



<http://portaildoc.univ-lyon1.fr>

Creative commons : Paternité - Pas d'Utilisation Commerciale -  
Pas de Modification 2.0 France (CC BY-NC-ND 2.0)



<http://creativecommons.org/licenses/by-nc-nd/2.0/fr>



Université Claude Bernard Lyon 1  
Institut des Sciences et Techniques de Réadaptation  
Département Orthophonie

---

**N° de mémoire 2042**

Mémoire d'Orthophonie

présenté pour l'obtention du

**Certificat de capacité d'orthophoniste**

Par

**BRAULT Marie**

**Inventaire et analyse des critères de validité et de fidélité  
mentionnés par les outils d'évaluation utilisés pour le bilan  
orthophonique en langage écrit en France**

Directeur de Mémoire :

**HILAIRE-DEBOVE Géraldine**

Membres du jury

**BOULENGER Véronique**

**POZARD Prescillia**

**HILAIRE-DEBOVE Géraldine**



Directeur ISTR  
**Xavier PERROT**

**Équipe de direction du département d'orthophonie :**

Directeur de la formation  
**Agnès BO**

Coordinateur de cycle 1  
**Claire GENTIL**

Coordinateur de cycle 2  
**Solveig CHAPUIS**

Responsables de l'enseignement clinique  
**Claire GENTIL**  
**Ségolène CHOPARD**  
**Johanne BOUQUAND**

Responsables des travaux de recherche  
**Lucie BEAUVAIS**  
**Nina KLEINSZ**

Responsable de la formation continue  
**Johanne BOUQUAND**

Responsable du pôle scolarité  
**Rachel BOUTARD**

Secrétariat de scolarité  
**Anaïs BARTEVIAN**  
**Constance DOREAU KNINDICK**  
**Patrick JANISSET**  
**Céline MOULARD**

## 1. UNIVERSITÉ CLAUDE BERNARD LYON 1

Président  
**Pr. FLEURY Frédéric**

Vice-président CFVU  
**Pr. CHEVALIER Philippe**

Président du Conseil Académique  
**Pr. BEN HADID Hamda**

Vice-président CS  
**M. VALLEE Fabrice**

Vice-président CA  
**Pr. REVEL Didier**

Directeur Général des Services  
**M. VERHAEGHE Damien**

### 1.1 Secteur Santé :

U.F.R. de Médecine Lyon Est  
Doyen **Pr. RODE Gille**

Directrice **Pr. SEUX Dominique**

U.F.R de Médecine et de  
maïeutique - Lyon-Sud Charles  
Mérieux  
Doyenne **Pr. BURILLON Carole**

Institut des Sciences Pharmaceutiques  
et Biologiques  
Directrice **Pr. VINCIGUERRA Christine**

Comité de Coordination des  
Etudes Médicales (C.C.E.M.)  
Président **Pr. COCHAT Pierre**

Institut des Sciences et Techniques de  
la Réadaptation (I.S.T.R.)  
Directeur **Dr. PERROT Xavier**

U.F.R d'Odontologie

Département de Formation et Centre  
de Recherche en Biologie Humaine  
Directrice **Pr. SCHOTT Anne-Marie**

### 1.2 Secteur Sciences et Technologie

U.F.R. Faculté des Sciences et  
Technologies  
Directeur **M. DE MARCHI Fabien**

Institut des Sciences Financières et  
d'Assurance (I.S.F.A.)  
Directeur **M. LEBOISNE Nicolas**

U.F.R. Faculté des Sciences  
Administrateur provisoire  
**M. ANDRIOLETTI Bruno**

Observatoire Astronomique de Lyon  
Directeur **Mme DANIEL Isabelle**

U.F.R. Biosciences  
Administratrice provisoire  
**Mme GIESELER Kathrin**

Ecole Supérieure du Professorat et  
de l'Education (E.S.P.E.)  
Administrateur provisoire  
**M. Pierre CHAREYRON**

U.F.R. de Sciences et Techniques  
des Activités Physiques et Sportives  
(S.T.A.P.S.)  
Directeur **M. VANPOULLE Yannick**

POLYTECH LYON  
Directeur **M. PERRIN Emmanuel**

Institut Universitaire de Technologie  
de Lyon 1 (I.U.T. LYON 1)  
Directeur **M. VITON Christophe**

## Résumé

Une multitude d'outils permet la passation du bilan orthophonique du langage écrit en France. Les orthophonistes français les utilisent souvent par popularité, et non pour leurs qualités psychométriques. Un questionnement sur les critères psychométriques les plus utilisés pour la construction d'un test a alors émergé. Pour répondre à cela, des manuels de tests ont été étudiés et une fiche descriptive a été complétée à destination des orthophonistes. Ces fiches ont ensuite été analysées, ce qui a permis de découvrir que la normalisation des échantillons des tests analysés, la représentation de la population générale, la sensibilité ainsi qu'une partie des validités disponibles sont développées dans la plupart des outils. Néanmoins, les seuils pathologiques ainsi que tous les critères de fidélité sont absents de beaucoup de tests. Ces critères, retrouvés dans des études précédemment menées en pays francophones ou anglophones, marquent le début d'une rigueur psychométrique progressivement recherchée par les soignants pour leurs outils de diagnostic. Les fiches, mises à disposition des orthophonistes français, leur permettront d'utiliser un test en ayant un regard éclairé sur ses qualités psychométriques.

Mots-clés : orthophonie – langage écrit – bilan – évaluation – validité – fidélité

## **Abstract**

A multitude of tools are available for the speech and language assessment of written language in France. French speech therapists often use them by popularity, not for their psychometric qualities. A questioning on the psychometric criteria most used for the construction of a test then emerged. To answer this, test manuals were studied and a descriptive sheet was completed for use by speech therapists. These sheets were then analysed, which revealed that the standardisation of the samples of the tests analysed, the representation of the general population, the sensitivity and some of the available validities are developed in most of the tools. Nevertheless, pathological thresholds and all the criteria of fidelity are absent from many tests. These criteria, found in studies previously conducted in French- or English-speaking countries, mark the beginning of a psychometric rigour progressively sought after by caregivers for their diagnostic tools. The cards, made available to French speech therapists, will enable them to use a test with an enlightened view of its psychometric qualities.

Key-words : speech and language therapy – written language – evaluation – assessment – validity – fidelity

## Remerciements

Je tiens tout d'abord à remercier ma directrice de mémoire, Géraldine Hilaire-Debove, orthophoniste, Docteur en Sciences du Langage, chargée d'enseignement à l'ISTR de Lyon et Directrice du Laboratoire UNADREO de Recherche clinique en Orthophonie, de m'avoir permis de prendre part au projet de l'UNADREO, merci pour son encadrement et sa disponibilité.

Je remercie également tous ceux qui ont contribué à ce mémoire, en particulier les éditions ECPA, qui auraient voulu me mettre à disposition leur matériel.

Je remercie mes différents maitres de stage et professeurs qui ont su attiser l'envie de devenir orthophoniste, j'espère pouvoir m'y enrichir autant qu'eux.

Je remercie également mes amies, devenues ma famille lyonnaise, pour ces cinq années d'échanges, d'apprentissage, de découvertes, d'instant précieux d'amitié, qui vont perdurer au-delà des kilomètres.

Merci à Clémentine, restée en terres bordelaises, dont la distance n'a jamais altéré son soutien et son amitié.

Merci à ma famille : Alexandra, ma mère, mon père, mes grands-parents, qui ont toujours su trouver de l'intérêt à mes projets ainsi que les mots apaisants quand ceux-ci ne se présentaient pas comme ils le devaient.

Enfin, merci à Loïc d'avoir traversé la France il y a cinq ans pour devenir la personne sans qui ma vie n'aurait pas la même saveur aujourd'hui.

## Sommaire

<b>I</b>	<b>Partie théorique .....</b>	<b>1</b>
1	<i>Le début des tests dans l'histoire de l'orthophonie.....</i>	<i>2</i>
2	<i>Les propriétés d'un test pertinent.....</i>	<i>3</i>
2.1	La standardisation .....	3
2.2	Les normes .....	3
2.3	La validité.....	4
2.3.1	La validité de contenu.....	4
2.3.2	La validité empirique.....	4
2.3.3	La validité écologique.....	5
2.3.4	La validité théorique.....	5
2.3.5	La sensibilité.....	6
2.3.6	La spécificité.....	6
2.4	La fidélité.....	7
2.4.1	La fidélité inter-observation.....	7
2.4.2	La fidélité inter-observateurs.....	7
2.4.3	La cohérence interne.....	7
3	<i>Les travaux réalisés à l'étranger .....</i>	<i>8</i>
3.1	Outils anglophones.....	8
3.2	Outils francophones .....	8
4	<i>Pistes de travail et problématique .....</i>	<i>9</i>
<b>II</b>	<b>Méthode .....</b>	<b>11</b>
1	<i>Échantillon.....</i>	<i>11</i>
1.1	Constitution de l'échantillon .....	11
1.2	Exclusion.....	11
2	<i>Grille d'analyse .....</i>	<i>13</i>
2.1	La grille d'évaluation (Annexe A).....	13
2.1.1	L'intérêt de ce matériel.....	13
2.1.2	Les différentes parties de la grille.....	13
2.1.2.1	La standardisation.....	13
2.1.2.2	La sensibilité.....	14



2.1.2.3	La fidélité.....	14
2.1.2.4	La validité.....	14
2.2	La fiche finale diffusée aux orthophonistes français (Annexe B).....	14
3	<i>Procédure</i> .....	15
3.1	Complétion de la grille.....	15
3.2	Complétion de la fiche.....	15
3.3	Comparaison des différents outils étudiés.....	15
<b>III</b>	<b>Résultats</b> .....	<b>16</b>
1	<i>Les conditions de normalisation</i> .....	16
1.1	L'échantillon de test.....	16
1.2	Données démographiques disponibles .....	17
1.2.1	Critères d'exclusion. ....	17
1.2.2	La période de passation. ....	18
1.2.3	La répartition géographique. ....	18
1.2.4	Représentation de la population. ....	19
1.3	Les mesures disponibles.....	19
2	<i>Sensibilité</i> .....	20
2.1	Présence d'études de cas-témoins.....	20
2.2	Les effets constatés .....	20
2.3	Le seuil pathologique .....	21
3	<i>Validité</i> .....	21
3.1	Validité théorique.....	21
3.2	Validité de contenu .....	22
3.3	Validité empirique.....	22
3.4	Validité écologique .....	23
4	<i>Fidélité</i> .....	23
4.1	Fidélité inter-observations .....	23
4.2	Fidélité inter-observateurs.....	24
4.3	Cohérence interne .....	24
<b>IV</b>	<b>Discussion et conclusion</b> .....	<b>25</b>
1	<i>Discussion des résultats</i> .....	25

1.1	Analyse de la normalisation et de la sensibilité .....	25
1.2	Critères psychométriques présents dans la majorité des tests .....	27
1.3	Critères psychométriques absents des tests.....	29
2	<i>Limites et perspectives</i> .....	30
3	<i>Conclusion</i> .....	30
<b>V</b>	<b>Références</b> .....	<b>32</b>
<b>VI</b>	<b>Annexe A : Grille de labellisation</b> .....	<b>I</b>
<b>VII</b>	<b>Annexe B : Fiche-site</b> .....	<b>II</b>
<b>VIII</b>	<b>Annexe C : Tableau récapitulatif des résultats</b> .....	<b>IV</b>
<b>IX</b>	<b>Annexe D : Fiches-site complétées</b> .....	<b>IV</b>

## I Partie théorique

La prise en charge orthophonique débute obligatoirement par un temps de bilan qui se divise en quatre temps : l'anamnèse, ou le temps de découverte du patient, la passation d'épreuves de bilan à l'aide d'un matériel approprié, la rédaction du compte-rendu orthophonique à destination du médecin prescripteur et le retour au patient ou à ses parents si celui-ci est mineur.

Une multitude de batteries d'évaluation et de tests sont disponibles pour les orthophonistes, le choix du matériel dépend d'un choix personnel et professionnel, certains sont choisis grâce à leur popularité, et pas toujours pour leurs qualités psychométriques (Betz et al., 2013). Le Collège français d'Orthophonie (CFO) a lancé une étude concernant l'évaluation orthophonique des troubles de langage écrit, les résultats montrent que la moitié des orthophonistes (49%) utilisent une seule batterie standardisée alors que 39% se servent plutôt de différents tests ne constituant pas une batterie. Seulement 4% des orthophonistes utilisent des tests qui ne sont pas standardisés (Helloin, 2019). Ainsi, les batteries standardisées deviennent les outils privilégiés des orthophonistes sur le territoire français.

Ce mémoire s'inscrit au sein d'une étude qui a débuté en 2016 par l'UNADREO (Union Nationale pour le Développement de la Recherche et de l'Évaluation en Orthophonie), société savante en orthophonie. Cette recherche vise à établir un état des lieux des tests utilisés lors du bilan afin d'en présenter leurs qualités psychométriques auprès des professionnels (Hilaire-Debove, 2017).

L'interrogation des orthophonistes rencontrés en stage, quant au choix du matériel à utiliser, a amené à un questionnaire sur la labellisation du matériel d'évaluation en orthophonie. Dans un premier temps, une partie théorique proposera une explication des différents termes psychométriques utilisés durant les travaux de recherche, ainsi qu'un regard sur ce qui a été fait dans les pays étrangers. L'étude des manuels des différents tests sera alors développée dans les parties suivantes grâce à l'analyse des fiches produites lors des recherches menées pour cette étude. Une discussion suivra.

## 1 Le début des tests dans l'histoire de l'orthophonie

Le terme d'orthophonie a été utilisé pour la première fois par le Docteur Marc Colombat de l'Isère en 1829, lors de l'ouverture de l'institut orthophonique de Paris pour la prise en soin du bégaiement. Ce terme ne renverra alors qu'à « l'étude de ce trouble et des vices de la parole » (Kremer & Lederlé, 2016). En 1903, le médecin André Catex, ORL à Paris, ouvre des consultations d'orthophonie dans une institution de sourds-muets. Par la suite, il va diriger, avec l'aide du Docteur Robert Jouet, la publication du premier Traité de l'orthophonie en 1920 (Heral, 2007, p37).

La naissance de l'orthophonie en France interviendra véritablement en 1925, par la rencontre de Suzanne Borel-Maisonny, phonéticienne et grammairienne, et du Professeur Veau, chirurgien, lorsque ce dernier l'interroge sur la récupération de la parole et de l'articulation des enfants opérés de fentes palatines (Kremer & Lederlé, 2016, p7). Après un développement des prises en charge du langage oral au sens large du terme dans les années 1930, l'année 1942 marque un tournant dans l'histoire de l'orthophonie puisque Madame Borel-Maisonny rencontre le Docteur Simon, associé de Binet, créateurs de l'échelle d'intelligence Binet-Simon. C'est de cette rencontre que les premières évaluations en orthophonie naissent, publiées exclusivement par Suzanne Borel-Maisonny dans le *bulletin de la société Binet (Histoire de l'orthophonie et de la FNO, s. d.)*, la première verra le jour en 1946, il s'agit des « tests Borel » (Hilaire-Debove, 2017). D'abord utilisées de façon qualitative avec des épreuves « maison » créées par les pionnières de la profession, ces évaluations ont toujours eu vocation à amener l'orthophoniste au diagnostic et à la construction du projet thérapeutique (Ferrand, 2002).

D'abord tournée vers le langage oral, l'orthophonie s'intéresse au langage écrit dès 1952, grâce à Claude Chassagny, futur psychanalyste, qui se concentre sur l'apprentissage de la lecture chez l'enfant. En 1958, il créait « l'École de formation des rééducateurs de la dyslexie » (Heral, 2007). C'est entre les années 1950 et 1970 que va se développer le courant pédagogique avec la recherche de « bon lecteur » et donc la création des premières évaluations de la lecture (Hilaire-Debove, 2017). Le bilan orthophonique du langage écrit va permettre de déterminer les compétences du sujet en lecture, la compréhension qui en découle, les dysfonctionnements et leurs répercussions (Touzin, 2002). S'ensuit, dans les années 1970, l'élaboration de batteries d'évaluation pour « normer le développement à partir de repères d'âge », en

particulier pour les tests de langage écrit, dont le nombre est exponentiel entre 2006 et 2016, pour atteindre, à ce jour, les 25 outils évaluatifs (Hilaire-Debove, 2017).

D'après l'Avenant 16 de la Convention Nationale des Orthophonistes datant d'octobre 2017, « le bilan orthophonique constitue un des outils principaux d'information, de coordination et d'amélioration de la qualité de prise en charge, à la disposition des orthophonistes ». Ainsi, l'orthophoniste français a, à sa disposition, une large gamme de matériel pour dresser son bilan orthophonique.

Mais quels éléments prendre en compte pour les différencier et en choisir un particulièrement ?

## **2 Les propriétés d'un test pertinent**

Pour qu'un test soit valide scientifiquement, il doit répondre à une standardisation, des normes, ainsi qu'à un certain nombre de critères psychométriques.

### **2.1 La standardisation**

La standardisation va déterminer les conditions de passation de l'évaluation. Elle est essentielle puisqu'elle cherche à réduire les biais de l'observateur (Marin-Curtoud et al., 2010), en particulier la subjectivité, les erreurs de mesure ainsi que les biais d'interprétation (Lafay & Cattini, 2018). Cela va être rendu possible grâce à des consignes claires énoncées dans le manuel d'utilisation de l'outil évaluatif : ainsi, l'évaluation va être considérée comme objective (Huteau & Lautrey, 2006). Dans leur article, Bouchard et al. (2009), font de la standardisation une « police d'assurance » d'un test. Cette standardisation permettra la reproductibilité de l'évaluation dans le temps et suivant l'examineur (fidélité inter-observation et inter-juges développées plus bas).

### **2.2 Les normes**

La norme est une « distribution des scores obtenus par un échantillon de personnes, représentatif d'une population définie, à un instrument qui a été administré dans des conditions standardisée » (Lafay & Cattini, 2018).

En premier lieu, le manuel du test doit rendre compte de la population d'échantillonnage sélectionnée. D'après les différentes lectures (Ivanova & Hallowell, 2013; Leclerc & Veys, 2014), chaque sous-groupe doit être constitué au minimum de 100 personnes pour avoir une bonne représentation. De plus, chaque individu doit être

décrit précisément : il est nécessaire d'avoir l'âge, le sexe, le niveau scolaire, la répartition géographique, les caractéristiques ethniques ainsi que le niveau socio-culturel de la personne testée (ou des parents si les individus de l'échantillon sont des enfants), cela pour s'assurer que le patient évalué sera représenté dans son sous-groupe (Leclerc & Veys, 2014). En effet, les résultats bruts des participants vont être étalonnés pour qu'ils puissent former des sous-groupes (Huteau & Lautrey, 2006). Cela amène aux tendances centrales, indispensables pour produire l'étalonnage. Il s'agit des moyennes et des écart-types, si la distribution est normale, ou des percentiles si l'étalonnage ne suit pas une courbe de Gauss (Huteau & Lautrey, 2006). La distribution normale est une « distribution théorique d'une variable continue au sein d'une population infinie ». Sa courbe de représentation, en forme de cloche, est symétrique, ce qui signifie que sa distribution l'est également. C'est grâce à cette normalisation qu'un seuil de pathologie peut être fixé à -2 écart-types car il représente moins de 5% de la population totale (Laveault & Grégoire, 2010b).

Chaque sous-groupe constitué va représenter un certain niveau de réussite, le patient testé sera alors comparé, le plus précisément possible, aux individus de ce sous-groupe pour le situer en termes de capacités (Lafay & Cattini, 2018).

Viennent ensuite les critères psychométriques à évaluer.

## **2.3 La validité**

La validité d'un test est son intérêt et sa capacité à mesurer ce pour quoi il est utilisé (Huteau & Lautrey, 2006; Lafay & Cattini, 2018). Différentes formes de validités peuvent être développées pour évaluer un outil de diagnostic.

### **2.3.1 La validité de contenu.**

Elle consiste à évaluer à quel degré les aspects du test sont représentatifs de ce que le test est censé observer (Laveault & Grégoire, 2010a), autrement dit, va être mesurée la pertinence de contenu du test (Lafay & Cattini, 2018). Cela sous-tend que les domaines à évaluer aient été rigoureusement définis en amont. Les épreuves évaluatives choisies devront donc être en adéquation avec le domaine à mesurer (Ivanova & Hallowell, 2013).

### **2.3.2 La validité empirique.**

Aussi appelée validité critérielle, elle rend compte de la corrélation entre tout critère mesurant le même domaine ou caractéristique (Laveault & Grégoire, 2010a), soit une

bonne évaluation par rapport à un critère de référence, qui sera externe au test (Lafay & Cattini, 2018; Marin-Curtoud et al., 2010).

Deux types de validité empirique sont décrits : la validité concomitante, aussi appelée concourante (Leclerc & Veys, 2014), qui présente une forte corrélation entre le score obtenu au test et d'autres outils permettant de mesurer le même concept, soit le critère externe, au même moment de mesure. Si corrélation il y a, les deux mesures auront la même caractéristique (Laveault & Grégoire, 2010a). C'est cette partie de la validité empirique qui sera développée dans la présente étude. Un autre type de validité empirique existe, il s'agit de la validité prédictive, qui est « un indicateur de la performance que l'on cherche à prédire » (Laveault & Grégoire, 2010, p178), en d'autres mots, elle est le reflet de la prédiction du score au test. (Ivanova & Hallowell, 2013). Appliquée au quotidien, Leclerc & Veys (2014), font état d'une corrélation entre la performance dans une activité quotidienne et celle du test.

