
Évaluation de la lecture au CP : mise en œuvre d'une approche multiple

Reading assessment in the first grade : a multi-tests approach

Marc Vantourout et Sylvette Maury



Édition électronique

URL : <http://journals.openedition.org/educationdidactique/2649>

DOI : [10.4000/educationdidactique.2649](https://doi.org/10.4000/educationdidactique.2649)

ISSN : 2111-4838

Éditeur

Presses universitaires de Rennes

Édition imprimée

Date de publication : 20 juin 2017

Pagination : 45-62

ISBN : 978-2-7535-6460-2

ISSN : 1956-3485

Référence électronique

Marc Vantourout et Sylvette Maury, « Évaluation de la lecture au CP : mise en œuvre d'une approche multiple », *Éducation et didactique* [En ligne], 11-1 | 2017, mis en ligne le 20 juin 2019, consulté le 08 décembre 2020. URL : <http://journals.openedition.org/educationdidactique/2649> ; DOI : <https://doi.org/10.4000/educationdidactique.2649>

Tous droits réservés

ÉVALUATION DE LA LECTURE AU CP : MISE EN ŒUVRE D'UNE APPROCHE MULTIPLE

Marc Vantourout

EDA, université Paris Descartes, maître de conférences, [marc.vantourout@parisdescartes.fr]

Sylvette Maury

EDA, université Paris Descartes, professeure émérite, [sylvette.maury@parisdescartes.fr]

Dans cet article, nous présentons un dispositif d'évaluation de la lecture en fin de CP (cours préparatoire) dont l'élaboration repose sur la prise en compte de deux dimensions du savoir-lire, le décodage-identification et la compréhension. Nous analysons la validité des épreuves qui le composent en référence à des modèles du décodage et de l'identification de mots ainsi que de la compréhension de texte. Le dispositif est testé, en passation individuelle, auprès de soixante-huit élèves de CP. Nous montrons par des analyses quantitatives et qualitatives des données l'intérêt d'une approche combinant plusieurs épreuves, dont deux consacrées à la compréhension autonome d'un texte écrit, pour évaluer en lecture des élèves et pour identifier différents types d'élèves en difficulté.

Mots-clés : évaluation, compréhension, décodage-identification, approche multiple, cours préparatoire.

Reading assessment in the first grade : a multi-tests approach

In this paper we present an evaluation method to assess the level of reading of end of first-grade children (cours préparatoire), which takes into account two reading skills dimensions: decoding-identifying and understanding. The validity of the method is analyzed through both word decoding/identifying models and text understanding models. On the basis of quantitative and qualitative data analyses of the tests passed by sixty-eight first grade children, we show the interest of an evaluation method that combines several tests, including two for autonomous understanding of a written text, in order to assess children reading skills and to identify the different types of difficulties they can encounter.

Keywords: assessment, comprehension, decoding-identifying, multi-tests approach, first grade.

ÉVALUER LA LECTURE AU CP ?

Éléments contextuels

Du « savoir déchiffrer » à la compréhension de textes écrits, les définitions du « savoir-lire », y compris celles concernant le lecteur débutant, sont multiples. Aujourd'hui, cependant, un large consensus se dégage pour reconnaître que celui-ci ne se résume pas au déchiffrement et que le but de la lecture est la compréhension. Avec ce point de vue, la lecture revêt deux composantes, décodage-identification de mots et compréhension de phrases et de textes, renvoyant respectivement à la distinction opérée en psychologie cognitive entre mécanismes de bas niveau et de haut niveau (Ecalte & Magnan, 2010) avec, pour le haut niveau, les compétences en compréhension telles les capacités à utiliser des connaissances en matière de structure de récit, à produire des inférences, à auto-évaluer le produit de la compréhension en cours (Blanc, 2009). Alors que ce point de vue est largement partagé en psychologie cognitive et en didactique, la place occupée par l'enseignement de la compréhension en cycle 2¹ est jugée « inexistante » dans les programmes par Bianco (2014, p. 20) et « indigente » par des inspecteurs généraux² (MEN, 2010, p. 12).

Dans les classes de CP (cours préparatoire), son évaluation semble par ailleurs mal réalisée, certainement en raison des difficultés rencontrées par les maîtres lorsqu'ils doivent choisir ou fabriquer des épreuves d'évaluation répondant aux injonctions institutionnelles. Ces difficultés ont été soulignées dans des rapports (MENESR, 2005 ; MENJVA-MESR, 2011) ainsi que par Gonthier-Maurin (2012) qui parle de « pression évaluative » (p. 56) et qui dénonce, plus généralement, une « inflation des prescriptions en dehors de tout cadre cohérent, conjuguée à la sous-prescription des moyens à mettre en œuvre pour les respecter, [qui] perturbe l'activité des enseignants » (p. 33). Notre travail s'inscrit dans ce contexte général préoccupant puisque l'évaluation représente une part importante du travail des enseignants (MEN-MESR, 2012). Pourtant, l'Éducation nationale a entrepris ces dernières années un effort de clarification vis-à-vis de l'évaluation, comme l'indique la publication du « Livret scolaire de l'école primaire » (SCEREN, 2012) qui permet d'attester progressivement des compétences et connaissances acquises par l'élève dans les divers domaines ensei-

gnés³. S'agissant de l'évaluation de la lecture au CP, ce livret contient neuf « items », chacun correspondant à une ligne où est inscrite une « compétence » ou une « connaissance ». Un seul item porte sur la compréhension alors que les huit autres portent sur les habiletés de base (celles correspondant à des mécanismes de bas niveau). Ce manque d'attention accordée à la compréhension est accentué par la formulation de cet item dans le livret, « Dire de quoi et de qui parle le texte lu », qui laisse à penser que, pour attester de la compréhension, il suffit de savoir dire de quoi parle un texte et qui est impliqué. Ainsi, contrairement à ce qui se passe dans l'énoncé même des programmes (MEN, 2008), « Dire de qui ou de quoi parle le texte lu ; trouver dans un texte ou son illustration la réponse à des questions concernant le texte lu ; reformuler son sens », il n'y a pas, dans l'item d'évaluation, de référence explicite au « sens du texte », associé aux mécanismes de haut niveau. Tout se passe donc, avec le livret, comme si l'enseignant était invité à renoncer à l'évaluation de ce qui fait le propre de la compréhension, habileté figurant pourtant dans les programmes et considérée comme de première importance dans les travaux consacrés à la maîtrise de la lecture.

Dans cet article consacré à l'évaluation de la lecture, nous nous proposons de bâtir et de tester un dispositif que nous qualifions d'« ingénierie évaluative courte ». Concrètement, nous avons fabriqué un texte et trois épreuves, une de lecture orale et deux de compréhension. L'ingénierie est conforme aux programmes, tout en se démarquant de la prescription secondaire que constitue le livret car nous privilégions l'évaluation de la compréhension. L'objectif fondamental est de parvenir à identifier les élèves en difficultés en lecture en fin de CP. Toutefois, sur la base d'un travail initié au sein d'une circonscription⁴, cette recherche a pour visée complémentaire, à moyenne échéance, d'outiller les maîtres afin de contribuer à l'évolution de leurs pratiques. En effet, nous rejoignons la conception de la formation professionnelle défendue par Goigoux & Cèbe (2013a et b) : « il [est] plus efficace d'agir directement sur les pratiques des enseignants pour modifier leurs conceptions et leurs connaissances que l'inverse ».

Évaluer la lecture dans le cadre des APDE

Les épreuves, du point de vue de leur conception, relèvent des approches psycho-didactiques des évaluations (APDE) (Vantourout & Goasdoué, 2014). Se référer aux APDE signifie, ici⁵, que l'on souhaite articuler au mieux didactique et psychologie cognitive de la lecture d'un côté, et évaluation de l'autre. Avec cette approche, nous nous centrons sur le système « tâche-activité » et retenons que les réponses du sujet, autrement dit ce qu'il livre à un observateur extérieur de ses compétences et acquis, sont fonction des situations auxquelles il est confronté et doivent être examinées comme telles. Ainsi, afin de comprendre au mieux la réponse du sujet, nous prenons en compte la dimension cognitive de l'activité, ce qui favorise grandement l'obtention d'un diagnostic de qualité, et permet de s'assurer de la validité psycho-didactique des épreuves⁶ (VPD) (*ibid.*) et, en corollaire, de la validité des jugements.

En conséquence, afin de s'assurer au mieux de la validité des épreuves, nous avons pris en compte, lors de leur élaboration, un certain nombre d'éléments théoriques et de résultats relatifs au fonctionnement cognitif des élèves dans le domaine de la lecture.

La littérature scientifique consacrée à la lecture est riche de modèles et de résultats. Notons toutefois qu'il existe un décalage entre, d'une part, l'état actuel des connaissances scientifiques relatives aux processus en jeu dans la lecture et son apprentissage et, d'autre part, les apports de ces connaissances pour l'enseignement (Martinet & Rieben, 2010) et, par extension, pour un projet d'évaluation comme le nôtre. L'évaluation figure en bonne place au sein de cette littérature, mais avec comme finalité la validation de modèles, sur la base d'analyses statistiques impliquant leurs constituants (voir Berger & Desrochers, 2011 ; Blanc & Brouillet, 2005 ; Ecalle & Magnan, 2010). Que pouvons-nous donc retenir de cette littérature alors que nous voulons évaluer, dans le cadre scolaire, les compétences et les acquis en lecture d'élèves à la fin du CP ?

