



**HAL**  
open science

# Étude des effets retest à deux épreuves de lecture chez des normolecteurs de CM2

Camille Arnaud

► **To cite this version:**

Camille Arnaud. Étude des effets retest à deux épreuves de lecture chez des normolecteurs de CM2. Médecine humaine et pathologie. 2021. dumas-03470520

**HAL Id: dumas-03470520**

**<https://dumas.ccsd.cnrs.fr/dumas-03470520v1>**

Submitted on 8 Dec 2021

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

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

---

MÉMOIRE DE RECHERCHE présenté pour l'obtention du  
**CERTIFICAT DE CAPACITE D'ORTHOPHONISTE  
ET DU GRADE DE MASTER 2 D'ORTHOPHONIE**

Présenté et soutenu le 10 juin 2021 par Camille ARNAUD

Né(e) le 12/06/1997

**ETUDE DES EFFETS RETEST  
A DEUX EPREUVES DE LECTURE  
CHEZ DES NORMOLECTEURS DE CM2.**

Directeur de Mémoire : Gilles Leloup

Co-directeur(s) de Mémoire : Karine Eskinazi-Harrar

Nice

2021



# Remerciements

Je tiens à remercier toutes les personnes ayant contribué à la réussite de ces études, et plus particulièrement à l'élaboration et la rédaction de ce mémoire.

Je remercie tout d'abord mes deux directeurs de mémoire, Mr Gilles Leloup et Mme Karine Eskinazi-Harrar qui ont su m'aiguiller, m'encourager et qui m'ont soutenu dans la réalisation de ce projet. Merci de m'avoir aidée à repousser mes limites pour donner le meilleur de moi-même.

Je remercie également Mme Auriane Gros, et Mme Valeria Manera, qui m'ont donné de nombreux et précieux conseils et qui ont très largement contribué au bon déroulé de ce mémoire.

Je dois également beaucoup aux personnes qui m'ont soutenue moralement pendant toute la durée de l'étude. Je pense notamment à mon fiancé, à mes collègues de promotion ainsi qu'à mes très proches amies. Merci notamment à Julie, Manon, Audrey, Ludivine, Marie et Vanessa, mes piliers niçois ainsi que Marine, Servane, Ophélie et Chloé, mes camarades de galère.

Merci à mes maîtres de stage, qu'elles soient niçoises ou corses, pour m'avoir apporté la formation et le soutien nécessaires pour devenir orthophoniste.

Merci à ma famille de m'avoir encouragée dans cette voie, qui j'en suis sûre maintenant est la mienne.

Merci à ma belle-famille et à ma nièce Jade de m'avoir aidée à garder le sourire jusqu'au bout. Et bien sûr, merci à Mimi qui a été présente à chaque instant.

Je remercie enfin les écoles élémentaires ayant participé à l'étude, ainsi que leurs élèves de CM1-CM2 qui ont été accueillants et disponibles tout au long des expérimentations.

# Sommaire

INTRODUCTION.....	3
<b>PARTIE THEORIQUE</b> .....	4
<b>1. La lecture</b> .....	4
1.1 Devenir lecteur expert.....	4
1.1.1 La lecture : définition .....	4
1.1.2 Apprendre à lire .....	4
1.1.3 Les spécificités du français.....	5
1.2 Les compétences cognitives sous-jacentes .....	6
1.2.1 Compétences phonologiques .....	6
1.2.2 Compétences visuelles et visuo-attentionnelles .....	7
<b>2. L'évaluation de la lecture</b> .....	8
2.1 Les modèles de lecture.....	8
2.1.1 Les modèles connexionnistes à double voie.....	8
2.1.2 Les modèles connexionnistes en triangle .....	8
2.1.3 Le modèle multi-traces .....	9
2.2 Tests évaluant la lecture.....	9
2.2.1 Les épreuves d'identification de mots isolés.....	9
2.2.2 Les épreuves de leximétrie et de compréhension écrite. ....	10
2.2.3 Les épreuves évaluant les compétences cognitives sous-jacentes.....	10
<b>3. Propriétés métriques des tests et leurs limites</b> .....	11
3.1 Les propriétés métriques d'un test : validité, fidélité, sensibilité et spécificité .....	11
3.1.1 La validité d'un test.....	11
3.1.2 La fidélité d'un test.....	12
3.1.3 La sensibilité et spécificité d'un test .....	12
3.2 Effets test retest.....	13
3.2.1 Définition.....	13

3.2.2	Données sur les effets retest .....	13
3.2.3	Effets retest et tests de lecture .....	14
<b>PARTIE PRATIQUE.....</b>		<b>16</b>
<b>1.</b>	<b>Problématique et objectifs .....</b>	<b>16</b>
<b>2.</b>	<b>Matériel et méthodes .....</b>	<b>17</b>
2.1.	Participants .....	17
2.1.1.	Critères d'inclusion .....	17
2.1.2.	Critères d'exclusion .....	17
2.2.	Matériel.....	17
2.2.1.	Les épreuves de l'étude .....	17
2.2.2.	Prétests.....	19
2.3.	Procédure .....	19
<b>3.</b>	<b>Stratégie d'analyse des données .....</b>	<b>20</b>
<b>4.</b>	<b>Résultats .....</b>	<b>20</b>
4.1.	Analyse des effets retest à Evalec et Eval2m .....	20
4.1.1.	Evalec .....	20
4.1.2.	Eval2m.....	21
4.1.3.	Comparaison qualitative entre les gains à l'épreuve de lecture de mots (Evalec-Primaire) et l'épreuve de lecture Eval2m (Evaléo 6-15).....	22
4.1.4.	Répartition des enfants en pourcentage d'évolution aux deux épreuves de lecture de mots	22
<b>DISCUSSION.....</b>		<b>24</b>
<b>CONCLUSIONS ET PERSPECTIVES.....</b>		<b>29</b>
<b>BIBLIOGRAPHIE .....</b>		<b>30</b>
<b>ANNEXES.....</b>		<b>38</b>

# INTRODUCTION

La validation des protocoles de soin des troubles du langage écrit nécessite des évaluations répétées, afin de mesurer l'évolution des performances du patient. Les propriétés métriques des tests utilisés ne mentionnent pas systématiquement les valeurs des effets retest. Même si l'aspect test retest est contrôlé, les délais entre ces passations répétées ne sont que rarement indiqués. Cette problématique se retrouve dans les batteries Evalec-Primaire et Evaléo 6-15. Cela peut induire des biais dans l'interprétation des résultats : l'évaluateur ne peut alors faire la différence entre ce qui relève d'un réel progrès dû à l'entraînement et ce qui n'est en fait qu'un simple effet retest. Les études portant spécifiquement sur les effets retest sont rares et celles concernant les effets retest à des tests de lecture le sont encore plus.

L'objectif primaire de cette étude est de mesurer l'effet retest à deux tâches de lecture évaluant l'identification de mots écrits, issues des batteries Evalec-Primaire et Evaléo 6-15, chez des normolecteurs de CM2. Le premier objectif secondaire est de comparer la sensibilité des différentes tâches de lecture aux effets retest, notre hypothèse étant que l'effet retest est similaire pour les deux types d'épreuves. Le deuxième objectif secondaire est de comparer les effets retest entre les tâches de lecture de mots sémantisés et les tâches de lecture de pseudomots de la batterie Evalec-Primaire. Des effets retest plus prononcés sur les mots sémantisés sont attendus. Le troisième objectif secondaire est d'évaluer les éventuelles différences inter-individuelles de ces effets retest. Afin de vérifier ces hypothèses, nous effectuons une première passation de ces deux tests auprès d'enfants normolecteurs scolarisés en classe de CM2, puis une deuxième passation de ces mêmes tests deux mois plus tard.

La partie théorique abordera l'apprentissage de la lecture et son évaluation, la description des propriétés métriques d'un test pour finalement s'intéresser aux données de la littérature concernant les effets retest. La partie pratique présentera le design de l'étude, les différents tests de lecture utilisés et les analyses statistiques. Une analyse quantitative regroupera dans un tableau l'ensemble des scores obtenus (précision et vitesse) pour chacune des deux épreuves. Une analyse qualitative présentera les pourcentages de gain entre les deux passations. Dans ces deux analyses, les scores obtenus pour chaque tâche de lecture de la batterie d'Evalec-Primaire seront analysés individuellement. Puis une répartition des enfants en fonction de leur évolution sera proposée afin de marquer les différences inter-individuelles. Enfin, plusieurs hypothèses concernant l'explication de la présence ou non d'effets retest seront émises.

# **PARTIE THEORIQUE**

## **1. La lecture**

### 1.1 Devenir lecteur expert

#### 1.1.1 La lecture : définition

La lecture est une tâche complexe, dont la finalité est de comprendre ce qui est lu. Lire et comprendre nécessitent l'intégration d'informations orthographiques, phonologiques, visuelles et sémantiques (Ziegler et al., 2019). Selon le modèle simple de lecture décrit par Gough et Tunmer (1986), lire dépend de deux variables : l'identification de mots écrits, spécifique à la lecture, et la compréhension. Ce modèle permet d'identifier 4 groupes de lecteurs : les hyperlexiques, les bons lecteurs, les lecteurs en difficulté et les dyslexiques. La compréhension écrite est sous-tendue par le produit de deux variables essentielles : la reconnaissance de mots écrits et la compréhension orale (Castles et al., 2018). La maîtrise de ces deux compétences permet de consacrer les ressources cognitives à la compréhension écrite (Bentolila & Germain, 2019).

#### 1.1.2 Apprendre à lire

Lors de l'apprentissage de la lecture, deux procédures se développent. Ces procédures correspondent à des mécanismes d'identification des mots écrits qui permettent de décoder les mots par une conversion graphophonologique (procédure sublexicale), autrement dit l'association d'une lettre (graphème) à un son (phonème) ; et de reconnaître globalement un mot (procédure lexicale) par un accès à la mémoire orthographique. L'apprentissage de ces procédures est le fruit d'un enseignement explicite, permis par la scolarisation (Sprenger-Charolles & Ziegler, 2019). Le traitement des mots est d'abord sériel, avec un décodage lent et laborieux afin de mettre en place la conversion graphophonologique. Lors de la classe préparatoire, l'enfant doit apprendre un nombre suffisant de règles de correspondances graphèmes phonèmes ainsi que leur systématisme (Ziegler, 2018). En appliquant ces règles, et en s'appuyant sur son lexique phonologique (association d'un référent à sa forme sonore), l'enfant est capable de décoder de nouveaux mots présents dans son lexique mental (Ziegler, 2019). Ziegler et collaborateurs (2014) considèrent le décodage comme un mécanisme d'auto-apprentissage, sans l'aide d'un tiers, qui permettra d'alimenter la mémoire orthographique. La représentation orthographique est ensuite renforcée par la lecture répétée du même mot, et permettra finalement une reconnaissance globale des mots écrits (Sprenger-Charolles &

Ziegler, 2019). L'identification devient alors une tâche de moins en moins complexe et moins coûteuse en charge cognitive, signature de l'automatisation de la lecture.

Sprenger-Charolles et collaborateurs (2003) et Ziegler (2018) montrent que des compétences précoces en décodage prédisent le niveau de la procédure orthographique. D'autres études longitudinales, dont celle de Ziegler et Goswami (2005) valident cette évolution, et précisent l'importance de la régularité et de la fréquence dans la construction des procédures de lecture.

Ainsi, au début de l'apprentissage, les mots réguliers fréquents ne sont pas mieux lus que les pseudomots. Puis, les mots réguliers fréquents seront mieux lus que les pseudomots. Enfin, l'enfant tout venant accède plus tardivement à tous les mots réguliers et irréguliers peu fréquents (Sprenger-Charolles et al., 2018). Un lecteur expert, à partir du CE2, n'est plus sensible aux effets de fréquence, de consistance et de longueur. Cependant, la régularité des mots est dépendante du niveau de transparence de la langue écrite (Ziegler & Goswami, 2005). Ainsi, la mise en place des deux procédures lexicale et sublexicale de lecture varie d'une langue à l'autre.