### **2.3.3 La validité écologique.**

Elle va rendre compte des caractéristiques d'utilisation d'un test pour qu'il soit compatible avec une activité de diagnostic médical. En effet, si la passation d'un test est trop longue, un praticien ne pourra envisager de l'utiliser, même si ce dernier a de bonnes propriétés psychométriques (Betz et al., 2013). De même, la randomisation d'une batterie évaluative, à savoir laisser au praticien le libre choix de l'ordre de présentation des différentes épreuves, qui ne sera donc pas fixé par un ordre pré-établi, lui laisse la possibilité de cibler au mieux les compétences à évaluer, et non de devoir présenter la totalité des épreuves, au risque de connaître par avance les bons résultats de certains domaines. De plus, cela permet de contourner l'effet de fatigabilité, souvent perçu en fin de protocole (Helloin & Thibault, 2006) ainsi que d'éviter les effets inter-épreuves que peut produire une passation en ordre identique (Launay et al., 2018a).

### **2.3.4 La validité théorique.**

Aussi appelée validité conceptuelle par Laveault & Grégoire (2010), ou encore validité de construit. Elle va permettre de rendre compte de la validité d'un test avec une mesure des théories scientifiques, normalement bases de construction du test (Huteau & Lautrey, 2006; Lafay & Cattini, 2018). Cette analyse se doit d'être partagée dans le manuel d'un test puisque c'est d'elle que découlent les objectifs du test. Elle ne se

suffit cependant pas à elle seule puisqu'un modèle théorique sera le reflet d'une théorie particulière, elle doit donc être considérée comme relative (Perdrix, 2018).

Deux autres notions font également partie de la validité, elles garantissent toutes deux le pouvoir diagnostique d'un test (Lafay & Cattini, 2018).

### **2.3.5 La sensibilité.**

Selon Perdrix (2018), la sensibilité « désigne la plus ou moins bonne aptitude d'un outil de discriminer les sujets entre eux ». En effet, la sensibilité va s'occuper de discriminer les individus en difficulté des individus sans difficulté apparente, ces derniers sont appelés les « vrais positifs ». Un test sensible ne devrait pas mettre à jour de « faux positifs ». Il y aura une bonne précision de discrimination, donc une bonne sensibilité, si 90% des individus avec des pathologies connues sont détectés, la précision de discrimination sera moyenne s'il n'y a que 89% et 80% d'individus pathologiques discriminés, et la précision sera inacceptable si moins de 80% des individus ayant une pathologie sont détectés (Friberg, 2010; Plante & Vance, 1994). La sensibilité se mesure dans un test grâce à la présence d'études de cas de sujets pathologiques (Hilaire-Debove, 2017). Les résultats de ces derniers sont comparés à la norme de référence, ils doivent correspondre à des résultats pathologiques pour que le test puisse être caractérisé de sensible (Betz et al., 2013).

Pour ce faire, un test nécessite un certain nombre d'individus dans l'échantillonnage mais aussi un nombre d'items suffisant, et de difficulté croissante, pour déduire les résultats.

### **2.3.6 La spécificité.**

La spécificité est la capacité d'un test à classer une personne saine alors que le test est négatif (Perdrix, 2018), il s'agit alors d'un « vrai négatif ». Peu, voire pas, de « faux positifs » doivent apparaître si un test est dit spécifique. Un seuil de discrimination est établi entre 80 et 95% de personnes saines bien classées. S'il est en-dessous pour un test, ce dernier n'est pas considéré comme spécifique (Lafay & Cattini, 2018).

La validité permet donc à un test de rendre compte de ce qu'il mesure, et plus précisément, qu'il respecte les enjeux fixés lors de la construction du test, que ce soit au travers des apports théoriques ou encore en le comparant aux outils mesurant sensiblement le même domaine. De plus, la sensibilité et la spécificité, classent les



individus en deux catégories : ceux avec des scores normaux et ceux avec des scores pathologiques (Perdrix, 2018).

## **2.4 La fidélité**

La fidélité va renseigner sur la différence qu'il existe entre une note obtenue et la note vraie du test (Laveault & Grégoire, 2010a). Elle va donc regarder la cohérence, la reproductibilité dans le temps ainsi que la stabilité du test. Il existe trois types de fidélité différentes.

### **2.4.1 La fidélité inter-observation.**

Elle est également appelée fidélité temporelle ou fidélité test-retest. Elle permet de mesurer la stabilité de la mesure dans le temps grâce à une administration du même test au même groupe de participants à quelques mois d'intervalle (entre un et trois). Il est évident que les résultats devront être semblables sur ces deux temps d'évaluation, ainsi les résultats ne seront pas considérés comme hasardeux (Lafay & Cattini, 2018). Un coefficient de corrélation est calculé lors de deux examens, le test est considéré comme fiable à partir de 0.70 (Huteau & Lautrey, 2006; Marin-Curtoud et al., 2010), toutefois, un coefficient de 0.90 est recommandé (Leclerc & Veys, 2014).

### **2.4.2 La fidélité inter-observateurs.**

Cette dernière peut aussi se faire appeler fidélité inter-juges (Ivanova & Hallowell, 2013). Elle sous-tend que l'examineur du test ne doit pas avoir de conséquence sur la passation, et que, si celui-ci change, les résultats de l'individu devront se trouver inchangés, ceci sous réserve que les administrateurs du test soient formés de la même façon pour rendre l'évaluation répliquable et fiable. Ici encore, un coefficient peut objectiver les résultats : il ne doit pas être inférieur à 0.60 pour une valeur préférée à 0.90 (Leclerc & Veys, 2014).

### **2.4.3 La cohérence interne.**

Encore appelé fidélité interne (Ivanova & Hallowell, 2013), ou consistance interne (Marin-Curtoud et al., 2010), elle va s'assurer que tous les composants du test réagissent bien entre eux, vers une même réponse. Elle va être mesurée grâce au calcul de l'alpha de Cronbach, la valeur s'étend de 0 à 1, 1 étant la valeur pour laquelle les éléments sont le plus cohérents entre eux, le seuil d'acceptation est fixé à 0.70 (Lafay & Cattini, 2018).

La fidélité vise donc à objectiver les capacités du test, que ce soit au niveau temporel ou de la consistance interne.

Les critères psychométriques précédemment définis remplissent un cahier des charges contraignant qui assure à un outil des bases solides pour permettre un diagnostic orthophonique, puisque le praticien pourra avoir un regard critique sur les résultats qui sortiront d'une passation d'un patient (Spaulding et al., 2006).

« Un outil qui ne présente pas dans son manuel de données quant à ses qualités métrologiques ne devrait en principe jamais être utilisé par un clinicien conséquent » (Perdrix, 2018).

### **3 Les travaux réalisés à l'étranger**

#### **3.1 Outils anglophones**

En 1994, Plante & Vance, étudient les critères psychométriques de 21 tests de langage oral disponibles pour les pays anglophones. La validité prédictive, ou validité empirique, n'apparaît dans aucun test étudié et seuls deux tests décrivent la normalisation de leur échantillon de test. La moitié des tests étudiés décrivent une qualification des testeurs, ainsi qu'une analyse des articles. En revanche, 20 tests mentionnent une description adéquate des consignes de passation et de notation des outils. Seuls 38% des tests de cette étude répondent à la moitié des critères psychométriques, ce qui paraît très peu pour la validation psychométrique d'un test.

En 2017, Denman et al., se concentrent à nouveau sur les bilans de langage oral anglophones. Grâce à une recherche bibliographique, 76 tests ont pu être trouvés mais uniquement 15 tests seront évalués pour leurs données psychométriques. Il résulte de cette étude qu'aucun test sélectionné ne fait état d'étude de la cohérence interne, ou de la sensibilité ou spécificité. Deux tests sur les 15 évalués mentionnent une étude de fidélité, et quatre sur les 15 contiennent une étude de la validité de contenu.

#### **3.2 Outils francophones**

En 2009, Bouchard et al., mènent une revue de littérature dans le but de générer un répertoire d'outils d'évaluation disponible au Canada pour le diagnostic des troubles du langage. 110 tests ont été trouvés puis classés selon le domaine évalué. Une sélection de 31 des outils, sur les 110 trouvés, a été faite pour mener une analyse statistique plus fine. Ces outils utilisés au Canada sont souvent issus d'outils

initialement étalonnés pour la population française, ce qui peut fausser les diagnostics posés outre-Atlantique. Sur 16 critères utilisés pour les comparer entre eux, un test remplit 6.81 critères, soit moins de la moitié. Les critères psychométriques les mieux représentés sont les objectifs des tests clairement annoncés, la validité théorique ainsi que des informations sur l'année de publication et la standardisation de l'échantillon de test. Cependant, certains critères psychométriques comme la fidélité inter-observateurs ou la représentation des extrêmes, ne sont jamais retrouvés dans les tests, l'étude des effets plancher ou plafond est retrouvée dans deux tests uniquement.

Pour poursuivre ce travail et le faire évoluer avec le temps, en 2016, est sorti un nouveau répertoire des outils franco-qubécois pour le diagnostic des troubles du langage et de la parole, étude réalisée par Monetta et al. Cent quarante-sept tests ont été recensés après un sondage fait auprès des orthophonistes concernant leur pratique, ainsi que des recherches bibliographiques. Sur ce nombre important d'outils trouvés, uniquement une petite partie (23) a subi un processus de vérification des données de mesure de normalisation. Cette étude n'a pas porté sur l'étude de la validité ou fidélité des tests sélectionnés. Il en est ressorti que peu de tests franco-qubécois étaient normés ainsi que vérifiés, et que, comme pour la précédente étude, beaucoup étaient finalement des outils créés pour la population française.

En France, Anne Lafay, dans deux études distinctes, a dressé un état des lieux des outils d'évaluation des habiletés mathématiques. Une première en 2014 (Lafay et al.) avec un répertoire exhaustif des outils mathématiques disponibles en fonction des domaines évalués. S'en suit une nouvelle en 2018 (Lafay & Cattini), qui elle, fait une analyse psychométrique de 23 tests en mathématiques, sélectionnés parmi les 31 outils répertoriés. Les critères psychométriques les mieux respectés sont la qualification du testeur, une bonne description des consignes de passation et cotation ainsi que du ou des objectif(s) des tests, une description complète de l'échantillon d'étalonnage et la mention de mesures de tendance centrale. Au contraire, les critères psychométriques les moins respectés sont la validité empirique, la sensibilité et la spécificité de l'outil, la fidélité inter-observations et celle inter-observateurs, ainsi que la cohérence interne.

#### **4 Pistes de travail et problématique**

Suite à ces travaux réalisés dans les pays nord-américains, ainsi qu'en France pour l'évaluation des habiletés mathématiques, la question s'est posée de savoir ce qu'il en

était des critères remplis par les outils destinés aux orthophonistes français pour l'évaluation du langage. L'UNADREO a alors entrepris de dresser un répertoire de ces outils disponibles, en mettant en lumière les différentes données psychométriques nécessaires pour faire d'un test un véritable outil de diagnostic orthophonique. Il est nécessaire de pouvoir comparer la construction psychométrique de chaque test afin que les orthophonistes français puissent utiliser, avec un regard éclairé, le test qui sera au plus près de la réalité clinique du patient. Le but n'est donc pas de dire qu'un test est « bien » ou « pas », mais uniquement de donner une appréciation objective selon certains critères scientifiques reconnus et des intérêts cliniques.

Ainsi, il est possible de se demander quelles sont les données psychométriques présentes dans un manuel de test orthophonique français, et si une évolution est perceptible en comparaison des différentes études menées précédemment.

Le présent travail se concentre sur les outils évaluant le langage écrit, ceux spécialisés dans le langage oral faisant partie d'une étude parallèle menée par l'UNADREO.

## II Méthode

### 1 Échantillon

#### 1.1 Constitution de l'échantillon

Ce mémoire porte sur les tests de langage écrit utilisés lors du bilan orthophonique. L'échantillon est issu d'un inventaire de ces tests, disponibles en France, présenté dans le tableau 1. Cette liste exhaustive a été constituée à partir des sites suivants : ECPA (*Outils d'évaluation langage oral et écrit*, s. d.), Orthoédition (*Evaluations*, s. d.) et Espace orthophonie (*Bilans*, s. d.). Les résultats obtenus ont été confrontés aux différents outils d'évaluation présentés durant les enseignements théoriques et pratiques de la formation initiale reçue à l'université Claude Bernard Lyon 1, lors de la quatrième année d'études.

#### 1.2 Exclusion

Ont été exclus de l'étude les tests publiés avant 2000, jugés comme trop anciens et non représentatifs des courants de pensées actuels, ainsi que ceux se rapportant au dépistage ou à l'évaluation de la dysgraphie. Ce travail de recherche se concentre sur les outils évaluant la lecture et l'orthographe, tant en compréhension qu'en expression, ainsi que les épreuves expliquant les troubles, généralement présentes dans les batteries d'évaluation. Le tableau 1 présente les 40 tests sélectionnés.

**Tableau 1 : Liste des tests**

Nom	Année de parution	Auteur(s)	Éditeur	Cause d'exclusion
<b>INCLUS</b>				
<b>ALOE</b>	2009	M. Plaza, A.M. Robert-Jahier	Adeprio	
<b>Alouette-R</b>	2005	P. Lefavrais	ECPA	
<b>Bale</b>	2010	M. Jacquier-Roux, C. Lequette, G. Pouget, S. Valdois, M. Zorman	Cogni-sciences	
<b>BELO</b>	2006	C. Pech-Georgel, F. George	Boeck Solal	
<b>BLCR</b>	2018	C. Alexandre-Beaumont, C. Delehaye-Boidin, L. Le Hay	Editions ortho et logo	
<b>BLI</b>	2003	A. Khomsi & J. Khomsi	ECPA	
<b>Chronodictées</b>	2006	B. Baneath, C. Boutard et C. Alberti	Orthoédition	
<b>CLéA</b>	2014	F. Pasquet, A. Parbeau-Gueno, E. Bourg	ECPA	
<b>Ecla 16+</b>	2014	C. Gola-Asmussen, C. Lequette, G. Pouget, C. Rouyer, M. Zorman	Cogni-Sciences	
<b>ELFE</b>	2008	C. Lequette, G. Pouget, M. Zorman	Cogni-sciences	
<b>Emilie</b>	2010	A. Duchêne	Happy Neuron	
<b>EPELE</b>	2002	A. Roch Lecours, C. Soum, J.-L. Nespoulous	Orthoédition	
<b>Epcy</b>	2011	A. Zanga	ECPA	

<b>EVALAD</b>	2011	C. Pech-Georgel, F. George	Boeck Solal	
<b>EVALEC</b>	2010	P. Colé, G. Leloup, A. Piquard-Kipffer, L. Sprenger-Charolles	Orthoédition	
<b>EVALéo 6-15</b>	2018	L. Launay, C. Maeder, J. Roustit, M. Touzin	Orthoédition	
<b>Exalang 11-15</b>	2009	M. Lenfant, M.P. Thibault, M.C. Helloin	Happy Neurone	
<b>Exalang 5-8</b>	2010	M.P. Thibault, M.C. Helloin	Happy Neurone	
<b>Exalang 8-11</b>	2012	M.P. Thibault, M.C. Helloin, M. Lenfant	Happy Neurone	
<b>Exalang LyFac</b>	2014	M.P. Thibault, M. Lenfant	Happy Neurone	
<b>Kodjo</b>	2019	C. Poitou	Boeck Solal	
<b>L2MA-2</b>	2010	C. Chevrié-Muller, C. Maillart, A.M. Simon, S. Fournier	ECPA	
<b>La conscience phonologique</b>	2001	D. Delpech, F. George, E. Nok	Boeck Solal	
<b>La forme noire</b>	2010	C. Maeder	Orthoédition	
<b>Le petit poucet</b>	2003	M.F. Arsicaud, C. De Plazaola, F. Gauhier, C. Pech-Georgel	Boeck Solal	
<b>Le vol du PC</b>	2006	C. Boutard, I. Claire et L. Gretchanovski	Orthoédition	
<b>Logator</b>	2009	G. Boudes, C. Boutard	Orthoédition	
<b>Maxence</b>	2013	C. Bessis, F. George, C. Pech-Georgel, R. Sallantin	Boeck Solal	
<b>Morphorem</b>	2012	C. Dufayard, G. Leloup, P. Colé, S. Casalis	Orthoédition	
<b>Orthographe au collège</b>	2008	M.P. Thibault	Orthomotus	
<b>Phonolec</b>	2008	P. Gatignol, M. Oudry, A.M. Robert-Jahier	ADEPRIO	
<b>Phonolec - collège</b>	2011	P. Gatignol, M. Oudry, M. Plaza, A.M. Robert-Jahier	Adeprio	
<b>Profil cognitif et linguistique</b>	2011	Laboratoire de recherche	Gérip	
<b>TCS-r</b>	2019	C. Maeder	Orthoédition	
<b>TCT</b>	2012	S. Cheneau	Mot à mot	
<b>TIME 2</b>	2003	J. Ecalle	Mot à mot	
<b>TIME 3</b>	2006	J. Ecalle	Mot à mot	
<b>Tinfolec</b>	2014	J. Ecalle, P. Revy	Gérip	
<b>TinHaLec</b>	2018	P. Auphan, J. Ecalle, A. Magnan	Adeprio	
<b>Vitesse en lecture</b>	2005	A. Khomsi	ECPA	
<b>EXCLUS</b>				
<b>Analec et la dyslexie</b>	1998	C. Chevrier-Müller, A.M. Simon, S. Fournier	ECPA	Trop ancien
<b>Belec</b>	1994	P. Mousty, J. Leybaert, J. Alegria, A. Content, J. Morais	Labo cognition	Trop ancien
<b>BHK</b>	2004	M. Charles, R. Soppelsa, J.M. Albaret	ECPA	Dysgraphie
<b>BMT-a</b>	2019	C. Billard, A. Mirassou, M. Touzin	Orthoédition	Dépistage
<b>BMT-i</b>	2019	C. Billard, A. Mirassou, M. Touzin	Orthoédition	Dépistage
<b>BP/BL</b>	2000	A. Inizan	EAP	Trop ancien
<b>Dictée des 66 mots</b>	1999	S. Valdois	X	Trop ancien
<b>DOR-LEC</b>	1998	M. Lobrot	X	Trop ancien
<b>Échelle ADE</b>	2014	A. Gavazzi-Eloy	Boeck-Solal	Dysgraphie
<b>ECL-Collège</b>	2005	A. Khomsi, I.Nanty, A.Parbeau-Guéno, F.Pasquet	ECPA	Dépistage
<b>ECOSSE</b>	1999	P. Lecocq	Septentrion	Trop ancien
<b>EDA</b>	2012	C. Billard, M. Touzin	Orthoédition	Dépistage

<b>LMC</b>	1990	A. Khomsi	ECPA	Trop ancien
<b>LMC-R</b>	1999	A. Khomsi	ECPA	Trop ancien
<b>LUM</b>	1999	A. Khomsi	ECPA	Trop ancien
<b>Médial</b>	1995	A. Ouzoulis	X	Trop ancien
<b>NBNO</b>	1992	J.C. Ravard, F. Ravard	X	Trop ancien
<b>Odedys</b>	2005	M. Jacquier-Roux, S. Valdois, M. Zorman	Cogni-sciences	Dépistage

## 2 Grille d'analyse

Ce mémoire a pour objet de dresser un portrait de chaque test de langage écrit français pour, ensuite, les comparer entre eux et donner un regard éclairé aux orthophonistes.

### 2.1 La grille d'évaluation (Annexe A)

#### 2.1.1 *L'intérêt de ce matériel.*

Un groupe de chercheurs pluridisciplinaires internationaux (étude COSMIN), s'est formé en 2005 aux États-Unis pour réfléchir à une meilleure sélection des outils de mesure dans la pratique clinique (Mokkink et al., 2016). Il en résulte une grille d'évaluation pour tous les professionnels utilisant des tests afin d'établir un diagnostic. Cette grille permet un consensus communautaire, pour que tous les outils soient évalués puis élaborés sur une base commune (Mokkink et al., 2010).

L'UNADREO, aux prémices de l'étude des tests en orthophonie, a également établi une grille d'évaluation en se référant à la littérature, dont plusieurs points reprennent ceux de l'étude COSMIN. Cette grille de labellisation, orientée vers l'orthophonie, va permettre d'exploiter au mieux et de façon identique les manuels des différents tests. Cette grille se compose de plusieurs parties reprenant les différents critères nécessaires pour l'évaluation d'un test.

#### 2.1.2 *Les différentes parties de la grille.*

D'après l'étude COSMIN, trois domaines sont nécessaires pour évaluer un outil de mesure : la fidélité, la validité ainsi que la sensibilité (Mokkink et al., 2010; Terwee et al., 2012). Ces trois critères sont également présents dans la grille utilisée pour la présente étude. Elle est complétée par la présentation générale de l'outil : la mention du nom complet, la date de parution, le nom du ou des auteurs. S'insère également la présentation de la normalisation de l'échantillon de population utilisé pour standardiser le test. Les différentes parties sont présentées ci-dessous.

##### 2.1.2.1 *La normalisation.*

La normalisation du test présente la description de l'étalonnage comprenant le nombre de sujets inclus et leur répartition suivant différentes classes (l'âge ou le niveau scolaire entre autres), la représentativité de la population testée avec le quadrillage du territoire national, le niveau socio-culturel ou le niveau d'éducation des sujets inclus dans la population, le sexe, l'âge, ainsi que les critères d'inclusion : ils se doivent d'être représentatifs de la population française générale (Hilaire-Debove, 2017).