Nous l'avons évoqué plus haut, il existe chez les cognitivistes français, en tout cas ceux auxquels nous nous référons, un consensus assez large quant à la présence de deux composantes-clés qui contribuent à la lecture, d'une part le décodage-identification de mots et d'autre part la compréhension du langage entendu, ce qui renvoie au modèle « *Simple View of Reading* » (Gough & Tunmer, 1986). Ainsi, il est

admis, sur la base d'explications en termes de « coût cognitif » en lien avec l'hypothèse d'efficacité verbale de Perfetti (2007), que la compréhension est affectée négativement lorsque l'activité de décodage est trop lente et qu'il y a un défaut d'automatisation de l'identification des mots (Goigoux & Cèbe, 2013a ; Martinet & Rieben, 2010). Les travaux éclairant les relations entre ces deux composantes donnent, quant à eux, des résultats nuancés. Pour certains, l'efficacité des mécanismes de bas niveau est ce qui différencierait les bons lecteurs des moins bons, « c'est le développement de procédures automatiques de reconnaissances des mots écrits qui permettra à l'apprenti lecteur d'atteindre un niveau de compréhension écrite égal à celui de l'oral en le dégageant des nécessités du décodage qu'implique cette reconnaissance » (Sprenger-Charolles & Colé, 2003, p. 46). Pour d'autres, qui défendent une vision simultanée du développement des habiletés de compréhension et des habiletés langagières de base, « la compréhension d'un texte ne se développe pas nécessairement dès lors que la lecture des mots est automatique mais dépend plutôt de différentes compétences dites de haut niveau » (Blanc, 2009, p. 14).

Sans entrer dans le débat sur l'importance de ces différentes composantes, nous retenons quelques résultats qui sont en lien avec les questions que nous nous posons, notamment ceux relatifs aux profils de lecteurs :

- chez les jeunes lecteurs, la contribution de l'identification des mots à la compréhension de l'écrit est particulièrement élevée (voir Cain, 2010, se référant à Catts, Hogan & Adlof, 2005) ;
- parmi les élèves qualifiés de « faibles compreneurs » en lecture, on trouve ceux qui le sont uniquement parce qu'ils rencontrent des difficultés lors de l'identification des mots, sans présenter de problèmes de compréhension générale ou spécifique à l'écrit ;
- à l'inverse, on trouve, parmi les élèves qui rencontrent des difficultés en compréhension de la lecture, des élèves présentant une capacité de lecture précise et fluide relativement à leur âge (Martinet & Rieben, 2010, citant Yuill & Oakhill, 1991). Ces élèves qui représenteraient 10 % d'une population d'enfants entre 7 et 11 ans, ont notamment des difficultés pour produire des récits cohérents (Ecalte & Magnan, 2010).

Ces résultats, qui dévoilent des liens complexes entre décodage-identification et compréhension, justifient, bien que nous nous préoccupions prioritairement de la compréhension autonome d'un texte écrit en fin de CP, que nous proposons également une épreuve de décodage-identification. Ce choix nous permet en outre d'aller dans le sens des attentes des programmes (MEN, 2008). Plus globalement, il est en accord avec des considérations, formulées par Martinet & Rieben (2010), relatives à l'enseignement et à l'apprentissage initial de la lecture :

« Beaucoup de malentendus résident encore d'une confusion de ces deux aspects [identification de mots et compréhension écrite de textes lus pas le lecteur lui-même ou par autrui] ou, pour le moins d'un manque de spécification laissant entendre qu'il est possible de parler de lecture en général. Or, concernant l'apprentissage, le cœur du problème réside dans le paradoxe suivant : les deux facettes doivent être à la fois théoriquement distinguées et exercées chacune pour elles-mêmes à travers des activités différentes tout en étant intégrées dans le but d'atteindre l'objectif ultime de la lecture qui est d'accéder au sens et de communiquer » (p. 190).

Reste à expliciter notre manière d'appréhender la compréhension. Nous adoptons le point de vue classique des psychologues cognitivistes résumé par Blanc & Brouillet (2005) : comprendre un texte, c'est, pour le lecteur, construire une représentation mentale qui intègre en un tout cohérent la situation décrite et qui « diffère de, et dépasse, l'information fournie par le texte » (p. 14). On trouve attachées à cette conception la distinction entre « mémoire du texte » et « mémoire de la compréhension du texte⁷ » ainsi que l'existence de différentes strates de représentation en mémoire « censées représenter des niveaux de compréhension et d'abstraction différents » (Bianco, 2010). Actuellement, on s'accorde pour désigner cette représentation, selon qu'elle inclut peu ou un volume conséquent de connaissances antérieures, respectivement sous les termes de base de texte ou de modèle de situation (Blanc & Brouillet, 2005). La base de texte est constituée d'éléments et de relations directement dérivés du texte. Il s'agit du niveau qui contient l'ensemble des propositions sémantiques lues ou inférées lors de la lecture. Le modèle de situation contient, quant à lui, les connaissances générales pour comprendre et

interpréter le texte ; il constitue une représentation schématique et intégrée de l'ensemble du texte et de l'ensemble des états, personnages et événements (voir Bianco, 2010 ; Blanc & Brouillet, 2005). Les modèles de compréhension évoqués impactent peu les programmes actuels mais alimentent les propositions des didacticiens (Goigoux & Cèbe, 2013a). Les niveaux de représentation méritent d'être pris en compte, et ils le seront dans notre travail, quant à la façon d'envisager l'évaluation car ils renvoient à des niveaux de compréhension plus ou moins aboutis.

Afin de s'assurer de la validité des épreuves que nous proposons, c'est-à-dire de leur adéquation à la conception sous-jacente du savoir-lire précédemment explicitée, nous prendrons en compte les activités de décodage-identification et de compréhension qu'elles engendrent. Les analyses *a priori* des épreuves reposeront sur des apports théoriques relatifs à ces activités qui seront explicités au fil du texte, en particulier dans le paragraphe *Présentation de l'expérimentation* et le paragraphe *Analyse du texte de Léa*. Les données recueillies seront confrontées à ces analyses, ce qui permettra d'analyser les réponses des élèves à chacune des épreuves (paragr. *L'épreuve de lecture à haute voix (E1)*, *Le rappel de texte (E2)* et *Le questionnaire (E3)*) et de mettre en évidence, le cas échéant, la diversité des comportements individuels. Les relations entre les trois épreuves seront interrogées dans le paragraphe *La prise en compte des trois épreuves* et, dans le paragraphe *Conclusion – discussion*, nous reviendrons sur les principaux résultats et discuterons de l'intérêt de ce dispositif d'évaluation.

PRÉSENTATION DE L'EXPÉRIMENTATION

Les épreuves

Les trois épreuves sont courtes, afin qu'une classe entière puisse être évaluée par une personne sur l'équivalent au plus d'une journée de classe⁸. Pour respecter cette contrainte de durée, un texte unique, que nous avons élaboré, sert de support aux trois épreuves. Elles ont été conçues en vue de faire l'objet de passations individuelles, de type « clinique standardisée », ce qui permet de recourir à l'oral pour les consignes, les questions et les réponses, et donc, de ne pas alourdir la dimension « lecture » de la tâche, le texte – que nous désignons par *Léa* (Fig. 1) – étant le seul écrit présenté à l'élève. En termes de validité,

ce moindre recours à l'écrit nous met à l'abri d'une série de « compétences écran » qui peuvent apparaître lorsque les élèves, jeunes de surcroît, doivent lire des consignes, des questions, des réponses (QCM), voire produire une réponse écrite.

La seconde épreuve de compréhension (E3) est un « questionnaire » portant sur le texte lu. Bien que les modalités d'usages et d'élaboration de ce type d'épreuves aient été fréquemment remises en cause (Cèbe, Goigoux, & Thomazet, 2004 ; Giasson, 1990 ;

Figure 1.
Texte Léa

Léa prépare un gâteau avec son papa. Elle casse les œufs. Elle les mélange avec le chocolat fondu dans le saladier. Papa verse ensuite le mélange dans le moule et le met dans le four. Léa est contente. Demain elle fête ses 6 ans avec ses amis !
« Mais, quelle est cette odeur de gâteau brûlé ? » dit la maman de Léa en entrant dans la cuisine. « Zut ! Le four était trop chaud, il va falloir tout recommencer ».

La première épreuve (E1) repose sur une tâche classique « *la lecture à haute voix* » d'un texte⁹ ; elle renvoie au « savoir-déchiffrer¹⁰ » et à un item du livret : « Lire à haute voix un texte court dont les mots ont été étudiés ». Cet item s'apparente à un « objectif terminal d'intégration » au sens de De Ketele (2001), c'est-à-dire à une macro-compétence impliquant la mobilisation de plusieurs des habiletés de base portées par les autres items du livret relatifs à l'identification des mots. Les résultats à cette première épreuve nous renseignent donc globalement sur le niveau de maîtrise du décodage textuel, le bon décodeur étant celui qui est capable de décoder rapidement et sans erreur le texte qu'il a en face de lui (Martinet & Rieben, 2010).

La première des deux épreuves de compréhension (E2) est un « *rappel de texte* ». Bien que l'usage de ce type d'épreuve soit recommandé pour développer les compétences narratives (Goigoux & Cèbe, 2013a et b ; Kucer, 2014), il est rarement utilisé pour l'évaluation, en particulier, à notre connaissance, à l'école élémentaire en France. À première vue, dans son principe, il paraît simple à mettre en œuvre, mais il s'avère en fait délicat à manier autant pour ceux qui proposent l'épreuve que pour ceux qui y répondent (Blanc & Brouillet, 2005 ; Giasson, 1990). Il contribue à la « reformulation du sens » tel que préconisé par le programme. La réussite à cette épreuve atteste d'une bonne compréhension, en lien avec la qualité de la représentation mentale que le lecteur a pu construire, en particulier sur le plan des relations causales entre les événements (voir Blanc, 2009 ; Goigoux & Cèbe, 2013b).

