du texte et de la bonne compréhension de la consigne. La réponse d'élève reproduite ici montre un personnage sur un ponton près d'une étendue d'eau dans laquelle se trouve un crocodile. Le personnage, vêtu d'un pagne, brandit un bâton au-dessus de sa tête en direction d'un reptile menaçant. La tâche réalisée par l'élève semble ainsi conforme à l'attente, ce qui conduit l'enseignant à évaluer favorablement la compétence visée « *comprendre le texte lu* » : « 1 (point) ». Une bonne réponse, mais une bonne compréhension ?

Deux indices nous conduisent à douter du jugement et à suspecter l'irruption de connaissances antérieures. Le lecteur projette un schéma de défense face à une agression implicite qu'un autre lecteur pourrait tout à fait contester. Il dessine un ponton, mot absent de la base de mots typiquement rencontrés par un enfant de 6 ans. Le traduction sous forme de dessin de ce mot invite à s'intéresser au manuel utilisé, et à comparer la tâche d'évaluation avec les textes au sujet desquels l'élève doit manifester de la compréhension par le dessin. Nous reproduisons ci-dessous une page du manuel utilisé :

¹⁰ Les bases MANULEX constituent une description complète de la langue française écrite adressée à l'enfant à l'école primaire. Elles sont à la disposition des chercheurs qui travaillent sur l'acquisition de l'écrit et ses dysfonctionnements mais également des enseignants qui pourront extraire de ces tables de fréquence, des listes de mots adaptées aux enfants de cycle 2 et 3.

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.




Figure 4 : Extrait de manuel (A l'école des albums, CP Retz, p.30)

La comparaison apporte des éléments nouveaux pour comprendre la réponse de l'élève. La proximité de l'épreuve avec le manuel est évidente. Le dessin du garçon, le dessin du ponton, les mêmes mots employés conduisent à penser que les réponses produites dépendent, dans cette situation précise, et pour une large part, de la mémoire de la situation présentée par le manuel ou de la mémoire de la compréhension de cette situation lors de l'étude, et non de la compréhension en elle-même de ce qu'il y a à lire, laquelle reposerait sur la capacité des élèves au décodage. En fin de premier trimestre, il est permis de douter de la capacité de tous les élèves à décoder le mots « crocodile » (d'après Manulex, ce mot se rencontre environ tous les 10 000 mots en CP et « ponton » ne figure pas dans la base CP).

On peut estimer que les preuves de validité didactique de cette épreuve sont insuffisantes : la tâche prescrite ne permet pas d'évaluer la compréhension d'un texte mais plutôt la mémoire de la situation générale de l'histoire et/ou de l'illustration d'un texte déjà étudié laquelle se fonde sur un modèle de situation déjà élaboré, partageant de nombreux points communs et qui n'est pas remis en cause (on pourrait concevoir, par exemple, que le crocodile soit sur le ponton et que Léon soit dans l'eau).

Dessiner constitue une piste intéressante – moyennant par exemple de compléter par un échange oral argumentatif avec l'élève pour éviter une interprétation trop à charge de l'évaluateur ou trop proche des tâches d'apprentissage – mais répondre à des questions reste la pratique majoritaire comme dans le cas de l'épreuve n°2 que nous analysons ci-après.

4.2.2 Épreuve 2 : évaluer la compréhension au moyen d'un questionnaire

En fin de premier trimestre, un enseignant souhaite évaluer la compétence « *Dire de quoi et de qui parle le texte lu* » et choisit pour ce faire l'unique épreuve suivante constituée d'un texte et d'un questionnaire :

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 5 : Épreuve de Lecture du 1er trimestre avec réponses d'élève

Nous focaliserons notre analyse sur le support permettant à l'évaluateur de recueillir des indices prouvant la compréhension de l'élève : le questionnaire considéré comme un texte sur le texte doté de caractéristiques propres.