### 1.1.3 Les spécificités du français

La maîtrise du décodage dépend de la consistance de la langue écrite, c'est-à-dire de la régularité des correspondances graphème/phonème de la langue (Ziegler, 2018). Les graphèmes transcrivent des phonèmes, mais ne sont pas vecteurs de sens. Or, le système du français est relativement régulier dans le sens de la lecture mais pas dans le sens de l'écriture. Ziegler et collaborateurs (2018) donnent l'exemple du phonème /o/ qui admet plusieurs graphèmes (au, o, eau, ot, op). De plus, les lettres de l'alphabet français peuvent correspondre à un phonème (o) ou s'associer pour former de nouveaux phonèmes (on, ou, oi). Ainsi, le français se compose de 37 phonèmes pour plus de 130 graphèmes, ce qui induit un apprentissage plus long de l'ensemble des correspondances graphèmes/phonèmes.

Plusieurs études montrent les différences inter-linguistiques dans l'apprentissage de la lecture. Parmi ces études, Goswami et collaborateurs (1998) ont étudié les capacités de décodage de pseudomots dans plusieurs langues. Le niveau de développement de la lecture est dépendant du niveau de transparence des relations graphème/phonème. De ce fait, le niveau de transparence d'un système d'écriture régulier permet aux apprenants d'avoir une expertise en décodage plus rapide que les apprenants français, où le niveau d'inconsistance en lecture est plus marqué. Le degré d'irrégularité de la langue prédit donc le niveau de lecture en fin de

CP et détermine la facilité avec laquelle le mécanisme de décodage se met en place (Ziegler, 2018).

La consistance d'une langue est donc un facteur indéniable dans l'apprentissage de la lecture, mais elle ne permet pas à elle seule d'expliquer les contraintes d'apprentissage de l'identification des mots écrits. Des compétences cognitives sous-jacentes sont également essentielles et prédisent le futur niveau de lecture, et cela quelle que soit la langue.

## 1.2 Les compétences cognitives sous-jacentes

### 1.2.1 Compétences phonologiques

La conscience phonologique et les procédures d'identification de mots écrits entretiennent des relations d'influence réciproques. Des études montrent que la conscience phonologique en maternelle est prédictive de la capacité à lire (Anthony & Francis, 2005; Hulme et al., 2012; Melby-Lervåg et al., 2012). La conscience phonologique, qui est la capacité à manipuler les unités phonologiques, n'est pas une compétence homogène, mais un ensemble de compétences. Morais (1991) la définit selon quatre niveaux de compétences : la conscience de chaînes phonologiques, la conscience des rimes de mots, la conscience syllabique et la conscience phonémique.

Selon Mann et Liberman (1985), les compétences en conscience syllabique (capacité à compter et manipuler les syllabes) en fin de maternelle prédisent le décodage ultérieur. Néanmoins, le prédicteur le plus puissant est la conscience phonémique. Des études longitudinales (Perfetti et al., 1987; Piquard & Sprenger-Charolles, 2013; Sprenger-Charolles et al., 1998, 2003) montrent que la capacité de segmentation phonémique est un prédicteur du futur niveau de lecture. À l'inverse, l'apprentissage de la lecture favorise le développement de la segmentation phonémique. Les études récentes s'accordent donc pour dire que la conscience phonémique est une des composantes de l'apprentissage de la lecture, plutôt qu'un pré-requis (Bentolila & Germain, 2019). Cependant, en accord avec le développement de la procédure orthographique, le poids de la capacité de segmentation phonémique sur la lecture de mots diminue avec le niveau scolaire.

La mémoire à court terme est la capacité à maintenir temporairement des unités phonologiques en mémoire (Christopher et al., 2012; Locascio et al., 2010). Cette mémoire transitoire phonologique joue un rôle prédictif dans les capacités à développer la procédure de conversion graphophonologique et lors des tâches de conscience phonologique (Engel de

Abreu et al., 2011). Les liens causaux d'un déficit phonologique avec un trouble du langage écrit constitue, à ce jour, une des hypothèses les plus robustes (Saksida et al., 2016).

Enfin, la vitesse d'accès lexical prédit fortement le futur niveau de lecture (Casalis et al., 2019). Cette compétence associée aux compétences de segmentation phonémique et de mémoire phonologique peut expliquer les troubles d'identification des mots écrits : Wolf et collaborateurs (2000) parlent de l'hypothèse d'un double déficit.

### 1.2.2 Compétences visuelles et visuo-attentionnelles

La lecture engage avant tout la modalité visuelle. La capacité à lire dépend entre autres de la qualité de l'analyse des traits constitutifs des lettres, des traitements séquentiels, de l'oculomotricité ou encore de la taille de l'empan visuo-attentionnel. (Casalis et al., 2019).

D'un point de vue anatomo-fonctionnel, les systèmes parvocellulaires et magnocellulaires constituent deux voies de transmission des informations de la rétine au cortex visuel. Le système parvocellulaire permet la reconnaissance des formes, des détails et des couleurs en vision centrale.

Le système magnocellulaire permet la reconnaissance des informations visuelles rapides (clignotement, mouvement) en vision périphérique. Ce système joue un rôle dans la détection du mot dans sa globalité et dans les mouvements oculaires (Vahine, 2017).

En lecture, l'identification des lettres et le codage de leur position dépendent de facteurs visuels (Pammer & Vidyasagar, 2005; Sprenger-Charolles & Ziegler, 2019; Stein, 2001). L'empan visuo-attentionnel qui correspond au nombre de lettres que l'enfant peut traiter simultanément dans la séquence d'un mot, doit être suffisant (Saksida et al., 2016; Valdois et al., 2013; Zoubinetzky et al., 2016). De même, lors de l'identification des mots écrits, un équilibre entre analyse globale et locale d'une scène visuelle complexe est nécessaire (Bedoin, 2017). Ces compétences visuelles et visuo-attentionnelles recourent donc un ensemble de compétences à la fois perceptives et attentionnelles qui jouent un rôle tant dans le développement de la procédure lexicale que sublexicale.

## **2. L'évaluation de la lecture**

### **2.1 Les modèles de lecture**

#### **2.1.1 Les modèles connexionnistes à double voie**

Les modèles de lecture actuels sont mathématiques, et implantés sur ordinateur. Ces modèles computationnels rendent compte des compétences du lecteur par simulation de son comportement de lecture (Phenix et al., 2016; Ziegler et al., 2019).

Le modèle à double voie ou DRC est le modèle de référence (Coltheart et al., 2001). Lorsqu'un mot écrit est présenté, le lecteur analyse les traits visuels (non spécifique à la lecture) et reconnaît ensuite les lettres. Avant d'accéder à la forme sonore du mot et de le prononcer, deux voies peuvent être empruntées. La voie phonologique ou sublexicale consiste à identifier les graphèmes et de les associer aux phonèmes correspondants. Coûteuse en ressources cognitives, elle est impliquée dans la lecture de mots nouveaux et de pseudomots. La voie orthographique ou lexicale permet l'analyse de la forme orthographique globale du mot écrit, qui permet l'activation des représentations orthographique et sémantique en mémoire. Elle est impliquée dans la lecture de mots réguliers connus et de mots irréguliers.

D'après ce modèle, ces deux voies agiraient en parallèle.

Même si ces modèles sont établis à partir d'une population d'adultes cérébro-lésés et ne simulent pas l'apprentissage de la lecture, ils peuvent néanmoins montrer ou simuler des données neuro-développementales (Valdois et al., 2007; Ziegler, 2019).

Le modèle CDP+ (Connexionist Dual Process) (Perry et al., 2007) complète le modèle DRC en simulant les effets de consistance et en mettant en avant les mécanismes d'auto-apprentissage (Phenix et al., 2016). Ce modèle tient compte d'une nouvelle composante au sein de la voie sublexicale : le buffer graphémique, qui code les graphèmes en fonction de leur position syllabique. Les conversions graphophonologiques se font par le biais d'un réseau d'assemblage, le TLA (two-layer assembly network) et sont renforcées à chaque fois que le lecteur les rencontre. Le mot lu contient l'ensemble des phonèmes les plus activés sur chaque position, les mots fréquents sont donc lus plus rapidement. A cela, le modèle CDP++ ajoute des connexions bidirectionnelles entre les unités du lexique phonologique, et permet ainsi de simuler la lecture d'items multi-syllabiques en français (Perry et al., 2014).

#### **2.1.2 Les modèles connexionnistes en triangle**

Contrairement aux modèles à double voie, les modèles en triangle rendent compte des liens fonctionnels qui existent entre les compétences de lecture (Casalis et al., 2018).

Initialement décrit par Seidenberg et McClelland (1989), le modèle en triangle ou modèle PDP (Parallel Distributed Processing) propose un traitement interactif avec activations réciproques entre les unités orthographiques, sémantiques et phonologiques. L'activation distribuée de ces différentes unités au sein du réseau conduit à la reconnaissance des mots écrits (Harm & Seidenberg, 2003). Le poids des connexions entre les différentes unités varie en fonction de l'exposition aux mots : plus le lecteur débutant est exposé au mot, plus les connexions sont renforcées (Phenix et al., 2016). Ainsi, ce réseau s'auto-structure avec la confrontation à l'écrit. Chaque mot se compose de plusieurs unités orthographiques et phonologiques (triplet de lettres et triplet de phonèmes), et chacune de ces unités est représentée dans plusieurs mots. Ce modèle met également en évidence l'interaction entre la fréquence et la régularité des mots (Ziegler et al., 2019). La voie phonologique traite les mots consistants ou fréquents alors que la voie sémantique traite les mots peu fréquents et peu consistants. (Harm & Seidenberg, 2003)

### 2.1.3 Le modèle multi-traces

Le modèle multi-traces (Ans et al., 1998) postule l'existence d'une fenêtre attentionnelle dont la taille varie en fonction de la longueur du mot. A travers cette fenêtre attentionnelle, le lecteur extrait l'identité des lettres. Une grande fenêtre attentionnelle suggère que le lecteur lit la totalité du mot alors qu'une fenêtre attentionnelle plus étroite suggère que le lecteur segmente le mot en unités plus petites (graphème ou syllabe). Ce modèle prend en compte l'hypothèse de la qualité de l'empan visuo-attentionnel dans l'apprentissage et le développement de la lecture en fonction du niveau de consistance dans les écritures alphabétiques. Cela renvoie à la théorie de la taille du grain (Ziegler & Goswami, 2005).

## 2.2 Tests évaluant la lecture

### 2.2.1 Les épreuves d'identification de mots isolés

Les différents modèles présentés servent de base pour construire les tests de lecture. Les tests évaluant l'identification de mots écrits sont construits en référence au modèle à double voie.

Les tests d'identification de mots écrits se déclinent généralement en 3 parties : la lecture de mots réguliers (e.g. minute), irréguliers (e.g. seconde) et de pseudomots (e.g. aspouche). Selon les modèles de lecture présentés précédemment, la lecture de mots réguliers inconnus ou de pseudomots respectant les règles phonotactiques permet d'objectiver le

fonctionnement de la procédure graphophonologique ou voie sublexicale ; la lecture de mots irréguliers ou de mots réguliers connus par le lecteur permet d'objectiver l'accès à la mémoire orthographique ou voie lexicale (Coltheart et al., 2001). Outre les effets de régularités et de fréquence, ces tests peuvent également prendre en compte l'effet de longueur du mot (mots courts versus mots longs).