Goigoux & Cèbe, 2013a et b ; Joole, 2008), le questionnaire reste cependant l'outil privilégié par les enseignants en matière d'évaluation de la compréhension (Fijalkow, 2003). Il répond d'ailleurs explicitement à l'un des objectifs inscrits dans le programme : « Trouver dans un texte ou son illustration la réponse à des questions concernant le texte lu ».

Population concernée, déroulement de la passation

Nous avons interrogé individuellement les soixante-huit enfants de trois classes de CP¹¹, à la fin du mois de juin, donc à une période où les élèves sont, en principe, entrés en phase d'automatisation de la lecture. La passation se déroule de la manière suivante :

- l'élève est d'abord invité à lire le texte (épreuve E1), la consigne donnée étant : « Tu vas lire le texte à voix haute, le mieux que tu peux ». Il ne bénéficie d'aucun délai de préparation et le temps de lecture est chronométré. Les erreurs de lecture (y compris les omissions, ajouts et substitutions) sont notées par l'expérimentateur sur une feuille individuelle de recueil ;
- puis, après avoir relu le texte s'il le souhaite¹², il est soumis à l'épreuve E2 (rappel de texte). Pour la passation, nous avons voulu une consigne compréhensible et « neutre », qui ne privilégie ni l'orientation vers la base de texte (appel explicite à la mémoire), ni vers le modèle de situation (appel explicite à

la compréhension) : « Tu peux relire le texte. Ensuite tu raconteras l'histoire comme si tu la racontais à quelqu'un qui ne la connaîtrait pas. À ce moment-là, je retournerai la feuille et tu ne pourras plus voir le texte. » ;

- dans un troisième temps pendant lequel le texte est remis à sa disposition, l'élève doit répondre au questionnaire (épreuve E3) que nous avons conçu (Fig. 2). Il comprend six questions, posées à l'élève par l'expérimentateur, les questions et les réponses étant exclusivement orales. Les cinq premières questions sont suivies d'une demande de justification dont le recueil facilite l'accès au processus de pensée et au niveau de représentation sollicité lors de la production de la réponse. Cette demande est adressée à l'élève, immédiatement après sa réponse à la question, en reprenant les termes qu'il a lui-même utilisés¹³. La question Q6 n'est pas suivie d'une telle demande car elle constitue en elle-même un appel à justification.

Figure 2.
Le questionnaire

<p>Q1 : Où se passe cette histoire ? Q2 : Qui verse le mélange dans le moule ? Q3 : Quand est-ce que Léa va fêter son anniversaire ? Q4 : Dans cette histoire, est-ce que Léa est contente ? Q5 : Dans cette histoire, qui aide Léa à faire le gâteau ? Q6 : Pourquoi la maman de Léa dit que le four était trop chaud ?</p>

Les productions verbales de l'élève lors des épreuves E2 et E3 (réponses et justifications) sont enregistrées puis transcrites avant d'être analysées.

ANALYSE DU TEXTE LÉA

Globalement, d'un point de vue quantitatif, on peut remarquer que ce texte est court (soixante-quinze mots), composé de neuf phrases courtes et de mots courts : trente-deux mots d'une seule syllabe, dix-huit mots de deux syllabes, deux mots de trois syllabes et un seul mot de quatre syllabes, « recommencer¹⁴ ».

Pour apprécier le niveau de difficulté que pose l'identification des mots du texte, nous avons pris en considération deux facteurs :

- le premier, basé sur l'exposition à l'écrit, permet d'estimer si les élèves ont une chance plus ou moins élevée de reconnaître « directement » un mot, donc de l'identifier automa-

tiquement via une procédure orthographique. La base eManulex¹⁵ (Lété, Sprenger-Charolles & Colé, 2004) permet d'estimer cette caractéristique du texte en procurant des mesures de fréquences de mots (SFI, *Standard Frequency Index*) basées sur un corpus d'écrits adressés à l'enfant. À l'exception du prénom « Léa », tous les mots du texte sont répertoriés dans la base ;

- le deuxième facteur fait appel à d'autres caractéristiques linguistiques des mots relatives au recodage phonologique auquel les élèves recourent lorsqu'ils n'identifient pas automatiquement un mot. Dans un tel cas, les procédures sollicitées sont de nature grapho-syllabiques et/ou grapho-phonémiques (Ecalte & Magnan, 2002), les difficultés liées à la mise en œuvre de telles procédures dépendant respectivement de la complexité des structures syllabiques ou des graphèmes.

À partir de ces deux facteurs, nous avons calculé pour chacun des mots un score de difficulté (Vantourout & Blanc, 2015) qui permet de considérer que sur les cinquante-quatre mots différents du texte, seuls dix mots, « ensuite, œufs, prépare, verse, brûlé, entrant, mélange, odeur, recommencer, contente », pourraient, *a priori*, poser des problèmes d'identification-décodage à une partie de la population évaluée¹⁶.

Pour appréhender plus spécifiquement les aspects liés à la compréhension du texte *Léa*, nous l'avons soumis à une analyse propositionnelle, méthode qui se rattache aux modèles de compréhension selon lesquels la compréhension est un traitement basé sur la mémoire (Blanc & Brouillet, 2003 ; 2005). Cette analyse permet de repérer dans le texte vingt-quatre propositions et quatorze arguments. La compréhension du texte, à plus forte raison si elle est évaluée en dehors de la présence du texte comme c'est le cas lors d'une tâche de rappel, laisse supposer une charge en mémoire non négligeable. Elle nécessite la production de sept inférences concentrées dans la deuxième partie du texte (à partir de « Léa est contente »), dont quatre inférences causales¹⁷. Elle est facilitée par l'intervention de schémas en partie imbriqués, « fabrication d'un gâteau » et « fête d'anniversaire », mais suppose l'inhibition d'un stéréotype relatif au rôle de chaque parent.

En définitive, si l'on se réfère à la typologie de Goigoux & Cèbe (2006), le texte *Léa* est un texte de

« type 2 » : peu chargé culturellement et symboliquement, il renvoie à un univers familier aux enfants, est rédigé dans une langue relativement simple, mais pose des problèmes de compréhension. Ceux-ci concernent « le repérage des personnages de l'histoire, l'explicitation de leurs intentions et de leurs émotions, le déroulement chronologique des différentes actions et la compréhension des relations implicites entre les événements » (p. 50-51).

L'ÉPREUVE DE LECTURE À HAUTE VOIX (E1)

Nous retenons deux indicateurs de performance en ce qui concerne l'épreuve E1 :

- le premier est un score, le MCLM (Mots Correctement Lus par Minute, voir Zorman, 2008). Il correspond à une moyenne relative à la durée de lecture : nombre de mots correctement lus ÷ durée de la lecture (en secondes) × 60. Le nombre de mots correctement lus 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 ;
- le second est le « niveau de lecture », tel qu'estimé par Giasson (2012) qui en distingue trois : le niveau d'autonomie (96 % au moins de mots correctement reconnus, soit ici moins de 3 erreurs), le niveau d'apprentissage (de 90 % à 95 %, de 4 à 7 erreurs) et le niveau de frustration (moins de 90 %, 8 erreurs et plus).

Par construction, ces indicateurs sont liés puisqu'ils sanctionnent tous les deux les mots échoués et ne retiennent que les mots correctement lus. Le MCLM peut être considéré comme une grandeur bidimensionnelle, la « vitesse de lecture » à haute voix. Le niveau de lecture nous informe sur la qualité de la lecture du point de vue de sa précision, c'est une grandeur unidimensionnelle. En recourant à ces deux indicateurs, nous disposons d'information sur la vitesse du décodage textuel ainsi que d'un « zoom » sur sa précision, ces deux qualités étant des critères d'automatisation de l'identification des mots composant le texte.

On peut se demander si ces deux indicateurs se différencient effectivement sur la population testée. En effet, si tel n'était pas le cas, on pourrait se limiter à préconiser le calcul du niveau de lecture plus simple que celui du MCLM puisqu'il n'exige pas le recueil du temps de lecture, ce qui représenterait une simplification importante en vue de l'outillage des enseignants. Le tableau 1, qui met en relation le MCLM et le niveau de lecture, permet de mieux appréhender la relation entre ces deux facteurs et montre qu'il est intéressant de croiser vitesse et précision.

Sur la population testée, la moyenne du MCLM s'élève à 59 avec un écart type de 28, valeurs qui confirment celles proposées en fin de CP par Giasson (2012) et Zorman (2008). Pour construire le tableau 1, nous avons réparti la population en quatre groupes de scores MCLM, construits en fonction de la moyenne et de l'écart type.

Tableau 1.

Effectifs et répartition de la population en pourcentages selon les scores MCLM et les niveaux de lecture

Niveaux de lecture	Mots Correctement Lus par Minute				Totaux
	≤ 30	[31 ; 59[[59 ; 88[≥ 88	
Autonomie	15	21	11	47	
	22 %	31 %	16 %	69 %	
Apprentissage	5	4	1	10	
	7,5 %	6 %	1,5 %	15 %	
Frustration	11			11	
	16 %			16 %	
Totaux	16	19	22	11	68
	23,5 %	28 %	32,5 %	16 %	100 %

De l'examen du tableau 1, il ressort que :

- seuls des élèves « autonomes » atteignent les vitesses de lecture les plus rapides (≥ 88) ;
- seuls des élèves très lents (≤ 30) sont dans le niveau de frustration ;
- une grande majorité d'élèves (69 %) se situe dans le niveau d'autonomie. Mais parmi eux, on observe des vitesses de lecture très dispersées (seuls les scores très bas de MCLM (≤ 30) ne sont pas représentés). Dans cette tranche de la population, on trouve les élèves qui lisent « bien » – du point de vue du décodage-identification – mais plus ou moins vite.