Parmi ces caractéristiques, distinguons :

- des inversions sujet-verbe absentes du texte support. Ces structures spécifiques de l'écrit ne sont pas au programme de CP ;
- des phrases interro-négatives non employées dans le texte support (« pourquoi ne va-t-elle pas à l'école ? »). Ces structures grammaticales ne figurent pas non plus au programme du CP ;
- 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.

A ceci s'ajoutent des amorces de réponse : 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... ».

La consigne. Après plusieurs lectures du texte, l'élève doit répondre aux questions « que pose la maîtresse » (130 mots). Qu'elles soient oralisées pas à pas par l'enseignant (ce qui supposerait que tous les élèves avancent au même rythme au cours de la passation collective et simultanée de l'épreuve) ou lues de façon autonome par l'élève (ou encore une combinaison des deux possibilités), elles nécessitent d'être mémorisées et fréquemment rappelées au cours de l'activité de réponse, ce qui est susceptible d'entrer en conflit avec la mémoire de la compréhension du texte ou la mémoire du texte lui-même (lequel se compose de bien moins de mots : 54 mots).

Les questions. Les questions n'engageant pas un échange direct avec l'élève, se pose celle de l'interprétation des réponses, correctes comme erronées :

Cas 1 : interpréter des *bonnes réponses* (BR)

Q1 : La première question interroge l'élève sur l'identification du motif pour lequel le personnage de l'histoire ne peut pas aller à l'école. Elle interroge *a priori* la compréhension d'une inférence de type cause-effet nécessaire pour élaborer le modèle de situation. Bien répondre signifierait ainsi bien avoir compris cette inférence causale (« car »). Mais une autre hypothèse sur l'activité de réponse *a priori* (pour en être certain, il faudrait pouvoir interroger l'élève directement sur la réponse produite au moment où il répond) peut être réalisée. La mise en relation de l'effet avec la cause est prise en charge par l'amorce de réponse : « car ». La réponse apportée par l'élève « elle est malade » est de ce point de vue partielle. L'élève pourrait avoir identifié ce lien entre la cause (malade) et les effets (ne pas aller à l'école) mais il pourrait aussi avoir répondu à la question en recopiant la partie de texte située après « car » en ligne 3 du texte, « car » étant le dernier mot entendu.

Q5 : Les réponses apportées aux trois premières questions permettent de répondre à la cinquième et de révéler un phénomène particulier agissant dans la relation entre questions. Des éléments fournis par le 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 ayant seulement lu les questions ce qui conduirait à évaluer dans ce cas un traitement stratégique du questionnaire remarquable mais ne correspondant pas avec le but recherché : évaluer le modèle de situation du texte élaboré par l'élève. Notons en outre que

L'on peut cocher la deuxième modalité de réponse (MR) sur la seule base de la lecture du titre donc sans avoir lu le texte.

Dans les deux cas analysés, les BR ne sont pas nécessairement synonymes de bonne compréhension du texte, laquelle renvoie au modèle mental dynamique de la situation évoquée par le texte que le lecteur élabore à partir de la lecture des mots du texte (ce qui implique les habiletés de déchiffrement/identification).

Cas 2 : interpréter des réponses erronées

Q4 : L'élève répond « Caroline a mal à la tête », « Caroline est dans son lit » « Caroline a les joues toutes rouges » réponses attendues. Les BR sont des reprises textuelles explicites du texte (phrases 2 et 4). Mais il répond également « Caroline a mal à la gorge » réponse considérée comme erronée. Une traduction de cette réponse erronée pourrait être que l'élève n'a pas compris la situation dans laquelle se trouve « Caroline » (point commun des MR proposées). Cela ne semble pas cohérent avec les réponses écrites 1 et 2 qu'il a déjà fournies. On pourrait concevoir qu'il ait suivi un cheminement menant à dire que « Caroline à mal à la gorge » parce que « Caroline est malade », « qu'elle tousse » et « que sa maman lui donne du sirop » ce qui, lié à sa propre expérience du monde, renverrait à un mal de gorge (et non une boisson sucrée traduisant une autre compréhension du sens du mot « sirop »). Cette réponse sensée et considérée comme non valide pourrait conduire l'évaluateur à penser que l'élève n'a pas bien compris le texte alors qu'elle signerait plutôt une très bonne compréhension de la situation.