### 2.2.2 Les épreuves de leximétrie et de compréhension écrite.

Les épreuves de leximétrie mesurent le niveau d'automatisme des procédures d'identification en proposant une lecture de textes ou de mots dans un temps limité (généralement entre 1 et 3 minutes : principe du « one minute test »). Les textes choisis peuvent être avec ou sans signification. L'association de ces deux types de tests est recommandée, la lecture de texte avec signification permettant d'évaluer le degré d'utilisation des indices contextuels. Ces épreuves prennent en compte à la fois un indice de précision et un indice de vitesse d'identification (Casalis et al., 2018).

Les épreuves de compréhension écrite consistent généralement à répondre à des questions (orales ou écrites) ou rappeler un texte après l'avoir lu. Ces épreuves peuvent aussi se présenter sous la forme de phrases écrites à compléter où le lecteur doit choisir un mot parmi plusieurs possibilités (closure de phrases), ou encore associer une phrase écrite à une image (Casalis et al., 2018).

### 2.2.3 Les épreuves évaluant les compétences cognitives sous-jacentes

Ces épreuves ont pour objectif de mesurer les déficits causaux d'un trouble d'identification. L'hypothèse d'un déficit phonologique, considérée comme la plus robuste, est mesurée le plus généralement par des épreuves de conscience phonologique et phonémique telles que la suppression d'une syllabe (e.g. varéla → réla) ou du premier phonème d'un pseudomot (e.g. zak → ak), par le jugement de rimes (e.g. oiseau-piano), ou par la recherche d'intrus sur le phonème initial (sale, ver, sire, scie). L'épreuve de répétition de pseudomots permet d'évaluer les compétences de la mémoire phonologique à court terme (Casalis et al., 2019). Cette évaluation de la mémoire phonologique est complétée par la répétition de chiffres à l'endroit qui évalue l'efficacité de la mémoire verbale à court terme, et la répétition de chiffres à l'envers qui apporte des indices sur l'efficacité de la mémoire de travail. Puis, la vitesse d'accès lexical est évaluée par une tâche de dénomination rapide d'images d'objets, de couleurs ou de chiffres.

L'hypothèse d'un déficit visuo-attentionnel est en lien avec une altération de la fenêtre attentionnelle. Son évaluation consiste à proposer des tâches de report partiel et global. (Casalis et al., 2019). Les déficits visuo-attentionnels peuvent également s'expliquer par un déséquilibre entre les modes d'analyse locale et globale d'une scène visuelle complexe (Bedoin et al., 2010).

En conclusion, l'ensemble de ces tests sont toutefois imparfaits : tous ont leurs limites respectives qui doivent être clairement définies. Prendre connaissance des propriétés métriques des tests avant de les utiliser est essentiel. Cela permet au praticien d'interpréter correctement les résultats obtenus. Ces indices sont généralement décrits dans les manuels des tests.

### **3. Propriétés métriques des tests et leurs limites**

#### **3.1 Les propriétés métriques d'un test : validité, fidélité, sensibilité et spécificité**

##### **3.1.1 La validité d'un test**

Un test est dit valide lorsqu'il regroupe un certain nombre de preuves. L'interprétation des résultats doit se baser sur des données théoriques et empiriques cliniques telles que des arguments rationnels, des corrélations avec d'autres mesures, des courbes de développement, ou des comparaisons de groupes cliniques (Grégoire, 2009). Une épreuve n'étant jamais pure, elle ne mesure jamais parfaitement ce qu'elle prétend mesurer. Les résultats obtenus doivent donc toujours être nuancés, en tenant compte de trois types de variance, décrits dans les manuels des tests. Tout d'abord la mesure des caractéristiques de plusieurs tests (dont le test étudié) détermine la variance partagée. Puis, les caractéristiques du test lui-même définissent la variance spécifique. Enfin, la variance de l'erreur correspond aux erreurs de mesure évoquées précédemment.

La validité est également déclinée en trois parties (Fermanian, 2005). La première, la validité d'apparence, est l'appréciation subjective de la validité par le testeur. Pour cela, le testeur analyse de manière globale le test et juge la valeur apparente du test. La seconde, la validité de contenu correspond à la validité des différents items du test. Puis la troisième, la validité critériée, est une évaluation faite par rapport à des critères extérieurs au test étudié. Pour cela, il faut se référer à un test de référence (« gold standard ») déjà validé et aux caractéristiques similaires, (Grégoire, 2009).

### 3.1.2 La fidélité d'un test

La fidélité d'un test correspond à sa précision, sa reproductibilité. Si un test est fidèle, alors à chaque passation, les résultats de l'individu à ce même test devraient toujours être identiques. Fermanian (2005) parle de fidélité intra-juges. La fidélité inter-juges correspond quant à elle à l'obtention de scores similaires par deux testeurs cotant le test en même temps.

Or, obtenir exactement les mêmes scores lors de passations répétées est impossible car un score ne représente jamais le niveau réel de compétences d'un individu (Grégoire, 2009). Des erreurs de mesure liées au participant (fatigue, état psychologique, distractions), liées à l'examineur (non-respect des consignes, cotations) ou à l'environnement (luminosité, bruit) sont inévitables. Ainsi, d'une passation à l'autre, les scores sont fluctuants et peuvent favoriser ou au contraire défavoriser l'individu.

Considérer cette erreur de mesure renvoie à déterminer un intervalle de confiance autour du score obtenu : le score réel de l'individu a de fortes chances de se trouver dans cet intervalle. Lors de passations répétées, pour qu'un test soit fidèle, les scores peuvent être différents mais doivent toujours tomber dans cet intervalle de confiance (Grégoire, 2009). Un test a une réelle valeur pratique lorsque sa fidélité est supérieure ou égale à 0.80 (Bartels et al., 2010).

### 3.1.3 La sensibilité et spécificité d'un test

Si un test doit être précis, il doit également permettre d'obtenir des scores différenciés. La sensibilité et la spécificité expriment la capacité informative du test, autrement dit la capacité du test à discriminer dans une population, les individus porteurs d'une maladie évaluée par le test. L'expression du trouble est évaluée en termes de probabilité et par un pourcentage d'individus sur un échantillon. Ceci nécessite un intervalle de confiance qui est généralement de 95% ; ce qui revient à dire qu'il est attendu qu'au moins 5% des individus d'un échantillon représentatif soient porteurs de cette maladie.

La sensibilité est déterminée sur une population de patients dont le statut « malade » est connu grâce à un test de référence (gold standard) et permet de catégoriser les vrais positifs. La spécificité est déterminée sur une population de patients dont le statut « non malade » est connu en référence au même test de référence que celui utilisé dans la population porteuse de la maladie étudiée, permettant ainsi de catégoriser les faux négatifs.

Le test idéal se caractérise par une sensibilité et une spécificité égale à 1. Aucun test n'étant parfait, cela est impossible. Un compromis entre sensibilité et spécificité doit être

trouvé en déterminant la valeur seuil du test. Cette valeur permet de distinguer précisément les groupes d'individus, tout en limitant les faux positifs et les faux négatifs. (Bertrand et al., 2010). Autrement dit, ces deux indices mesurent la puissance discriminante d'un test (Grégoire, 2009). Dans le cas des textes de lecture, la valeur seuil permet de discriminer les mauvais lecteurs des bons lecteurs, tout en limitant les faux négatifs et les faux positifs.

### 3.2 Effets test retest

#### 3.2.1 Définition

Bartels et al (2010) définissent l'effet retest comme une amélioration des scores à un test donné entre deux passations de ce même test, en l'absence d'intervention entre les deux passations.

L'effet retest participe à l'évaluation de la fidélité d'un test. Or, cet effet est rarement évalué et décrit dans les tests (Grégoire, 2009).

#### 3.2.2 Données sur les effets retest

De nombreux facteurs sont mis en avant pour expliquer cet effet retest : une familiarisation avec l'environnement du test (connaissance du déroulé du test, de l'évaluateur) reconnaissance des items du test ou encore apprentissage procédural (Bartels et al., 2010; Hausknecht et al., 2007). Cet effet retest peut également dépendre des propriétés du test comme sa complexité ou ses modalités de présentation (Benedict & Zgaljardic, 1998 ; Beglinger et al., 2005), ou des caractéristiques individuelles comme le QI, l'âge, la personnalité, la motivation, l'humeur (Bartels et al., 2010 ; Reeve & Lam, 2007) qui peuvent constituer des facteurs d'une meilleure adaptation à un test.

La méta-analyse de Scharfen et collaborateurs (2018) montre des effets retest significatifs sur des tests évaluant des capacités cognitives. Ainsi, les effets retest sont plus marqués sur des tâches complexes (comme celles incluant la mémoire de travail) qui sont plus anxiogènes au départ et qui nécessitent la mise en place de stratégies. Les temps de réponse sont généralement plus longs lors de la première passation, puis plus rapides lors des suivantes du fait de la mise en place de stratégies de réponses aux items du test. Ces stratégies s'améliorent au fil des passations d'autant que les tests recrutent des ressources cognitives ou attentionnelles. Les performances à ce type d'épreuves peuvent donc se stabiliser plus tardivement et être plus sensibles à l'effet retest. Les tests évaluant les compétences cognitives qui incluent des processus cognitifs généraux, nécessitent un apprentissage plus

long. Ces tests génèrent moins d'effets retest que les tests de connaissance (culture générale) déjà reliés à des stratégies d'apprentissage. Lievens et collaborateurs (2005) soulignent que les tests de connaissance sont plus ciblés, et induisent donc plus facilement un apprentissage. Par ailleurs, l'étude de Dough et collaborateurs (2010) révèle qu'un outil d'évaluation utilisé à plusieurs reprises, devient de fait un outil d'apprentissage. L'apprentissage est compris ici par la répétition de la tâche qui renforce la connaissance de l'outil.

L'intervalle de temps entre les passations est aussi à prendre en compte (Bartels et al., 2010 ; Donovan & Radosevich, 1999). La méta-analyse de Scharfen et collaborateurs (2018) relève un effet retest plus marqué lors de la seconde passation d'un test que lors de la troisième. Néanmoins, cet effet semble dépendant de l'intervalle entre chaque passation et de la nature de la tâche. Bartels et collaborateurs (2010) ont étudié les effets retest d'une batterie de tests neuropsychologiques proposés à sept reprises à des adultes sains. Ces passations étaient plus ou moins espacées sur une période d'un an. Les effets retest sont les plus significatifs dans la phase initiale, lorsque les tests sont passés sur des intervalles de temps courts (3 mois). Après trois mois et sur des intervalles de temps plus longs, les auteurs observent une stabilisation des performances.

Ainsi, les principaux facteurs qui peuvent expliquer la présence d'effets retest sont les caractéristiques propres de ce que mesure le test et sa construction (choix des items et modalité de passation) ainsi que l'intervalle de temps entre les passations.

### 3.2.3 Effets retest et tests de lecture

Dans la littérature, les effets retest ont été principalement étudiés dans le cadre d'outils d'évaluation psychologique de compétences cognitives telles que l'attention, la mémoire, les fonctions exécutives, les épreuves de fluence ou encore les fonctions visuo-spatiales (Bartels et al., 2010; Scharfen et al., 2018). Les études sur les effets retest d'outils d'évaluation de la lecture sont réduites.

L'étude de Reynolds et collaborateurs (2003) a étudié l'effet d'un programme (le DDAT : Dyslexia Dyspraxia attention-deficit treatment) chez les enfants dyslexiques. Cette étude démontre des gains significatifs en lecture de pseudomots et de segmentation phonémique chez le groupe contrôle, à qui aucun entraînement n'a été proposé.