Ainsi le MCLM permet-il, au moins pour la partie de la population – largement majoritaire – qui se trouve dans le niveau d'autonomie, une description plus fine des performances individuelles lors de l'épreuve de lecture à haute voix. Nous retiendrons donc préférentiellement cet indicateur.

LE RAPPEL DE TEXTE (E2)

Méthode d'analyse des rappels et résultats

Étant donné le nombre important de textes de rappel que nous avons à analyser (68), il n'était pas

envisageable de les soumettre à une analyse propositionnelle, qui aurait été de toute façon mal adaptée aux productions orales d'élèves de CP. En effet, il est connu qu'à ce niveau, les élèves produisent généralement des rappels peu structurés qui se prêtent mal à ce type d'analyse (voir Blanc & Brouillet, 2005).

Nous nous sommes inspirés de Giasson (1990) pour élaborer une grille basée sur la division en quinze unités du texte *Léa*, unités dont les limites correspondent aux pauses que feraient des adultes lors d'une lecture à voix haute¹⁸ (Fig. 3).

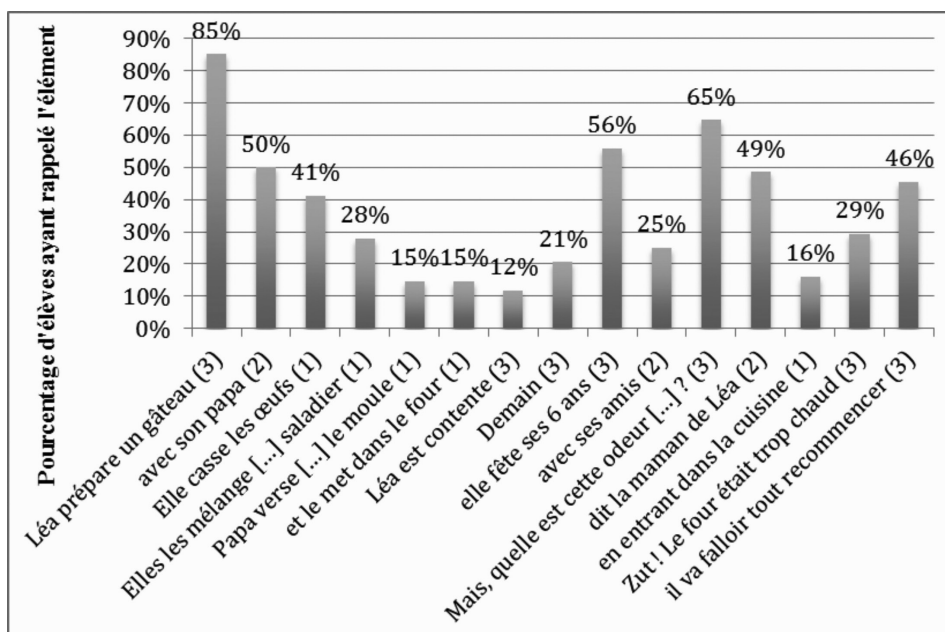
À ces unités sont attribués trois, deux ou un point(s), selon leur importance dans la construction d'une représentation valide de la situation. Nous distinguons ainsi les éléments d'importance élevée (« Léa prépare un gâteau »), moyenne (« avec ses amis ») ou faible (« elle casse les œufs »).

La figure 3 présente le pourcentage d'élèves ayant rappelé chacune des unités du texte (sur l'axe horizontal, à côté de chaque unité figure le nombre de points qui lui est attribué).

Parmi les résultats saillants, certains sont conformes à l'idée que le pourcentage de rappel serait lié positivement à l'importance de l'unité rappelée :

- la majorité des éléments d'importance élevée ou moyenne (6 sur 10) est bien rappelée ($\geq 46\%$) ;

Figure 3.
Pourcentages d'élèves ayant rappelé chacune des unités du texte



– la grande majorité (4 sur 5) des éléments de faible importance est peu rappelée ($\leq 28\%$).

D'autres résultats se démarquent de cette idée et des éléments d'importance élevée sont très mal rappelés. Ainsi :

– le seul élément permettant de saisir la temporalité « demain » n'est rappelé que par 21 % des élèves ;

– « Léa est contente » est le moins rappelé de tous les éléments (12 %), résultat qui va dans le sens des travaux selon lesquels les émotions sont peu restituées par des enfants de six ans (Blanc, 2009) ;

– la « fête des 6 ans », qui, avec « la préparation du gâteau », paraît centrale pour construire une représentation valide de la situation évoquée, est rappelée par moins de la moitié des élèves.

D'un point de vue qualitatif, nous relevons que ces rappels sont très majoritairement peu structurés : il s'agit, le plus souvent et indépendamment des scores, d'une succession-juxtaposition d'éléments, dont très peu attestent d'une prise en compte acceptable des marques temporelles et des liens de causalité entre les différentes composantes du récit, ce qui confirme les observations de Makdissi & Boisclair (2004, cités par Blanc, 2009, p. 34).

Score de rappel

Pour pouvoir apprécier de manière synthétique les rappels produits, nous attribuons à chaque élève un score compris entre 0 et 10, défini de la manière suivante : si n est la somme des points affectés aux unités qu'il a rappelées, le score est égal à $n/32 \times 10$ (32 étant le maximum possible de points). Un rappel qui contiendrait tous les éléments rendant possible l'élaboration d'une représentation intégrant en un tout cohérent la situation évoquée par le texte, c'est-à-dire les sept unités d'importance élevée, aurait un score de 6,6. Seulement six élèves (9 %) atteignent ou dépassent ce score et quarante et un élèves (60 %) ont un score ≤ 5 . Les résultats montrent que les rappels sont majoritairement lacunaires, la moyenne des scores se situant à 4,1 avec un écart type de 2.

La faiblesse des scores pourrait être liée au manque d'éléments de repère dont disposent les élèves quant au « produit attendu », ces derniers ayant été rarement, voire jamais, confrontés à une tâche de rappel¹⁹. Comme le rappelle Blanc (2009)

pour les tâches de production orale, proches du rappel de texte, « l'absence de certaines informations ne reflète pas obligatoirement que l'enfant ne les a pas comprises, mais soulève la possibilité qu'il ne les juge pas importantes à la compréhension de l'histoire qu'il produit » (p. 19). C'est pourquoi, cette épreuve de production, très ouverte, est suivie d'un questionnaire, épreuve plus classique à laquelle les élèves sont davantage habitués.

LE QUESTIONNAIRE (E3)

Méthode d'analyse des réponses et des justifications des élèves

On ne peut valablement analyser, du point de vue de la compréhension, les réponses à une question portant sur un texte sans analyser la nature des liens que la question entretient avec le texte. Parmi les typologies qui prennent en compte cette dimension, en plus de la distinction classique entre questions littérales et inférentielles (Cèbe, Goigoux & Thomazet, 2004 ; Giasson, 1990), nous retiendrons la typologie « Relation Question Réponse » (RQR) (Pearson & Johnson, 1978, cités par Giasson, 1990) qui distingue :

- trois types de relations (R) possibles, attachées au couple « question / réponse correcte » : « explicite (Q) et textuelle (R) », « implicite (Q) et textuelle (R) » ou « implicite (Q) et fondée sur les schémas du lecteur (R) » ;
- deux types de relations concernant les réponses erronées, l'« intrusion textuelle » et l'« intrusion d'un schéma », correspondant, respectivement, à un recours inapproprié au texte ou à l'activation d'un schéma non adapté.

L'analyse *a priori* du questionnaire (les réponses sont considérées comme incluant les justifications) est alors la suivante²⁰ :

- Q1 (Où se passe cette histoire ?) admet plusieurs réponses correctes (la maison de Léa, la cuisine, etc.) qui exigent une inférence de lieu, et la RQR renvoie au type « implicite et fondée sur les schémas du lecteur²¹ » ;
- Q2 (Qui verse le mélange dans le moule ?), Q3 (Quand est-ce que Léa va fêter son anniversaire ?) et Q5 (Dans cette histoire, qui aide Léa à faire le gâteau ?) se rattachent au type de relation « explicite et textuelle », faisant *a priori* appel à la mémoire du texte et à la base de texte ;

- la question Q4 (Dans cette histoire, est-ce que Léa est contente ?) qui supporte toutes les réponses (Léa est contente, ou ne l'est pas, ou les deux, selon que l'on tient compte de la fête et/ou du gâteau brûlé) renvoie au type « explicite et textuel ». Toutefois, le choix d'une justification fondée sur le sens, et non sur un retour au texte, implique la réalisation d'une inférence causale ;
- la question Q6 (Pourquoi la maman de Léa dit que le four était trop chaud ?) sollicite *a priori* le modèle de situation, la réponse correcte impliquant la réalisation d'une inférence causale et l'adoption du point de vue de la maman « elle dit que le four était trop chaud car le gâteau est brûlé ».

Sur la base de cette analyse *a priori*, on peut attendre la réussite la plus faible à Q6 et les plus élevées à Q2, Q3 et Q5.

Analyse des réponses et résultats

Les tableaux 2 et 3 présentent une synthèse des résultats et des analyses portant respectivement sur les réussites et les échecs.