Dans tous les cas, la relation entretenue entre la question et la réponse est de nature textuelle et explicite : elle est clairement sous-tendue par des indices formulés tels quels dans le texte lui-même ou dans le questionnaire, lequel n'est donc pas neutre.

En conclusion, le questionnaire se suffit presque entièrement à lui seul pour répondre à toutes les questions : on peut déduire des questions 1 et 5 de quoi y répondre mutuellement, la connaissance et l'expérience du monde permet de cocher tout ce qui se rapporte à la maladie et, par recoupement entre les question 3 et 5, déterminer qu'un autre personnage Nicolas est cité dans la situation. Seule la réponse à la question 2 nécessite de lire le texte. Cela suffit à invalider l'usage de cette tâche pour évaluer la compréhension du texte.

On peut ajouter plusieurs remarques :

- nous ne disposons pas d'éléments permettant de savoir si le texte est connu des élèves préalablement à l'évaluation parce qu'il aurait été étudié tel quel comme dans le cas précédemment analysé.
- le traitement du seul questionnaire complexifie l'activité de lecture et de restitution de la compréhension. Il engendre une tâche de lecture supérieurement difficile à celle de la lecture du texte dont il est censé permettre d'évaluer la compréhension. Cette difficulté est de plusieurs ordres : volume du texte, déchiffrement, compréhension, type de texte et syntaxe de nature différente, mémoire de la tâche...
- aucune question n'interroge le modèle de situation élaboré par l'élève : 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.

- il n'est pas si simple de déterminer si une bonne réponse signe une bonne compréhension ni, même à l'inverse, qu'une réponse erronée traduise une compréhension erronée de la situation décrite par le texte.
- la méthodologie adoptée (une analyse individuelle tâche par tâche) ne permet pas de généraliser. Elle permet simplement de souligner qu'un certain nombre de problèmes se posent dans le cas particulier de l'épreuve considérée, choisie en raison de la présence de multiples cadres de réponses (formulation de réponse, formulation partielle de réponse, QCM-C) dont nous pouvons identifier pour chacun certaines limites. Les remarques formulées sont également vraies pour d'autres épreuves fondées sur ces mêmes modalités de questionnement, pour les mêmes raisons que celles évoquées ici.

Les preuves de validité didactique de cette tâche d'évaluation de la compréhension sont insuffisantes : la tâche prescrite ne permet pas d'évaluer la compréhension d'un texte mais plutôt la compréhension d'un système composé d'un texte et d'un questionnaire sur le texte qui le dépasse, ne s'appuyant en réalité que très peu sur celui-ci.

5. Discussion

La démarche qualitative employée pour analyser livrets et épreuves sous l'angle de la validité nous conduit à pouvoir discuter deux ensembles de résultats.

5.1. *L'analyse des livrets*

Les processus individuels de référentialisation des enseignants les conduisent à élaborer des critères et des référentiels pour évaluer, disparates, hétérogènes, plutôt influencés par les items des Progressions, nous conduisant à évoquer un pilotage institutionnel insuffisamment clarifié de l'évaluation par ces derniers. Leur grand nombre a pour effet de conduire les enseignants à utiliser un grand nombre d'épreuves pour évaluer, voire à procéder à une évaluation sur la base de simples observations, non accessibles *a posteriori* à un observateur extérieur.

Le statut de référentiel des Progressions est remis en cause dans le discours par certains enseignants qui l'utilisent en réponse à ce qu'ils perçoivent davantage comme une commande exogène formelle (l'évaluation revêt dans ce cas une nouvelle fonction de conformité) qu'une nécessité pédagogique. Ces résultats conservent une forme d'actualité aujourd'hui encore : les enseignants choisissent toujours les items du livret qu'ils font figurer dans le livret scolaire qui n'a véritablement d'« unique » que le nom et le support.