#### 2.1.2.2 *La sensibilité.*

Comme décrit plus haut, elle correspond à la bonne discrimination des individus avec une pathologie, soit un « vrai positif ». Elle s'établit grâce à la présence de groupes témoins notamment.

#### 2.1.2.3 *La fidélité.*

Dans la fidélité, vont être présentées la cohérence interne (Mokkink et al., 2010), mesurée par l'alpha de Cronbach, la fidélité inter-observations, ou la mesure de la fiabilité dans le temps d'un test, ainsi que la fidélité inter-observateurs, qui assure la reproduction des résultats suivant l'examineur.

#### 2.1.2.4 *La validité.*

La validité se divise en quatre sous-parties : la validité de contenu, mesure de la pertinence de contenu du test ; la validité empirique, qui mesure la corrélation entre différents outils mesurant le même domaine ; la validité écologique, qui perçoit la compatibilité d'un test pour le diagnostic médical ; et pour finir la validité théorique, base du test car c'est de cette dernière que doit partir la construction d'un test.

La grille se termine par une conclusion et l'intérêt clinique que présente le test étudié.

## **2.2 La fiche finale diffusée aux orthophonistes français (Annexe B)**

Dans l'étude menée par l'UNADREO, tous les tests sont vus par deux groupes distincts qui remplissent, individuellement, une grille d'évaluation par test. Les deux grilles obtenues par test sont ensuite comparées et utilisées pour remplir une fiche-site qui sera publiée sur le site de l'UNADREO à condition que les deux évaluations concordent. Dans le cas contraire, le test est revu par un tiers.

La grille d'analyse est remplie à l'aide des éléments collectés via le manuel et autres publications. Ainsi, seront retrouvés le nom de l'éditeur de l'outil ; les références bibliographiques permettant la construction du test, en particulier son étalonnage ; les



différentes composantes évaluées par l'outil, soit les compétences évaluées pour le langage écrit ici ; la population ciblée pour le test ; une brève description de l'outil ainsi qu'une rapide explication des consignes de passation.

Le but de la création de ces grilles et fiches est de permettre une uniformisation des critères de mesure avec une taxinomie, une terminologie, et une définition de propriétés communes (Mokkink et al., 2010).

### **3 Procédure**

Pour cette étude, l'approche de base sera celle de l'UNADREO, mais une seule grille d'analyse sera réalisée et synthétisée sur une fiche à partir des manuels sélectionnés.

#### **3.1 Complétion de la grille**

Les différentes parties de la grille seront remplies grâce à l'étude du manuel pour chaque test. En effet, une partie « normalisation » est présente dans la plupart des manuels de test, souvent dans les premiers chapitres sous le nom de « population et étalonnage ». Les éléments de sensibilité, fidélité et validité se trouvent, en principe, dans un chapitre détaillant les critères psychométriques de chaque test, ce chapitre peut être absent suivant le test.

#### **3.2 Complétion de la fiche**

La fiche, qui sera diffusée sur le site de l'UNADREO, se remplit à partir de la grille renseignée au préalable. Des informations supplémentaires seront cependant à notifier comme les références bibliographiques, disponibles en fin de manuel, les différentes composantes évaluées seront notées dans la description des différentes activités proposées, tout comme la population-cible du test, la description de l'outil ainsi que l'explication des consignes de passation.

#### **3.3 Comparaison des différents outils étudiés**

À la suite de la rédaction des fiches, les données récoltées ont été centralisées pour être comparées entre elles et ainsi dresser un premier état des lieux des critères psychométriques actuellement utilisés pour la validité des tests orthophoniques sous la forme d'une analyse descriptive.

L'un des objectifs de ce travail, comme énoncé plus haut, est également d'établir une fiche qui sera disponible auprès des professionnels souhaitant connaître les qualités ou faiblesses psychométriques des outils qu'ils utilisent.

### III Résultats

Les grilles et les fiches de chaque test, présentées précédemment, ont été complétées, elles sont disponibles en annexe D. Les conditions sanitaires liées à la pandémie du Covid-19 n'ont pas permis d'étudier certains tests présents dans la liste exhaustive de la méthode. Ainsi, les tests suivants ne figureront pas dans les résultats : le protocole EPELE (Roch-Lecours et al., 2002), Morphorem (Dufayard et al., 2012) et EVALEC (Colé et al., 2010) de chez Orthoédition ; la BLI (Khomsi & Khomsi, 2003), Vitesse en lecture (Khomsi, 2005), Epocy (Zanga, 2011) et L2MA-2 (Chevrié-Muller et al., 2010) des ECPA ; Maxence (Bessis et al., 2013), Le petit poucet (Arsicaud et al., 2003) et EVALAD (Puech-Georgel & George, 2011) de Boeck Solal ; Phonolec (Gatignol et al., 2008), TinHaLec (Auphan et al., 2018) et ALOE (Plaza & Robert-Jahier, 2009) d'Adeprio et BLCR (Alexandre-Beaumont et al., 2018) des éditions ortho et logo.

Vingt-cinq tests sont inclus dans ce travail, y compris le Timé-3 (Ecalte, 2006) et La forme noire (Maeder, 2010), dont les fiches avaient déjà été établies par le groupe de travail de l'UNADREO.

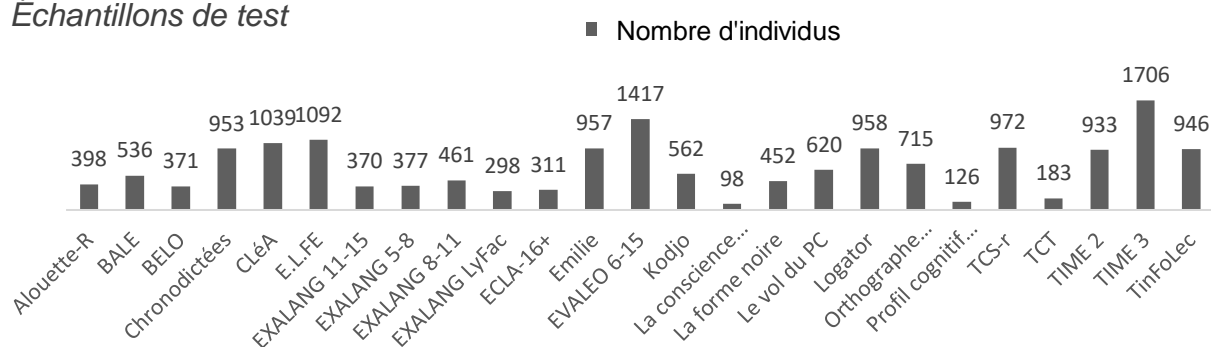
Un tableau récapitulatif des résultats est présent en Annexe C.

## 1 Les conditions de normalisation

### 1.1 L'échantillon de test

**Figure 1**

*Échantillons de test*



L'échantillon de population le plus grand se trouve dans le Timé-3 avec 1706 individus inclus dans la normalisation, le plus petit est dans La conscience phonologique (Delpech et al., 2001), avec 98 sujets.

La répartition des sujets se fait le plus souvent par la classe s'il s'agit d'enfants (dans 92% des tests), ou par le niveau d'études supérieures pour les adultes (dans 8% des

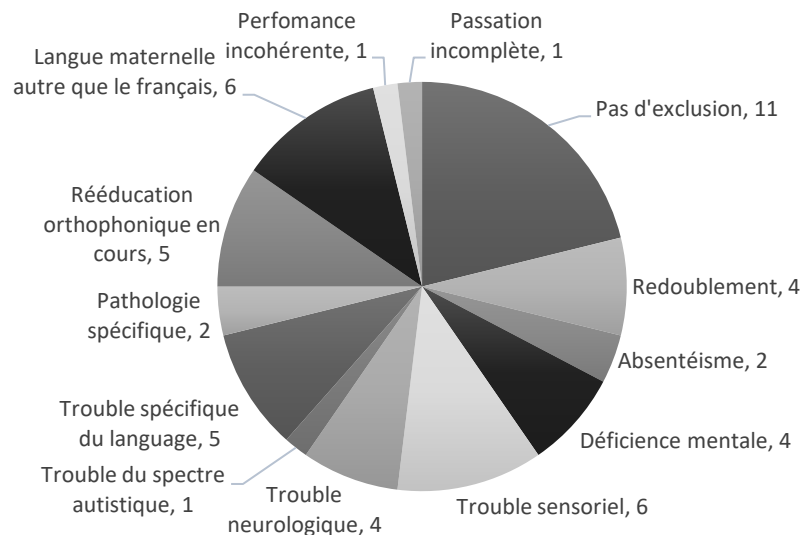
tests), par l'âge (dans 28% des tests), par le sexe (dans 56% des tests), et de façon moins courante par le niveau socio-culturel (dans 12% des tests), et par répartition géographique (dans 4% des tests).

## 1.2 Données démographiques disponibles

### 1.2.1 Critères d'exclusion.

**Figure 2**

*Critères d'exclusion*



L'échantillon des tests est défini, entre autres, par les critères d'exclusion, exposés dans la majorité des manuels (14 sur 25). Six manuels sur les 25 étudiés excluent les sujets dont le français n'est pas la langue maternelle ou principale et ceux ayant des troubles sensoriels. Les patients ayant une rééducation orthophonique et, par conséquent, des troubles du langage sont exclus de cinq manuels sur les 25 étudiés. Quatre manuels sur les 25 étudiés excluent les individus ayant déjà redoublé, les individus ayant une pathologie avérée de déficience mentale ainsi que ceux ayant des troubles neurologiques. Les élèves avec un fort taux d'absentéisme et les sujets avec une pathologie spécifique, telle qu'un handicap, seront exclus de deux manuels. La présence de trouble du spectre autistique, de performance incohérente avec le reste de la population de test ou encore la passation incomplète seront des critères d'exclusion retrouvés indépendamment dans certains manuels.

### 1.2.2 La période de passation.

**Figure 3**

*Présence de la période de passation*

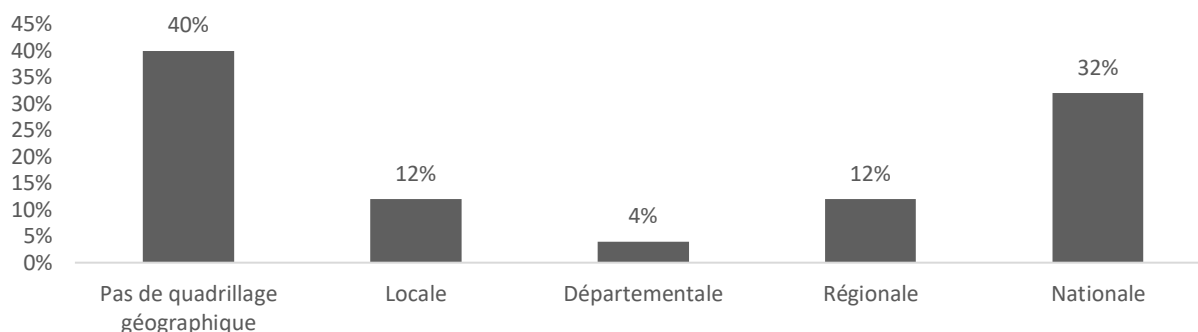


La période de passation des épreuves est mentionnée dans 11 manuels sur les 25 étudiés, soit dans 44% des manuels. Si elle est présente, elle se concentre généralement au milieu de l'année scolaire, soit entre novembre et mars. Peu de tests proposent deux périodes d'étalonnage, seulement le Timé-2 (Ecalte, 2003) qui comprend deux périodes de passation : en septembre-octobre et en janvier-février. 56% des tests (soit 14) ne mentionnent pas la période de passation de l'étalonnage.

### 1.2.3 La répartition géographique.

**Figure 4**

*Description de la répartition géographique*

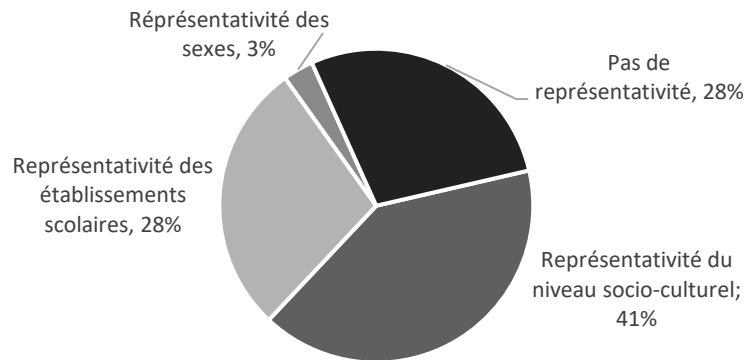


La répartition géographique des différents individus utilisés pour constituer l'échantillon se fait différemment suivant les tests. Elle peut être locale, à savoir à l'échelle d'une ville, ainsi que régionale dans 12% des tests, départementale pour 4% des tests et nationale pour 32% des tests. Un total de 40% des tests, soit 10 sur les 25 étudiés, ne donne aucune information sur le quadrillage géographique.

### 1.2.4 Représentation de la population.

**Figure 5**

*Représentation de la population nationale*

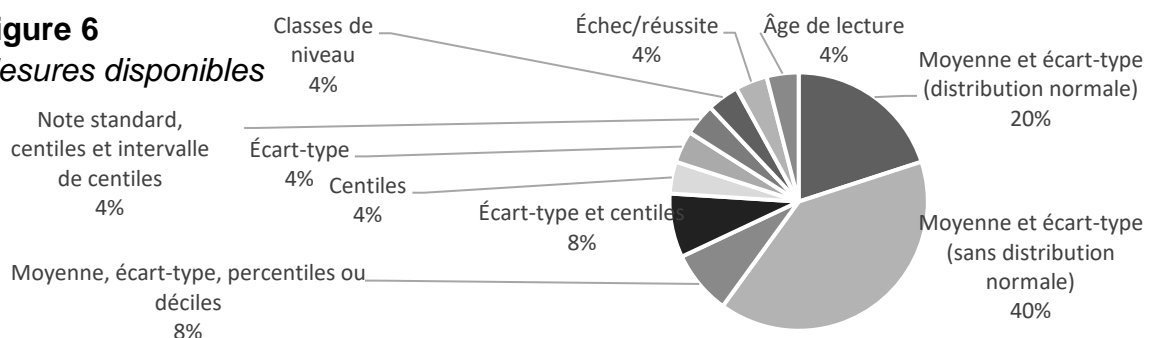


Le niveau socio-culturel de la population française est représentatif dans 41% des tests étudiés. La représentativité des établissements scolaires français, à savoir l'inclusion d'établissements publics et privés, dans des zones rurales ou urbaines, favorisées ou défavorisées et l'inclusion des zones d'éducation prioritaire (ZEP), est citée dans 28 % des tests. La représentativité des sexes est mentionnée dans 3% des tests. Il est à relever qu'Orthographe au collège (Thibault, 2008) mentionne que les élèves dont les parents sont agriculteurs ou cadres supérieurs sont moins représentés que dans la moyenne nationale. De son côté, le TinFoLec (Ecalte & Révy, 2014) révèle un pourcentage des élèves de ZEP inclus dans la normalisation inférieur à la moyenne nationale. Un ensemble de 28% des outils étudiés, soit neuf sur 25, ne présente pas d'éléments analysant la représentativité de la population nationale dans leur échantillon de population.

### 1.3 Les mesures disponibles

**Figure 6**

*Mesures disponibles*



D'après la figure 6, 20% des tests présentent une distribution normale, ce qui leur permet de donner les résultats en moyenne et écart-type. Les scores des patients sont donnés en moyenne et écart-type mais sans qu'il n'y ait eu une normalisation de la distribution, dans 40% des tests. D'autres tests proposent différentes mesures, 4%

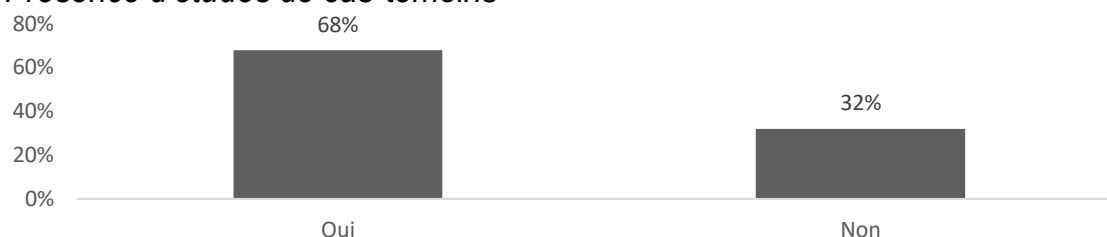
pour chaque mesure, notamment les résultats en centiles (BELO (Pech-Georgel & George, 2006)), en écart-type (TinFoLec), en note standard et intervalles de centiles (CLéA (Pasquet et al., 2014)), en classes de niveau (EVALéo 6-15 (Launay et al., 2018b)), en échec/réussite (Profil cognitif et linguistique (Le Lutin et al., 2011)) ou encore en âge de lecture (Timé-2).

## 2 Sensibilité

### 2.1 Présence d'études de cas-témoins

**Figure 7**

*Présence d'études de cas-témoins*

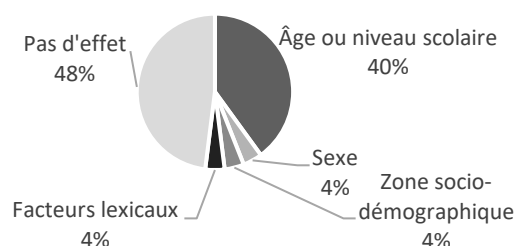


Pour étudier la sensibilité, 68% des tests, soit 17 sur 25, incluent une étude cas-témoins, qui correspond à un sujet ayant une pathologie du langage avérée, dans leur étude. La cohorte la plus importante se situe dans EVALéo 6-15 avec 107 cas-témoins présentés. Il est à noter que le manuel d'Orthographe au collège mentionne une sélection de critères spécifiques dans sa population pour rendre les résultats plus sensibles. Sur les 25 tests analysés, il en reste 8, soit 32%, qui ne mentionnent aucune étude de cas-témoin dans leur recherche.

### 2.2 Les effets constatés

**Figure 8**

*Les effets constatés*



Il est rapporté un effet de l'âge et du niveau scolaire sur les résultats dans 40% des tests étudiés, soit dix sur 25. Un ensemble de test représentant 4% montre un effet du sexe, des zones socio-démographiques ainsi que des facteurs lexicaux. Le Kodjo (Béra & Poitou, 2019) met en avant qu'il n'y a pas d'effet provoqué par le niveau scolaire. Au total, 48% des outils de l'étude ne font aucune mention d'un quelconque effet.

## 2.3 Le seuil pathologique

**Figure 9**

*Présence d'un seuil pathologique*



Sur les 25 tests étudiés, neuf, ou 36%, fixent un seuil de pathologie. Ainsi, pour Exalang 8-11 (Thibault et al., 2012) et Exalang 11-15 (Lenfant et al., 2009), un score au-dessous de -1,65 écart-type (ou au-dessous du percentile 5) est considéré pathologique, La forme noire fixe son seuil pathologique au cinquième percentile, pour CLéA, cela sera sous le percentile 10 et pour l'E.L.FE (Lequette et al., 2008), sous le percentile 15. Le seuil de pathologie pour le test Kodjo est fixé à -1,5 écart-type, celui de Émilie (Duchene, 2010) est à -1 écart-type. Les scores pathologiques d'EVALéo 6-15 se situeront dans la classe 1, et pour Timé-2, un patient aura un score pathologique s'il a un retard supérieur à 18 mois. Une majorité des manuels, 64%, ne mentionnent aucun seuil pathologique.

## 3 Validité

### 3.1 Validité théorique

**Figure 10**

*Présence de la validité théorique*



Sur les 25 outils étudiés, 18, soit 72%, présentent une validité théorique comme base de construction des épreuves. Les théories utilisées sont plus ou moins détaillées. Il est à noter que sept outils étudiés, soit 28% des tests de l'étude, ne mentionnent pas l'approche théorique dans laquelle le test a été construit.

### 3.2 Validité de contenu

**Tableau 2 : Présence de validité de contenu**

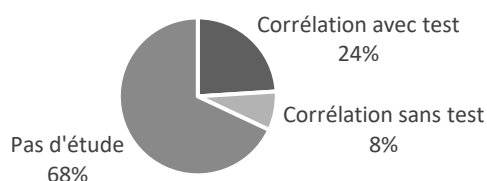
Caractéristiques	Nombre de tests	Pourcentage
Régularité des mots (ou pseudo-mots), leur fréquence et leur longueur	6	20%
Évaluation du temps et de la qualité de lecture	5	17%
Construction des textes en fonction de leur longueur, compréhension des actions, champ lexical pertinent, questionnaire après lecture	4	13%
Évaluation de l'orthographe lexicale, morphosyntaxique et segmentation	4	13%
Utilisation de pseudo-mots	3	10%
Batteries d'évaluation sans détail de construction	2	7%
Nombre de mots lus et erreurs produites	1	3%
Évaluation des compétences lexicales, morphosyntaxiques et narratives	1	3%
Utilisation des graphies complexes	1	3%
Pas de validité de contenu	3	10%

Une très grande majorité des outils évalués dans cette étude présente une validité de contenu, ils sont 22 au total, soit 90% des tests évalués, ils peuvent contenir une seule ou plusieurs caractéristiques. Ils sont 20% à justifier le choix des items par la régularité des mots (et des pseudo-mots s'ils sont présents), leur fréquence et leur longueur. L'évaluation du temps et de la qualité de lecture se retrouve dans 17% des tests étudiés. Pour évaluer la compréhension de texte, 13% des tests ont construit leurs textes en fonction de leur longueur, la compréhension des actions et d'un champ lexical pertinent, ils proposent tous un questionnaire après la lecture du texte. La transcription, évaluée par la dictée, se fera grâce au contrôle de l'orthographe phonétique lexicale, morphosyntaxique et de la segmentation pour 13% des outils. 7% des outils, qui sont des batteries, mentionnent la validité de contenu de leurs épreuves, sans toutefois détailler la construction de chaque épreuve. 3% des tests utiliseront le nombre de mots lus et les erreurs produites, une évaluation des compétences lexicales, morphosyntaxique et narratives ainsi que l'utilisation des graphies complexes. Les trois outils qui ne font pas état de la validité de contenu représentent 10% des outils évalués.