Du point de vue des processus de réponses et du fonctionnement cognitif des élèves, nous distinguons

trois types de justifications acceptables. Le premier, qui fait appel au modèle de situation, constitue le niveau de représentation de la situation évoquée par le texte et de compréhension le plus abouti, sa construction reposant sur la mobilisation de connaissances propres au lecteur, dont parfois des schémas. Par exemple, pour Q1, « dans la cuisine... parce qu'ils font un gâteau, et le four, il est souvent dans la cuisine » ou « chez la petite fille... parce que c'est son anniversaire et elle va inviter ses copines ». Le deuxième renvoie à la base de texte car il s'appuie largement sur des éléments figurant dans le texte ou directement dérivés de celui-ci²². Par exemple, pour Q1, « dans la cuisine... parce qu'à un moment, ils le disent ». Le « retour au texte » est une procédure qui consiste à montrer dans le texte l'endroit où se trouve la réponse. La procédure reposant sur la base de texte est parfois proche du retour au texte, mais seule la première repose sans équivoque sur la mémoire du texte. Les justifications non-acceptables et les réponses erronées sont dues très majoritairement à des intrusions textuelles ou de modèles de situation. Par exemple, à l'origine des échecs à Q2, nous trouvons très fréquemment la réponse « Léa » pour laquelle nous inférons une intrusion textuelle en lien avec une mémoire du texte imprécise.

Tableau 2.
Réussites aux questions et analyse des réussites

Questions	Réussites Réponses correctes	Justifications acceptables	Types de justifications acceptables			Justifications non-acceptables ou absence de réponse	Types de justifications non-acceptables		
			Modèle situation – schéma	Base de texte	Retour au texte		Intrusion modèle situation, schéma	Intrusion textuelle	Autre ou absence de réponse
Q1	87 %	72 %	63 %	7 %	2 %	15 %		7 %	8 %
Q2	68 %	62 %	21 %	29 %	12 %	6 %	1 %		5 %
Q3	72 %	68 %	10 %	34 %	24 %	4 %	1 %		3 %
Q4	Sans objet	94 %	21 %	66 %	7 %	6 %	1 %	3 %	2 %
Q5	91 %	85 %	18 %	50 %	17 %	6 %	1 %		5 %
Q6	54 %	54 %	54 %			Sans objet			

Tableau 3.
Échecs aux questions et analyse des échecs

Questions	Échecs Réponses erronées	Types d'échecs				
		Intrusion modèle situation, schéma	Intrusion textuelle	Intrusion autre	Autre	Absence de réponse
Q1	13 %		3 %	3 %	3 %	4 %
Q2	32 %	6 %	23 %		3 %	
Q3	28 %	9 %	12 %	1 %	3 %	3 %
Q5	9 %	5 %	4 %			
Q6	46 %	28 %	6 %		3 %	9 %

Confrontation des résultats à l'analyse *a priori* :

- la question Q6 est effectivement la moins bien réussie. La réussite, qui renvoie généralement à un niveau de compréhension élevé, repose sur un modèle de situation plus ou moins élaboré ; toutefois, dans un certain nombre de cas, celui-ci se démarque peu de la base de texte. Les questions Q2, Q3 et Q5 bénéficient effectivement d'un bon taux de réussite et sont le plus souvent justifiées par un recours à la base du texte ou un retour au texte ;
- contrairement à nos attentes, la question Q1 est quasiment la mieux réussie (2^e rang, 87 %). L'analyse des justifications permet de comprendre : seules 63 % sollicitent le modèle de situation, alors que les autres sollicitent la mémoire du texte et la base de texte, ou font un retour au texte. Plus généralement, une question peut finalement s'avérer de type

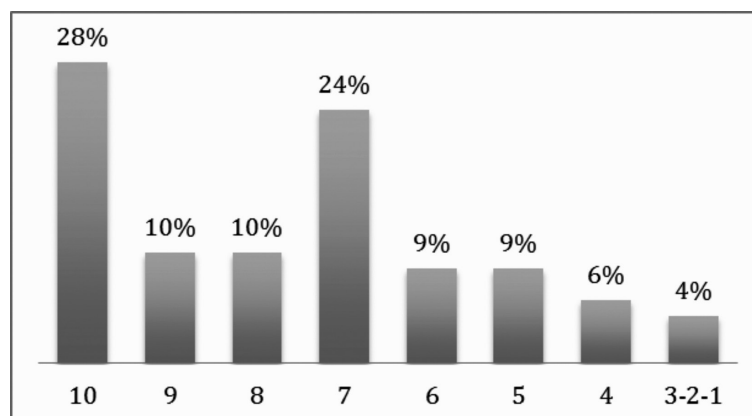
littéral et inférentiel, même si, lors de l'analyse *a priori* (qui peut mettre en avant la procédure la plus efficace), elle semblait relever largement, voire exclusivement, de l'un des deux types.

Score au questionnaire

Nous avons attribué à chaque élève un score sur 10 points : un point par bonne réponse et justification acceptable, à l'exception de Q4 pour laquelle seule la justification est prise en compte et de Q6 qui ne comporte pas de demande de justification.

Sur la figure 4, on constate une réussite élevée pour 48 % des élèves (scores 10, 9, 8), moyenne pour 42 % (scores 7, 6, 5), alors que 10 % des élèves sont en difficulté, voire en grande difficulté (scores 4, 3, 2, 1) pour comprendre ce texte de type 2.

Figure 4.
Répartition de la population (en %) selon le score au questionnaire



LA PRISE EN COMPTE DES TROIS ÉPREUVES

Les relations entre les trois épreuves

Pour appréhender les éventuelles relations entre les trois variables constituées par les scores à chacune des épreuves, nous avons calculé les coefficients de corrélation entre les variables prises deux à deux. Les résultats saillants sont les suivants :

- les deux indicateurs de la compréhension que sont les scores au rappel et au questionnaire sont positivement corrélés ($r = 0.53$). On note cependant que, au niveau global, le questionnaire (E3) est mieux réussi que le rappel (E2), les écarts types des deux distributions de notes ayant des valeurs voisines (2,2 pour E3 et 2 pour E2) alors que les moyennes sont respectivement égales à 7,5 (questionnaire) et 4,1 (rappel). Cette différence, en faveur du questionnaire, peut s'expliquer car les deux épreuves de compréhension évaluent *a priori* des compétences différentes, la majorité des questions de E3 pouvant être réussies par un simple recours à la base de texte, voire un retour au texte (cf. paragr. *Analyse des réponses et résultats* : les questions explicites et textuelles représentent 60 % du score), alors que le rappel, pour être bien réussi, exige le recours au modèle de situation (cf. paragr. *Score de rappel* : les 7 unités de texte les plus importantes dans la construction d'une représentation valide de la situation représentent 66 % du score). En outre, l'épreuve de rappel, contrairement au questionnaire, ne s'inscrit pas dans les habitudes scolaires, ce qui peut dérouter certains élèves ;
- le MCLM (E1) est positivement corrélé à chacune des épreuves de compréhension. La corrélation statistique semble plus forte entre les scores MCLM et questionnaire ($r = 0.58$) qu'entre les scores MCLM et rappel ($r = 0.39$). Cela pourrait s'expliquer par le côté « scolaire » commun aux deux épreuves de lecture à haute voix (E1) et de questionnaire (E3), alors que l'épreuve de rappel (E2), elle, rompt avec les habitudes scolaires.

Les évaluations individuelles

Pour nous, dont le projet est d'évaluer individuellement les élèves, la relative simplicité des résultats globaux ne doit pas faire oublier la complexité des

comportements individuels. Afin d'illustrer celle-ci, nous avons considéré les observations correspondant aux soixante-huit élèves comme autant de points dans un espace à trois dimensions dont les coordonnées sont les scores à chacune des trois épreuves. Nous avons construit les figures 5, 6 et 7, projections de ce nuage de points sur les trois plans correspondant respectivement aux épreuves MCLM et questionnaire (Fig. 5), MCLM et rappel (Fig. 6), rappel et questionnaire (Fig. 7). Nous avons matérialisé dans chacune des figures l'axe vertical et l'axe horizontal correspondant aux moyennes respectives des deux variables représentées, afin de visualiser, pour chaque observation, la « qualité » des performances aux deux épreuves concernées (la moyenne constituant le seuil entre les bons et les mauvais scores de chaque distribution).

La diversité des comportements individuels est illustrée par la dispersion des points représentant les observations. Elle appelle quelques commentaires²³ :

- certains élèves ont un bon score en MCLM et des scores faibles dans les épreuves de compréhension (un cas typique correspond à Cat. 3). Ce profil de lecteur est connu des enseignants : il s'agit des élèves pour lesquels « la vitesse de déchiffrage n'améliore pas nécessairement la qualité de la compréhension » (Goigoux & Cèbe, 2013a, p. 9) ;
- plus paradoxalement, on trouve quelques élèves qui, confrontés à ce texte de type 2 – considéré donc *a priori* comme plutôt facile à déchiffrer et présentant des difficultés de compréhension –, ont un excellent niveau en compréhension (très bons scores à E2 et E3) avec une valeur du MCLM plutôt faible (Cat. 4). Mais, ces élèves, lents pour décoder, font relativement peu d'erreurs ; ce serait la précision de leur décodage qui leur permettrait de réussir les épreuves de compréhension. Nous retrouvons là également un profil de lecteur déjà repéré²⁴ (voir Adolf, Catts & Little, 2006).

