

Analyse critique des caractéristiques psychométriques des outils francophones évaluant la parole chez les enfants d'âge préscolaire

J. Cattini^{1,2}

C. Maillart³

¹ Orthophoniste, Luxembourg-ville, Luxembourg

² Collaboratrice scientifique, CPLU, RUCHE, Université de Liège, Liège, Belgique

³ Professeure ordinaire, Département de logopédie, RUCHE, Université de Liège, Liège, Belgique

Correspondance :

Julie Cattini

juliecattini@hotmail.com

Conflits d'intérêt : Les autrices déclarent être membres du conseil d'administration de Tool2care mentionné dans l'article. Elles n'en tirent toutefois aucun bénéfice financier.

Résumé

L'identification et la prise en charge du Trouble des Sons de la Parole sont primordiales pour diminuer les risques de difficultés académiques et socio-affectives. Cette étude identifie et analyse les données psychométriques de 32 outils conçus pour les enfants d'âge préscolaire. Les résultats montrent un faible niveau de preuve des caractéristiques psychométriques pour tous les outils. Des recommandations sont proposées au regard des données disponibles.

Mots-clefs : Psychométrie, Orthophonie, Parole, Enfant d'âge préscolaire

Abstract:

Identifying and treating Speech Sound Disorder is crucial to reducing the risk of academic and socio-emotional difficulties. This study identifies and analyses the psychometric data of 32 tools designed for preschool children. The results show a low level of evidence for the psychometric properties of all the tools. Recommendations are proposed based on the available data.

Keywords: Psychometry, Speech and Language Therapy, Speech, Preschool child

Remerciements :

Nous tenons à remercier Jonathan Burnay et Sylvie Willems pour leur soutien constant lors de l'analyse des données ainsi que les chercheurs et chercheuses ayant eu l'amabilité de répondre à toutes les questions et à fournir les informations nécessaires pour compléter les analyses.

Note du 02.09.2024 :

La référence bibliographique concernant le test Isadyle a été corrigée dans cette version du manuscrit.

Note du 03.10.2024 :

Le nom des épreuves suivantes a été modifié à la demande des auteurs :

Ancienne dénomination	Nouvelle dénomination
Epreuve de dénomination d'images de Masson	Epreuve de dénomination d'images de Martinez Perez
Epreuve de répétition d'images de Masson	Epreuve de répétition d'images de Martinez Perez
Epreuve de diadocosinésies de Masson	Epreuve de diadocosinésies de Martinez Perez
Epreuve de discrimination phonémique de Masson	Epreuve de discrimination phonémique de Martinez Perez
Epreuve de jugement phonologique de Masson	Epreuve de jugement phonologique de Martinez Perez

Les auteurs de l'épreuve de diadococinésies de Martinez Perez nous ont contacté pour nous informer que les données présentes dans la référence principale présentaient des erreurs de calcul. Les données concernant les mesures de tendance centrale et le pouvoir discriminant sont invalides. L'ensemble des corrections est apporté en rouge dans le manuscrit.

Introduction

En orthophonie, l'évaluation du patient est le premier acte réalisé par le clinicien afin de répondre à la demande initiale. Par la suite, elle constitue un processus de collecte de données pour soutenir les prises de décision clinique à réaliser au cours du suivi (Duboisdindien & Cattini, sous presse). L'utilisation d'outils d'évaluation peut avoir plusieurs fonctions (Youngstrom et al., 2017) comme la détection de difficultés, l'aide au diagnostic, la compréhension du profil du patient (par exemple, au regard d'un modèle théorique ou en vue de sélectionner une approche thérapeutique) ou encore le suivi de l'évolution du patient dans le temps. Chacune de ces fonctions exigent des qualités psychométriques différentes (dans ce même numéro, Burnay et al., 2024) et doivent être adaptées à la littérature spécifique du domaine d'évaluation considéré. Dans cet article, le domaine ciblé est celui de l'évaluation de la parole chez l'enfant.

L'évaluation du Trouble des Sons de la Parole

Le Trouble des Sons de la Parole (TSP) regroupe des profils variés de personnes présentant des difficultés de perception ou de production de la parole impactant leur intelligibilité (ASHA, 2020). Cette condition de haute prévalence (Eadie et al., 2015) constitue un domaine de la communication pour lequel les orthophonistes francophones souhaitent obtenir des informations fiables pour sélectionner leurs outils d'évaluation (Cattini et al., in prep.).

Les recommandations actuelles (Bates & Titterton, 2017) documentent principalement les données à récolter lors de l'évaluation initiale pour affiner le profil du patient (par exemple, le relevé des erreurs de parole, l'estimation de la stimulabilité) lors de la production de la parole mais peu d'informations existent concernant le diagnostic, la détection ou encore l'évaluation de l'évolution. Par ailleurs, ces recommandations sont générales et dépendantes des outils disponibles dans la langue de l'utilisateur. Idéalement, le répertoire des outils d'évaluation de la parole devrait comprendre des épreuves variées et adéquates sur le plan psychométrique afin de permettre l'évaluation initiale et continue du patient.