Ces données plaident pour des effets retest dont les pourcentages de gains varient d'une épreuve à l'autre. En effet, selon la revue de littérature de McArthur (2007), les tests les plus susceptibles de générer des effets retest entre deux passations sont ceux qui testent : la

compréhension écrite (100% de gains), la lecture de mots (50%), l'écriture (50%), la conversion graphophonologique (33%), la conscience phonémique (33%), le langage expressif et réceptif (25%), la mémoire phonologique (17%), la dénomination rapide (17%) et l'épellation (15%). Ces gains sont d'autant plus importants que les délais entre deux passations sont courts (Bartels et al., 2010).

# PARTIE PRATIQUE

## 1. Problématique et objectifs

Les compétences en langage écrit d'un patient et leur évolution sont fréquemment évaluées par la passation des mêmes tests en pré et post-intervention. Ces passations, parfois espacées d'un intervalle de temps assez court, peuvent induire un effet retest, qui n'est pas systématiquement mesuré et décrit dans les propriétés métriques du test. L'absence de prise en compte de l'effet retest peut biaiser l'analyse des résultats, en relevant des gains ou une normalisation des performances à ces tests, alors que l'évolution des performances serait imputable à cet effet retest.

L'objectif primaire de notre étude exploratoire est de mesurer un possible effet retest chez des enfants normolecteurs de CM2, à deux épreuves d'identification de mots écrits très utilisées dans l'évaluation clinique : l'épreuve de lecture d'Evalec-Primaire (Sprenger-Charolles et al., 2010) et l'épreuve de lecture Eval2m de la batterie Evaléo 6-15 (Launay et al., 2018). Les critères de mesure sont les scores bruts obtenus à la première passation des tests (T0) comparés à ceux de la deuxième passation (T1) à deux mois d'intervalle. Les scores de précision et de vitesse de lecture sont retenus pour l'épreuve de lecture d'Evalec-Primaire ; les scores de mots lus (ML) et de mots correctement lus (MCL) pour Eval2m.

Le premier objectif secondaire est de comparer les effets retest mesurés pour ces deux tâches de lecture. Les critères de mesure sont les mêmes que pour l'objectif primaire.

Le deuxième objectif secondaire est de comparer les effets retest mesurés pour les tâches de lecture de mots sémantisés (mots réguliers et irréguliers- MIMR) aux effets retest mesurés pour les tâches de lecture de pseudomots (PM) de la batterie Evalec-Primaire. Les critères de mesure sont les scores bruts obtenus en précision et le temps de lecture moyen par item pour chaque catégorie de mots étudiés (MIMR et PM).

Un troisième objectif secondaire vise à déterminer s'il existe des différences inter-individuelles concernant la présence ou non de ces effets retest. Les critères de mesure sont donc les pourcentages d'évolution entre T0 et T1 aux épreuves de lecture à haute voix d'Evalec-Primaire et à Eval2m pour chaque enfant.

## **2. Matériel et méthodes**

### **2.1. Participants**

#### **2.1.1. Critères d'inclusion**

Dans le cadre de l'étude, 45 enfants ont été inclus, dont 22 filles et 23 garçons. Ces enfants sont scolarisés en classe de CM2 dans deux écoles primaires du secteur privé.

#### **2.1.2. Critères d'exclusion**

L'étude portant sur des enfants normolecteurs, la passation de prétests a conduit à l'exclusion de 3 enfants (2 filles et 1 garçon). Ces enfants ont obtenu au moins un score pathologique à une des deux épreuves de leximétrie (Alouette-R ou Le Pingouin) et/ou à l'épreuve de compréhension écrite (L3 de Orlec-D) ET au moins un score pathologique à une des épreuves évaluant les compétences cognitives sous-jacentes à la lecture (compétences phonologiques et visuo-attentionnelles). L'ensemble de ces épreuves seront décrites dans la section « Matériel ».

De plus, la fermeture d'une classe liée à la COVID-19 a conduit à l'interruption d'une passation et donc à l'exclusion d'un autre garçon. Finalement, l'étude porte sur 41 enfants (20 filles, 21 garçons). Tous ces enfants sont normolecteurs, en accord avec les critères d'inclusion du DSM 5 (version française : Crocq et al., 2015) et d'exclusion.

### **2.2. Matériel**

#### **2.2.1. Les épreuves de l'étude**

Deux tests d'identification de mots écrits ont été utilisés dans notre étude afin de mesurer les effets retest.

- L'épreuve « Lecture à voix haute »

Cette épreuve issue de la batterie informatisée Evalec-Primaire (Sprenger-Charolles et al., 2010) est construite en référence au modèle à double voie, permettant ainsi de mesurer l'efficacité des deux voies de lecture. Les données relevées sont la précision et la vitesse d'identification, les effets de régularité, de lexicalité et de fréquence. Cette épreuve est, selon les auteurs, spécifiquement destinée à diagnostiquer un trouble d'identification des mots écrits de type dyslexie dès la fin du CE1. Elle se compose de 4 listes de mots :

- La première liste contient 48 mots réguliers et irréguliers fréquents. Ces mots sélectionnés ont trois niveaux de régularité : 24 mots totalement réguliers avec une

lettre par phonème (12 mots, e.g. pilote), ou un digraphe (12 mots e.g., soupe), 12 mots partiellement réguliers comprenant des graphies contextuelles comme « c » ou « g » (e.g fragile) et 12 mots irréguliers (e.g sept). Les items de chaque sous-liste sont appariés en longueur et en fréquence.

- La seconde liste contient 36 pseudomots appariés aux 36 mots réguliers (e.g. pitode apparié à pilote) cités ci-dessus.
- La troisième et la quatrième liste se composent respectivement de 20 mots irréguliers et de 20 pseudomots. Chacune de ces deux listes contient 10 items courts (e.g. « dix » pour les mots irréguliers ; « dul » pour les pseudomots) et 10 items longs (« sculpture » pour les mots irréguliers ; « scaltoure » pour les pseudomots). Ces items ont tous été choisis selon les critères évoqués précédemment.

Une fois la consigne lue, le participant lit les mots qui s'affichent un à un au centre de l'écran. Pour chaque liste, les deux ou trois premiers mots sont des essais. Dès que l'enfant a lu le mot, l'examineur appuie sur une touche du clavier pour passer au mot suivant. Cette épreuve prend en compte les réponses correctes et leur temps de latence en milliseconde (délai entre le début d'affichage du mot et le début de la réponse vocale de l'individu, mesuré par détection vocale). Ce temps de latence, enregistré uniquement pour les mots lus correctement est calculé automatiquement par le logiciel. La prise en compte de ces deux indices (précision et vitesse) renseigne l'examineur sur le degré d'automatisation des procédures d'identification des mots écrits. Les productions sont enregistrées dans le logiciel et les productions erronées de l'enfant sont reportées à l'écrit, dans le cahier de passation. Le temps de passation de cette épreuve est d'environ 10 minutes.

- L'épreuve « Eval2m » (lecture de mots en 2 minutes)

Cette épreuve, issue de la batterie Evaléo 6-15 (Launay et al., 2018) permet d'évaluer la vitesse d'identification de mots écrits. Elle se compose de 263 mots réguliers et irréguliers de complexité croissante présentés en colonne. Les premiers mots sont classés du plus court au plus long, et pour chaque longueur de mot, ils sont classés du plus fréquent au moins fréquent. Le participant doit lire le plus de mots possible à haute voix en deux minutes. L'évaluateur coche les mots oubliés ou mal lus dans le logiciel et note les productions erronées de l'enfant sur le cahier de passation. A la fin de ces deux minutes, trois indices sont relevés : un indice de vitesse (le nombre de mots lus), un indice de précision (le taux de mots correctement lus) et un indice de vitesse et précision confondues (nombre de mots correctement lus, mesurant l'efficacité de l'identification du mot écrit).

### 2.2.2. Prétests

La passation de ces épreuves de lecture de mots isolés s'accompagne de la passation de prétests, visant à mettre en évidence les enfants dyslexiques dans le but de les exclure de l'étude. Ces épreuves comprenaient des épreuves de lecture (Alouette-R (Lefavrais, 1967), L3 d'ORLEC de Lobrot (2004), le Pingouin (Launay et al., 2018)) ainsi que des épreuves de compétences phonologiques (Répétition de pseudomots et analyse Phonologique (Sprenger-Charolles et al., 2010), répétition de chiffres endroit/ envers (Launay et al., 2018)) et visuo-attentionnelles (Empan visuo-attentionnel (Launay et al., 2018)).

### 2.3. Procédure

L'étude est décrite oralement aux participants et par écrit à leurs responsables légaux. Tous les enfants et leurs parents ayant participé à cette étude ont rempli un formulaire de consentement éclairé (cf. Annexe 1). Deux passations sont proposées.

Lors de la première passation (T0), les épreuves d'analyse phonologique, de répétition de pseudomots, de répétition de chiffres et d'empan visuo-attentionnel d'Evaléo sont proposés dans l'ordre énoncé. Ces épreuves sont suivies de la passation des différents tests de lecture : épreuves de lecture à haute voix d'Evalec-Primaire, Eval2m, Alouette-R, Le Pingouin et l'épreuve L3 d'ORLEC-D. La durée de cette passation est d'environ 1h.

Lors de la deuxième passation (T1), ces épreuves de lecture sont proposées dans l'ordre inverse (voir figure 1). La durée de cette deuxième passation est d'environ 30 minutes. Les passations entre T0 et T1 sont espacées de 2 mois (+/- 4 jours). Ces intervalles de temps entre les tests sont fixés en accord avec la littérature sur les effets retest. Entre chaque passation, aucune intervention ou entraînement n'est proposé.

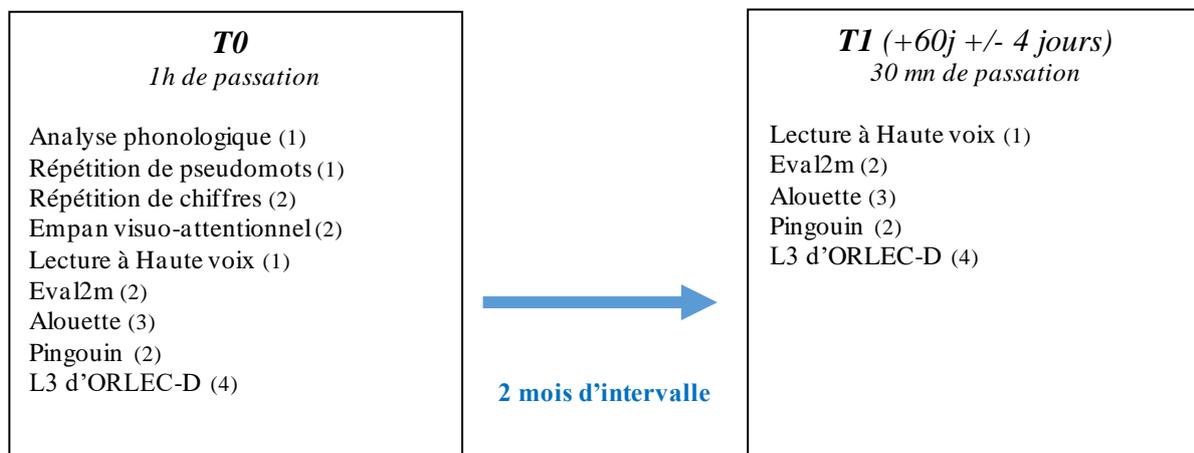


Figure 1: Description de l'ordre de passation des épreuves en T0 et T1. (1) Evalec-Primaire (2) Evaléo 6-15 (3) Lefavrais (4) Lobrot

Ces deux passations se sont déroulées au sein des écoles primaires, dans des salles individuelles. Quatre étudiantes ont participé à cette étude, mais l'évaluatrice était la même pour chaque enfant à T0 et T1.