Ces deux premiers commentaires confirment l'existence de « profils atypiques » au sens de Braibant (2004), c'est-à-dire d'élèves pour lesquels on constate un décalage ou encore une supériorité relative dans l'un des types d'épreuves. Ils montrent combien il est important d'évaluer la compréhension autonome de texte écrit dès le CP. En effet, il apparaît que, en se privant de cette évaluation, on court

Figure 5.

Pour chaque élève : valeurs du MCLM (en abscisse) et du score au questionnaire (en ordonnée)

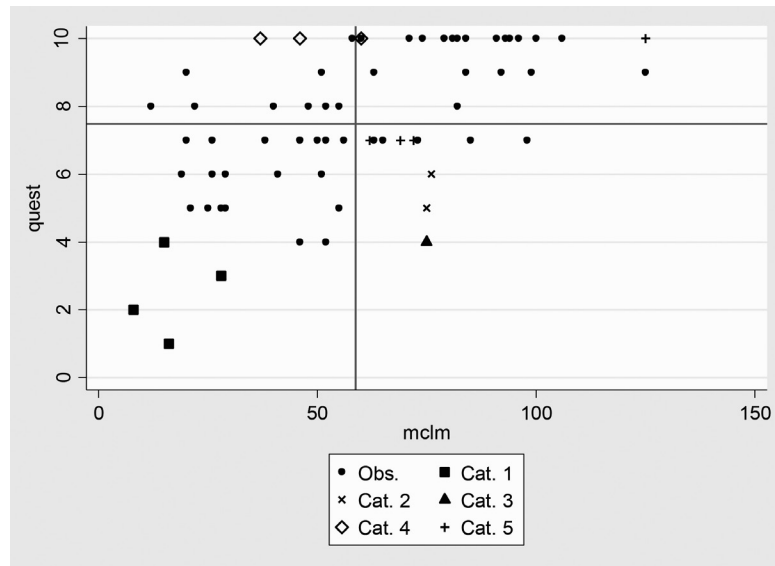
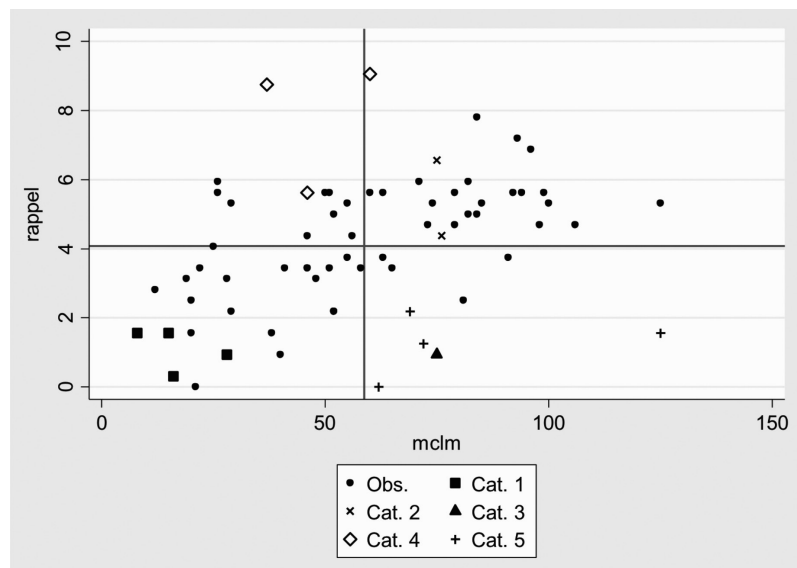


Figure 6.

Pour chaque élève : valeurs du MCLM (en abscisse) et du score au rappel (en ordonnée)

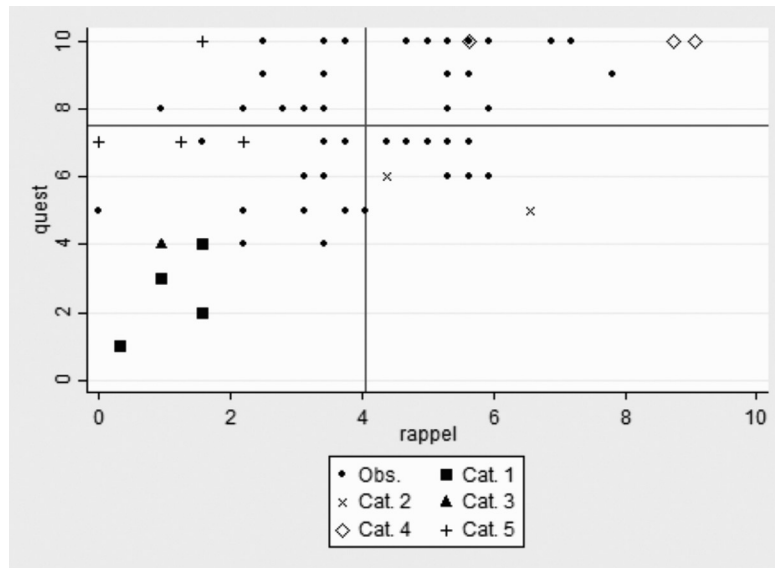


un double risque : avec les élèves concernés par le premier commentaire, c'est-à-dire pour lesquels une relative aisance en décodage peut dissimuler des difficultés en compréhension (voir Yuil & Oakhill, 1991), le risque de créer des « faux positifs », d'imaginer, par une sorte d'effet de halo, compétents des élèves qui ne le sont pas ; à l'opposé, avec les élèves visés par le deuxième commentaire, qui, malgré de

faibles performances en décodage, font preuve de performances élevées en compréhension, celui de créer des « faux négatifs », de dénier toute compétence à des élèves pourtant compétents dans ce domaine. Ce dernier point doit inciter à la prudence quant à la non-évaluation de la compréhension, assez systématiquement pratiquée en classe, pour les élèves considérés comme de mauvais décodeurs en CP.

Figure 7.

Pour chaque élève : valeurs des scores au rappel (en abscisse) et au questionnaire (en ordonnée)



On trouve des élèves performants, voire très performants, dans les épreuves E1 et E2 qui réalisent de mauvais scores à E3 (Cat. 2). Respectivement, on trouve des élèves ayant de mauvais scores à E2 alors qu'ils sont performants à E1 et E3 (Cat. 5). Ceci montre que, lorsque la perspective est de produire des évaluations individuelles, ce qui est notre but, les épreuves E2 et E3 ne sont pas substituables. On repère enfin quatre élèves (Cat. 1) qui sont en très grande difficulté : ils ont des scores très bas aux trois épreuves et sont, selon la classification de Giasson (2012), au niveau de frustration en lecture (paragr. *L'épreuve de lecture à haute voix (E1)*).

Résultats globaux

Au niveau global, la population testée se répartit de la manière suivante :

- vingt et un élèves, soit 31 % de la population testée réalisent des scores supérieurs ou égaux à la moyenne des scores à chacune des trois épreuves ;
- dix-sept élèves, soit 25 % de la population, obtiennent trois scores inférieurs aux moyennes respectives.

Ces deux résultats permettent uniquement de dire que, au sein de cette population issue très majoritairement de milieux favorisés (cf. note 11), dix-sept

élèves obtiennent de moins bons résultats que leurs camarades, à ces épreuves reposant sur un même texte. Bien que ces moyennes ne puissent constituer un seuil révélateur de la difficulté²⁵, on peut estimer que, parmi les dix-sept élèves mentionnés ci-dessus, quatre élèves, ceux de la Cat. 1 dont les trois scores sont extrêmement bas, sont en difficulté en lecture en fin de CP.

Sept élèves (ceux des Cat. 3 et 4), soit 10 % de la population, ont des performances comparables aux deux épreuves de compréhension (toutes les deux bonnes / toutes les deux mauvaises) et qui contrastent avec la performance à E1 (respectivement : mauvaise / bonne). Seuls vingt-trois élèves, soit 34 % de la population, réalisent des performances différentes aux deux épreuves de compréhension (on trouve parmi eux les élèves des Cat. 2 et 5). Ils se répartissent à parts pratiquement égales entre ceux qui obtiennent un meilleur score à E2 (11 élèves) ou à E3 (12 élèves).

Bien entendu, ces résultats quantitatifs doivent être relativisés : ils dépendent *a priori* du texte choisi comme support aux trois épreuves (ici le texte *Léa*, cf. Fig. 1), mais aussi de la conception du questionnaire. Nous apportons cependant dans ce travail des méthodes d'analyse du texte de référence (paragr. *Analyse du texte Léa*), des questions (paragr. *Méthode d'analyse des réponses et des justifications des élèves*), des rappels (paragr. *Méthode d'analyse des rappels et*

résultats), ainsi qu'une définition des divers scores – notamment une définition originale du score de rappel (paragr. *Score de rappel*) – qui doivent permettre la mise en œuvre d'évaluations conçues à partir d'autres textes. En s'appuyant sur ces outils fonctionnels, il s'agira alors, via l'utilisation d'ingénieries évaluatives courtes, d'être en mesure, au sein des classes et sur le temps scolaire, d'identifier les forces et faiblesses des élèves en lecture.

CONCLUSION – DISCUSSION

Le dispositif expérimental mis en place, destiné à évaluer le savoir-lire en fin de CP, prend en compte deux dimensions du savoir-lire, le décodage-identification et la compréhension autonome de texte écrit. Il prévoit la passation individuelle et orale de trois épreuves, bâties autour d'un même texte et qui font l'objet d'analyses *a priori* auxquelles sont confrontées les données expérimentalement recueillies.