### 3.3 Validité empirique

**Figure 11**

*Étude de la validité empirique*



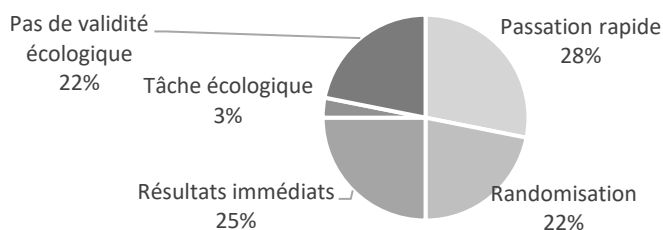


Dans cette étude, 24% des tests rendent compte d'une étude de la validité empirique, soit une corrélation positive avec des tests évaluant les mêmes compétences, qui sont nommés. 8% des manuels font état d'une corrélation positive mais sans citer les tests utilisés pour l'étude. Une grande majorité des manuels étudiés, 17 tests sur 25, ne mentionnent pas d'étude de validité empirique.

### 3.4 Validité écologique

**Figure 12**

*Étude de la validité écologique*



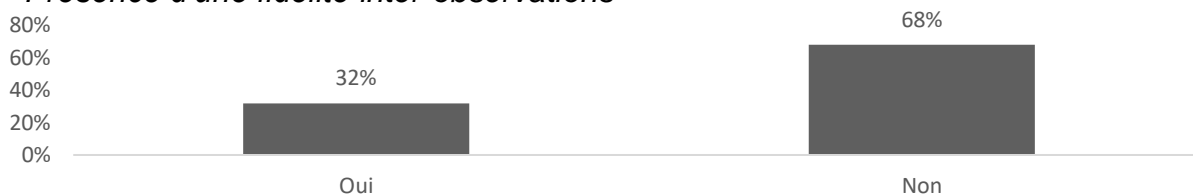
Une grande partie des manuels étudiés fait état d'une validité écologique. Ainsi, 28% des manuels mentionnent une passation rapide, 22% des outils déclarent que la batterie a été randomisée, 25% proposent des résultats disponibles dès la fin de la passation, 4% déclarent que les tâches proposées durant la passation sont proches de celles rencontrées en situation scolaire. Sept manuels, soit 22% des outils étudiés, ne mentionnent aucune validité écologique.

## 4 Fidélité

### 4.1 Fidélité inter-observations

**Figure 13**

*Présence d'une fidélité inter-observations*

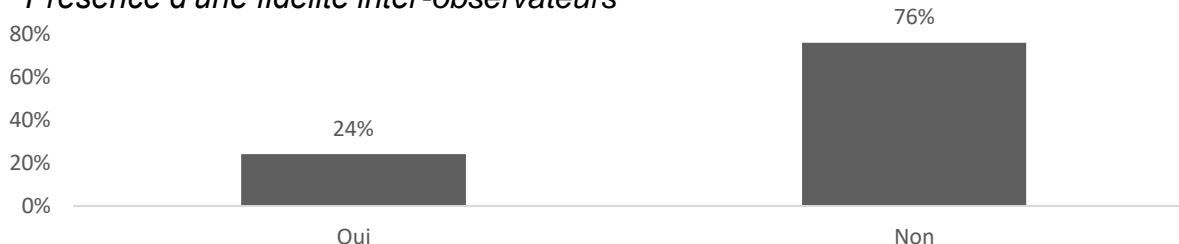


La fidélité inter-observations est citée dans huit tests sur les 25, représentant 32% de l'échantillon étudié. Elle se retrouve notamment dans Chronodictées (Baneath et al., 2006), avec une corrélation auprès de 18 études de cas, dans Exalang 5-8 (Thibault & Helloin, 2010) avec une corrélation comprise entre 0.8 et 0.95 suivant les épreuves, dans Timé-2, dont les coefficients de corrélation sont proches de un avec 49 études de cas, et dans Timé-3 qui présente une mesure sur 26 cas. Aucune analyse de cette fidélité n'est retrouvée pour 68% des outils évalués.

## 4.2 Fidélité inter-observateurs

**Figure 14**

*Présence d'une fidélité inter-observateurs*

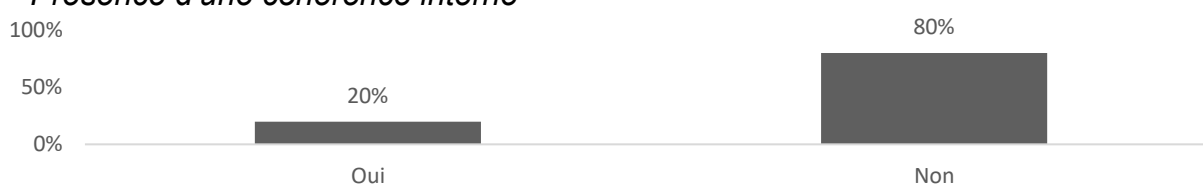


Six tests, soit 24% des outils évalués, comportent la fidélité inter-observateurs : EVALéo 6-15 avec 808 testeurs, Exalang 5-8 avec 22 testeurs, Exalang 8-11 avec 50 testeurs, Exalang 11-15 avec 30 testeurs et Exalang LyFac (Thibault & Lenfant, 2014) avec 23 testeurs. Pour chacun de ces manuels, la différence entre les testeurs n'est pas significative. Le Logator (Boudes & Boutard, 2009) note que, pour l'échantillon adulte, plusieurs orthophonistes libéraux ont servi de testeurs. Les 19 autres tests, soit 76% des outils évalués, ne mentionnent pas d'étude de fidélité inter-observateurs.

## 4.3 Cohérence interne

**Figure 15**

*Présence d'une cohérence interne*



L'analyse de la cohérence interne n'est retrouvée que rarement puisqu'elle est présente dans cinq manuels uniquement, soit dans 20% des tests évalués. Ainsi, EVALéo 6-15, le TCS-r (Maeder, 2019), le Timé-2, le Timé-3 et le TinFoLec relèvent une bonne cohérence interne avec un coefficient alpha de Cronbach supérieur à 0.7, valable pour toutes les modalités du test. Les 21 autres manuels, soit 80% des tests de cette étude, ne font état d'aucune étude de la cohérence interne et donc ne donnent pas de coefficient alpha de Cronbach.

Pour conclure, les différents tests évalués dans ce mémoire montrent des conditions de normalisation respectées, une bonne représentation de la population, des études de sensibilité, de validité théorique, de contenu et écologique assez bien développées. Cependant, peu de seuils pathologiques sont exprimés, la validité empirique n'est que très peu étudiée, tout comme la fidélité, quels que soient ses aspects.

## IV Discussion et conclusion

Ce mémoire a pour objet l'analyse des outils d'évaluation du langage écrit disponibles pour les orthophonistes français à partir d'une grille et d'une fiche, afin d'étudier leur pertinence psychométrique.

### 1 Discussion des résultats

Certains critères vus dans les études menées en amont de ce présent travail, en Amérique du Nord ou en France notamment, se retrouvent dans les tests de langage écrit étudiés. Ces critères, tels que la qualification des testeurs, la description adéquate des consignes de passation ou standardisation, les informations sur l'année de passation, présents dans les études précédentes ne figurent pas dans les critères de notation des tests de cette étude car ces informations existent, aujourd'hui, pour tous les outils d'évaluation et sont visibles sur les fiches.

#### 1.1 Analyse de la normalisation et de la sensibilité

Les résultats présentés en troisième partie, montrent que l'échantillon de normalisation le plus grand se trouve dans le Timé-3 (Ecalte, 2006), le plus petit est celui de La conscience phonologique (Delpech et al., 2001). Alors qu'en 2010, Friberg fixait un seuil de représentativité à 100 individus, une étude menée par Piovesana & Senior en 2018, montre qu'à partir de 85 individus, les moyennes et écart-types, trouvés pour la population de test, sont représentatifs. Ainsi, La conscience phonologique qui n'était pas représentative de la population générale en 2010, se voit finalement entrer dans les attentes d'une normalisation correcte, dont les moyennes et écart-types pourront être utilisés significativement. Les autres outils présentent tous une normalisation représentative de la population, puisque supérieure à 85 individus.

Cette normalisation dépend de plusieurs critères dont les critères d'exclusion : il est possible de penser qu'ils vont permettre de rendre le test plus sensible, donc de déterminer des « vrais positifs » puisque, comme les individus ayant une pathologie ne sont pas comptabilisés, ils seront plus facilement discriminés lors de la passation du bilan. D'après les résultats trouvés, il y a 11 tests sur les 25 étudiés qui ne font état d'aucune exclusion. Cependant, une étude approfondie de la normalisation de ces tests serait nécessaire puisque, la population générale présente entre 3,3% et 24,2% de sujets pathologiques, selon l'environnement socio-économique de l'individu (Fluss et al., 2009), l'exacte représentativité de la population générale résiderait dans le fait

de retrouver une même proportion de sujets pathologiques dans l'échantillon du test. Ainsi, la batterie d'évaluation CLÉA (Pasquet et al., 2014) ne présente pas de critère d'exclusion mais elle a bénéficié d'une étude de sensibilité lors de sa construction avec une inclusion de quatre études de cas, un effet de l'âge ou du niveau scolaire, et un seuil de pathologie, fixé au percentile 10. Ce test prouve que, malgré le manque de critères d'exclusion exprimés pour la normalisation, les résultats d'étude de sensibilité confirment que l'inclusion de sujets pathologiques dans la population du test n'a pas d'impact sur les résultats donnés puisque, si la population incluse dans la normalisation comprend le pourcentage de sujets pathologiques trouvés dans la population générale, elle sera d'autant plus représentative de cette population.

La normalisation d'un outil évaluatif dépend également de la représentation de la population. Ainsi, un test doit avoir une bonne représentation du niveau socio-culturel (McCauley & Swisher, 1984), à savoir que la proportion d'individus dans les différentes classes sociales que comporte le territoire doit se retrouver dans l'échantillon du test. De même, il est de bonne augure de constater une représentation des établissements scolaires correcte : l'inclusion d'établissements privés ou publics, des zones rurales ou urbaines, favorisées ou défavorisées, et des zones d'éducation prioritaires, ainsi qu'une bonne représentation des sexes en fonction de la répartition nationale. Un ensemble de 72% des tests étudiés présente un ou plusieurs critères de représentation de la population, cela montre que ces critères sont nécessaires pour l'établissement d'une bonne normalisation. Peut également être incluse dans la représentation de la population, la répartition géographique des sujets participant à la normalisation. Ainsi, une certaine vigilance sera demandée pour s'assurer que l'outil utilisé présente un bon quadrillage géographique et ne se contente pas d'une petite portion de la population nationale (McCauley & Swisher, 1984). En effet, les résultats scolaires et le vocabulaire des élèves peuvent différer d'une région à l'autre, il est donc plus juste d'inclure toutes les régions françaises, y compris celles d'Outre-mer.

La normalisation, plutôt bien étudiée dans les différents tests de cette étude, est dans la lignée des résultats trouvés par Bouchard et al. (2009) et Lafay & Cattini, (2018). A l'inverse, on trouve une normalisation de l'échantillon mieux étudiée ici que dans l'étude de Plante & Vance, (1994) où seulement deux outils détaillent le nombre de sujets dans leur normalisation ou donnent des exemples de données démographiques disponibles ; ou celle de Monetta et al., (2016) qui reproche aux tests disponibles pour

les orthophonistes canadiens de ne présenter que des normes françaises et non franco-québécoises, ou d'inclure très peu de sujets dans la normalisation.

La suite de cette discussion présente les points psychométriques se concentrant sur la validité et la fidélité, retrouvés dans la plupart des tests, et ceux absents des tests.

## **1.2 Critères psychométriques présents dans la majorité des tests**

Tous les tests étudiés pour cette étude présentent les mesures disponibles pour évaluer le patient. La différence réside dans la normalisation du test qui a été faite lors de sa construction. Comme le mentionnent Laveault & Grégoire (2010), la normalisation de la distribution permet d'utiliser une tendance centrale pour donner les résultats, à savoir les moyennes et écart-types des résultats bruts. Les tests n'ayant pas une distribution normalisée doivent traduire leurs résultats en percentiles. Les résultats de cette présente étude montrent que 40% des tests utilisant les moyennes et écart-types comme unités de mesure n'ont pas subi d'étude de normalisation. Les résultats trouvés pour un patient avec ce type d'outils devront donc être analysés avec prudence puisque leur représentativité est remise en question. Pour pallier cette nécessité de normalisation de la distribution, certains outils comme EVALéo 6-15 (Launay et al., 2018b), le Profil cognitif et linguistique (Le Lutin et al., 2011) et le Timé-2 (Ecalte, 2003), ont pris le parti de rendre compte des résultats de façon plus qualitative en donnant des classes de niveau, des résultats en termes d'échec-réussite ou encore en âge de lecture. Ces mesures permettent de rendre compte de façon rapide et visuelle des résultats d'un patient.

La sensibilité est étudiée par la présence d'études de cas-témoins, visibles dans 68% des tests. Plus cette cohorte sera grande, plus les résultats seront fiables puisque cela signifie que le test discrimine bien les patients avec des pathologies, ce pour quoi il est construit. La sensibilité se traduit également par les effets qui peuvent être constatés lors de la construction de l'outil, présents dans 52% des tests étudiés : une grande majorité de ces tests fait mention d'un effet de l'âge ou du niveau scolaire, ce qui signifie que les résultats des sujets de l'échantillon s'améliorent avec les acquisitions académiques, cela marque une bonne détection des capacités de chaque individu par le test. De façon moins importante, le sexe, la zone socio-démographique dans laquelle le sujet évolue, ainsi que les facteurs lexicaux peuvent être des effets rencontrés dans les tests.

La validité est, de façon générale, plutôt bien présente dans les tests. D'abord, 72% des tests présentent une validité théorique plus ou moins détaillée. Cette validité est sûrement l'une des plus importantes car elle rend compte de la théorie utilisée pour la construction du test (Laveault & Grégoire, 2010a), cela paraît indispensable pour la validité scientifique d'un test. Elle doit s'appuyer sur des théories récentes de préférence. Il est alors naturel de se questionner sur les tests ne présentant pas de validité théorique puisqu'ils ne présentent donc aucune théorie sur laquelle baser leur construction. Deuxièmement, 90% des tests étudiés présentent une validité de contenu, directement dépendante de la validité théorique puisque c'est à partir du domaine défini comme étant à évaluer que les items de passation seront déterminés (Ivanova & Hallowell, 2013). Cette validité mesurée surtout pour les tests ciblant une compétence particulière, à la différence des batteries d'évaluation, doit être au plus proche des éléments théoriques pour évaluer une compétence cible. Ainsi, plusieurs tests inclus dans cette étude présentent les conditions selon lesquelles les items ont émergé : la régularité, la fréquence et la longueur des mots semblent indispensables à maîtriser pour des items couvrant l'évaluation lexicale, tout comme l'évaluation du temps ou de la qualité de lecture, entre autres, sont nécessaires pour l'évaluation de la leximétrie et de l'identification des mots écrits. En fonction du domaine évalué, chaque item aura donc, dans 90% des tests, une justification scientifique du choix de l'auteur. Tout comme pour la validité théorique, les tests ne présentant pas de validité de contenu clairement définie peuvent poser question quant à leur efficacité de diagnostic. Troisièmement, la dernière validité la mieux représentée est la validité écologique. Elle permet de rendre compte de la praticité et de la qualité du support en tant que test diagnostique à l'usage des soignants dans 78% des manuels de tests. Il est à mentionner la passation rapide : un orthophoniste français, devant les nombreuses demandes de bilan peut difficilement accepter une passation de plusieurs heures, des résultats donnés immédiatement, ainsi qu'une randomisation évitant la fatigabilité du patient, et qui permet une sélection pertinente des épreuves proposées en fonction des résultats obtenus aux premières épreuves.

En comparaison avec les différentes études nord-américaines ou françaises, les tests français de langage écrit présentent de meilleurs résultats en ce qui concerne l'étude des validités, en particulier celle de contenu en comparaison avec l'étude menée par

Denman et al., (2017). La sensibilité paraît également mieux étudiée comparée aux études de Denman et al., (2017) ou de Lafay & Cattini, (2018).

### **1.3 Critères psychométriques absents des tests**

Le seuil pathologique est un critère de sensibilité très peu retrouvé dans les tests étudiés, 36% uniquement en mentionnent un. Mais, bien que le seuil pathologique mette sur la voie du diagnostic orthophonique, ce n'est pas le seul élément à devoir entrer en compte dans la pose du diagnostic puisque le patient est à considérer dans l'entièreté des éléments cliniques disponibles (Perdrix, 2018). Il semblerait que l'usage soit d'utiliser le seuil de -2 écart-types, ou le percentile 10, qui semblent arbitraires et qui ne peuvent, en aucun cas, être les seuls juges de la présence ou absence d'une pathologie. Ainsi, un seuil pathologique devrait être propre à chaque outil pour être représentatif de sa significativité (Perdrix, 2018).

Si trois critères de validité sont majoritairement décrits dans les tests inclus, la validité empirique est, elle, sous-représentée pour 68% des tests étudiés. Il n'y a donc pas d'étude de comparaison faite avec des tests déjà existants, et validés, pour ces outils. Ce critère de validité, indissociable de ceux de la validité théorique et de contenu, pose des interrogations quant à la véracité de ces tests dont aucune preuve de corrélation entre leurs résultats et ceux d'un test publié n'est avancée.

La fidélité n'est que très peu représentée, quelle que soit sa forme : 68% des tests étudiés ne présentent pas de fidélité inter-observations, cela interroge sur la signification des résultats d'une deuxième passation du test, comme lors d'un bilan orthophonique de renouvellement, 76% des outils inclus ne mentionnent aucune fidélité inter-observateurs, cela laisse entrevoir un biais de passation suivant l'examineur, et 80% des tests ne présentent pas d'étude de cohérence interne. Ce critère, qui détermine la cohérence des composantes évaluées, permet d'avoir une vision globale de l'outil, cela est remis en question par le très faible taux de cohérence interne exprimé par les tests étudiés.

Pour établir un parallèle avec les études citées précédemment, la fidélité n'a jamais été développée dans les manuels, quel que soit le domaine évalué ou l'année de publication de l'étude. Il est alors possible de se questionner sur la raison de ces absences d'analyses : les calculs et corrélations engagés sont-ils trop complexes lors de l'élaboration d'un test orthophonique ? Seulement, bien que ces critères ne fassent

pas partie de bon nombre de tests, ils sont présents dans des outils tels que Timé-2 (Ecalte, 2003) et le Timé-3 (Ecalte, 2006), qui ont le point commun d'avoir été développés par des psychologues cognitivistes. Ainsi, la technicité de la psychologie pourrait apporter à l'orthophonie la rigueur psychométrique nécessaire à la construction d'un test.

## **2 Limites et perspectives**

La première limite présentée par cette étude est que tous les tests sélectionnés, n'ont pas pu être étudiés, faute de mise à disposition des manuels en raison de la situation sanitaire de la pandémie du Covid-19. Ce travail n'est donc pas exhaustif et tend à être finalisé dans un second temps pour pouvoir rendre compte de l'ensemble des outils disponibles en France pour l'évaluation du langage écrit. La seconde limite est qu'une seule personne a dressé les fiches et a donc analysé les éléments psychométriques des différents manuels. Une confrontation avec une nouvelle fiche faite par un tiers serait d'autant plus objective et proche de la réalité psychométrique du test étudié. De plus, ce travail présente uniquement une analyse descriptive de la validité et de la fidélité des tests. Pour aller plus loin, des analyses de corrélation pourraient mettre en évidence des liens entre les critères psychométriques qui montreraient les éléments nécessaires pour l'acquisition et la vérification des différents critères présentés au fil de cette étude. Pour finir, les fiches, mises en annexe D, permettent aux orthophonistes d'avoir un premier aperçu des propriétés psychométriques des tests étudiés dans ce mémoire.

## **3 Conclusion**

Cette étude dresse un état des lieux des critères de normalisation et psychométriques présents pour la construction des tests de langage écrit français. Les éléments les mieux représentés sont l'étude de la normalisation, de la sensibilité, bien que le seuil de pathologie ne soit pas disponible dans chaque outil, de la validité théorique, socle de la construction d'un outil, de la validité écologique et celle de contenu. La validité empirique, quant à elle, n'est que très peu étudiée, tout comme les différentes formes de la fidélité. L'évaluation des outils disponibles en France permet à tout orthophoniste les utilisant d'avoir un regard objectif sur ces derniers. Grâce aux travaux menés par l'UNADREO, il peut en connaître les limites et les avantages et ainsi garder son esprit critique lors de la passation et de l'analyse des épreuves. Bien que ce travail mette en lumière les critères psychométriques nécessaires à la réalisation d'un outil d'évaluation



pertinent, il n'en faut pas moins négliger l'apport clinique des tests. En effet, même si un test ne remplit que peu de critères psychométriques, l'important est l'analyse qu'en fera le clinicien, même si les résultats ne sont que qualitatifs (Monetta et al., 2016), ils sont pertinents pour la pose du diagnostic ainsi que la prise en soin.