Les différents organisateurs du travail de l'enseignant en matière d'évaluation (programmes, Progressions, livret SCEREN ...) s'accordant mal, la prescription ne lève pas les obstacles que rencontrent les enseignants au moment d'opérationnaliser la commande d'évaluation, au moment de construire leur référent : que convient-il d'évaluer ? comment communiquer le résultat de l'évaluation ?

5.2. *L'analyse des épreuves de compréhension en lecture*

Deux analyses de tâches d'évaluation de la compréhension ont été présentées. Elles sont représentatives de deux des trois modalités d'évaluation préférées par les enseignants formant la population étudiée. La troisième modalité qui engage des réponses écrites à des questions écrites est minoritaire et pose un autre problème évident d'évaluation d'une compétence

écran (la capacité à formuler une réponse par écrit) qui ne peut être pensée comme allant de soi en CP. Cela permet de mettre au jour une autre problématique concrète de référentialisation : quel référent (didactique) pour évaluer la compréhension en lecture ? Comprendre un texte est-il synonyme de « dessiner ce qui est écrit » ou de « répondre à des questions écrites » ? La validité didactique de l'épreuve reposant sur la construction par l'élève d'un modèle de situation du texte, comment l'apprécier pour chaque élève si on ne l'interroge pas directement ?

La première épreuve analysée permet de poser une autre question : faudrait-il évaluer ce qui a été enseigné ou évaluer le transfert ou la généralisation de ce qui a été enseigné ?

L'usage de questionnaires engageant une réponse dessinée ou écrite pour évaluer la compréhension en lecture est constaté dès le CP. Dans le premier cas, les exemples présentés ont permis d'identifier qu'outre la difficulté de représenter par le dessin la compréhension attendue d'un texte, le risque était grand en tout début d'apprentissage d'évaluer une mémoire de la situation et/ou de l'illustration d'un texte déjà étudié. Dans le deuxième cas, l'usage de questionnaires permet sans doute d'éviter ces obstacles mais pose un nouveau double problème constitué, d'une part, par le traitement nécessaire d'un nouveau texte sur le texte dont il est censé évaluer la compréhension, et d'autre part, par le crédit ou discrédit à apporter aux réponses. Si bien que l'évaluation porte sur un ensemble hétérogène : le texte, des questions sur le texte. Outre qu'ils attestent d'une évaluation de la mise en mémoire du texte support, davantage qu'un véritable travail de compréhension, les questionnaires occasionnent une charge attentionnelle accrue, engagent une charge de lecture supplémentaire (difficultés de déchiffrage, difficultés de compréhension, changement de type de texte, difficultés liées à la syntaxe, à la polysémie des termes utilisés), une mobilisation de stratégies variées (de contrôle de l'activité par exemple) que l'on ne peut pas ne pas considérer. Et l'on n'est pas certain que les réponses apportées, correctes ou non, traduisent une bonne compréhension du texte. Quoiqu'il en soit, le modèle de situation élaboré par l'élève ne semble pas être approché puisque l'élève doit davantage trouver des bonnes réponses attendues.

Dans une recherche précédente (Blanc, 2015), nous avons pu montrer que les élèves pouvaient répondre à un questionnaire sans avoir lu le texte support, comme dans l'exemple analysé ici, et obtenir des résultats parfaitement semblables à ceux d'un échantillon national. On peut élaborer une représentation du texte assez fine, simplement à partir des informations sur le texte, fournies par les questions, lesquelles comportent en outre des éléments de réponse à d'autres questions. Voici qui ferait apparaître, aux quatre domaines de compétences mobilisés en lecture décrits par Goigoux, (2003), Ecalle et Magnan (2010), Bianco (2015)¹¹, un cinquième domaine : celui de savoir traiter un questionnaire.