### **3. Stratégie d'analyse des données**

Après avoir été regroupées dans un tableau Excel, les données obtenues ont fait l'objet d'un test de normalité réalisé à partir du test Kolmogorov-Smirnov. Ce test permettant de vérifier la distribution normale des scores obtenus à l'épreuve de lecture d'Evalec et d'Eval2m montre une distribution gaussienne pour tous nos résultats. Le test paramétrique T-test pour échantillon apparié issu de BiostaTGV est ainsi sélectionné pour l'analyse statistique. Ce test permet une comparaison inter-groupe (groupe T0/ groupe T1).

L'analyse de ces données est complétée par la taille de l'effet mesurée à partir de la formule d'un d de Cohen. La taille d'effet est une mesure de la force de l'effet observé d'une variable sur une autre et plus généralement d'une inférence. De plus, à titre d'analyse qualitative, les pourcentages de gains pour chaque épreuve seront calculés. Enfin, une répartition des enfants en fonction de leur pourcentage de réussite entre T0 et T1 sera proposée afin d'apprécier les différences inter-individuelles.

### **4. Résultats**

#### **4.1. Analyse des effets retest à Evalec et Eval2m**

##### **4.1.1. Evalec**

Concernant la première analyse, pour chaque enfant, les scores en précision obtenus aux 4 listes (mots sémantisés et pseudomots) sont additionnés, soit un total de 123 mots. La même méthode est appliquée pour les résultats en vitesse de lecture. Les scores en précision à T1 ( $\bar{x}_{T1} = 117.17$  ;  $ET = 3.67$ ) ne sont pas significativement supérieurs à ceux obtenus à T0 ( $\bar{x}_{T0} = 116.32$  ;  $ET = 4.50$ ) ( $t_{40} = -1.79$  ;  $p = 0.08$ ). Les scores ont augmenté de 1% entre T0 et T1 et la taille de l'effet est de 0.21.

En revanche, la vitesse de lecture obtenue à T1 ( $\bar{x}_{T1} = 5644.66$  ;  $ET = 717.88$ ) est très significativement inférieure à celle obtenue à T0 ( $\bar{x}_{T0} = 6250.32$  ;  $ET = 748.15$ ) ( $t_{40} = 7.07$  ;  $p < 0.001$ ). Le délai de réponse a ainsi diminué de 10 % entre les deux passations. La taille de l'effet est de 0.83. L'ensemble de ces résultats sont visibles dans le tableau 1.

La seconde analyse, pour chaque enfant, différencie la somme des scores en précision (soit un total de 68 mots) et la vitesse de lecture obtenus en lecture de mots sémantisés (mots

réguliers, irréguliers, mots irréguliers courts et longs) ; puis la somme des scores en précision (soit un total de 55 mots) et le temps de lecture par item obtenus en lecture de pseudomots (pseudomots, pseudomots courts et longs). Le nombre de mots sémantisés correctement lus à T1 ( $\bar{x}_{T1}= 66.54$  ;  $ET= 1.36$ ) est significativement supérieur à celui de T0 ( $\bar{x}_{T0}= 66.05$  ;  $ET= 1.60$ ) ( $t_{40}= -2.29$  ;  $p < 0.05$ ). Les scores ont augmenté de 1 %. En revanche, le nombre de pseudomots correctement lus à T1 ( $\bar{x}_{T1}= 50.63$  ;  $ET= 2.72$ ) n'est pas significativement supérieur à celui de T0 ( $\bar{x}_{T0}= 50.27$  ;  $ET=3.57$ ) ( $t_{40}= -0.90$  ;  $p = 0.374$ ). Les scores ont également augmenté de 1 %. La taille de l'effet est plus importante pour la lecture de mots sémantisés ( $d=0.33$ ) que pour celle des pseudomots ( $d=0.11$ ).

Par item, le temps moyen de lecture de mots sémantisés à T1 ( $\bar{x}_{T1}= 43.18$  ;  $ET= 5.50$ ) est très significativement inférieur à celui de T0 ( $\bar{x}_{T0}= 48.26$  ;  $ET= 5.73$ ) ( $t_{40}= 9.09$  ;  $p < 0.001$ ). Le temps moyen de lecture de mots sémantisés a ainsi diminué de 11 % entre les deux passations. Concernant les pseudomots, le temps moyen de lecture à T1 ( $\bar{x}_{T1}= 49.24$ ;  $ET= 7.80$ ) est aussi très significativement inférieur à celui de T0( $\bar{x}_{T0}= 53.98$   $ET= 7.90$ ) ( $t_{40}=3.99$  ;  $p < 0.001$ ). Le temps de lecture a diminué de 9 %. La taille de l'effet est plus importante en ce qui concerne la vitesse de lecture des mots sémantisés ( $d= 0.90$ ) que des pseudomots ( $d=0.60$ ).

#### 4.1.2. Eval2m

Le nombre de mots lus et le nombre de mots correctement lus sont pris en compte selon le relevé des scores proposé dans le test. Le nombre de mots lus à T1 ( $\bar{x}_{T1}= 189.15$  ;  $ET= 26.71$ ) est significativement supérieur au nombre de mots lus à T0 ( $\bar{x}_{T0}= 182.29$  ;  $ET= 24.20$ ) ( $t_{40}= -3.23$  ;  $p = 0.002$ ). Ainsi, les scores ont augmenté de 4 % entre T0 et T1 et la taille de l'effet est de 0.27.

Le nombre de mots correctement lus à T1 ( $\bar{x}_{T1}= 186.66$  ;  $ET= 27.06$ ) est très significativement supérieur à celui de T0 ( $\bar{x}_{T0}= 178.93$  ;  $ET= 25.24$ ) ( $t_{40}= -3.68$  ;  $p < 0.001$ ). Les scores ont augmenté de 4 % entre T0 et T1 et la taille de l'effet est de 0.30. Ces résultats sont répertoriés dans le tableau 2.

N=41	T0		T1		p value	d de Cohen	Pourcentage de gain %
	M (SD)	M (SD)	M (SD)	M (SD)			
<u>Evalec</u>							
Précision tous mots	116.32 (4.50)	117.17 (3.67)	0.08	0.21	1		
Vitesse tous mots	6 250.32 (748.15)	5644.66 (717.88)	< 0.001	0.83	10		
Précision mots sémantisés (MIMR) /68	66.05 (1.60)	66.54 (1.36)	<0.05	0.33	1		
Vitesse mots sémantisés (par item en ms)	48.26 (5.73)	43.18 (5.50)	< 0.001	0.90	11		
Précision pseudomots)/55	50.27 (3.57)	50.63 (2.72)	0.374	0.11	1		
Vitesse pseudomots (par item en ms)	53.98 (7.90)	49.24 (7.80)	< 0.001	0.60	9		
<u>Eval2m</u>							
Nombre de mots lus (ML)	182.29 (24.20)	189.15 (26.71)	0.002	0.27	4		
Nombre de mots correctement lus (MCL)	178.93 (25.24)	186.66 (27.06)	< 0.001	0.30	4		

**Tableau 1 :** Analyse des effets retest à Evalec et Eval2m à T0 et T1. Eval2m comprend 263 mots, mais la lecture est limitée à 2 minutes. NORMES : 141 à 182 mots lus (ML) ; 135 à 180 mots correctement lus (MCL) pour des enfants de CM1-CM2.

#### 4.1.3. Comparaison qualitative entre les gains à l'épreuve de lecture de mots (Evalec- Primaire) et l'épreuve de lecture Eval2m (Evaléo 6-15).

La taille de l'effet la plus importante concerne les scores obtenus en vitesse de lecture à Evalec (d=0.83). La taille de l'effet la plus faible porte sur les scores en précision à Evalec (d=0.21). Enfin, la taille d'effet est similaire pour les scores de mots lus (d=0.27) et de mots correctement lus à Eval2m (d=0.30). A l'épreuve de lecture d'Evalec, en considérant tous les scores (vitesse et précision) des mots sémantisés et des pseudomots, la taille de l'effet est plus importante en vitesse de lecture (MIMR : d=0.90 ; PM : d=0.60) qu'en précision (MIMR : d=0.33 ; PM : d= 0.11).

En tenant compte de cet indice de vitesse de lecture (score ML pour Eval2m), le pourcentage de gains est plus important pour l'épreuve de lecture d'Evalec (11%) que pour Eval2m (4 %).

#### 4.1.4. Répartition des enfants en pourcentage d'évolution aux deux épreuves de lecture de mots

Le pourcentage d'évolution à l'épreuve de lecture d'Evalec et à Eval2m entre T0 et T1 est très variable d'un enfant à l'autre. Ces résultats sont détaillés dans le tableau 2.

Concernant la précision en lecture de mots sémantisés d'Evalec-Primaire, 44 % des enfants sont plus précis à T1, 34 % n'ont pas amélioré leurs scores entre T0 et T1 et 22 % ont observé des scores plus faibles à T1. Pour ce qui est de la vitesse de lecture, 93% des enfants lisent plus vite à T1 alors que 7 % d'entre eux lisent moins vite. Concernant les pseudomots, 44 % des enfants sont plus précis à T1, 39% d'entre eux ont obtenu des scores plus faibles à T1 et 17 % des enfants ont obtenu des scores identiques à T0 et T1.

Le nombre de mots lus à Eval2m a augmenté pour 73 % des enfants entre T0 et T1, mais a diminué pour 22% d'entre eux. 5% des participants obtiennent les mêmes scores à T0 et T1. Le nombre de mots correctement lus a augmenté pour 76 % des enfants, mais a diminué pour 24 % d'entre eux.

L'ensemble de ces scores est visible dans le tableau 2.

N = 41		Pourcentage d'évolution entre T0 et T1					
		De -30 à -15%	De -14,9 à -0,1%	0%	De 0,1 à 15 %	De 15,1 à 30%	De 30,1 à 45%
Pourcentage d' enfants (nombre d' enfants)	<i>Evalec-Primaire</i>						
	<u>En précision</u>						
	Mots sémantisés		22 (9)	34 (14)	44 (18)		
	Pseudomots		39 (16)	17 (7)	41 (17)	2 (1)	
	<u>En vitesse</u>						
	Mots sémantisés		7 (3)		66 (27)	27 (11)	
Pseudomots	5 (2)	24 (10)		39 (16)	24 (10)	7 (3)	
	<i>Eval2M</i>						
	<u>Nbr. de mots lus</u>		22 (9)	5 (2)	61 (25)	12 (5)	
	<u>Nbr. de mots correctement lus</u>		24 (10)		63 (26)	12 (5)	

Tableau 2 : Pourcentage d'enfants selon le pourcentage d'évolution des scores entre T0 et T1.

Dans la présentation de ces résultats, pour plus de lisibilité et pour la mise en évidence des effets retest, une vitesse de lecture plus importante à T1 qu'à T0 est considérée comme une perte (car l'enfant lit moins vite) et non pas comme un gain. Autrement dit, ici, une vitesse plus élevée à T1 correspond à une perte sur les scores en vitesse de lecture (représentée dans le tableau 2) à T1.

## DISCUSSION

En langage écrit, des protocoles de soin sont proposés aux patients afin d'améliorer leurs performances en identification de mots écrits. L'évolution de ces performances est mesurée par des passations répétées (pré et post entraînement) de tests d'identification de mots écrits. De nombreux outils existent, mais peu d'entre eux décrivent précisément leurs propriétés métriques telles que les effets retest. Cela peut induire un biais dans l'interprétation des scores obtenus par un individu après un entraînement. L'absence d'études sur les effets retest à des épreuves de lecture couplé à la nécessité de considérer ces effets retest dans la validation des protocoles de soins ont conduit à la réalisation de cette étude.