## V Références

- Alexandre-Beaumont, C., Delehay-Boidin, C., & Le Hay, L. (2018). *B.L.C.R. - Batterie de Lecture et de Compréhension de Récit*. Ortho & logo.
- Arsicaud, M.-F., De Plazaola, C., Gautier, F., & Puech-Georgel, C. (2003). *Test d'orthographe—Le petit poucet*. De Boeck Solal.
- Auphan, P., Ecalle, J., & Magnan, A. (2018). *TinHaLec*. Adeprio.
- Baneath, B., Boutard, C., & Albertini, C. (2006). *Chronodictées*. Orthoédition.
- Béra, C., & Poitou, C. (2019). *Kodjo : Évaluer la compréhension en lecture*. De Boeck Supérieur.
- Bessis, C., George, F., Puech-Georgel, C., & Sallantin, R. (2013). *Maxence*. De Boeck Solal.
- Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors Influencing the Selection of Standardized Tests for the Diagnosis of Specific Language Impairment. *Language, Speech, and Hearing Services in Schools, 44*(2), 133-146. [https://doi.org/10.1044/0161-1461\(2012/12-0093\)](https://doi.org/10.1044/0161-1461(2012/12-0093))
- Bilans*. (s. d.). Espace orthophonie. Consulté 8 novembre 2019, à l'adresse <https://www.espace-orthophonie.fr/18-bilans>
- Bouchard, M.-E. G., Fitzpatrick, E. M., & Olds, J. (2009). *Psychometric Analysis of Assessment Tools Used with Francophone Children*. 33(3), 11.
- Boudes, G., & Boutard, C. (2009). *Logator*. Orthoédition.
- Boutard, C., Claire, I., & Gretchanovsky, L. (2006). *Le Vol du PC*. Orthoédition.
- Cheneau, S. (2012). *TCT - Test de Compréhension de Texte*. Mot à mot Édition.
- Chevrié-Muller, C., Maillart, C., Simon, A.-M., & Fournier, S. (2010). *L2MA-2*. ECPA.
- Colé, P., Leloup, G., Piquard-Kipffer, A., & Sprenger-Charolles, L. (2010). *EVALEC*. Orthoédition.
- Delpech, D., George, F., & Nok, E. (2001). *La conscience phonologique*. Solal.
- Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y.-W., & Cordier, R. (2017). Psychometric Properties of Language Assessments for Children Aged 4–12

- Years : A Systematic Review. *Frontiers in Psychology*, 8, 1515.  
<https://doi.org/10.3389/fpsyg.2017.01515>
- Duchene, A. (2010). *EMILIE - Protocole de compréhension d'un texte écrit chez les collégiens*. HappyNeuron.
- Dufayard, C., Leloup, G., Colé, P., & Casalis, S. (2012). *Morphorem*. Orthoédition.
- Ecalle, J. (2003). *Timé-2*. Mot à mot Éditions.
- Ecalle, J. (2006). *Timé-3*. Mot à mot Éditions.
- Ecalle, J., & Révy, P. (2014). *TinFoLec*. Gérip.
- Évaluations*. (s. d.). Orthoédition. Consulté 8 novembre 2019, à l'adresse  
<https://www.orthoedition.com/evaluations/>
- Ferrand, P. (2002). Éditorial. *Rééducation orthophonique*, 212, 3-6.
- Fluss, J., Bertrand, D., Ziegler, J., & Billard, C. (2009). Troubles d'apprentissage de la lecture : Rôle des facteurs cognitifs, comportementaux et socio-économiques. *Développements*, 1(1), 21. <https://doi.org/10.3917/devel.001.0021>
- Friberg, J. C. (2010). Considerations for test selection : How do validity and reliability impact diagnostic decisions? *Child Language Teaching and Therapy*, 26(1), 77-92. <https://doi.org/10.1177/0265659009349972>
- Gatignol, P., Oudry, M., & Robert-Jahier, A.-M. (2008). *Phonolec*. Adeprio.
- Gola-Asmussen, C., Lequette, C., Pouget, G., Rouyer, C., & Zorman, M. (2011). *ECLA-16+—Évaluation des Compétences de Lecture chez l'Adulte de plus de 16 ans*. Cogni-Sciences.
- Helloin, M. C. (2019). Prise en soin des troubles du langage écrit. *L'orthophoniste*, 392, 15-23.
- Helloin, M. C., & Thibault, M.-P. (2006). *Exalang 3-6*. Orthomotus.
- Héral, O. (2007). *L'orthophonie avant l'Orthophonie*. Orthoédition.
- Hilaire-Debove, G. (2017). Pourquoi et comment évaluer les outils d'évaluation en orthophonie. *XVIIèmes Rencontres Internationales d'orthophonie*, 40-62.

- Histoire de l'orthophonie et de la FNO.* (s. d.). Fédération Nationale des Orthophonistes. Consulté 8 avril 2020, à l'adresse <https://www.fno.fr/ressources-diverses/histoire-de-lorthophonie/>
- Huteau, M., & Lautrey, J. (2006). *II. Qu'est-ce qu'un test ?* (p. 20-47). La Découverte ; Cairn.info. <https://www.cairn.info/les-tests-d-intelligence--9782707149992-p-20.htm>
- Ivanova, M. V., & Hallowell, B. (2013). A tutorial on aphasia test development in any language : Key substantive and psychometric considerations. *Aphasiology*, 27(8), 891-920. <https://doi.org/10.1080/02687038.2013.805728>
- Jacquier-Roux, M., Lequette, C., Pouget, G., Valdois, S., & Zorman, M. (2010). *BALE - Batterie Analytique du Langage Ecrit*. Cogni-Sciences.
- Khomsi, A. (2005). *Vitesse en lecture*. ECPA.
- Khomsi, A., & Khomsi, J. (2003). *BLI - Batterie prédictive de l'apprentissage de la lecture et Batterie de lecture*. ECPA.
- Kremer, J.-M., & Lederlé, E. (2016). *L'orthophonie en France* (8ème éd.). PUF.
- Lafay, A., & Cattini, J. (2018). Analyse psychométrique des outils d'évaluation mathématique utilisés auprès des enfants francophones. *Revue canadienne d'orthophonie et d'audiologie*, 42(2), 147-164.
- Lafay, A., Saint-Pierre, M.-C., & Macoir, J. (2014). *L'évaluation des habiletés mathématiques de l'enfant : Inventaire critique des outils disponibles*. 26.
- Launay, L., Maeder, C., Roustit, J., & Touzin, M. (2018a). EVALEO 6-15 – Batterie d'évaluation du langage oral et du langage écrit chez les sujets de 6 à 15 ans. *Rééducation orthophonique*, 273, 93-135.
- Launay, L., Maeder, C., Roustit, J., & Touzin, M. (2018b). *EVALéo 6-15*. Orthoédition.
- Laveault, D., & Grégoire, J. (2010a). La validité des résultats à un test. In *Introduction aux théories des tests, en psychologie et en sciences de l'éducation* (p. 377). De Boeck & Larcier.
- Laveault, D., & Grégoire, J. (2010b). Les scores et leur distribution. In *Introduction aux théories des tests, en psychologie et en sciences de l'éducation*. De Boeck &

Larcier.

Le Lutin, La Villette, Le Liris Lyon 1, & EMC Lyon 2. (2011). *Profil cognitif et linguistique*. Gériop.

Leclerc, A.-L., & Veys, E. (2014). Réflexions sur le choix de tests standardisés lors du diagnostic de dysphasie. *Approche Neuropsychologique des Apprentissages chez l'Enfant*, 26, 374-382.

Lefavrais, P. (2005). *Alouette-R*. ECPA.

Lenfant, M., Thibault, M.-P., & Helloin, M. C. (2009). *Exalang 11-15*. Orthomotus.

Lequette, C., Pouget, G., & Zorman, M. (2008). *E.L.FE - Évaluation de la Lecture en Fluence*. Cogni-Sciences.

Maeder, C. (2010). *La forme noire—Test de compréhension écrite de récits pour enfants de 9 à 12 ans*. Orthoédition.

Maeder, C. (2019). *TCS-r—Test de compréhension Syntaxique révisé*. Orthoédition.

Marin-Curtoud, S., Rousseau, T., & Gatignol, P. (2010). État des lieux sur « le test » Qu'appelle-t-on un test ? Qu'est-ce qu'évaluer ? Du test au testeur... Comment franchir le pas? *L'orthophoniste*, 296, 19-26.

McCauley, R. J., & Swisher, L. (1984). Psychometric Review of Language and Articulation Tests for Preschool Children. *Journal of Speech and Hearing Disorders*, 49(1), 34-42. <https://doi.org/10.1044/jshd.4901.34>

Mokkink, L. B., Prinsen, C. A. C., Bouter, L. M., Vet, H. C. W. de, & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy*, 20(2), 105-113. <https://doi.org/10.1590/bjpt-rbf.2014.0143>

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737-745.

<https://doi.org/10.1016/j.jclinepi.2010.02.006>

Monetta, L., Desmarais, C., MacLeod, A. A. N., St-Pierre, M.-C., Bourgeois-Marcotte, J., & Perron, M. (2016). *Recension des outils franco-qubécois pour l'évaluation des troubles du langage et de la parole*. 40(2), 11.

*Outils d'évaluation langage oral et écrit*. (s. d.). pearson clinical (ECPA). Consulté 8 novembre 2019, à l'adresse <https://www.pearsonclinical.fr/orthophonie/outils-d-evaluation/langage-oral-et-ecrit>

Pasquet, F., Parbeau-Gueno, A., & Bourg, E. (2014). *CLÉA* (Version 1.0.4) [Computer software]. ECPA.

Pech-Georgel, C., & George, F. (2006). *BELO Batterie d'évaluation de lecture et d'orthographe*. Solal.

Perdrix, R. (2018). Pour une contribution éclairée et raisonnée de l'évaluation standardisée et normalisée aux diagnostics de troubles développementaux du langage. Quelques éléments pour choisir, comprendre et exploiter les tests en orthophonie. *Rééducation orthophonique*, 273, 48-70.

Piovesana, A., & Senior, G. (2018). How Small Is Big : Sample Size and Skewness. *Assessment*, 25(6), 793-800. <https://doi.org/10.1177/1073191116669784>

Plante, E., & Vance, R. (1994). Selection of Preschool Language Tests : A Data-Based Approach. *Language, Speech, and Hearing Services in Schools*, 25, 15-24.

Plaza, M., & Robert-Jahier, A.-M. (2009). *ALOE*. Adeprio.

Puech-Georgel, C., & George, F. (2011). *EVALAD*. De Boeck Solal.

Roch-Lecours, A., Soum, C., & Nespoulous, J.-L. (2002). *EPELE*. Orthoédition.

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility Criteria for Language Impairment : Is the Low End of Normal Always Appropriate? *Language, Speech, and Hearing Services in Schools*, 37, 61-72.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M., & de Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties : A scoring system for the COSMIN checklist. *Quality of Life Research*, 21(4), 651-657.

<https://doi.org/10.1007/s11136-011-9960-1>

Thibault, M.-P. (2008). *Orthographe au collège*. Orthomotus.

Thibault, M.-P., & Helloin, M. C. (2010). *Exalang 5-8*. Orthomotus.

Thibault, M.-P., Helloin, M. C., & Lenfant, M. (2012). *Exalang 8-11*. Orthomotus.

Thibault, M.-P., & Lenfant, M. (2014). *Exalang LyFac*. Orthomotus.

Touzin, M. (2002). Évaluation du langage écrit. *Rééducation orthophonique*, 212, 43-51.

Zanga, A. (2011). *Epocy*. ECPA.

### **Textes officiels**

*Avis relatif à l'avenant n° 16 à la convention nationale organisant les rapports entre les orthophonistes et l'assurance maladie, signée le 31 octobre 1996 (JORF n°0251 du 26 octobre 2017, texte n°127).*

## VI Annexe A : Grille de labellisation

### ÉVALUATION TEST en ORTHOPHONIE

Nom de l'outil :

Auteur(s) :

Qualités du test	Avantages	Limites
Standardisation		
<i>Présence des étalonnages, de la stratification</i>		
<i>Présence des variables linguistiques</i>		
<i>Représentativité de la population testée</i>		
<i>Présence de groupes témoins</i>		
Sensibilité (Permet de situer le résultat de chaque sujet dans un groupe de référence / de discriminer finement les sujets)		
Fidélité (Le degré de précision des mesures qu'il permet est évalué)		
<i>Fidélité inter-observations</i>		
<i>Fidélité inter-observateurs</i>		
<i>Cohérence interne</i> <i>Coefficient de l'alpha de Cronbach</i>		
Validité (La signification théorique ou pratique des mesures est précisée)		
<i>Validité de contenu</i>		
<i>Validité empirique / critérielle</i> <i>Analyse de corrélation</i>		
<i>Validité écologique</i>		
<i>Validité théorique</i>		

Conclusion :

Intérêt clinique :



## VII Annexe B : Fiche-site



### Référence du test

*(titre, auteurs, année)*

### Acquisition de l'outil

*(éditeur, téléchargement)*

### Références (articles/études/mémoire)

*(articles/études/mémoire réalisés sur le test et sa validité)*

### Composantes évaluées

### Population ciblée

### Description de l'outil

### Passation

*(matériel et durée)*

### Normalisation

***(Condition de réalisation)***

**Données démographiques disponibles : *(sexe, milieu, etc...)***

**Mesures disponibles : *(moyenne, ET, rangs centiles etc...)***

### Sensibilité

### Validité

- Validité théorique :
- Validité de contenu :
- Validité empirique :
- Validité écologique :

## Fidélité

- *Fidélité inter-observations :*
- *Fidélité inter-observateurs :*
- *Cohérence interne :*

## Conclusion

## Intérêt clinique

# VIII Annexe C : Tableau récapitulatif des résultats

Critères psychométriques	Normalisation					Sensibilité			Validité			Fiabilité		
	Echantillon de population	Critères exclusions	Données de passation	Répartition géographique	Représentation population	Mesures disponibles	Etude de cas	Effets	Théorème de l'écoulement	Empirique	Écologique	Inter-observations	Inter-observateurs	Confiabilité interne
Alouette-R	398 sujets	Non	Non	Non	Moyenne, E.T.	60 cas	Effet âge/nv. scolaire	Oui	Non	Passation rapide	Non	Non	Non	Non
BAE	536 sujets	Non	Non	Non	Moyenne, E.T.	Non	Non	Non	Non	Passation rapide	Non	Non	Non	Non
BEL	371 sujets	Trouble sensoriel (laudif) / Rééducation orthophonique en cours	Non	Marseille	Représentation des établissements scolaires / Niveau socio-culturel	Centiles	Non	Non	Oui	Corrélation positive avec l'Alouette	Non	Non	Non	Non
Chromodictée	953 sujets	Non	Non	Région parisienne	Non	E.T., centile	38 cas	Effet âge/nv. scolaire	Oui	Non	Bonne fiabilité temporelle	Non	Non	Non
CGA	10395 sujets	Non	Non	Non	Notes standard, centiles et intervalles de centiles	P10	Effet âge/nv. scolaire	Oui	Oui	Corrélation positive avec BILIO	Non	Non	Non	Non
ECLA-16+	311 sujets	Trouble sensoriel (laudif) / Performances mixtes avec la population	Non	Non	Moyenne, E.T.	Non	Non	Non	Oui	Non	Non	Non	Non	Non
ELFE	10923 sujets	Non	Non	Isère/Grenoble	Moyenne, E.T., percentiles	1 cas	Non	Non	Non	Non	Non	Non	Non	Non
EVALEO 6-15	14175 sujets	Troubles neurologiques, sensoriels, déficiences intellectuelles, troubles du spectre autistiques et troubles des apprentissages / Rééducation orthophonique	Oui	Oui	7 classes de niveau	107 cas	Effet âge/nv. scolaire	Oui	Oui	Non	Bonne fiabilité temporelle	Bonne fiabilité temporelle	608 testeurs : bonne corrélation	> 0,7 pour toutes les modalités
EXALANG 11-15	370 sujets	Langue maternelle autre que le français / Absentisme / Troubles sensoriels ou déficiences intellectuelles / Redoublements / Rééducation orthophonique	Non	Non	Moyenne, E.T. (distribution normale)	30 cas	Effet âge/nv. scolaire	Oui	Oui	Corrélation positive avec Chronodictée et orthographe au collège	Randomisation	Bonne fiabilité temporelle	30 testeurs : bonne corrélation	Non
EXALANG 5-8	377 sujets	Troubles neurologiques, troubles du langage / Rééducation orthophonique en cours	Oui	Non	Moyenne, E.T. (distribution normale)	Plusieurs cas	Effet âge/nv. scolaire	Oui	Oui	Corrélation positive avec NEEL, Bellec, DMA, Ecoole et UNCL	Randomisation	Bonne fiabilité temporelle	22 testeurs : bonne corrélation	Non
EXALANG 8-11	461 sujets	Langue maternelle autre que le français / Troubles neurologiques, troubles du langage	Non	Non	Moyenne, E.T. (distribution normale)	Plusieurs cas	Effet âge/nv. scolaire	Oui	Oui	Corrélation positive avec d'autres tests	Randomisation	Bonne fiabilité temporelle	50 testeurs : bonne corrélation	Non
EXALANG lycé	298 sujets	Langue maternelle autre que le français / Trouble sensoriel, mental ou trouble du langage	Oui	Non	Moyenne, E.T. (distribution normale)	Plusieurs cas	Non	Non	Oui	Corrélation positive avec d'autres tests	Randomisation	Non	23 testeurs : bonne corrélation	Non
La conscience phonologique	562 sujets	Non	Non	Marseille	Moyenne, E.T.	7 cas	Non	Non	Oui	Non	Non	Non	Non	Non
La forme noire	452 sujets	Langue maternelle autre que le français / Handicap / Rééducation orthophonique en cours	Non	Oui	Moyenne, E.T.	Non	Non	Non	Non	Non	Non	Non	Non	Non
Le vol du PC	620 sujets	Non	Non	Région parisienne et Guadeloupe	Moyenne, E.T.	Non	Non	Non	Oui	Non	Non	Non	Non	Non
Logobor	938 sujets	Non	Non	Paris	Moyenne, E.T.	Plusieurs cas	Non	Non	Oui	Non	Non	Plusieurs testeurs	Non	Non
Orthographe au collège	715 sujets	Langue maternelle autre que le français / Absentisme / Troubles sensoriels ou déficiences intellectuelles / Redoublements	Non	Non	Moyenne, E.T.	Sélection de critères spécifiques	Non	Non	Non	Non	Non	Non	Non	Non
Perfi cognitif et linguistique	126 sujets	Pathologies spécifiques	Non	Non	Échec/réussite	Non	Non	Non	Non	Non	Non	Non	Non	Non
Entile	657 sujets	Langue maternelle autre que le français / Dyslexie ou retard scolaire / Passation non complète	Non	Rhône	Moyenne, E.T.	1 E.T.	Effet âge/nv. scolaire	Oui	Oui	Non	Non	Non	Non	Non
TCS-r	972 sujets	Non	Non	Oui	Moyenne, E.T. (distribution normale)	97 cas	Non	Non	Oui	Non	Bonne fiabilité temporelle	Bonne fiabilité temporelle	Non	> 0,7 pour toutes les modalités
TCT	183 sujets	Dyslexie ou troubles neurologiques	Non	Non	Moyenne, E.T., percentiles	22 cas	Non	Non	Oui	Non	Non	Non	Non	Non
TIME 2	933 sujets	Non	Non	Oui	Age de lecture	1 cas	Non	Non	Oui	Non	Bonne fiabilité temporelle	Bonne fiabilité temporelle	Non	> 0,7 pour toutes les modalités
TIME 3	17064 sujets	Redoublements	Oui	Oui	Moyenne, E.T., dévils	Non	Effet des sexes, ages socio-démographiques, facteurs lexicaux	Oui	Oui	Corrélation positive avec l'Alouette, la pile et le cart et E20	Passation rapide	Bonne fiabilité temporelle	Non	> 0,7 pour toutes les modalités
TimeSalc	946 sujets	Redoublements	Oui	Oui	E.T.	5 cas	Effet âge/nv. scolaire	Non	Non	Non	Randomisation	Non	Non	> 0,7 pour toutes les modalités

## Annexe D : Fiches-site complétées

### Annexe D-a : Fiche de l'Alouette-r



#### Référence du test

Alouette-R (2005)

Pierre LEFAVRAIS

#### Acquisition de l'outil

ECPA

#### Références (articles/études/mémoire)

Non renseignées

#### Composantes évaluées

- La lecture : le temps de lecture, le nombre de mots correctement lus, indice de précision et de vitesse

#### Population ciblée

Enfants entre 6 et 16 ans

#### Description de l'outil

- Manuel
- Planche imprimée illustrée
- Feuille de protocole
- Fiche récapitulative individuelle

#### Passation

Chronométrer et analyser la lecture d'un texte non-signifiant (3 minutes)

#### Normalisation

#### Données démographiques disponibles :

- 398 enfants répartis selon l'âge et le sexe
- Différentes régions françaises sont représentées
- La réévaluation de l'étalonnage date de 2005 (supérieur à 5 ans), la période de passation n'est pas notifiée.