6. Conclusion

Cette recherche s'intéressait à la qualité de l'évaluation interne instituée et particulièrement à la validité, plus exactement une validité plurielle (didactique, pédagogique et/ou curriculaire) de deux éléments du référent de l'enseignant : les livrets et les épreuves. Le principal résultat obtenu est que cette évaluation se caractérise par un défaut de validité didactique, en particulier pour ce que nous avons discuté ici concernant l'évaluation de la compréhension

¹¹ <http://centre-alain-savary.ens-lyon.fr/CAS/education-prioritaire/ressources/theme-1-perspectives-pedagogiques-et-educatives/lire-ecrire-parler-pour-apprendre-dans-toutes-les-disciplines/dossier-lire-ecrire/comprehension-partager-les-references>

en lecture, interrogeant directement la validité de conséquence de celle-ci et des jugements évaluatifs transmis aux familles.

Si ce résultat est utile pour prendre la mesure d'un phénomène peu étudié concernant l'activité évaluative de l'enseignant autour d'un prescrit formé par le couple *évaluer/informer*, il convient d'insister sur quelques éléments de contexte et limites afin de ne pas le traduire en injuste procès.

Les enseignants auprès de qui les données ont été collectées ne manquent pas de compétence. Plus d'un quart des enseignants sont formateurs et tous sont en réflexion sur ce sujet, ce qui tend à justifier le terme de « défi » employé en introduction. En revanche, plusieurs ont signifié clairement que produire un livret relevait avant tout d'une injonction formelle ce qui pourrait les conduire à utiliser une sorte de double évaluation : une évaluation « de conformité » et une évaluation proche d'une évaluation soutien d'apprentissage (ESA) (Laveault et Allal, 2016 ; Allal et Laveault, 2009), correspondant au fait d'impliquer l'élève dans le processus d'évaluation et d'intégrer des évaluations formatives formelles et informelles, ainsi que les différents types de régulations pouvant découler d'une évaluation sommative critériée.

Les enseignants ont été sollicités pour fournir les livrets et les épreuves ayant servi, et notre analyse ne se fonde que sur ces traces. Il est donc possible que d'autres modalités d'évaluation, relevant de la « *simple observation* » telle que formulée dans la préface du livret SCEREN, soient employées, notamment l'appréciation de réponses orales à des questions orales, posées dans le quotidien de la classe, permettant d'évaluer la compréhension en lecture. Elles échappent à notre analyse. Une partie de l'évaluation, implicite, difficilement observable et quantifiable ne se fonde pas, ou pas entièrement, sur des réponses d'élèves à des épreuves. Nous en voulons pour preuve l'exemple de l'enseignant K de notre corpus (cf. fig. 3) qui produit un livret scolaire renseignant les compétences des élèves mais ne se fonde sur aucune épreuve tangible. Rien ne permet d'affirmer ou d'infirmer que cette modalité d'évaluation ne participe pas d'une ESA.

Mais la démarche adoptée, qualitative et didactique, visait surtout à examiner :

- l'effet des processus de référentialisation conduisant les enseignants à déterminer l'objet de l'évaluation et à choisir des tâches, et à
- analyser ces tâches sous un angle didactique et du point de vue de l'adéquation de ce pourquoi elles sont utilisées à l'activité de réponse des élèves.

Les résultats permettent de contribuer à documenter l'activité évaluative des enseignants, à mettre l'accent sur le contenu des épreuves et les processus de réponse mis en jeu par les élèves, et à pointer des éléments problématiques relatifs aux processus de référentialisation, lesquels apparaissent nettement comme complexes. Blanchouin *et al.* (2022) se sont attachés à étudier les gestes évaluatifs de professeurs des écoles en mathématiques. Notre démarche est voisine. L'usage de questionnaires est un fait documenté et ancien (Pearson et Gallagher, 1983 ; Giasson, 1990 ; Joole 2008 et Leclercq, 1986 et 2006 pour les QCM). Toutefois, à l'heure de protocoles standardisés y ayant massivement recours (Évaluation nationale Repères CP), il est essentiel de pouvoir analyser finement les tâches du point de vue de l'activité de réponse sollicitée chez les élèves, au risque d'utiliser des tâches souffrant d'un manque de validité didactique et de faire de telles modalités d'évaluation un modèle pour évaluer constituant une référence pour les enseignants. Comme Lepage *et al.* (2022) qui pointent « des manques évidents à cet égard » (p.8), nous ne pouvons que souscrire au constat