L'objectif primaire de cette étude était d'évaluer les effets retest à l'épreuve de lecture d'Evalec et à Eval2m, deux tests d'identification de mots écrits. Ces deux tests ont ainsi été proposés à deux reprises (T0 et T1), avec 2 mois d'intervalle entre chaque passation, à 41 normolecteurs scolarisés en classe de CM2. Les analyses montrent qu'il existe un effet retest à ces deux épreuves, tant en précision qu'en vitesse de lecture. Ainsi, à Evalec, le pourcentage d'effet retest est estimé à 1% concernant les scores en précision et à 10% concernant la vitesse de lecture. Le pourcentage d'effet retest à Eval2m est de 4%, tous scores confondus. Ces résultats rejoignent ceux de la méta-analyse de Mc Arthur (2007) qui indique que la lecture de mots génère des effets retest à hauteur de 50%, bien que les pourcentages de gains observés dans notre étude ne soient pas aussi spectaculaires. Plusieurs hypothèses peuvent être émises pour expliquer ces gains. Tout d'abord, les tests présentés à T0 et T1 étaient identiques, tant sur le contenu que sur la forme (modalités de présentation, consignes). La méta-analyse de Scharfen et collaborateurs (2018) révèle que les effets retest entre deux passations sont plus prononcés lorsque le test est présenté sous une forme identique à T0 et T1, en comparaison avec la passation d'une forme alternative du test à T1 (même contenu mais présentation différente). L'intervalle de temps entre ces deux passations est une autre hypothèse explicative : des études, dont celles de Bartels et coll.(2010) et Hausknecht et coll.(2007) ont montré que plus les délais entre chaque passation sont courts, plus les effets retest sont grands. Dans notre étude, l'intervalle de temps séparant T0 et T1 était assez court (2 mois). Certains éléments des tests ont donc pu être récupérés facilement en mémoire, d'autant plus que les modalités de présentation de ces deux épreuves sont peu habituelles (lecture en colonne pour Eval2m et épreuve informatisée pour Evalec-Primaire). Les enfants se souvenaient très bien de la forme des tests, mais aussi des consignes. Ainsi, un temps d'adaptation n'a pas été nécessaire à T1.

De plus, ils étaient tous très motivés à l'idée de participer à cette étude. L'implication de ces facteurs dans la présence d'effets retest a également été décrite dans d'autres études (Beglinger et al., 2005; Reeve & Lam, 2007). La familiarisation avec les modalités de présentation, avec les consignes ou avec l'évaluatrice constituent donc d'autres explications possibles de la présence d'effets retest. L'étude de Randall et Villado (2017) montre que l'anxiété liée à la passation d'un test, le manque de compréhension des consignes, ou le manque de familiarité avec l'environnement du test diminuent lors d'une deuxième passation. Grissom et Bhatnagar (2009) parlent du concept d'« habitude », qui permet à l'individu de libérer davantage de ressources cognitives à la tâche à effectuer lors d'une seconde passation.

Le premier objectif secondaire était de comparer les effets retest à l'épreuve de lecture d'Evalec à ceux d'Eval2m. Concernant les scores en précision d'Evalec, et les deux scores d'Eval2m (Mots lus « ML » et mots correctement lus « MCL »), la taille de l'effet est faible, ce qui suggère que la différence de l'ensemble de ces scores est négligeable entre T0 et T1. Néanmoins, concernant les scores en vitesse de lecture à Evalec, la taille de l'effet est grande, ce qui indique que l'effet retest est plus important en vitesse de lecture qu'en précision. Autrement dit, à T1 les enfants lisent beaucoup plus vite qu'à T0. Tous scores confondus, l'effet retest est donc similaire pour les deux tests de lecture. En revanche, en tenant compte des indices de vitesse de lecture (score ML pour Eval2m), le pourcentage de gains est plus important, ce qui suggère que l'épreuve de lecture d'Evalec est plus sensible aux effets retest qu'Eval2m.

Le deuxième objectif secondaire consistait à comparer les effets retest à l'épreuve de lecture d'Evalec en fonction du type de mots : mots sémantisés et pseudomots. Concernant les mots sémantisés, la taille de l'effet est grande en vitesse et faible/moyenne en précision. Pour les pseudomots, la taille de l'effet est moyenne en vitesse et très faible en précision. Ces résultats sont discutés de deux manières : d'une part en comparant les scores aux mots sémantisés à ceux aux pseudomots, indices de précision et de vitesse confondus ; d'autre part en comparant les scores en précision avec les scores obtenus en vitesse de lecture, tous types de mots confondus.

Concernant la première analyse, la taille de l'effet est plus grande sur les mots sémantisés que sur les pseudomots, ce qui suggère que l'effet retest est plus important sur les mots sémantisés. Cela confirme l'hypothèse d'un effet de sémantisation qui augmente l'effet retest. L'étude ayant été réalisée chez des enfants de CM2, nous émettons l'hypothèse que les mots sémantisés rencontrés dans les tests étaient déjà connus des enfants, et donc associés à la fois à une

représentation orthographique et à une représentation sémantique stockées en mémoire, comme cela est décrit dans les modèles de lecture. Les auteurs du modèle de lecture à double voie (Coltheart et al., 2001) et du modèle CDP+ (Perry et al., 2007) expliquent que le lecteur utilise la voie lexicale pour lire les mots sémantisés, ce qui entraîne l'activation des représentations sémantiques et orthographiques. Le modèle en triangle (Seidenberg & McClelland, 1989) décrit quant à lui un traitement interactif entre unités phonologiques, orthographiques et sémantiques. Sprenger-Charolles et Ziegler (2019) expliquent qu'à partir du CE2, les mots sémantisés sont mieux lus que les pseudomots : alors que les mots sémantisés sont reconnus dans leur globalité, la lecture de pseudomots nécessite l'intervention de la procédure de décodage, qui requiert davantage de ressources cognitives. Pour ces différentes raisons, il est possible que la récupération en mémoire des mots sémantisés à T1 soit meilleure que celle des pseudomots et donc que l'effet retest soit plus prononcé pour les mots sémantisés que pour les pseudomots. Cependant, la lecture de pseudomots, qui nécessite beaucoup de ressources cognitives (Ziegler, 2019), peut être considérée comme une tâche complexe. D'après les données de la littérature (Scharfen et al., 2018), plus la tâche est complexe, plus les effets retest sont grands. Ces effets retest auraient donc dû être plus importants sur les pseudomots que sur les mots sémantisés. Or, cela ne s'observe que chez une petite minorité d'enfants. Nous émettons l'hypothèse que, du fait des performances souvent supérieures aux normes, cette tâche n'était pas si complexe pour les enfants de cette étude.

La seconde analyse révèle que quel que soit le type de mots, les indices de vitesse de lecture obtiennent une taille d'effet plus importante, ce qui suggère que l'effet retest est supérieur en vitesse de lecture. Cela peut s'expliquer par le fait qu'aucun entraînement ou correction des items erronés n'ait été proposé entre T0 et T1. Concernant les pseudomots, les enfants n'ont pas pu avoir de feedback sur leurs productions, ces items étant propres à Evalec. Pour ce qui est des mots sémantisés mal lus (et notamment les mots irréguliers qui sont ceux ayant généré le plus d'erreurs), si les enfants n'ont pas rencontré ce mot en contexte, ou que personne ne leur a donné la bonne forme phonologique entre T0 et T1, ils n'ont pas pu l'intégrer à leur lexique orthographique (Ziegler et al., 2014). Cette interprétation doit cependant être considérée avec précaution, étant donné que beaucoup d'enfants avaient déjà atteint le score plafond en précision à T0 (particulièrement sur les mots sémantisés). Ces enfants n'avaient donc aucune marge de progression.

Le troisième objectif secondaire visait à déterminer si la présence d'effets retest était variable d'un enfant à l'autre. Nos résultats montrent que les scores des enfants sont très

hétérogènes. La majorité des enfants a obtenu des gains entre T0 et T1, tant en vitesse qu'en précision, à l'épreuve de lecture d'Evalec comme à Eval2m. Cependant, une partie d'entre eux a obtenu des scores identiques aux deux passations, voire des scores moins bons à T1. Autrement dit, alors que certains enfants obtiennent de meilleurs scores en précision et une lecture plus rapide, d'autres obtiennent l'effet inverse avec une lecture plus lente et une diminution du nombre de mots correctement lus. Aussi, certains enfants lisent plus vite, mais ont des scores en précision plus faibles à T1, ou à l'inverse certains enfants lisent moins vite mais obtiennent de meilleurs scores en précision. L'étude de Reeve et Lam (2007) et celle de Bartels et collaborateurs (2010) indiquent que les caractéristiques propres à chaque individu (motivation, humeur, personnalité) peuvent faire varier le pourcentage d'effet retest observé, ce qui pourrait en partie expliquer ces différences inter-individuelles. Nous pouvons également émettre l'hypothèse que les facteurs expliquant la présence d'effets retest décrits plus haut ont une influence variable d'un enfant à l'autre.

Par ailleurs, Grégoire (2009) explique que lors de la passation d'un test, des erreurs de mesure aléatoires sont inévitables. Ces erreurs peuvent être propres au sujet comme la fatigue, les distractions, ou les variations dans la mise en œuvre des procédures. Dans le cadre de notre étude, toutes ces erreurs de mesure liées au participant lui-même peuvent se retrouver : bien que les procédures de passation étaient les mêmes à T1, l'ordre de passage des enfants était différent. Ainsi nous pouvons émettre l'hypothèse que les enfants ayant passé les tests en fin de journée, après une journée d'école, ont eu un effet de fatigabilité plus important que ceux les ayant passés en début de journée. Cette fatigabilité a pu être renforcée par le fait que l'épreuve de lecture d'Evalec et Eval2m n'étaient pas les seules épreuves proposées le jour des passations. Les distractions ont été variables d'un enfant à l'autre, certaines passations ayant été gênées par des nuisances sonores propres au milieu scolaire (bruits de couloir, passage, cour de récréation) ou par des travaux. Ces erreurs de mesure aléatoires ont ainsi pu favoriser certains enfants et en défavoriser d'autres. Ces différences inter-individuelles rappellent donc que les tests ne sont que des outils d'évaluation, qui n'ont aucun sens sans interprétation et observations cliniques (Grégoire & Laveault, 2014).

Ces résultats restent toutefois intéressants, tant pour la pratique clinique que pour la recherche. L'existence des effets retest n'étant pas systématiquement prise en compte, l'interprétation des résultats est possiblement biaisée. Concernant la pratique clinique des orthophonistes, les gains observés pour un patient après un entraînement rééducatif peuvent être biaisés, les gains pouvant en partie s'expliquer par un effet retest. En recherche expérimentale,

il n'est pas rare de constater que les effets d'une intervention ne sont mesurés que sur le groupe à qui on propose cet entraînement. La présence d'un groupe contrôle (à qui aucun entraînement n'est proposé) n'est pas systématique. Cela est pourtant primordial, notamment pour évaluer le pourcentage d'effets retest à prendre en compte dans l'interprétation des résultats.

## CONCLUSIONS ET PERSPECTIVES

Cette étude est issue d'un double constat : le manque de description des effets retest dans les manuels des tests, et l'absence d'études évaluant spécifiquement les effets retest à des tests de lecture. L'hypothèse principale de cette étude postulait l'existence d'un effet retest à l'épreuve de lecture d'Evalec-Primaire et à Eval2m. Pour vérifier cette hypothèse, nous avons proposé deux passations de ces tests, espacées de 2 mois, à 41 normolecteurs scolarisés en CM2.