**Mesures disponibles** : calcul des moyennes et écart-types pour déterminer l'âge de lecture

### Sensibilité

- Évolution des indices quantitatifs en fonction de l'âge et du niveau scolaire
- Inclusion de 60 enfants suivis en orthophonie : différence significative entre les scores entre le groupe témoin et la population testée.

### Validité

- *Validité théorique* : modèles théoriques définis (double axe de lecture)
- *Validité de contenu* :
  - o Test de lecture en lisant un texte non signifiant,
  - o Variables utilisées pour l'évaluation de la lecture : temps de lecture, nombre de mots lus, nombres d'erreurs produites, indice de précision et de vitesse de lecture
- *Validité empirique* : non renseignée
- *Validité écologique* : passation très rapide (3 minutes)

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Peu d'étude de validité psychométrique, pas d'étude de fidélité. Cependant, bonne sensibilité apparente.

### Intérêt clinique

Passation rapide, profil de lecteur rapidement établi

## Annexe D-b : Fiche de la BALE



### Référence du test

BALE (Batterie Analytique du Langage Écrit), 2010

Monique JACQUIER-ROUX, Christine LEQUETTE, Guillemette POUGET, Sylviane VALDOIS, Michel ZORMAN

### Acquisition de l'outil

Disponible gratuitement sur Cogni-sciences.

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

- Langage oral
- Lecture : fluence en lecture, liste de mots, conversion graphème-phonème, compréhension écrite
- Orthographe
- Mémoire
- Capacités phonologiques
- Traitements visuels

### Population ciblée

Élève du CE1 au CM2

### Description de l'outil

- Manuel de passation
- Livret de l'élève
- Planches d'épreuves

### Passation

Consignes différentes suivant le domaine évalué : 40 subtests sont présents. Durée de passation entre 2 et 4 heures.

### Normalisation

#### Données démographiques disponibles :

- 536 élèves du CE1 au CM2 ont été évalués puis répartis suivant leur classe :

- 143 enfants au CE1
- 127 enfants au CE2
- 116 enfants au CM1
- 150 enfants au CM2
- L'échantillonnage a été fait entre novembre et février.
- Il date cependant de plus de 5 ans (réalisation en 1999/2000)

**Mesures disponibles** : résultats donnés en moyenne et écart-types, mais pas d'étude de normalité.

### Sensibilité

- Pas d'inclusion d'enfants ayant une pathologie avérée pour le test de sensibilité.
- Pas de seuil explicite de pathologie

### Validité

- *Validité théorique* : développement des critères diagnostiques de la CIM-10, mais très peu de détails sur les théories d'acquisition de la lecture.
- *Validité de contenu* :
  - Les différentes épreuves mesurent bien les compétences voulues :
    - Lecture et orthographe : listes de mots réguliers et irréguliers et des pseudo-mots de haute et basse fréquence, étude de la correspondance graphème-phonème.
- *Validité empirique* : non renseignée
- *Validité écologique* : la batterie est randomisée, cependant, le temps de passation pour toutes les épreuves peut être long : entre 2 et 4 heures.

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Batterie complète mais peu de critères psychométriques sont remplis

### Intérêt clinique

Possibilité de dresser un profil complet des difficultés du patient

## Annexe D-c : Fiche de la BELO



### Référence du test

BELO : batterie d'évaluation de lecture et d'orthographe (2006)

Catherine PECH-GEORGEL et Florence GEORGE

### Acquisition de l'outil

Boeck Solal

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

- La lecture de mots et de texte
- La conscience phonologique
- La mémoire auditivo-verbale
- L'orthographe : dictée de mots et de phrases
- Dénomination rapide (accès au stock lexical)

### Population ciblée

Enfants de CP ou CE1 (5 ans 8 mois à 8 ans 6 mois)

### Description de l'outil

- Cahiers de passation : CP et CE1
- Fiches de protocole
- Livre de l'utilisateur

### Passation

Consigne différente suivant le domaine évalué par la batterie

30 à 45 minutes de passation pour l'ensemble de la batterie

### Normalisation

#### Données démographiques disponibles :

- 371 enfants répartis suivant leur classe, sexe et âge :
  - o 100 enfants en CP 1<sup>er</sup> trimestre,
  - o 69 enfants en CP 2<sup>ème</sup> trimestre,



- 100 enfants en CP 3<sup>ème</sup> trimestre,
- 102 enfants pour le CE1
- Inclusion des écoles de quartiers favorisés et défavorisés, de niveau socio-culturels variés
- Exclusion des enfants avec des troubles auditifs et suivant une rééducation orthophonique
- La passation se fait uniquement à Marseille, donc tout le territoire national n'est pas représenté
- Étalonnage datant de 2006 (supérieur à 5 ans)

**Mesures disponibles** : évaluation par centile (pas de distribution normale donc pas de possibilité de détailler les moyennes et écart-types)

### Sensibilité

- Pas d'étude de sensibilité avec un groupe témoin mentionné.

### Validité

- *Validité théorique* : détail des théories utilisées (modèles de la lecture, de l'orthographe, les pré-requis cognitifs sous-tendant l'apprentissage de l'écrit pour la construction de la batterie.
- *Validité de contenu* : les épreuves proposées testent bien les niveaux de lecture et d'orthographe du patient :
  - Déterminer les stratégies de lecture, évaluation de l'exactitude et de la vitesse de lecture
- *Validité empirique* : comparaison avec l'Alouette : corrélation positive et significative pour tous les niveaux et les épreuves
- *Validité écologique* : 30 à 45 minutes de passation

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Pas d'évaluation de la fidélité mais les autres critères sont bien détaillés

### Intérêt clinique

Évaluation de la lecture et de l'orthographe dans les premières années d'entrée dans l'écrit

## Annexe D-d : Fiche de CLÉA



### Référence du test

CLÉA (2014)

PASQUET Frédéric, PARBEAU-GUENO A., BOURG E.

### Acquisition de l'outil

ECPA, logiciel à installer

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

- Compréhension du langage oral
- Compréhension du langage écrit : lecture de mots ou de phrases ; de texte
- Production du langage oral : phonologie, lexique et morphosyntaxe
- Production du langage écrit : transcription, orthographe lexicale et grammaticale
- Jugement du langage oral et du langage écrit : phonologie, lexique et grammaire

### Population ciblée

Enfants de 2 ans 6 mois à 14 ans 11 mois

### Description de l'outil

- Logiciel avec 12 épreuves :
  - o 4 de compréhension
  - o 4 de production
  - o 4 de jugement

### Passation

Consignes différentes suivant les domaines évalués par la batterie

Pas de durée spécifique, passation en une seule fois.

### Normalisation

**Données démographiques disponibles :**

- 1039 enfants répartis
  - Selon l'âge :
    - 2 ;6 - 2 ;11 ans : 49 enfants
    - 3 - 3 ;5 ans : 67 enfants
    - 3 ;6 – 3 ;11 ans : 86 enfants
    - 4 - 4 ;5 ans : 89 enfants
    - 4 ;6 – 4 ;11 ans : 73 enfants
    - 5 ans : 88 enfants
    - 6 ans : 87 enfants
    - 7 ans : 83 enfants
    - 8 ans : 78 enfants
    - 9 ans : 71 enfants
    - 10 ans : 86 enfants
    - 11 – 12 ans : 94 enfants
    - 13 – 14 ans : 88 enfants
  - Le sexe :
    - 525 filles
    - 514 garçons
  - La classe :
    - PSM : 162 enfants
    - MSM : 153 enfants
    - GSM : 92 enfants
    - CP : 89 enfants
    - CE1 : 77 enfants
    - CE2 : 87 enfants
    - CM1 : 67 enfants
    - CM2 : 83 enfants
    - 6<sup>ème</sup> : 45 enfants
    - 5<sup>ème</sup> : 51 enfants
    - 4<sup>ème</sup> : 38 enfants
    - 3<sup>ème</sup> : 44 enfants
- Il y a également une répartition selon la catégorie socio-professionnelle ainsi que du niveau d'études des parents
- Enfants testés entre novembre et mars.
- Pas d'information sur la répartition géographique des enfants testés
- Échantillon à plus de 5 ans

**Mesures disponibles** : résultats donnés en note standardisées, centiles et intervalles de centiles.

### Sensibilité

- Étude de 4 cas-témoins, bonne discrimination de leurs difficultés.
- Les scores des différentes épreuves (compréhension, production et jugement) augmentent avec l'âge, et ce de façon significative

- Zone pathologique fixée au-dessous du centile 10.

### Validité

- *Validité théorique* : Théories largement développées dans les fondements théoriques du manuel
- *Validité de contenu* : les items mesurent bien ce qu'ils sont censés mesurer, et ce, pour toute la batterie :
  - o Lecture et compréhension : compétences lexicales, morphosyntaxiques et narratives, fluence en lecture
  - o Production orthographique : transcription, orthographe lexicales et grammaticales
- *Validité empirique* : Comparaison avec le test BILO : les résultats sont corrélés en fonction des bonnes réponses et du temps mis pour répondre (qui diminue avec le temps).
- *Validité écologique* : résultats disponibles dès la fin de la passation. Cependant, pas de randomisation de la batterie.

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Bel échantillon de population pour situer un patient.

### Intérêt clinique

Permet de dresser un profil complet de patient.

## Annexe D-e : Fiche de la Conscience phonologique



### Référence du test

La conscience phonologique (2001)

Damienne DELPECH, Florence GEORGE, Evelyne NOK

### Acquisition de l'outil

Solal

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

La conscience phonologique

### Population ciblée

Élèves de CP (soit autour de 6 ans)

### Description de l'outil

- Manuel de passation
- Livret de passation

### Passation

- Différentes épreuves de mémoire : répétition de rythmes, phrases, chiffres
- Décomposition des mots : l'examineur dit un mot, le patient doit le découper en syllabe, puis en phonèmes
- Conscience phonologique : donner des mots qui riment, l'examineur donne des mots, le patient doit supprimer des syllabes puis des phonèmes suivant leur position dans le mot.

### Normalisation

#### Données démographiques disponibles :

- 98 enfants de CP testés (moins de 100 donc pas de représentativité de l'échantillon)
- Différents milieux socio-culturels sont représentés
- Pas d'information sur la période scolaire durant laquelle a été fait le test

- Pas de représentation de l'ensemble du territoire nationale (uniquement Marseille et ses environs)

**Mesures disponibles :** calcul de moyennes et écart-types (mais pas d'étude de normalisation).

### Sensibilité

- Pas d'étude de sensibilité
- Pas de détermination d'une zone pathologique

### Validité

- *Validité théorique* : brèves descriptions des théories comme celles de Morais et al, (1987), Tunmer ou encore Lecocq
- *Validité de contenu* : mesure de la compétence phonologique selon les différentes épreuves proposées :
  - o Utilisation de graphies complexes, de mots réguliers et irréguliers, des phrases courtes à structure syllabique très simple et des phrases plus longues à structure complexe, cela pour la lecture
- *Validité empirique* : non renseignée
- *Validité écologique* : non renseignée

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Moins de 100 individus testés, pas de validité des paramètres psychométriques.

### Intérêt clinique

Test de la conscience phonologique mais elle sera uniquement qualitative au vu de la population d'échantillonnage.

## Annexe D-f : Fiche de Ecla-16+



### Référence du test

ECLA-16+, 2010

Céline Gola-Asmussen, Christine Lequette, Guillemette Pouget, Catherine Rouyer, Michel Zorman

### Acquisition de l'outil

Disponible en téléchargement sur Cogni-sciences

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

- Lecture
- Orthographe
- Mémoire
- Capacités métaphonologiques
- Traitement visuel

### Population ciblée

Jeunes adultes de 16 à 25 ans

### Description de l'outil

- Livret de passation
- Livret consignes
- Manuel de l'utilisateur

### Passation

Consignes différentes suivant les domaines évalués par la batterie.

La passation dure entre 30 et 45 minutes.

### Normalisation

#### Données démographiques disponibles :

- 311 jeunes sont testés
- Répartition selon le sexe :
  - o 153 garçons
  - o 158 filles
- Répartition selon l'établissement fréquenté :
  - o 167 en lycée général et technologique,

- 73 en lycée professionnel,
- 29 en formation d'apprentis,
- 3 en collège,
- 14 en enseignement supérieur
- Représentation de toutes les catégories socio-professionnelles des parents, en pourcentage représentatif de la population française
- Critères d'exclusion : individus porteurs d'une pathologie de l'audition ainsi que ceux dont les performances sont incohérentes par rapport à la population (3 sujets)
- Pas d'informations sur la représentation nationale, ni sur la période de passation
- L'étalonnage date de plus de 5 ans

**Mesures disponibles** : résultats donnés en moyenne et écart-types (mais pas de normalisation)

### Sensibilité

- Inclusion d'un groupe de sujets dyslexiques
- Significativité des différences de résultats entre les individus dyslexiques et les normo-lecteurs (pour l'épreuve de dictée)
- Les sujets dyslexiques sont bien discriminés par le test

### Validité

- *Validité théorique* : de nombreux éléments sur l'apprentissage de la lecture sont présents
- *Validité de contenu* : les items de la batterie mesurent bien les compétences-cibles, en particulier :
  - La liste de lecture de mots qui a été constituée en fonction de la fréquence d'apparition des mots dans la langue, leur caractère régulier, irrégulier, ou de pseudo-mots
- *Validité empirique* : non renseignée
- *Validité écologique* : la batterie n'est pas randomisée

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignées

### Conclusion

L'étalonnage est assez fourni et représentatif de la population ciblée.

### Intérêt clinique

Ce test permet de situer un jeune adulte dans ses capacités de lecture pour éventuellement, poursuivre la construction d'un profil diagnostique.



## Annexe D-g : Fiche de l'E.L.FE



### Référence du test

E.L.FE (Évaluation de la Lecture en Fluence), juin 2008

Christine Lequette, Guillemette Pouget, Michel Zorman

### Acquisition de l'outil

Cogni-sciences, disponible en téléchargement

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

Évaluation rapide du niveau de déchiffrage des élèves

### Population ciblée

Élèves du CE1 à la 5<sup>ème</sup>

### Description de l'outil

- Feuille de cotation
- 2 textes différents à faire lire

### Passation

Passation rapide : lecture d'un texte de compréhension facile pendant 1 minute

### Normalisation

#### Données démographiques disponibles :

- 783 enfants scolarisés du CE1 au CM2 dans les écoles de l'Isère
- 184 élèves de 6<sup>ème</sup> et 125 élèves de 5<sup>ème</sup> scolarisés autour de Grenoble.
- Échantillon d'élèves représentatif des élèves scolarisés (en termes de sexe, pourcentage d'élèves en réseau d'éducation prioritaire, catégorie socio-professionnelle des parents)
- Période de l'année et conditions de passation non précisées

**Mesures disponibles :** Moyenne, écart-type (mais pas d'étude de normalisation) et percentile

### Sensibilité

- Une étude de cas discriminante
- Zone de pathologie en dessous du 15<sup>ème</sup> percentile

### Validité

- *Validité théorique* : non renseignée
- *Validité de contenu* : niveau de compréhension des textes semble adapté aux élèves de la tranche d'âge-cible, mais pas d'information sur ce point
- *Validité empirique* : non renseignée
- *Validité écologique* : passation très rapide

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Test très rapide à faire passer, qui s'apparente plus à une évaluation préliminaire pour une éventuelle recherche de trouble grâce à un outil d'évaluation plus élaboré.

### Intérêt clinique

Vitesse de lecture du sujet évaluée très rapidement, mais ne se suffit pas à lui-même pour un diagnostic.

## Annexe D-h : Fiche de Émilie



### Référence du test

EMILIE, protocole de compréhension d'un texte écrit chez les collégiens, 2010.

Annick DUCHENE

### Acquisition de l'outil

Test disponible sur Happy Neuron

### Références (articles/études/mémoire)

4 mémoires d'orthophonie à Lyon et Besançon :

- Le premier en 2004 : pré-test des élèves de 6<sup>ème</sup>.
- Le second en 2005, par Claire Guyot : premier étalonnage des élèves de 6<sup>ème</sup>, 5<sup>ème</sup> et 4<sup>ème</sup>.
- Le troisième en 2006, par Corinne Monin : comparaison des épreuves des cahiers d'évaluation de l'Éducation Nationale avec le protocole Émilie, pour les élèves de 6<sup>ème</sup>
- Le quatrième en 2007, par Anne-Lise Chalamel et Claire Gonin : détection des mauvais compreneurs

### Composantes évaluées

- Connaissances lexico-sémantiques
- Gestions des liens sémantiques
- Gestion de la cohérence, de la pertinence, des inférences
- La mémoire de travail
- La flexibilité mentale
- La théorie de l'esprit

### Population ciblée

Collégiens : 6<sup>ème</sup>-3<sup>ème</sup>

### Description de l'outil

- Un manuel
- Un livret de passation
- Les planches d'épreuves
- Plateforme en ligne pour le chronomètre et les résultats

### Passation

Faire lire un texte au patient, puis il va répondre à différentes questions (sous forme de QCM ou plus ouvertes) pour évaluer sa compréhension du texte.

Pas de durée spécifique donnée.

## Normalisation

### Données démographiques disponibles :

- 957 enfants répartis selon leur classe :
  - o 347 élèves de 6<sup>ème</sup>
  - o 306 élèves de 5<sup>ème</sup>
  - o 217 élèves de 4<sup>ème</sup>
  - o 87 élèves de 3<sup>ème</sup>
- L'étalonnage date de 2007 (donc est supérieur à 5 ans)
- Inclusion de collèges privé, public, semi-rural ou urbain.
- Différents niveaux socio-culturels sont représentés
- Exclusion des enfants dyslexiques ou en retard scolaire, les enfants dont la langue française n'est pas la langue maternelle, ceux dont la passation n'est pas complète.
- L'échantillon est fait uniquement avec des élèves du Rhône (pas de quadrillage national)
- La période de passation de test n'est pas mentionnée.

**Mesures disponibles :** calcul des moyennes et écart-types (mais pas d'étude de normalisation).

### Sensibilité

- Les résultats s'améliorent avec le niveau scolaire
- Pas d'étude faite avec un groupe témoin
- Zone pathologique fixée à 1 écart-type

### Validité

- *Validité théorique* : beaucoup d'éléments pour la théorie de développement du récit, la pragmatique, la psycholinguistique, la psychologie cognitive et la neuropsychologie
- *Validité de contenu* :
  - o Les différentes épreuves testent bien la compréhension écrite : questions posées après lecture d'un texte, visent les informations explicites, implicites, les distracteurs, les questions importantes et le clou de l'histoire.
  - o Les élèves dont une dyslexie ou des troubles de l'écrit sont évoqués ne répondent pas à ce test, il n'y a donc bien que la compréhension qui est testée.
- *Validité empirique* : non renseignée
- *Validité écologique* : non renseignée

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Bel échantillon de population mais peu de données psychométriques évaluées.

### Intérêt clinique

Test conçu pour évaluer la compréhension en lecture uniquement.

## Annexe D-i : Fiche de EVALéo 6-15



### Référence du test

*EVALéo 6-15 (avril 2018)*

Laurence LAUNAY, Christine MAEDER, Jacques ROUSTIT, Monique TOUZIN

### Acquisition de l'outil

Orthoédition

### Références (articles/études/mémoire)

19 mémoires pour l'étalonnage ainsi que 7 articles entre 2015 et 2018

### Composantes évaluées

- Pour le langage oral: phonétique & phonologie, métaphonologie, lexique & sémantique, métalexique & métaphonologie, morphosyntaxe, métasyntaxe, récit, pragmatique
- Pour le langage écrit: compréhension en lecture, identification de mots en lecture, écriture & graphisme, orthographe, récit
- Autres compétences évaluées: gnosies, visuo-attentionnel, contrôle de l'inhibition, mémoire à court terme, praxies, raisonnement logique

### Population ciblée

Enfants de 6 à 15 ans.

### Description de l'outil

- Un livret utilisateur
- Un livret de consignes et de cotation
- Un livret planche d'épreuves
- Cahiers de passation : langage écrit et langage oral
- Livret d'étalonnage
- Plateforme de résultats en ligne
- Matériel : personnages et objets en 3D, plan, languettes, jetons, cartes, images séquentielles, planche cache

### Passation

Consignes différentes suivant les domaines évalués par la batterie

Pas de durée spécifique

## Normalisation

### Données démographiques disponibles :

- 1417 enfants répartis en 11 classes : CP 1<sup>er</sup> trimestre, CP 3<sup>ème</sup> trimestre, CE1 1<sup>er</sup> trimestre, CE1 3<sup>ème</sup> trimestre, CE2, CM1, CM2, 6<sup>ème</sup>, 5<sup>ème</sup>, 4<sup>ème</sup>, 3<sup>ème</sup>
- Périodes de passation données par le manuel
- Étalonnage ayant moins de 5 ans
- Zones géographiques variées (couverture du territoire national)
- Tous les niveaux socio-économiques sont représentés
- Exclusion des enfants avec des troubles neurologiques, sensoriels, déficience intellectuelle, du spectre autistique, des apprentissages, ou ayant un suivi orthophonique

**Mesures disponibles** : répartition en 7 classes de niveau

## Sensibilité

- 107 sujets pathologiques, avec des troubles du langage oral et écrit sont intégrés à l'étalonnage : bonne discrimination des sujets témoins
- Les scores et le temps évoluent positivement et de façon significative avec l'âge et la classe
- La classe 1 représente la zone pathologique.