d'un manque de formation des enseignants à l'évaluation. Peu de programmes universitaires préparant les étudiants aux métiers de professeur font une place pleine et entière à l'évaluation¹². Et pourtant la 5^e compétence professionnelle commune aux professeurs inscrite au référentiel métier¹³ s'intitule précisément : « *Evaluer les progrès et les acquisitions des élèves* ». Une perspective consiste donc en la participation / formation d'enseignants à des ingénieries d'évaluation sous-tendues par ce type de travaux (Blanc et Vantourout, 2018 ; Vantourout et Maury, 2017 ; Vantourout, 2020). L'enjeu majeur de cette démarche est double : participer à la conception d'outils d'évaluation valides du point de vue didactique, fondés sur une compréhension fine par l'enseignant des processus de réponse des élèves à une épreuve d'évaluation, et conduire les enseignants à une meilleure identification des processus de référentialisation intervenant dans la définition des référentiels pour évaluer, inhérents à chacun d'eux.

¹² Ainsi on ne retrouve pas de place dédiée à l'évaluation dans les maquettes des masters MEEF des INSPE de Paris et Bordeaux <https://www.inspe-paris.fr/sites/default/files/maquetteglobale.inspe201920.pdf>
https://www.inspe-bordeaux.fr/application/files/3316/6490/0816/Maquette_MEEF_1er_degre_M1-M2.pdf

¹³ Arrêté du 1-7-2013 - J.O. du 18-7-2013

7. Références bibliographiques

- Allal, L. et Laveault, D. (2009). Assessment for Learning: Évaluation-soutien d'Apprentissage. *Mesure et évaluation en éducation*, 32(2), 99–106.
- 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. doi:10.3917/lse.422.0011
- Barbier, J. M. (1985). *L'évaluation en formation*. Presses universitaires de France.
- Bastien, C. et Bastien Tonniazzo, M. (2004). *Apprendre à l'école*. Armand Colin.
- Baudouin, J. et Pita, J. (2011). Récit et formation : reconnaissance et biographie évaluative. Dans Anne Jorro (dir). *La professionnalité émergente : quelle reconnaissance* (p. 83-97). De Boeck Supérieur. doi-org.ezproxy.u-paris.fr/10.3917/dbu.ketel.2011.01.0083
- Blanc, C. (2015, janvier). Une épreuve psycho-didactique pour évaluer en lecture en fin de CP [Communication orale]. 27e Colloque international de l'ADMEE-Europe, Liège, Belgique. <http://orbi.ulg.ac.be/handle/2268/192246>.
- Blanc, C. (2017). *Étude de la validité de l'évaluation interne conduite en Cours Préparatoire, en Français et en Mathématiques* [Thèse de doctorat publiée]. Université Paris Descartes.
- Blanc, C. et Vantourout, M. (2018). Une approche psycho-didactique pour évaluer la lecture au Cours Préparatoire ? *Repères*, 56. <https://doi.org/10.4000/reperes.1224>
- Blanc, C., Vantourout, M. et Maury, S. (2018). Validité d'épreuves d'évaluation mettant en jeu des illustrations au CP : mise en œuvre d'une méthodologie d'analyse de quelques épreuves. *Évaluer. Journal international de recherche en éducation et formation*, 4(1), 21-36. <http://admee.ulg.ac.be/journal/index.php/ejiref/article/view/144>
- Blanc, C., Vantourout, M. et Maury, S. (2018). Validité d'épreuves d'évaluation mettant en jeu des illustrations au CP : mise en œuvre d'une méthodologie d'analyse de quelques épreuves. *Évaluer. Journal international de recherche en éducation et formation*, 4(1), 21-36. <https://doi.org/1048782/e-jiref-4-1-21>
- Blanc, N. (2010). *Lecture et habiletés de compréhension chez l'enfant*. Paris : Dunod.
- Blanchouin, A., Grapin, N. et Mounier, E. (2022). Documenter l'activité évaluative des professeurs des écoles à partir de leurs gestes évaluatifs. Étude de cas en mathématique. *Évaluer. Journal international de recherche en éducation et formation*, 8(1), 3-28. <https://doi.org/10.48782/e-jiref-8-1-3>
- Cèbe, S. et Goigoux, R. (2009). *Lector et Lectrix : apprendre à comprendre les textes narratifs*. Retz.
- Cèbe, S., Goigoux, R. et Thomazet, S., (2004). *Enseigner la compréhension : principes didactiques, exemples de tâches et d'activités*. https://cache.media.eduscol.education.fr/file/education_prioritaire_et_accompagnement/54/4/lire_ecrire_enseigner_comprehension_115544.pdf
- Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck.
- Crahay, M. (2011) Curriculum et théories de l'apprentissage. Dans quelle mesure les curriculums prennent-ils en compte ce que l'on sait des apprentissages scolaires. Dans M. De Kesel, J.-L. Dufays et A. Meurant (p. 9-29). *Le curriculum en questions : La progression et les ruptures des apprentissages disciplinaires de la maternelle à l'université*. Presse Universitaires de Louvain.
- Eccles, J. et Magnan, A. (2010). *L'apprentissage de la lecture et ses difficultés*. Dunod.
- Figari, G. (2006). Les référentiels. Dans G. Figari et L. Mottier Lopez (dir.) *Recherches sur l'évaluation en Education*. (p. 101-108). L'Harmattan.
- Fijalkow, E. (2003). *L'enseignement de la lecture-écriture au cours préparatoire, entre tradition et innovation*. L'Harmattan.