Nos résultats montrent l'existence d'un effet retest similaire à ces deux épreuves, bien que les effets retest en vitesse de lecture soient plus marqués dans l'épreuve de lecture d'Evalec. La familiarisation avec l'environnement du test, la motivation et l'intervalle de temps court entre les deux passations sont les principales hypothèses retenues pour expliquer ces effets retest. Cette étude révèle également que les mots sémantisés (mots réguliers et irréguliers) sont plus sensibles à l'effet retest que les pseudomots, tant en vitesse qu'en précision, même si encore une fois, les effets retest sont plus marqués en vitesse de lecture. Enfin, nos résultats montrent que les effets retest sont très variables d'un enfant à l'autre : alors que certains d'entre eux obtiennent des gains plus ou moins importants entre les deux passations, d'autres n'améliorent pas leurs scores, voire obtiennent une dégradation de ces scores lors de la deuxième passation.

Afin de limiter cet effet retest, la proposition de formes alternatives d'un même test, c'est-à-dire avec le même contenu mais avec des modalités de présentation différentes est une piste à explorer dans de futures études. Les deux épreuves de lecture étudiées ne sont qu'un échantillon des tests utilisés en orthophonie. Des études similaires, portant sur d'autres tests orthophoniques seraient ainsi pertinentes. Proposer des intervalles de temps différents entre chaque passation, ou proposer plus de deux passations serait également intéressant.

Ainsi, ces résultats sont pertinents tant pour la pratique clinique que pour la recherche. Dans le premier cas, les gains d'un patient observés après un entraînement rééducatif peuvent être biaisés si les effets retest ne sont pas pris en compte, ces gains pouvant en partie s'expliquer par l'effet retest. Dans le second cas, ces résultats rappellent à quel point il est important d'inclure des groupes contrôles, dans les études qui visent notamment à prouver l'efficacité d'une intervention afin de différencier les gains liés à l'intervention proposée et ceux liés aux effets retest. Enfin, l'observation de différences inter-individuelles rappelle également qu'un clinicien ou un chercheur ne doit pas s'appuyer uniquement sur les performances obtenues à un test, l'interprétation clinique étant tout aussi importante.

## BIBLIOGRAPHIE

- Ans, B., Carbonnel, S., & Valdois, S. (1998). A Connectionist Multiple-Trace Memory Model for Polysyllabic Word Reading. *Psychological review*, *105*, 678-723.  
<https://doi.org/10.1037//0033-295X.105.4.678-723>
- Anthony, J., & Francis, D. (2005). Development of Phonological Awareness. *Current Directions in Psychological Science*, *14*, 255-259. <https://doi.org/10.1111/j.0963-7214.2005.00376.x>
- Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults : Longitudinal study on frequent repetitive cognitive testing. *BMC neuroscience*, *11*, 118. <https://doi.org/10.1186/1471-2202-11-118>
- Bedoin. (2017). Rééquilibrer les analyses visuo-attentionnelles globales et locales pour améliorer la lecture chez des enfants dyslexiques de surface. *Approche Neuropsychologique des Apprentissages chez l'Enfant*, *29*(148).
- Bedoin, N., Kéïta, L., Roussel, L., Herbillon, C., & Launay, V. (2010). *Diagnostic et remédiation d'un déficit d'inhibition des détails dans la dyslexie de surface* (p. 179-210).
- Beglinger, L., Gaydos, B., Tangphao, O., Duff, K., Kareken, D., Crawford, J., Fastenau, P., & Siemers, E. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, *20*, 517-529.  
<https://doi.org/10.1016/j.acn.2004.12.003>
- Benedict, R., & Zgaljardic, D. (1998). Practice Effects During Repeated Administrations of Memory Tests With and Without Alternate Forms. *Journal of clinical and experimental neuropsychology*, *20*, 339-352. <https://doi.org/10.1076/jcen.20.3.339.822>
- Bentolila, A., & Germain, B. (2019). *L'apprentissage de la lecture*. Nathan.

- Bertrand, D., Fluss, J., Billard, C., & Ziegler, J. (2010). Efficacité, sensibilité, spécificité : Comparaison de différents tests de lecture. *L'Année psychologique*, *110*, 299-320.  
<https://doi.org/10.4074/S000350331000206X>
- Casalis, S., Bois Parriaud, F., Cavalli, E., & Chaix, Y. (2018). *Les dyslexies*. Elsevier Masson.
- Casalis, S., Leloup, G. orthophoniste, & Bois Parriaud, F. (2019). *Prise en charge des troubles du langage écrit chez l'enfant* (2e édition). Elsevier Masson.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars : Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, *19*, 5-51.  
<https://doi.org/10.1177/1529100618772271>
- Christopher, M., Miyake, A., Keenan, J., Pennington, B., DeFries, J., Wadsworth, S., Willcutt, E., & Olson, R. (2012). Predicting Word Reading and Comprehension With Executive Function and Speed Measures Across Development : A Latent Variable Analysis. *Journal of experimental psychology. General*, *141*, 470-488.  
<https://doi.org/10.1037/a0027375>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological review*, *108*, 204-256. <https://doi.org/10.1037/0033-295X.108.1.204>
- Crocq, M.-A., Guelfi, J.-D., Boyer, P., Pull, C.-B., & Pull, M.-C. (2015). *DSM-5—Manuel diagnostique et statistique des troubles mentaux* (5e édition). Elsevier Masson.
- Donovan, J., & Radosevich, D. (1999). A Meta-Analytic Review of the Distribution of Practice Effect : Now You See It, Now You Don't. *Journal of Applied Psychology*, *84*, 795-805. <https://doi.org/10.1037/0021-9010.84.5.795>
- Engel de Abreu, P., Gathercole, S., & Martin, R. (2011). Disentangling the relationship between working memory and language : The roles of short-term storage and

- cognitive control. *Learning and Individual Differences*, 569-574.  
<https://doi.org/10.1016/j.lindif.2011.06.002>
- Fermanian, J. (2005). Validation des échelles d'évaluation en médecine physique et de réadaptation : Comment apprécier correctement leurs qualités psychométriques. *Annales de Réadaptation et de Médecine Physique*, 48(6), 281-287.  
<https://doi.org/10.1016/j.annrmp.2005.04.004>
- Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency : Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, 19(1), 19-52. Cambridge Core.  
<https://doi.org/10.1017/S0142716400010560>
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education*, 7(1), 6-10.  
<https://doi.org/10.1177/074193258600700104>
- Grégoire, J. (2009). Chapitre 4. Propriétés métriques du WISC-IV. In *L'examen clinique de l'intelligence de l'enfant: Vol. 2e éd.* (p. 149-185). Mardaga; Cairn.info.  
<https://www.cairn.info/l-examen-clinique-de-l-intelligence-de-l-enfant--9782804700164-p-149.htm>
- Grégoire, J., & Laveault, D. (2014). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*.
- Grissom, N., & Bhatnagar, S. (2009). Habituation to repeated stress : Get used to it. *Special Issue: Neurobiology of Habituation*, 92(2), 215-224.  
<https://doi.org/10.1016/j.nlm.2008.07.001>
- Harm, M., & Seidenberg, M. (2003). *Computing the Meanings of Words in Reading:*

- Hausknecht, J., Halpert, J., Paolo, N., & Gerrard, M. (2007). Retesting in selection : A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373-385.
- Hulme, C., Bowyer-Crane, C., Carroll, J., Duff, F., & Snowling, M. (2012). The Causal Role of Phoneme Awareness and Letter-Sound Knowledge in Learning to Read. *Psychological science*, 23, 572-577. <https://doi.org/10.1177/0956797611435921>
- Launay, L., Maeder, C., & Roustit, J. (2018). *Evaléo 6-15 : Évaluation du langage écrit et du langage oral 6 - 15 ans*. Ortho édition.
- Lefavrais, P. (1967). *Manuel du test de l'alouette : [Test psychologique] : test d'analyse de la lecture et de la dyslexie*. Centre de psychologie appliquée.
- Lievens, F., Buyse, T., & Sackett, P. (2005). *Retest effects in operational selection settings : Development and test of a framework*.
- Locascio, G., Mahone, E., Eason, S., & Cutting, L. (2010). Executive Dysfunction Among Children With Reading Comprehension Deficits. *Journal of learning disabilities*, 43, 441-454. <https://doi.org/10.1177/0022219409355476>
- Mann, V., & Liberman, I. (1985). Phonological Awareness and Verbal Short-Term Memory. *Journal of learning disabilities*, 17, 592-599. <https://doi.org/10.1177/002221948401701005>
- McArthur, G. (2007). Test-retest effects in treatment studies of reading disability : The devil is in the detail. *Dyslexia (Chichester, England)*, 13, 240-252. <https://doi.org/10.1002/dys.355>
- Melby-Lervåg, M., Lyster, S.-A., & Hulme, C. (2012). Phonological skills and their role in learning to read : A meta-analytic review. *Psychological bulletin*, 138, 322-352. <https://doi.org/10.1037/a0026744>

- Morais, J. (1991). Phonological Awareness : A Bridge Between Language and Literacy. In D. J. Sawyer & B. J. Fox (Éds.), *Phonological Awareness in Reading : The Evolution of Current Perspectives* (p. 31-71). Springer New York. [https://doi.org/10.1007/978-1-4612-3010-6\\_2](https://doi.org/10.1007/978-1-4612-3010-6_2)
- Pammer, K., & Vidyasagar, T. (2005). Integration of the visual and auditory networks in dyslexia : A theoretical perspective. *Journal of Research in Reading*, 28, 320-331. <https://doi.org/10.1111/j.1467-9817.2005.00272.x>
- Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic Knowledge and Learning to Read are Reciprocal : A Longitudinal Study of First Grade Children. *Merrill-Palmer Quarterly*, 33(3), 283-319. <https://www.jstor.org/stable/23086537>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories : The CDP+ model of reading aloud. *Psychological Review*, 114(2), 273-315. <https://doi.org/10.1037/0033-295X.114.2.273>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2014). When silent letters say more than a thousand words : An implementation and evaluation of CDP++ in French. *Journal of Memory and Language*, 72, 98-115. <https://doi.org/10.1016/j.jml.2014.01.003>
- Phenix, T., Diard, J., & Valdois, S. (2016). *Les modèles computationnels de lecture* (p. 167).
- Piérart, B., & Grégoire, J. (2004). Déchiffrer et comprendre : Le test de closure en lecture revisitée—Étalonnage belge du L3 de Lobrot. *Le Langage et l'Homme*, XXXIX(2), 87-100.
- Piquard, A., & Sprenger-Charolles, L. (2013). Early predictors of future reading skills : A follow-up of French-speaking children from the beginning of kindergarten to the end of the second grade (Age 5 to 8). *L'Année psychologique*, 491-521. <https://doi.org/10.4074/S0003503313014012>

- Randall, J., & Villado, A. (2017). Take two : Sources and deterrents of score change in employment retesting. *Human Resource Management Review*, 27.  
<https://doi.org/10.1016/j.hrmr.2016.10.002>
- Reeve, C., & Lam, H. (2007). The Relation Between Practice Effects, Test-Taker Characteristics and Degree of g-Saturation. *INTERNATIONAL JOURNAL OF TESTING*, 7, 225-242. <https://doi.org/10.1080/15305050701193595>
- Reynolds, D., Nicolson, R. I., & Hambly, H. (2003). Evaluation of an exercise-based treatment for children with reading difficulties. *Dyslexia*, 9(1), 48-71.  
<https://doi.org/10.1002/dys.235>
- Rohrer, D., & Pashler, H. (2010). Reviews/Essays : Recent Research on Human Learning Challenges Conventional Instructional Strategies. *Educational Researcher*, 39, 406-412. <https://doi.org/10.3102/0013189X10374770>
- Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Demonet, J., Bricout, L., Billard, C., Nguyen-Morel, M.-A., Le Heuzey, M., Soares-Boucaud, I., George, F., Ziegler, J., & Ramus, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology*, 52.  
<https://doi.org/10.1037/dev0000184>
- Scharfen, J., Peters, J., & Holling, H. (2018). Retest effects in cognitive ability tests : A meta-analysis. *Intelligence*, 67, 44-66. <https://doi.org/10.1016/j.intell.2018.01.003>
- Seidenberg, M., & McClelland, J. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological review*, 96, 523-568.  
<https://doi.org/10.1037//0033-295X.96.4.523>
- Sprenger-Charolles, L., Colé, P., Piquard-Kipffer, A., & Leloup, G. (2010). *Une batterie informatisée d'évaluation diagnostique des troubles spécifiques d'apprentissage de la lecture*. Isbergues : Ortho-Edition.