## Validité

- *Validité théorique* : de nombreuses références théoriques sont détaillées dans le manuel
- *Validité de contenu* : les items de la batterie mesurent bien les compétences-cibles
- *Validité empirique* : non renseignée
- *Validité écologique* : randomisation de la batterie

## Fidélité

- *Fidélité inter-observations* : deux passations sont espacées de 6 mois pour CP-CE1, bonne corrélation entre les passations
- *Fidélité inter-observateurs* : 808 testeurs, bonne corrélation entre les testeurs
- *Cohérence interne* : tous les items évalués présentent un coefficient alpha de Cronbach supérieur à 0.7

## Conclusion

Batterie récente avec une validation psychométrique conséquente

## Intérêt clinique

Dressage du profil complet du patient avec une seule batterie d'évaluation

## Annexe D-j : Fiche d'Exalang 5-8



### Référence du test

EXALANG 5-8 (2010)

Marie-Pierre THIBAUT, Marie Christel HELLOIN

### Acquisition de l'outil

Précédemment délivré par Orthomotus, maintenant disponible sur la plateforme HappyNeuron

### Références (articles/études/mémoire)

Frey J., Jarry A-S, Exalang 5/8, une batterie informatisée de l'examen du langage oral et écrit chez l'enfant de 5 à 8 ans, mémoire d'orthophonie, Université de Lille, 2003.

Thibault M.-P. - Le bilan orthophonique : entre mesures et interprétations – Exalang 5/8, une batterie d'évaluation du langage oral et écrit. Thèse de doctorat en linguistique, Université de Rouen, 2006.

### Composantes évaluées

- Langage oral
- Traitement visuo-attentionnel
- Phonologie
- Entrées visuelle et auditive
- Mémoire
- Lecture
- Orthographe

### Population ciblée

Enfants entre 5 et 8 ans

### Description de l'outil

- Un logiciel avec 35 items d'évaluation
- Un cahier de passation

### Passation

Consignes différentes suivant les compétences analysées par la batterie.

La passation dure en moyenne 45 minutes, avec un choix d'items ciblés pour le patient.

### Normalisation

#### Données démographiques disponibles :

- 377 élèves répartis
  - o Selon la classe :

- GSM : 95 enfants
- Mi-CP : 80 enfants
- CP - CE1 : 81 enfants
- CE1 – CE2 : 121 enfants
- Selon le sexe :
  - Filles : 194
  - Garçons : 183
- Les périodes de passation sont notifiées dans le manuel
- Inclusion de plusieurs établissements scolaires, publics et privés, répartis en zone rurale, urbaine, favorisée et défavorisée
- Répartition des familles par niveau socio-professionnel
- Pas d'indication du quadrillage national
- Exclusion des troubles neurologiques et troubles sévère du langage en cours de rééducation
- Les passations ont eu lieu en 2002 et 2010, soit depuis plus de 5 ans

**Mesures disponibles :** écart-types, moyennes et percentiles, ainsi que des notes standards (échelle à 5 classes) car la distribution est normale.

### Sensibilité

- Inclusion d'élèves présentant des pathologies du langage écrit et oral, bonne discrimination des patients avec des troubles.
- Un score brut de temps diminue, et un score brut de réussite qui augmentent avec l'âge et la classe
- Pas de seuil pathologique explicité

### Validité

- *Validité théorique* : la batterie est basée sur des théories linguistiques détaillées
- *Validité de contenu* : les items du test évaluent bien les aptitudes considérées :
  - Plusieurs épreuves sont proposées en fonction de la classe de l'enfant
  - Construction des listes suivant la fréquence d'apparition du mot via la lecture, la caractéristique des mots : s'ils sont réguliers ou irréguliers.
- *Validité empirique* : bonne corrélation obtenue avec Exalang 3/6, N-EEL, Belec, L2MA, ECOSSE, LMC-R
- *Validité écologique* : la passation peut se faire en une ou plusieurs fois, chaque item est présenté une seule fois par testing, le retest est possible, il y a un étalonnage par item, et le test est randomisé

### Fidélité

- *Fidélité inter-observations* : les passations sont espacées de 2 à 6 semaines, avec un coefficient de corrélation qui varie de 0.8 à 0.95
- *Fidélité inter-observateurs* : 22 testeurs, pas de différences significatives entre les passations
- *Cohérence interne* : non renseignée

### Conclusion

Très bonne batterie d'évaluation au niveau psychométrique

### Intérêt clinique

Permet de dresser un profil complet de patient.



## Annexe D-k : Fiche d'Exalang 8-11



### Référence du test

EXALANG 8-11 (2012)

Marie-Pierre THIBAUT et Marie-Christel HELLOIN et Mickaël LENFANT

### Acquisition de l'outil

Précédemment délivré par Orthomotus, maintenant disponible sur la plateforme HappyNeuron

### Références (articles/études/mémoire)

ANCELOT J. (2008) : comment évaluer les habiletés pragmatiques cognitives et discursives en compréhension ? Pré-validation d'une épreuve pour adolescents. Mémoire d'orthophonie, Université de Paris, sous la direction de MC Helloin et MP Thibault.

AUDRY L., BENALET Y. (2012) : Pré-validation de neuf épreuves au sein de la batterie Exalang 8-11 ans, évaluant le langage oral, écrit et les compétences transversales, mémoire d'orthophonie, sous la direction de MC Helloin, université de Caen.

BRUNEL P (2011) : Pré-validation d'épreuves de compréhension de phrases, mémoire d'orthophonie sous la direction de MC Helloin, université de Caen.

GUINGANT S., MONTIGNY A., (2006): Le récit oral et écrit : construction d'épreuves de compréhension et de rappel, mémoire d'orthophonie, université de Lille, sous la direction de MC Helloin et MP Thibault.

THIBAUT M.-P., LENFANT M., HELLOIN M.-C. (2006) : Elaboration de certaines épreuves de compréhension au sein de batteries mesurant les capacités langagières chez des enfants entre 3 et 15 ans, Rééducation Orthophonique 227, Isbergues, Ortho-Edition.

### Composantes évaluées

- Attention, phonologie et mémoire
- Lexique
- Langage oral
- Lecture : lecture de mots, de non-mots, leximétrie, compréhension de phrases en images, compréhension de texte

- Orthographe
- Compétences transversales

### Population ciblée

Enfants entre 8 et 11 ans (soit en CE2, CM1 et CM2)

### Description de l'outil

- Un logiciel avec 31 épreuves d'évaluation
- Un livret de passation

### Passation

Consignes différentes suivant les compétences analysées par la batterie.

La passation dure entre 40 et 60 minutes, avec un choix d'items ciblés pour le patient.

### Normalisation

#### Données démographiques disponibles :

- 461 enfants testés répartis
  - o Selon leur sexe :
    - 233 garçons
    - 228 filles
  - o Selon leur classe :
    - CE2 : 105 enfants
    - CM1 : 118 enfants
    - CM2 : 130 enfants
- Période de passation notifiées dans le manuel (étalonnage fait en milieu d'année scolaire : entre février et avril)
- Inclusion de plusieurs établissements scolaires, publics et privés, répartis en zone rurale, urbaine, favorisée et défavorisée
- Répartition des familles par niveau socio-professionnel
- La répartition géographique n'est pas mentionnée
- Exclusion des troubles neurologiques, des troubles sévères spécifiques du langage ainsi que d'enfants dont le français n'est pas la langue majoritaire

**Mesures disponibles :** écart-types, moyennes et percentiles, ainsi que des notes standards (échelle à 5 classes) car la distribution est normale.

### Sensibilité

- Les enfants répertoriés comme présentant des troubles des apprentissages apparaissent bien dans la zone pathologique
- Un score brut de temps diminue, et un score brut de réussite qui augmentent avec l'âge et la classe
- Seuil de pathologie fixé à -1,65 écart-type ou percentile 5.

## Validité

- *Validité théorique* : la batterie basée sur des théories linguistiques évoquées succinctement.
- *Validité de contenu* : les items sont appropriés pour évaluer les aptitudes considérées :
  - o Lecture et orthographe : utilisation de mots fréquents et rares, réguliers et irréguliers
- *Validité empirique* : bonne corrélation avec d'autres épreuves testant les mêmes compétences
- *Validité écologique* : la passation peut se faire en plusieurs fois, la batterie est randomisée, le retest est possible, un étalonnage par item.

## Fidélité

- *Fidélité inter-observations* : corrélation positive et significative, bonne fiabilité temporelle globale
- *Fidélité inter-observateurs* : 50 testeurs différents, corrélation positive très forte, pas de différence notable entre les juges
- *Cohérence interne* : non renseignée

## Conclusion

Très bonne batterie d'évaluation au niveau psychométrique

## Intérêt clinique

Permet de dresser un profil complet de patient.

## Annexe D-I : Fiche d'Exalang 11-15



### Référence du test

EXALANG 11-15 (2009)

Mickaël LENFANT, Marie-Pierre THIBAUT, Marie-Christel HELLOIN

### Acquisition de l'outil

Précédemment délivré par Orthomotus, maintenant disponible sur la plateforme HappyNeurone

### Références (articles/études/mémoire)

ANCELOT J. (2008) : comment évaluer les habiletés pragmatiques cognitives et discursives en compréhension ? Pré-validation d'une épreuve pour adolescents. Mémoire d'orthophonie, Université de Paris, sous la direction de MC Helloin et MP Thibault.

GUINGANT S., MONTIGNY A., (2006) Le récit oral et écrit : construction d'épreuves de compréhension et de rappel, mémoire d'orthophonie, université de Lille, sous la direction de MC Helloin et MP Thibault.

LENFANT M., THIBAUT M.P., HELLOIN M.C. (2006) La compréhension chez les 3-15 ans au travers d'épreuves orales et écrites, Glossa, Isbergues, Ortho-Edition.

TANG-LECLERC M.-N., (2006) Exalang 11/15 : création de quelques épreuves en compréhension orale. Etudes de cas. Mémoire d'orthophonie, université de Lille, sous la direction de MP Thibault et MC Helloin.

THIBAUT M.-P. (2008) Orthographe au collège, 6 épreuves étalonnées pour mesurer l'orthographe, Mont-Saint-Aignan, Motus.

THIBAUT M.-P., LENFANT M., HELLOIN M.-C. (2006) : Elaboration de certaines épreuves de compréhension au sein de batteries mesurant les capacités langagières chez des enfants entre 3 et 15 ans, Rééducation Orthophonique 227, Isbergues, Ortho-Edition.

### Composantes évaluées

- Attention et mémoire

- Lexique et mémoire sémantique
- Lecture : lecture de mots, leximétrie, réduction d'énoncés, lecture-recherche
- Compétences orthographiques
- Compétences langagières orales
- Compétences transversales

### Population ciblée

Enfants de 11 à 15 ans (soit niveau collège, de la 6<sup>ème</sup> au la 3<sup>ème</sup>)

### Description de l'outil

- Un logiciel avec 28 items d'évaluation
- Un cahier de passation

### Passation

Consignes différentes suivant les compétences analysées par la batterie.

La passation dure en moyenne une heure, avec un choix d'items ciblés pour le patient.

### Normalisation

#### Données démographiques disponibles :

- 370 adolescents répartis
  - o Selon leur classe :
    - 6<sup>ème</sup> : 140 élèves
    - 5<sup>ème</sup> : 153 élèves
    - 4<sup>ème</sup> : 166 élèves
    - 3<sup>ème</sup> : 180 élèves
  - o Selon le sexe :
    - Filles : 188
    - Garçons : 182
- Les périodes de passations sont notifiées dans le manuel (entre janvier et mai)
- Inclusion de plusieurs établissements scolaires, publics et privés, répartis en zone rurale, favorisée et défavorisée
- Répartition des familles par niveau socio-professionnel
- Pas d'indication du quadrillage national
- Exclusion des enfants n'ayant pas le français comme langue maternelle, ceux ayant trop d'absentéisme, ceux avec des troubles sensoriels ou intellectuels avérés, des multi-redoublants, des enfants ayant eu des séances d'orthophonie mais dont la rééducation est terminée depuis au moins 3 ans.
- Les passations datent de plus de 5 ans

**Mesures disponibles :** écart-types, moyennes et percentiles, ainsi que des notes standards (échelle à 5 classes) car la distribution est normale.

### Sensibilité

XXX

- Inclusion d'élèves présentant des pathologies du langage écrit et oral (30), bonne discrimination des patients avec des troubles.
- Évolution réussite/temps croissante avec les années mais un effet plafond se dessine avec les années.
- Seuil de pathologie fixée à -1,65 écart-type ou percentile 5

### Validité

- *Validité théorique* : les modèles théoriques utilisés sont cités dans le manuel
- *Validité de contenu* : les items du test sont appropriés pour évaluer les aptitudes considérées
  - o Utilisation de mots réguliers, irréguliers, fréquents et moins fréquents
- *Validité empirique* : bonne corrélation avec d'autres tests ayant pour objectif d'évaluer les mêmes compétences (par exemple Chronodictées, Orthographe au collège, EVAC...)
- *Validité écologique* : la passation peut se faire en une ou plusieurs fois, chaque épreuve est étalonnée, le retest est possible, la batterie est randomisée.

### Fidélité

- *Fidélité inter-observations* : corrélation positive en test/retest, bonne sensibilité temporelle
- *Fidélité inter-observateurs* : 30 testeurs différents, pas de différences significatives entre eux
- *Cohérence interne* : non renseignée

### Conclusion

Très bonne batterie d'évaluation au niveau psychométrique

### Intérêt clinique

Permet de dresser un profil complet de patient.

## Annexe D-m : Fiche d'Exalang LyFac



### Référence du test

EXALANG LyFac (2014)

Marie-Pierre THIBAUT et Mickaël LENFANT

### Acquisition de l'outil

Précédemment délivré par Orthomotus, maintenant disponible sur la plateforme HappyNeuron

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

- Mémoire
- Langage élaboré : flexibilité lexicale, repérage d'anaphores, consignes orales et inférences
- Lecture : lecture de mots, de logatomes, leximétrie, compréhension de texte, repérage
- Orthographe : texte à choix multiple, complétion de phrases, synthèse orthographique, tableaux de synthèse

### Population ciblée

Les adolescents et adultes dès 15 ans

### Description de l'outil

- Un logiciel avec 14 items d'évaluation
- Un cahier de passation

### Passation

Consignes différentes suivant les compétences analysées par la batterie.

La passation dure en moyenne une heure.

### Normalisation

#### Données démographiques disponibles :

- 298 adolescents et jeunes adultes répartis
  - o Selon leur tranche d'âge :
    - <17 : 23%
    - 17-18 : 24%
    - 18-19 : 27%
    - >19 : 26%

- Selon la classe :
  - 2<sup>nde</sup>/1<sup>ère</sup> : 34%
  - Terminale : 31%
  - Bac >= : 35%
- Ils ont été testés au mois de juillet pour l'année de référence
- Quadrillage géographique réalisé avec une bonne représentation socio-culturelle des parents.
- Sont exclus les participants dont la langue maternelle n'est pas le français, avec un trouble sensoriel ou mental ou encore un trouble sévère du langage
- L'échantillonnage est supérieur à 5 ans.

**Mesures disponibles** : résultats donnés en notes brutes (moyennes et écart-types) mais également en notes standard grâce à la distribution normale.

### Sensibilité

- Inclusion des enfants diagnostiqués avec des troubles du langage, et qui sont bien dépistés
- Pas de seuil de pathologie défini

### Validité

- *Validité théorique* : présence des derniers modèles théoriques pour la construction de la batterie (neuropsychologique, linguistiques et pragmatiques)
- *Validité de contenu* : les épreuves du test sont appropriées pour évaluer les aptitudes considérées
- *Validité empirique* : bonne corrélation avec d'autres tests comparables en termes de compétences évaluées
- *Validité écologique* : la batterie est randomisée, la passation se fait en une heure en moyenne. Cependant, la passation ne peut se faire qu'en une seule fois.

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : 23 testeurs différents ont participé à l'étalonnage, il n'y a pas de différence significative entre eux qui est relevée
- *Cohérence interne* : non renseignée

### Conclusion

Très bonne batterie d'évaluation au niveau psychométrique

### Intérêt clinique

Permet de dresser un profil complet de patient de plus de 16 ans (choix de test très limité) pour les accompagner dans une demande d'aménagement scolaire par exemple.



## Annexe D-n : Fiche de Kodjo



### Référence du test

Kodjo : évaluer la compréhension en lecture (octobre 2019)

Céline BERA et Caroline POITOU

### Acquisition de l'outil

De Boeck supérieur (ouvrage imprimé et disponible sur internet)

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

La compréhension de la lecture

### Population ciblée

- Kodjo 1 : élèves du CE2 au CM2
- Kodjo 2 : collégiens de la 6<sup>ème</sup> à la 3<sup>ème</sup>

### Description de l'outil

- Un manuel de passation
- 7 planches pour les textes à lire (2 différents en fonction du niveau scolaire)
- Un logiciel

### Passation

Faire lire un texte au patient, puis il doit répondre aux différentes questions à l'écrit. L'épreuve est chronométrée, et ne doit pas dépasser 40 minutes.

### Normalisation

#### Données démographiques disponibles :

- 562 enfants au total
  - o 335 en cycle II, répartis selon leur classe
    - CE2 : 146 élèves
    - CM1 : 86 élèves
    - CM2 : 103 élèves
  - o 227 au collège, répartis selon leur classe
    - 6<sup>ème</sup> : 64 élèves

- 5<sup>ème</sup> : 54 élèves
- 4<sup>ème</sup> : 45 élèves
- 3<sup>ème</sup> : 64 élèves
- Étalonnage ayant moins de 5 ans
- Il n'y a aucune information sur la période scolaire de passation, la répartition géographique ou encore du niveau socio-culturel des familles des enfants constituant cet échantillon de population.

**Mesures disponibles :** résultats en moyenne et écart-types (mais pas de normalisation)

### Sensibilité

- Étude de 7 cas cliniques : patients ayant des troubles du langage écrit diagnostiqués, bonne discrimination lors de la passation du test.
- Seuil de pathologie fixé à -1,5 écart-type

### Validité

- *Validité théorique* : quelques explications des processus de lecture mais de façon très succincte.
- *Validité de contenu* : la lecture du texte ainsi que les questions qui en découlent évaluent les compétences de compréhension en lecture.
  - Textes construits selon sa longueur, le déroulement des actions qui permet une bonne compréhension, un champ lexical pertinent (vocabulaire choisi selon la fréquence des mots)
  - Pour Kodjo 2 : évaluation par les questions des connaissances lexicales, des micro et macroprocessus de lecture.
- *Validité empirique* : non renseignée.
- *Validité écologique* : non renseignée.

### Fidélité

- *Fidélité inter-observations* : non renseignée.
- *Fidélité inter-observateurs* : non renseignée.
- *Cohérence interne* : non renseignée.

### Conclusion

Peu de précisions statistiques (fidélité et validité), l'échantillon de population n'est pas décrit donc on peut se questionner sur la représentativité de l'échantillon utilisé pour l'étalonnage.

### Intérêt clinique

Évaluation unique de la compréhension en lecture mais peu de représentativité statistique.

## Fiche D-o : Fiche de Logator



### Référence du test

Logator (2009)

Géraldine BOUDES, Corinne BOUTARD

### Acquisition de l'outil

Orthoédition

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

- Utilisation de la voie phonologique
- Utilisation de la voie lexicale

### Population ciblée

Dès le début de l'apprentissage de la lecture (enfants à partir de 6 ans et adultes).

### Description de l'outil

Cahier de cotation et de passation, épreuves très rapides (de 10 à 15 minutes pour les jeunes lecteurs, sinon moins de 5 minutes).

### Passation

Première épreuve de lecture de 60 logatomes, puis dictée de 20 logatomes.

### Normalisation

#### Données démographiques disponibles :

- 958 sujets répartis
  - o Selon la classe (pour les enfants) :
    - CE1 : 78 élèves
    - CE2 : 63 élèves
    - CM1 : 69 élèves
    - CM2 : 86 élèves
    - 6<sup>ème</sup> : 61 élèves
    - 5<sup>ème</sup> : 48 élèves

- 4<sup>ème</sup> : 53 élèves
- 3<sup>ème</sup> : 51 élèves
- 2<sup>nde</sup> : 49 élèves
- 1<sup>ère</sup> : 89 élèves
- Terminale : 103 élèves
- Selon l'âge (pour les enfants) :
  - 7 ans : 76 enfants
  - 8 ans : 60 enfants
  - 9 ans : 69 enfants
  - 10 ans : 88 enfants
  - 11 ans : 56 enfants
  - 12 ans : 48 enfants
  - 13 ans : 52 enfants
  - 14 ans : 54 enfants
  - 15 ans : 52 enfants
  - 16 ans : 61 enfants
  - 17 ans : 96 enfants
  - 18 ans et plus : 245 sujets
- Selon le sexe :
  - Féminin : 544
  - Masculin : 414
- Selon le niveau d'études pour les adultes :
  - Certificat d'études : 20 sujets
  - BEPC : 18 sujets
  - CAP : 29 sujets
  - BEP/BTA : 17 sujets
  - BAC : 31 sujets
  - BAC+2 et plus : 93 sujets
- Selon le niveau socioprofessionnel des adultes :
  - 1 (agriculteurs exploitants) : 16 sujets
  - 2 (artisans, commerçants, chefs d'entreprise) : 19 sujets
  - 3 (cadres et professions intellectuelles supérieures) : 23 sujets
  - 4 (professions intermédiaires) : 47 sujets
  - 5 (employés) : 84 sujets
  - 6 (ouvriers) : 19 sujets
- Pour les enfants : sujets de Paris intra-muros uniquement
- Pas d'information sur la période de passation, étude datant de 2009 (supérieure à 5 ans)

**Mesures disponibles** : calcul de moyennes et d'écart-types (mais pas de normalisation).

### Sensibilité

- Analyse descriptive avec un calcul de moyenne et d'écart-type.