L'utilisation de l'oral présente de nombreux avantages, en particulier pour l'épreuve E3 (questionnaire) : elle permet d'éviter des compétences écran liées au traitement des questionnaires écrits et de demander des justifications, ce qui facilite l'accès à la représentation de la situation construite par le lecteur. Les très faibles taux de non-réponses, que l'on peut également attacher à la verbalisation des questions et des réponses, constituent de solides arguments pour justifier ce choix de l'oral.

Le recours à deux épreuves (le rappel de texte E2 et le questionnaire E3) pour évaluer la compréhension, s'il a évidemment un « coût », présente des avantages : en effet, outre que disposer de plusieurs épreuves améliore « mécaniquement » la qualité du jugement évaluatif porté sur l'élève, les analyses qualitatives ont montré que E2 et E3 ne sollicitaient pas forcément le même niveau de représentation²⁶. Par exemple, il est intéressant de confronter les 72 % de réponses correctes (« demain ») à la question Q3 lors de l'épreuve E2 (Quand est-ce que Léa va fêter son anniversaire ?), aux 21 % de rappel de l'élément « demain » lors de l'épreuve E2 (tableau 2 et Fig. 3). L'analyse qualitative montre que les justifications acceptables à Q3 renvoient largement à la base de texte, ou directement au texte (retour), c'est-à-dire très probablement à un niveau de représentation peu élaboré ; alors que, lors du rappel, l'emploi de « demain » atteste le plus souvent que

l'élève a saisi la temporalité différente dans laquelle s'inscrivent les situations « préparation du gâteau » et « fête d'anniversaire » – comme dans cet énoncé « et après, demain, le jour de ses 6 ans, elle fera la fête » – ce qui contribue à la cohérence du modèle de situation.

La relative simplicité des résultats globaux, qui indiquent des corrélations positives entre les variables, masque cependant une grande dispersion des résultats individuels (Fig. 5, 6, 7). Ainsi, si l'on veut être en mesure de savoir où chaque élève se situe relativement aux compétences en lecture, les analyses qualitatives que nous avons réalisées plaident en faveur de l'évaluation du décodage textuel et de la compréhension autonome de texte écrit, à partir d'une diversité d'épreuves.

L'utilisation d'une épreuve de rappel individuel de texte pour évaluer la compréhension en lecture, en milieu scolaire et avec des élèves de CP, constitue certainement l'aspect le plus innovant de ce travail. Cependant, la mise en œuvre de cette épreuve ne se fait pas sans difficulté :

- elle pose des problèmes méthodologiques que nous avons cherché à dépasser en nous dotant d'une grille de recueil des productions. Comme l'utilisation des grilles qualitative est délicate et coûteuse, nous avons opté, à regret car on sait « que la qualité d'un rappel de récit est plus révélatrice d'une bonne compréhension que la quantité d'événements relatés » (Goigoux & Cèbe, 2013a, p. 5), pour une grille quantitative qui valorise la mémoire ;
- si l'analyse des éléments rappelés par l'ensemble des élèves (Fig. 1) a produit des résultats intéressants, les rappels et résultats individuels (scores de rappel) sont décevants. La tâche s'est avérée compliquée pour la plupart des élèves, certainement, comme nous l'avons dit, en raison d'une rupture avec les tâches utilisées pour l'enseignement. Pourtant, l'intérêt de ce type d'épreuves, pour développer les compétences narratives en production dans l'apprentissage de la compréhension, ne fait pas de doute (Blanc, 2009 ; Goigoux & Cèbe, 2013a et b ; Kucer 2014).

De nombreux professeurs sont démunis quand ils doivent évaluer leurs élèves, notamment en lecture en CP. Ce constat n'a rien d'étonnant, puisque les chercheurs s'accordent sur l'extrême difficulté à évaluer la compréhension écrite²⁷ (Rocher, 2013), surtout chez les lecteurs débutants (Bianco, 2010).

L'un des buts de notre travail était d'élaborer et de tester un dispositif suffisamment explicite dans ses intentions et sa mise en œuvre pour qu'il puisse servir de base à une ingénierie évaluative courte qui serait, dans un second temps (travail en projet), testée auprès de professeurs. Concrètement, lors de la présente recherche, les modalités de passation retenues ont permis de respecter les contraintes de temporalité de la classe (la durée moyenne des entretiens individuels a été de 6-7 minutes). Bien sûr, le travail du professeur, lorsqu'il évalue, ne se limite pas au recueil des données : il doit ensuite exploiter ces données en vue de produire un jugement évaluatif. Notre recherche présente de manière explicite la façon dont les données recueillies sont qualitativement analysées ainsi que le mode de calcul des scores respectifs aux trois épreuves. De ce point de vue, le dispositif proposé constitue un outil d'évaluation communicable. Une partie importante du travail à faire dans le projet envisagé consistera donc à regarder comment les professeurs s'approprient (ou non) les méthodes d'analyse des données individuelles et comment ils traduisent les résultats des analyses en un jugement évaluatif porté sur le livret scolaire.

NOTES

1. Le cycle 2, au moment de la recherche puis de la rédaction de l'article, comprend les deux premières classes de l'école primaire, CP et CE1. À compter de la rentrée scolaire 2016, la classe de CE2 est intégrée au cycle 2.
2. Les programmes en vigueur lors de la réalisation de la recherche sont ceux de 2008 (MEN, 2008). Dans les programmes de 2015 (BO spécial du 26 novembre 2015) qui entreront en vigueur en septembre 2016, la place de la compréhension est désormais solidement établie. Elle est notamment affirmée dès la présentation générale des objectifs du cycle 2 : « La maîtrise du fonctionnement du code phonographique [...] constitue un enjeu essentiel de l'apprentissage du français au cycle 2. Cependant, l'apprentissage de la lecture nécessite aussi de comprendre des textes narratifs ou documentaires, de commencer à interpréter et à apprécier des textes, en comprenant ce qui parfois n'est pas tout à fait explicite » (p. 4).
3. En raison de son aspect fonctionnel et de sa légitimité institutionnelle, ce livret constitue un « outil » pour l'évaluation massivement utilisé dans les écoles.
4. Notre engagement dans cette recherche est lié à la participation de l'un d'entre nous à un stage de circonscrip-

tion consacré à l'évaluation de la lecture au cycle 2, dans le cadre de la mise en œuvre du livret scolaire de l'école primaire.

5. À l'origine, les APDE ont été développées en lien avec l'évaluation en mathématiques.

6. La définition classique et générale de la validité est souvent résumée par une question : l'épreuve mesure-t-elle bien (ou évalue-t-elle bien) ce qu'elle est censée mesurer (ou évaluer) ? La validité psycho-didactique est une approche qualitative de la validité qui conduit à s'intéresser aux contenus 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.

7. Le rôle de la mémoire est important dans les théories de la compréhension de texte, avec comme hypothèse que la compréhension est étroitement liée à une mémorisation efficace des éléments du texte déjà lus et au rappel de connaissances et d'expériences personnelles stockées en mémoire (Blanc & Brouillet, 2003).

8. Par cette contrainte de durée, nous voulions ménager la possibilité de proposer à des professeurs le dispositif mis en place afin qu'ils l'utilisent de manière autonome dans leur classe.

9. Il s'agit d'une épreuve classique, dite de « fluence », telle que popularisée en France par Zorman et son équipe (Pourchet & Zorman, 2013 ; Zorman, 2008). Cependant, elle ne peut pas être qualifiée d'« épreuve de fluence » au sens fort. En effet, la lecture fluente, qui requiert d'être *a minima* un bon décodeur, demande bien plus et s'avère une compétence experte pour un élève de CP (voir Martinet & Rieben, 2010).

10. 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.

11. Les trois classes appartiennent à la même école qui accueille principalement des élèves de milieux favorisés, ce qui interdit de considérer la population évaluée comme représentative de l'ensemble des élèves de CP.

12. Nous avons constaté que la majorité des élèves relit le texte. Parmi ceux qui ne le font pas dans un premier temps, quelques-uns demandent à pouvoir le faire ensuite lorsqu'ils sont confrontés à la difficulté de la tâche de rappel (nous le leur permettons).

13. Par exemple si à la question Q1 l'élève répond « dans la maison de Léa », alors l'expérimentateur enchaîne par « comment sais-tu que cette histoire se passe dans la maison de Léa ? ».

14. Des éléments présentés ici sont proches de ceux qui interviennent dans le calcul d'indices de lisibilité, tel celui de Flesch-De Landsheere. Nous n'avons pas utilisé cet indice car il ne serait pas valide pour les textes courts de l'école primaire (Henry, 1980).

15. eManulex : [<http://www.manulex.org/fr/home.html>]

16. Ce score pourrait entrer dans le calcul d'un indice de difficulté du décodage et permettre la comparaison de textes.

17. Par exemple, le lecteur doit inférer la causalité entre (1) « Léa est contente. » et (2) « Demain elle fête ses 6 ans avec ses amis ! » : (1) parce que (2).

18. Ces unités ne correspondent pas aux propositions de l'analyse propositionnelle.

19. Certains des comportements observés, une fois la consigne donnée, peuvent être interprétés comme des difficultés en matière de représentation ou de redéfinition de la tâche.

20. L'utilisation de la typologie RQR repose sur la réponse effectivement produite par l'élève, ce qui la démarque d'autres typologies reposant sur le type de réponse attendue par l'enseignant (par exemple, réponse littérale). Pour cette analyse *a priori*, nous utilisons cette typologie pour anticiper sur les différents processus que des élèves mettront en œuvre pour répondre.