- Sprenger-Charolles, L., Desrochers, A., & Gentaz, É. (2018). Apprendre à lire-écrire en français. *Langue française*, 199(3), 51-67. Cairn.info.  
<https://doi.org/10.3917/lf.199.0051>
- Sprenger-Charolles, L., Siegel, L., & Béchenec, D. (1998). Phonological Mediation and Semantic and Orthographic Factors in Silent Reading in French. *Scientific Studies of Reading*, 2, 3-29. [https://doi.org/10.1207/s1532799xssr0201\\_1](https://doi.org/10.1207/s1532799xssr0201_1)
- Sprenger-Charolles, L., Siegel, L. S., Béchenec, D., & Serniclaes, W. (2003). Development of phonological and orthographic processing in reading aloud, in silent reading, and in spelling : A four-year longitudinal study. *Journal of Experimental Child Psychology*, 84(3), 194-217. [https://doi.org/10.1016/S0022-0965\(03\)00024-9](https://doi.org/10.1016/S0022-0965(03)00024-9)
- Sprenger-Charolles, L., & Ziegler, J. (2019). *Apprendre à lire : Contrôle, automatismes et auto-apprentissage*.
- Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia*, 7(1), 12-36.  
<https://doi.org/10.1002/dys.186>
- Vahine, T. (2017). *Traitements visuels précoces du langage écrit : Études chez l'enfant et l'adulte jeune*.
- Valdois, S., Ans, B., & Carbonnel, S. (2007). *Principes de base pour l'élaboration d'un modèle connexionniste psychologiquement plausible de l'apprentissage de la lecture*.
- Valdois, S., Peyrin, C., Lassus, D., Lallier, M., Demonet, J., & Kandel, S. (2013). Dyslexia in a French-Spanish bilingual girl : Behavioural and neural modulations following a visual attention span intervention. *Cortex*, 53.  
<https://doi.org/10.1016/j.cortex.2013.11.006>
- Wolf, M., Miller, L., & Donnelly, K. (2000). Retrieval, Automaticity, Vocabulary Elaboration, Orthography (RAVE-O). *Journal of learning disabilities*, 33, 375-386.  
<https://doi.org/10.1177/002221940003300408>

- Ziegler. (2018). Différences inter-linguistiques dans l'apprentissage de la lecture. *Langue française*, 199(3), 35-49. Cairn.info. <https://doi.org/10.3917/lf.199.0035>
- Ziegler. (2019). *L'art de lire et d'enseigner la lecture*. 20.
- Ziegler, J., & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages : A Psycholinguistic Grain Size Theory. *Psychological bulletin*, 131, 3-29. <https://doi.org/10.1037/0033-2909.131.1.3>
- Ziegler, J., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching : Implications for dyslexia. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369, 20120397. <https://doi.org/10.1098/rstb.2012.0397>
- Ziegler, J., Perry, C., & Zorzi, M. (2019). *Modeling the Variability of Developmental Dyslexia*.
- Zoubinetzky, R., Collet, G., Serniclaes, W., Nguyen-Morel, M.-A., & Valdois, S. (2016). Relationships between Categorical Perception of Phonemes, Phoneme Awareness, and Visual Attention Span in Developmental Dyslexia. *PloS One*, 11(3), e0151015. PubMed. <https://doi.org/10.1371/journal.pone.0151015>

# ANNEXES

## Annexe 1 : Formulaire de consentement éclairé



Département d'Orthophonie  
28, av. Valombrose  
06107 Nice Cedex

*Recueil du consentement des parents et de l'enfant à l'étude d'effets retest de la lecture pour les enfants scolarisés en CMI et CM2.*

### Promoteur

Laboratoire Cobtek (EA 7276)  
Institut Claude Pompidou/CMRR  
10 rue Molière  
06100 Nice

### Secrétariat et Contact

04 93 37 76 30, \_\_\_\_\_

[Francoise.pasturaud@unice.fr](mailto:Francoise.pasturaud@unice.fr)

### Investigateur coordinateur

Gilles Leloup  
Cobtek (EA 7276) - CHU Nice (C.E.R.T.A)

### Contact

[gilles.leloup@univ-cotedazur.fr](mailto:gilles.leloup@univ-cotedazur.fr)

## **Chère Madame, Cher Monsieur,**

Bonjour et merci de lire ces quelques informations avant de donner votre accord pour que votre enfant participe à l'étude sur les effets retest des tests de lecture. A la suite de l'annonce que vous avez reçue, la loi n°2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine (dite loi Jardé) nécessite que vous donniez votre consentement pour que votre enfant participe à cette étude sur les effets retest de lecture.

Nous devons vous rappeler l'objectif, la durée, le lieu de passation, les tests de lecture, les mentions de récoltes des résultats, de l'archivage des données, les bénéfices et les risques de cette étude.

*Objectif* : en pratique clinique ou lors des études expérimentale sur les troubles de la lecture (dyslexie), les cliniciens et les chercheurs doivent prendre des mesures répétées de vitesse et de précision de lecture. Or, la répétition de ces mesures peut invalider les résultats si nous ne connaissons pas le gain moyen qu'obtiennent les enfants à ces passations répétées. C'est pour

cette raison que sollicitons votre enfant, son consentement et le votre pour obtenir ces informations qui nous permettront d'avoir des barèmes de mesures de lecture plus fiables.

*Durée* : au total d'1h30 répartie sur 5 passations. La première passation est de 30 minutes, les quatre suivantes de 20 minutes chacune réparties sur un an.

*Lieu de passation* : son établissement scolaire

*Tests* : 5 tests de lecture et 5 tests pour évaluer ses habiletés langagières et visuelles, et des questions sur ce que votre enfant pensera des tests qu'il va lire. Plus un questionnaire de renseignements destiné aux parents et un questionnaire destiné aux enseignants.

*Récoltes des résultats* : anonyme (nom de votre enfant codé en 4 lettres), stockage des données au CHU-Lenval (CERTA). Possibilité sur simple mail ou courrier de détruire toutes les données concernant les données sur votre enfant. Une copie des questionnaires et des résultats des passations sera conservée par l'investigateur pendant le temps de l'étude et de la rédaction d'une publication scientifique. A la suite, toutes les données seront archivées.

*Participation* volontaire : à tout moment votre enfant lors d'une passation ou vous-mêmes pouvez faire part par mail, courrier, le souhait de suspendre la participation à l'étude.

*Bénéfices et Risques* : les possibles bénéfices sont de vous informer et d'informer votre enfant sur son niveau de lecture. Dans le cas d'une suspicion de difficultés, nous vous informerons directement. A notre connaissance pas de risque, et si ces passations inquiètent trop votre enfant, elles seront de suite suspendues à sa demande et à la vôtre.

Dans le cadre de la validation des effets retest chez les enfants scolarisés en CM1 et CM2, j'autorise mon enfant (nom et prénoms).....,né(e) ....., scolarisé(e) en classe de ..... ,à l'école....., à participer à la passation du questionnaire.

Le

A

Signature d'un ou des deux parents précédée « je confirme que mon enfant participera à cette étude »

ARNAUD Camille

## **Etude des effets retest à deux épreuves de lecture chez des normolecteurs de CM2.**

### Résumé

**Contexte :** La mesure de l'évolution des performances entraînées d'un individu passe par des évaluations répétées qui peuvent induire des effets retest. Ces effets sont généralement décrits dans les manuels, mais pas de manière systématique. L'interprétation des résultats peut alors être biaisée, une partie des gains obtenus pouvant s'expliquer plus facilement par un effet retest que par l'efficacité d'un entraînement.

**Méthode :** L'objectif principal de cette étude est de vérifier s'il existe un effet retest à deux épreuves de lecture : Lecture à Haute Voix d'Evalec-Primaire et Eval2m d'Evaléo 6-15. Notre étude porte sur 41 enfants normolecteurs (20 filles, 21 garçons) scolarisés en CM2. Deux passations des mêmes tests de lecture (T0 et T1), espacées de 2 mois leur ont été proposées, sans entraînement entre les deux.

**Résultats :** Les scores en précision (P) à l'épreuve de lecture d'Evalec-Primaire sont meilleurs à T1 ( $\bar{x}_{T1} = 117.17$ ) qu'à T0 ( $\bar{x}_{T0} = 116.32$ ) tout comme les scores en vitesse de lecture (Tps) ( $\bar{x}_{T1} = 5644.66$  ;  $\bar{x}_{T0} = 6250.32$ ). Les scores à Eval2M sont également meilleurs à T1 (Mots lus :  $\bar{x}_{T1} = 189.15$  ; Mots correctement lus :  $\bar{x}_{T1} = 186.66$ ) qu'à T0 (ML :  $\bar{x}_{T0} = 182.29$  ; MCL :  $\bar{x}_{T0} = 178.93$ ). Les pourcentages de gain sont plus importants sur les mots sémantisés (P : 1% ; Tps : 11%) que sur les pseudomots (P : 1% ; Tps : 9%).

**Discussion :** Ces résultats montrent l'existence d'un effet retest globalement similaire à ces deux épreuves, bien qu'il soit plus marqué en vitesse de lecture. De plus, les mots sémantisés sont plus sensibles à l'effet retest que les pseudomots. Des études similaires, portant sur d'autres tests orthophoniques seraient pertinentes.

### Mots-clés

Effet retest, identification de mots écrits, normo-lecteurs, enfants, évaluation

## Abstract

**Background:** Speech and language therapy protocol validation requires repetitive assessments to measure trained performance evolution. These repetitive assessments can lead to retest effects. These effects are described in test manuals, but not systematically. Interpretation of results can be biased because a part of gains can be explained by a retest effect rather than the treatment efficiency.

**Methods:** The purpose of this study is to assess the existence of a retest effect at 2 reading tests : “Lecture à haute voix” (Evalec-Primaire) and Eval2m (Evaléo 6-15). In our study 41 children (20 girls and 21 boys) with good reading skills in their fifth year of primary school were recruited. Two sessions of these reading tests (T0 and T1) were conducted with a 2 months interval and without training between them.

**Results:** Second session (T1) reading accuracy scores (P) ( $\bar{x}_{T1}= 117.17$ ) at Evalec test are better than first session ones ( $\bar{x}_{T0}= 116.32$ ). We observed similar results for reading speed (Tps) ( $\bar{x}_{T1}= 5644.66$  ;  $\bar{x}_{T0}= 6250.32$ ). Second session Eval2m score (Read words:  $\bar{x}_{T1}= 189.15$  ; Words correctly read :  $\bar{x}_{T1}= 186.66$ ) are also better than first session ones (RW :  $\bar{x}_{T0}= 182.29$  ; WCR :  $\bar{x}_{T0}= 178.93$ ). Percentage gains are more important in semantics words (P : 1% ; Tps : 11%) than in pseudowords ones (P : 1% ; Tps : 9%).

**Discussion:** Our results show that there are similar retest effects at both studied tests, more substantial in reading speed. These retest effects are also more important in semantics words than in pseudowords. It'd be relevant to explore retest effects in other speech and language therapy tests with similar studies.

## Keywords

Retest effect, word recognition process, good reading skills, children, assessment