- Florin, A., Tricot, A., Chesné J.-F., Piedfer-Quêne, L. et Simonin-Kunerth, M. (2023). *Dossier de synthèse : L'évaluation en classe, au service de l'apprentissage des élèves*. Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Dossier-de-synthese.pdf
- Giasson, J. (1990). *La compréhension en lecture*. De Boeck.
- Goigoux, R. (dir.) (2016). *Étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages*. (Rapport de recherche remis à Madame la directrice générale de l'enseignement scolaire (DGESCO-MENESR), ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche). Université de Lyon. <https://ife.ens-lyon.fr/sites/default/files/2022-05/rapport-lire-et-ecrire.pdf>
- Goigoux, R. et Cèbe, S. (2014). *Enseignement de la lecture : réponses à cinq questions posées par le conseil supérieur des programmes*. http://cache.media.education.gouv.fr/file/CSP/06/3/GOIGOUX_Roland_PU_et_CEBE_Sylvie_MCF_-_CSP_Contribution_395063.pdf
- Goigoux, R. (2003). Quelques points de repère pour une didactique de la compréhension. *Langage & pratiques*, 31, 51-60.
- Grapin, N. et Grugeon-Allys, B. (2018). Approches psychométrique et didactique de la validité d'une évaluation externe en mathématiques : quelles complémentarités et quelles divergences ? *Mesure et évaluation en éducation*, 41(2), 37-66. <https://doi.org/10.7202/1059172ar>
- Grapin, N., Vantourout, M. et Grugeon-Allys, B. (2022). S'assurer de la validité didactique des évaluations en milieux scolaires. *Recherches en didactiques*, 33, 59-83. <https://doi.org/10.3917/rdid1.033.0059>
- 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*. Université de Liège. <http://www.aspe.ulg.ac.be/grisay/fichiers/APER03.pdf>
- Grugeon-Allys, B. et Grapin, N. (2016, novembre). *Validité des évaluations : une approche épistémologique* [Communication orale]. Colloque Évaluation en mathématiques. Université Paris Est Créteil.
- Hadji, C. (2018). *L'évaluation à l'école, pour la réussite de tous les élèves*. Nathan.
- Joole, P. (2008). *Comprendre les textes écrits*. Retz / CRDP de l'Académie de Versailles.
- Jorro, A. (2009). L'évaluation comme savoir professionnel. Dans L. Mottier et M. Crahay (dir.) (2009). *Évaluations en tensions*. (p. 219-231). De Boeck.
- de Landsheere, V. (1988). *Faire réussir, faire échouer : La compétence minimale et son évaluation*. Presses universitaires de France.
- Laveault, D. et Grégoire, J. (2014). *Introduction aux théories des tests en sciences humaines*. De Boeck.
- Laveault, D. et Allal, L. (2016). Implementing Assessment for Learning: Theoretical and Practical Issues. Dans D. Laveault et L. Allal (dir.), *Assessment for learning. Meeting the challenge of implementation* (p. 1-18). Springer International.
- Lepage, I., Potvin-Rosset, E. et Leduc, D. (2022). Évaluer : entre innovation et tradition. Résultats préliminaires d'une boussole interactive pour innover en évaluation des apprentissages. *Évaluer. Journal international de recherche en éducation et formation*, 8(3), 5-28. <https://doi.org/10.48782/e-jiref-8-3-5>
- Leplat, J. et Hoc, J.M. (1983). Tâche et activité dans l'analyse psychologique des situations. *Cahiers de psychologie cognitive*, 3(1), 49-63.