21. Les réponses et justifications observées font appel au schéma « préparation du gâteau » ou « fête d'anniversaire ».

22. Pour déterminer les niveaux de compréhension, nous ne distinguons que deux niveaux de représentation : la base de texte que nous différencions du modèle de situation (voir parag. « Évaluer la lecture dans le cadre des APDE »).

23. Pour assurer la lisibilité dans les figures, les observations sur lesquelles portent les commentaires ont été réparties en 5 catégories, chacune affectée d'un signe distinctif.

24. Parmi les élèves évalués, figure celui qui peut être considéré comme le deuxième meilleur compreneur de tous, avec un score de 10 au questionnaire et de 88 % au rappel de texte (2^e rang), alors qu'il se situe au 54^e rang du point de vue de la vitesse de lecture (MCLM = 37), mais en ne commettant que 4 erreurs. Le lien que nous supposons entre précision et compréhension nous semble pouvoir être rapproché d'un résultat établi par Adlof, Catts & Little (2006) : « *Results showed that fluency did not account for unique variance in reading comprehension and that few individuals had problems in fluency separate from word recognition accuracy or listening comprehension* » (p. 933).

25. La moyenne est retenue pour chaque épreuve comme seuil entre les « bons » et les « mauvais scores ». Les scores des élèves sont plus ou moins éloignés de ce seuil, avec deux catégories seulement (bon / mauvais), nous simplifions une réalité qui est bien plus complexe. Nous retrouvons notamment ici les limites déjà repérées dans des travaux sur l'échec scolaire qui montrent que « lors de l'évaluation des performances des élèves, l'enseignant est prisonnier du microcosme que constitue sa classe » (Crahay, 2003, p. 83) : le fait d'être « bon » ou « mauvais » dépend de la population à laquelle on appartient.

26. Le questionnaire peut être en grande partie réussi par un recours à la base de texte, voire un retour au texte (cf. parag. *La prise en compte des trois épreuves*), alors que le rappel de texte, qui est une tâche relativement ouverte, sollicite la mobilisation de compétences narratives en production. La réussite à cette dernière épreuve dépend directement de la qualité de la représentation mentale construite (Blanc, 2009 ; Goigoux & Cèbe, 2013a et b).

27. Comme le souligne Rocher (2013, en citant Fletcher, 2006), « l'évaluation de la compréhension est difficile car il ne s'agit pas d'un processus manifeste directement observable. Seuls les produits du processus de compréhension sont observés, et une inférence est alors réalisée sur la nature du processus et sur la qualité de la compréhension » (p. 69).

RÉFÉRENCES

- Adlof, S., Catts, H., & Little, T. (2006). Should the Simple View of Reading Include a Fluency Component? *Reading and Writing: An Interdisciplinary Journal*, 19, 933-958.
- Berger, M.-J., & Desrocher, A. (dir.). (2011). *L'évaluation de la littératie*. Ottawa : Presses de l'Université d'Ottawa.
- Bianco, M. (2014). *Conseil supérieur des programmes : Propositions pour une programmation de l'enseignement de la compréhension en lecture*. 24 pages. Repéré à [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 texte : peut-on l'apprendre et l'enseigner. Dans M. Crahay & M. Dutrévis, *Psychologie des apprentissages scolaires* (p. 229-254). Bruxelles : De Boeck.
- Blanc, N. (2009). *Lecture et habiletés de compréhension chez l'enfant*. Paris : Dunod.
- Blanc, N., & Brouillet, D. (2005). *Comprendre un texte – L'évaluation des processus cognitifs*. Paris : In Press.
- Blanc, N. & Brouillet, D. (2003). *Mémoire et compréhension – Lire pour comprendre*. Paris : In Press.
- Braibant, J.-M. (1994). Le décodage et la compréhension – Deux composantes essentielles de la lecture en 2^e primaire. Dans J. Grégoire & B. Piérart, *Évaluer les troubles de la lecture : les nouveaux modèles théoriques et leurs implications diagnostiques* (p. 173-194). Bruxelles : De Boeck.
- Cain, K. (2010). *Reading Development and Difficulties*. Oxford : Blackwell Publishing Ltd.
- Cèbe, S., Goigoux, R., & Thomazet, S. (2004). *Enseigner la compréhension*. Document non-publié. 38 pages.
- Chauveau, G. (2007). *Le savoir-lire aujourd'hui*. Paris : Retz.
- Crahay, M. (2003). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck.

- De Ketele, J.-M. (2001). Évaluation des compétences : une problématique émergente. Dans G. Figari & M. Achouche (dir.), *L'activité évaluative réinterrogée* (p. 39-43). Bruxelles : De Boeck.
- Ecalle, J., & Magnan, A. (2010). *L'apprentissage de la lecture et ses difficultés*. Paris : Armand Colin.
- Ecalle, J., & Magnan, A. (2002). *L'apprentissage de la lecture – Fonctionnement et développement cognitif*. Paris : Armand Colin.
- Fijalkow, E. (2003). *L'enseignement de la lecture-écriture au cours préparatoire, entre tradition et innovation*. Paris : L'Harmattan.
- Giasson, J. (2012). *La lecture : apprentissages et difficultés*. Bruxelles : De Boeck.
- Giasson, J. (1990). *La compréhension en lecture*. Bruxelles : De Boeck.
- Goigoux, R., & Cèbe, S. (2013a). *Lectorino & Lectorinette – Apprendre à comprendre des textes narratifs*. Paris : Retz.
- Goigoux, R., & Cèbe, S. (2013b). Enseigner la compréhension à l'école élémentaire : des résultats de recherches à la conception d'un outil didactique. *Revue Recherches*, 58, 29-45.
- Goigoux, R. & Cèbe, S. (2006). *Apprendre à lire à l'école*. Paris : Retz.
- Gonthier-Maurin, B. (2012). *Sénat – Rapport fait par la mission d'information sur le métier d'enseignant* (publication n° 601, session ordinaire de 2011-2012). Repéré à [https://www.senat.fr/notice-rapport/2011/r11-601-notice.html]
- Gough, P., & Tunmer, W. (1986). Decoding, Reading and Reading Disability. *Remedial and Special Education*, 7(1), 6-10.
- Henry, G. (1980). Lisibilité et compréhension. *Communication & langages*, 45(1)7-16.
- Joole, P. (2008). *Comprendre les textes écrits*. Paris : Retz.
- Kucer, S. B. (2014). What Retellings Can Tell us about the Nature of Reading Comprehension in School Children. *Australian Journal of Language and Literacy*, 37(1), 31-44.
- 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.
- Martinet, C., & Rieben, L. (2010). L'apprentissage initial de la lecture et ses difficultés. Dans M. Crahay & M. Dutrévis, *Psychologie des apprentissages scolaires* (p. 190-224). Bruxelles : De Boeck.
- MEN. (2010). *Note de synthèse sur la mise en œuvre de la réforme de l'enseignement primaire*. Repéré à [http://www.education.gouv.fr/archives/2012/refondonslecole/wp-content/uploads/2012/07/rapport_igen_igaenr_note_de_synthese_sur_la_reforme_de_l_enseignement_primaire_juillet_2010.pdf]
- MEN. (2008). *Les nouveaux programmes de l'école primaire*. Repéré à [http://cache.media.education.gouv.fr/file/02_fevrier/24/3/BOEcolePrimaireWeb_24243.pdf]
- MENESR. (2005). *Les acquis des élèves, pierre de touche de la valeur de l'école ?* Rapport. Repéré à [http://www.education.gouv.fr/archives/2012/refondonslecole/wp-content/uploads/2012/07/rapport_igen_igaenr_les_acquis_des_eleves_pierre_de_touche_de_la_valeur_de_l_ecole_2005.pdf]
- MEN-MESR. (2012). *Les composantes de l'activité professionnelle des enseignants outre l'enseignement dans les classes*. Rapport. Repéré à [http://cache.media.education.gouv.fr/file/2012/05/8/2012-070_rapport_227058.pdf]
- MENJVA-MESR. (2011). *Note de synthèse sur le suivi de la réforme de l'enseignement primaire : pilotage local et suivi des élèves*. Repéré à [http://www.education.gouv.fr/archives/2012/refondonslecole/wp-content/uploads/2012/07/rapport_igen_note_de_synthese_sur_le_suivi_de_la_reforme_de_l_enseignement_primaire_juillet_2011.pdf]
- Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, 11(4), 357-383.
- Pourchet, M., & Zorman, M. (2013). *Fluence de lecture CP / CE – Volume I*. Grenoble : La Cigale.
- Rocher, T. (2013). *Mesure de compétences : les méthodes se valent-elles ? Questions de psychométrie dans le cadre de l'évaluation de la compréhension de l'écrit* (Thèse de doctorat non-publiée, Université Paris Ouest Nanterre La Défense).
- SCEREN. (2012). *Le livret scolaire de l'école primaire*. 63 pages.
- Sprenger-Charolles, L., & Colé, P. (2003). *Lecture et dyslexie – Approche cognitive*. Paris : Dunod.
- Vantourout, M., & Goasdoué, R. (2014). Approches et validité psycho-didactiques des évaluations. *Éducation et Formation*, e-302, 139-156.
- Vantourout, M., & Blanc, C. (2015). *Une épreuve psycho-didactique pour évaluer en lecture en fin de CP*. Communication présentée au 27^e colloque de l'Admées-Europe : « L'évaluation à la lumière des contextes et des disciplines ». Liège : Belgique, janvier.
- Yuill, N., & Oakhill, J. V. (1991). *Children's Problems in Text Comprehension*. Cambridge: Cambridge University Press.
- Zorman, M. (2008). *Évaluation de la lecture en fluence*. Document non-publié. 12 pages.