- Analyse comparative avec un groupe-témoins ayant une pathologie développementale du langage écrit : moyennes et écart-types significativement inférieurs à ceux de la population d'étalonnage.
- Pas de seuil de pathologie défini

### Validité

- *Validité théorique* : quelques notions théoriques sur le modèle de la double voie, des confrontations théoriques.
- *Validité de contenu* : utilisation de pseudo-mots pour tester la voie phonologique mais les difficultés peuvent aussi être le reflet de troubles attentionnels.
- *Validité empirique* : Non renseignée
- *Validité écologique* : passation très rapide (moins de 5 minutes)

### Fidélité

- *Fidélité inter-observations* : Non renseignée
- *Fidélité inter-observateurs* : protocole passé par plusieurs orthophonistes en libéral pour l'échantillonnage de population adulte.
- *Cohérence interne* : Non renseignée

### Conclusion

Épreuve très rapide à faire passer, épreuve assez discriminante mais attention à bien mesurer la voie phonologique et ne pas être parasité par un trouble attentionnel.

### Intérêt clinique

Étalonnage adulte, vision très rapide d'un possible déficit de la voie phonologique.

## Annexe D-q : Fiche d'Orthographe au collège



### Référence du test

Orthographe au collège (2008)

Marie-Pierre THIBAUT

### Acquisition de l'outil

Orthomotus

### Références (articles/études/mémoire)

Non renseignés

### Composantes évaluées

- Phonologie
- Orthographe lexicale
- Orthographe grammaticale
- Compréhension après lecture de texte

### Population ciblée

Élèves de collège (6<sup>ème</sup>-3<sup>ème</sup>)

### Description de l'outil

Manuel de passation : textes à dicter, texte à lire et questions de compréhension

### Passation

- 3 dictées différentes
- Une épreuve de closure de texte
- Une épreuve d'orthographe grammaticale
- Une épreuve de lecture et de compréhension de texte

### Normalisation

#### Données démographiques disponibles :

- 715 élèves répartis
  - o Selon le sexe :

- Filles : 350 sujets
- Garçons : 365 sujets
- Selon le niveau scolaire :
  - 6<sup>ème</sup> : 319 sujets
  - 5<sup>ème</sup> : 197 sujets
  - 4<sup>ème</sup> : 107 sujets
  - 3<sup>ème</sup> : 92 sujets
- Les épreuves sont passées en milieu d'année scolaire.
- Pas d'étalonnage fait, ni de notion de répartition géographique
- Épreuves passées en 2008 (donc supérieur à 5 ans)
- Critères d'exclusion : la langue maternelle n'est pas le français, les enfants ayant un fort taux d'absentéisme, les enfants ayant des troubles sensoriels ou intellectuels avérés, les redoublants.
- Échantillon non représentatif dans les catégories socio-professionnelles des parents : agriculteurs et cadres supérieurs (étudiants et retraités sous-représentés mais logique suivant l'âge de la population-cible).

**Mesures disponibles** : résultats sous forme de moyenne et écart-types, mais pas de différence significative des moyennes.

### Sensibilité

- Mise en place de critères spécifiques car peu de sensibilité des résultats (35% des élèves avaient des troubles des apprentissages suivant les premières mesures)
- Pas de seuil de pathologie défini

### Validité

- *Validité théorique* : non renseignée
- *Validité de contenu* : test de l'orthographe grâce aux dictées, cotation suivant plusieurs critères de phonologie, orthographe lexicale et grammaticale
- *Validité empirique* : non renseignée
- *Validité écologique* : non renseignée

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Test très peu développé, servant plus de « dépistage » des troubles des apprentissages.

### Intérêt clinique

Intérêt limité car peu de sensibilité.

## Annexe D-r : Fiche du Profil cognitif et linguistique



### Référence du test

Profil cognitif et linguistique (2011)

3 laboratoires de recherche : la Villette, Lyon1 et Lyon 2

### Acquisition de l'outil

Disponible sur la plateforme Gérip

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

Évaluation du profil cognitif et linguistique par :

- Perception
- Attention
- Mémoire
- Fonctions visuo-spatiales
- Raisonnement logique
- Langage oral
- Langage écrit : lecture flash, lecture de mots, conversion grapho-phonémique, lecture de phrases, sémantique, compréhension de textes, orthographe grammaticale, orthographe lexicale
- Compétences transversales

### Population ciblée

Élèves de 8 à 18 ans

### Description de l'outil

- Logiciel avec plusieurs tâches différentes

### Passation

- Chaque domaine est divisé en plusieurs sous-domaines et chaque sous-domaine contient plusieurs exercices.
- Il y a différents niveaux d'évaluation en fonction du temps dont on dispose pour la passation et en fonction de la finesse du profil que l'on veut dresser (chaque épreuve est déclinée en 9 items de difficulté graduelle).



- Le temps de passation varie entre 10 et 30 minutes (selon la difficulté choisie).

## Normalisation

### Données démographiques disponibles :

- 126 sujets ont été évalués puis répartis
  - o Selon leur classe :
    - CE1 : 23 sujets
    - CE2 : 55 sujets
    - CM1 : 48 sujets
- L'échantillonnage est supérieur à 5 ans
- Pas d'information sur la période scolaire durant laquelle le test a été fait
- Pas d'information sur la répartition géographique des patients testés
- Exclusion des patients ayant des pathologies spécifiques.

**Mesures disponibles :** les résultats donnés sont décrits comme « échoué » ou « réussi »

## Sensibilité

- Utilisation de patients suivis en orthophonie (surtout pour l'échantillon des adolescents), mais pas d'information sur leurs résultats
- Pas de seuil de pathologie défini

## Validité

- *Validité théorique* : non renseignée
- *Validité de contenu* : les différents exercices proposés évaluent bien les champs de compétences cible mais pas plus de détails.
- *Validité empirique* : non renseignée
- *Validité écologique* : différents niveaux d'évaluation en fonction du temps de passation nécessaire et de la finesse du profil à dresser (chaque item est décliné en 9 items de difficulté graduelle)

## Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

## Conclusion

Peu d'étude psychométrique détaillant la validité, la fidélité et la sensibilité sont décrites dans le manuel. L'échantillon de population n'est pas très conséquent.

## Intérêt clinique

Ce bilan permet de dresser un profil assez complet en un temps très acceptable (au plus 30 minutes). L'évaluation du langage écrit ne prend pas en compte l'écriture.

## Annexe D-s : fiche TCS-r



### Référence du test

Test de Compréhension Syntaxique Révisé (TCS-r), 2019

Christine MAEDER

### Acquisition de l'outil

Orthoéditions

### Références (articles/études/mémoire)

### Composantes évaluées

- Compréhension syntaxique à l'oral
- Compréhension syntaxique à l'écrit

### Population ciblée

Enfants du CM1 à la terminale (10-18 ans)

### Description de l'outil

Un livret de consignes, un feuillet de cotation version orale et un feuillet version écrite.

### Passation

3 épreuves dans chaque version :

- Désignation d'images
- Appariement d'énoncé
- Mime par manipulation

Les épreuves seront chronométrées.

### Normalisation

### Données démographiques disponibles :

- 972 participants répartis
  - o Selon le niveau scolaire :
    - CM1 : 115 sujets
    - CM2 : 137 sujets
    - 6<sup>ème</sup> : 110 sujets
    - 5<sup>ème</sup> : 122 sujets

- 4<sup>ème</sup> : 102 sujets
- 3<sup>ème</sup> : 94 sujets
- 2<sup>nde</sup> : 92 sujets
- 1<sup>ère</sup> : 79 sujets
- Terminale : 121 sujets
- Selon le niveau d'études parentales
- Échantillon datant de 2019 (a moins de 5 ans), réparti sur l'ensemble du territoire national
- La période de passation n'est pas mentionnée

**Mesures disponibles :** calcul des moyennes et écart-types avec un échantillon normé.

### Sensibilité

- Différence significative entre les groupes « trouble du langage » (échantillon de 97 enfants présentant des troubles du langage) et « typique ».
- Pas de seuil de pathologie défini

### Validité

- *Validité théorique* : pas de théorie linguistique développée, uniquement la théorie psychométrique.
- *Validité de contenu* : test de la compréhension selon différentes modalités (désignation, appariement et mime)
- *Validité empirique* : non renseignée
- *Validité écologique* : non renseignée

### Fidélité

- *Fidélité inter-observations* : la stabilité des résultats est satisfaisante
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : toutes les modalités ont un coefficient de Cronbach supérieur à 0.7 (excepté la désignation d'images avec la présentation oral et écrite)

### Conclusion

Bonne représentation de la population de référence, ainsi qu'une bonne sensibilité.

### Intérêt clinique

Test qui se concentre uniquement sur la compréhension.

## Annexe D-t : Fiche TCT



### Référence du test

TCT – Test de Compréhension de Texte, 2012

Sophie CHESNEAU

### Acquisition de l'outil

Disponible chez Mot-à-Mot

### Références (articles/études/mémoire)

Thèse de doctorat de l'auteur pour la première phase d'étalonnage

### Composantes évaluées

Évaluation de la compréhension de texte grâce à la lecture de textes (chronométrée) puis à des questionnaires qui en découlent.

### Population ciblée

Adultes cérébro-lésés de 16 à 80 ans

### Description de l'outil

- Un livret de passation avec le cadre théorique, les consignes de passations et de notation ainsi que les tableaux des normes
- Des planches de passation avec les différents textes

### Passation

- Il est demandé au patient de lire 3 textes et de répondre à un questionnaire associé
- Passation assez rapide : entre 10 et 15 minutes

### Normalisation

#### Données démographiques disponibles :

- Échantillon de 183 sujets de 16 à 80 ans, répartis en deux échantillons suivant le niveau d'études :
  - Niveau d'études supérieur à 12 ans
    - 31 sujets de 21 à 40 ans (8 hommes et 21 femmes)
    - 30 sujets de 60 à 80 ans (9 hommes et 21 femmes)
  - Niveau d'études inférieur à 12 ans
    - 38 sujets de 16 à 25 ans (14 hommes et 24 femmes)
    - 30 sujets de 26 à 40 ans (11 hommes et 19 femmes)
    - 28 sujets de 41 à 65 ans (10 hommes et 18 femmes)
    - 26 sujets de 66 à 80 ans (11 hommes et 15 femmes)

- Critères d'inclusion : langue maternelle française et droitier, exclusion si le sujet a un diagnostic de dyslexie ou de troubles neurologiques ou psychiatriques.
- Pas d'étalonnage pour le groupe d'âge de 40-60 ayant fait des études supérieures : les résultats ne sont pas homogènes, l'auteur avance un effet protecteur du niveau de scolarité qui a une influence sur leurs performances en compréhension.
- Échantillon a plus de 5 ans
- Pas d'information sur le niveau socio-culturel des sujets testés
- Pas d'information sur les régions dans lesquelles ont été testés les sujets, ni sur le pays de sélection des sujets (France ou Canada ?) : on sait juste qu'ils sont francophones.

### Mesures disponibles :

- Moyennes et écart-types pour la mesure du temps (mais pas de normalisation).
- Percentiles et écart-types proposés

### Sensibilité

- Protocole testé auprès de 6 patients cérébro-lésé gauche (au Canada) et 6 patients traumatisés crâniens (à Paris) : objectivation de difficultés spécifiques pour chacun des patients mais non objectivables « par des tests habituellement utilisés en clinique d'aphasie ».
- Population clinique assez restreinte
- Pas de niveau de sensibilité établi
- Pas de définition du seuil de pathologie
- Beaucoup de scores qui plafonnent

### Validité

- *Validité théorique* : références anciennes, mais le test est basé sur le modèle de Kintsch (1988), avec l'indentification de 3 niveaux de représentation dans la compréhension de texte : surface, macro-structure, micro-structure
- *Validité de contenu* : non renseignée
- *Validité empirique* : non renseignée
- *Validité écologique* : test rapide à faire passer (10-15 minutes)

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### Conclusion

Permet de tester la compréhension en lecture chez l'adulte cérébro-lésé (peu d'outils disponibles à ce jour), sur des bases théoriques de traitement de texte solides, mais la validité de l'étalonnage pose question : population restreinte, effet plafond observé.

### Intérêt clinique

Ce test permet d'objectiver les plaintes en compréhension de lecture et de mettre en place des objectifs thérapeutiques.

## Annexe D-u : Fiche de Timé-2



### Référence du test

Timé2 (2003)

Jean ECALLE

### Acquisition de l'outil

Mot à mot éditions

### Références (articles/études/mémoire)

Ecalle, J. (2003a). Développement des processus d'identification des mots écrits : une étude transversale entre 6 et 8 ans. *Rééducation orthophonique*, 213, 77-96.

### Composantes évaluées

- Acquisition de la lecture avec un âge lecteur

### Population ciblée

- Enfants de 6 à 8 ans

### Description de l'outil

- Feuillet de passation
- Planches images

### Passation

3 consignes, choisir un mot dans un ensemble de 5 (le mot correct, un pseudo-mot ou mot homophone, un pseudo-mot visuellement proche, un voisin orthographique et un pseudo-mot non conventionnel) :

- Entourer le mot que le patient entend
- Entourer le mot qui va avec l'image présentée
- Entourer le mot qui va avec celui présenté dans un cadre au-dessus

### Normalisation

#### Données démographiques disponibles :

- 933 élèves de CP, CE1 et CE2
- Répartition sur 8 départements français, représentation de plusieurs catégories socio-professionnelles, inclusion d'établissements ZEP

- 2 périodes de passation : septembre- octobre et janvier-février
- Publication en 2003 (supérieure à 5 ans)

**Mesures disponibles** : mesure de l'âge lexique

### Sensibilité

- Bon pouvoir de différenciation (potentiel de différenciation) entre un sujet normo-lecteur et un dyslexique : une étude de cas.
- Diagnostic possible sur retard de plus de 18 mois

### Validité

- *Validité théorique* : explication du cadre théorique (les voies de lecture)
- *Validité de contenu* : différentes présentations des mots, il y a donc une nécessité de l'avoir dans son stock lexical, il ne s'agit pas uniquement de déchiffrage.
  - o Contrôle de la fréquence, régularité et longueur des mots, régularité orthographique
- *Validité empirique* : analyse avec le test « la pipe et le rat » (Lefavrais, 1986) : forte corrélation avec les CE1 mais coefficients faibles avec les CP.
- *Validité écologique* : non renseignée

### Fidélité

- *Fidélité inter-observations* : test-retest dans un intervalle de 15 jours sur 49 sujets : coefficients obtenus proches de 1 donc excellente fidélité
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : coefficient satisfaisant pour chacune des catégories de mots (alpha de Cronbach supérieur à 0.7), seul la liste d'homophone obtient un résultat inférieur à 0.70

### Conclusion

Ce test a de très bons critères psychométriques.

### Intérêt clinique

Il va définir un âge de lecture et nous allons donc pouvoir nous rendre compte de l'écart à la norme que peut avoir un jeune enfant. Un portrait de lecteur est dressé.

## Annexe D-v : Fiche du TinFoLec



### Référence du test

TinFoLec (2014)

ECALLE Jean, REVY Philippe

### Acquisition de l'outil

Matériel disponible sur la plateforme GERIP

### Références (articles/études/mémoire)

Non renseignées

### Composantes évaluées

La lecture, avec différentes épreuves : la connaissance des sons des lettres, la discrimination phonologique, le décodage, la discrimination orthographique, la décision lexicale.

### Population ciblée

Enfants du CE1 à la fin du CM2.

### Description de l'outil

- Logiciel avec 5 tâches de lecture différentes

### Passation

Les consignes et indications de passation sont données par le logiciel. Toutes les épreuves sont chronométrées.

Il n'y a pas de renseignement sur la durée de passation des différentes épreuves.

### Normalisation

#### Données démographiques disponibles :

- 946 élèves testés répartis
  - o Selon leur classe
    - CE1 : 249 élèves
    - CE2 : 181 élèves
    - CM1 : 248 élèves
    - CM2 : 268 élèves



- Selon leur sexe
  - Filles : 475
  - Garçons : 451
- Le test s'est déroulé lors du second semestre de l'année scolaire
- Huit départements de l'est et l'ouest de la France sont représentés.
- L'étalonnage a plus de 5 ans
- Il y a une exclusion des redoublants, le pourcentage des élèves en ZEP est inférieur à la moyenne nationale.

**Mesures disponibles :** les résultats sont donnés selon l'écart à la norme (écart-types).

### Sensibilité

- Bonne discrimination avec 5 cas de patients connus testés et identifiés.
- Effet significatif du niveau scolaire sur les performances, quelles qu'elles soient.

### Validité

- *Validité théorique* : peu de théories développées dans ce manuel
- *Validité de contenu* : corrélations testées pour prouver que les hypothèses selon lesquelles les compétences mesurées par le test mesurent bien les processus de lecture
  - Mots choisis selon leur fréquence (indice U G1-G5 de Manulex) et leur longueur
- *Validité empirique* : non renseignée
- *Validité écologique* : les épreuves sont randomisées, la mesure du temps de réponse ainsi que la précision de lecture sont établies. Cependant, il n'y a pas de renseignements sur la durée de passation des épreuves.

### Fidélité

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : la corrélation (autour de 0.7) est significative

### Conclusion

Une bonne représentativité de l'échantillon de passation ainsi qu'une bonne sensibilité.

### Intérêt clinique

Un profil de lecteur est dressé assez rapidement, les scores sont automatiques après la passation du test, tout est informatisé.

## Annexe D-w : Fiche du Vol du PC



### Référence du test

Le vol du PC (2006),

Corinne BOUTARD, Isabelle CLAIRE et Laurent GRETCHANOVSKY

### Acquisition de l'outil

Orthoédition

### Références (articles/études/mémoire)

Mémoire de Laurent Gretchanovsky, 1998, Élaboration d'un outil d'évaluation de la lecture proposé à 243 adolescents.

Mémoire d'Isabelle Claire, 1998-1999, Standardisation d'un outil d'évaluation de la lecture proposé à 620 adolescents.

### Composantes évaluées

- Temps et qualité de la lecture
- Nombres et types d'erreurs
- Compréhension du texte lu

### Population ciblée

- Collégiens et lycéens : de la 6<sup>ème</sup> à la terminale

### Description de l'outil

- Un feuillet de cotation
- Un texte à lire

### Passation

Lecture d'un texte puis série de questions pour déterminer la compréhension

### Normalisation

#### Données démographiques disponibles :

- 620 sujets répartis
  - o Selon le sexe
    - Féminin : 310 sujets
    - Masculin : 310 sujets

- Selon l'âge
  - 11 ans : 36 sujets
  - 12 ans : 44 sujets
  - 13 ans : 45 sujets
  - 14 ans : 51 sujets
  - 15 ans : 119 sujets
  - 16 ans : 120 sujets
  - 17 ans : 123 sujets
  - 18 ans et plus : 82 sujets
- Selon la classe :
  - 6<sup>ème</sup> : 48 sujets
  - 5<sup>ème</sup> : 47 sujets
  - 4<sup>ème</sup> : 52 sujets
  - 3<sup>ème</sup> : 54 sujets
  - 2<sup>nde</sup> : 142 sujets
  - 1<sup>ère</sup> : 121 sujets
  - Terminale : 80 sujets
  - 2<sup>nde</sup> pro : 26 sujets
  - 1<sup>ère</sup> pro : 32 sujets
  - BTS : 18 sujets
- Période de passation non précisée, étude publiée en 2006 (donc supérieure à 5 ans)
- Pas d'exclusion de redoublement, de bilinguisme familial ou de prise en charge en orthophonie
- Sujets de la région parisienne et de Guadeloupe uniquement

**Mesures disponibles** : moyennes et écart-types (mais pas de normalisation)

### Sensibilité

- Pas d'étude de la sensibilité, expérience clinique privilégiées dans ce test
- Corrélations faites entre les différents types d'erreurs de lecture, les critères de lecture (vitesse, qualité), les éléments de restitution (éléments narratifs, documentaires), la compréhension, en fonction de l'âge
- Pas de seuil de pathologie défini

### Validité

- *Validité théorique* : description et explication des théories utilisées pour la construction du test : description des voies et des processus de lecture.
- *Validité de contenu* :
  - Mesure du temps et de la qualité de lecture en corrélation avec l'acquisition des voies de lecture : insertion de toutes les difficultés du français, régularité, longueur et fréquence du mot, ainsi qu'une introduction de certains logatomes
  - Questionnaire pour la compréhension

- *Validité empirique* : non mentionnée
- *Validité écologique* : passation rapide

### **Fidélité**

- *Fidélité inter-observations* : non renseignée
- *Fidélité inter-observateurs* : non renseignée
- *Cohérence interne* : non renseignée

### **Conclusion**

Test rapide à faire passer et à analyser. Base de diagnostic pour poursuivre sur des épreuves plus spécialisées.

### **Intérêt clinique**

Permet un rapide état des lieux de la lecture du patient, en vue d'un aménagement de tiers-temps.