- Lété, B., Sprenger-Charolles, L. et Colé, P. (2004). MANULEX : A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, 36(1), 156–166. <https://doi.org/10.3758/BF03195560>
- Ministère de l'Éducation Nationale et ministère de l'Enseignement Supérieur et de la Recherche (2008). *Les nouveaux programmes de l'école primaire* (Bulletin officiel hors-série n° 3 du 19 juin 2008). <https://www.education.gouv.fr/bo/2008/hs3/default.htm>
- Pearson, P. D. et Gallagher, M. C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology*, 8(3), 317–344. [https://doi.org/10.1016/0361-476X\(83\)90019-X](https://doi.org/10.1016/0361-476X(83)90019-X)
- PIREF (2003). Recommandations du jury. Conférence de consensus sur l'enseignement de la lecture à l'école primaire. Des premiers apprentissages au lecteur compétent. <https://www.snuipp.fr/IMG/pdf/Piref.pdf>
- Rey, O. (2012). Le défi de l'évaluation des compétences. *Dossier d'actualité Veille et Analyses de l'Ifé*, 76. <https://edupass.hypotheses.org/118>
- 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.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. De Boeck.
- Services Culture Editions Ressources pour l'Éducation Nationale (2010). *Le livret scolaire de l'école primaire. Édition 2010. Cycle 1-Cycle 2-Cycle 3. Programmes 2008. Paliers 1 et 2 du Socle commun. Évaluations CE1 et CM2*. SCEREN-CRDP Académie Orléans-Tours.
- Vantourout, M. (2020). Évaluations et didactiques, validités didactiques: quelles synergies? Dans N. Younès, C. Gremion et E. Sylvestre (dir.), *Évaluations, sources de synergies*. (p. 93-121). Presses de l'ADMEE.
- Vantourout, M. et Goasdoué, R. (2014). Approches et validité psycho-didactiques des évaluations. *Revue Éducation & Formation*, e-302, 139-156. <https://shs.hal.science/halshs-01239551/document>
- Vantourout, M. et Maury, S. (2017). Évaluation de la lecture au CP : mise en œuvre d'une approche multiple. *Éducation et didactique*, (11-1), 45–62. <https://doi.org/10.4000/educationdidactique.2649>
- Vergnaud, G. (2006). Le rôle des disciplines de référence dans la conception de l'évaluation. 1.1 La psychologie cognitive. Dans G. Figari et L. Mottier Lopez (dir.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (p.167-175). L'Harmattan.