Détaillons les différentes fonctions de l'évaluation présentées ci-dessus en les appliquant au domaine de la parole. Une étape possible consiste à **détecter** des difficultés afin de réaliser des investigations plus approfondies. Pour cela, des épreuves rapides et particulièrement sensibles (c-à-d, capables de détecter une difficulté) sont nécessaires. Toutefois, en clinique, cette étape est généralement informelle et se résume à des inquiétudes des parents ou de l'enseignant. Pour l'éligibilité à un remboursement des soins, l'orthophoniste peut être amené à confirmer que l'enfant présente un décalage par rapport aux performances attendues pour

son âge chronologique. Poser le **diagnostic** de TSP nécessite de recueillir des données au niveau de l'intelligibilité et de documenter l'impact fonctionnel du trouble. Un diagnostic différentiel entre des sous-types de TSP (par exemple, dyspraxie verbale, trouble phonologique, dysarthrie) requiert de compléter les informations obtenues. **Comprendre le profil du patient et ses difficultés** en évaluant les aspects perceptifs (par exemple, discrimination auditive), moteurs (par exemple, contrôle moteur) et de représentations phonologiques (par exemple, analyse phonologique, stabilité des productions) permet, entre autres, de déterminer les cibles de la prise en charge et l'approche choisie. Par exemple, la liste des sons maîtrisés (c-à-d, inventaire phonétique) permet d'identifier les cibles à travailler et la prise en compte de la stimulabilité des sons guide à la fois les cibles de la prise en charge mais aussi la nature de l'approche thérapeutique à privilégier. Enfin, **l'évaluation de l'évolution** ou de l'effet d'une intervention est fondée sur des mesures sensibles liées aux résultats attendus pour le patient (par exemple, meilleure intelligibilité et/ou précision de la parole dans différents contextes). Le Tableau 1 résume les principales composantes à évaluer selon l'objectif ciblé.

Tableau 1. Proposition des composantes à évaluer selon l'objectif de l'évaluation

	Comprendre	Détecter	Diagnostiquer	Évaluer l'évolution
Estimation de l'intelligibilité		X	X	X
Précision de la parole en contexte isolé – analyse normée¹		X	X	X*
Discrimination auditive	X		X	
Jugement phonologique/ décision lexicale	X			
Contrôle moteur (e.g. diadococinésies)	X		X	
Fonctions oromotrices verbales et non-verbales	X			
Stabilité des productions	X		X	
Précision de la parole en contexte isolé – analyse descriptive²	X			X
Précision de la parole en contexte fonctionnel (i.e. échantillon de langage)	X			X

¹ Lors d'une analyse normée, les performances du patient seront comparées à celles d'enfant du même âge. Les mesures utilisées pour ce type d'interprétation sont, par exemple, le nombre de mots corrects ou le pourcentage de consonnes correctes.

² Lors d'une analyse descriptive d'une tâche de dénomination/répétition d'images, l'orthophoniste réalisera une analyse des erreurs de parole comme la présence de processus phonologiques simplificateurs ou l'analyse des structures syllabiques.

Inventaire phonétique	X	
Stimulabilité	X	
Conscience phonologique	X	X
Mémoire à court terme verbale	X	

Légende. *L'utilisation d'évaluations normées pour évaluer l'évolution d'un patient doivent être utilisées avec précaution en considérant les données de fidélité pour interpréter les résultats à l'aide de l'erreur standard de mesure (McCauley & Swisher, 1984).

L'analyse des qualités psychométriques des outils d'évaluation

Après avoir identifié les composants importants liés aux différentes fonctions de l'évaluation de la parole, les orthophonistes doivent disposer d'outils adéquats sur le plan psychométrique et pouvoir les utiliser correctement (Daub et al, 2021). Or, les études s'étant intéressées à la question des qualités psychométriques d'outils d'évaluation francophones (Lafay & Cattini, 2018 ; Leclercq & Veys, 2014 ; Gaul-Bouchard et al., 2019 ; Macchi et al., 2023) s'accordent sur la faible qualité générale des outils mis à disposition des cliniciens. Dans le domaine de la parole, Meloni et al. (2022) ont évalué la qualité des outils diagnostiques évaluant la perception de la parole. Les résultats montrent que les tests ne satisfont pas les critères psychométriques et que très peu d'études justifient le choix de leurs cibles en contrôlant les variables psycholinguistiques. Cependant, ce travail ne fournit pas d'information pour les outils évaluant la production de la parole et/ou visant à détecter, comprendre et évaluer l'évolution.

Notons également que les études citées précédemment réalisent une analyse exhaustive des caractéristiques psychométriques sans pour autant pouvoir proposer une analyse critique claire selon l'objectif de l'évaluation. Pourtant, il est recommandé de réaliser une analyse spécifique de l'outil selon la fonction évaluative recherchée. Il est ainsi préférable de considérer un outil d'évaluation pour un objectif précis même si d'autres caractéristiques psychométriques sont absentes pour un autre objectif (Daub et al., 2021). Le Tableau 2 résume le niveau d'importance des caractéristiques psychométriques au regard de l'objectif d'évaluation selon la grille développée par Tool2Care (voir pour plus de détails : Burnay et al., 2024).

Tableau 2. Récapitulatif des caractéristiques psychométriques essentielles, importantes ou secondaires selon l'objectif de l'évaluation.

		<i>Validité de la structure interne</i>	<i>Validité théorique et de contenu</i>	<i>Relations avec d'autres variables</i>	<i>Validité prédictive</i>	<i>Sensibilité</i>	<i>Spécificité</i>	<i>Sensibilité au changement</i>	<i>Fidélité test-retest</i>	<i>Fidélité interjuges</i>	<i>Consistance interne</i>	<i>Distribution des données</i>	<i>Erreur standard de mesure</i>
Comprendre	Essentiel		X										
	Important	X		X	X				X	X	X		X
	Secondaire					X	X	X				X	
Détecter	Essentiel			X	X	X						X	
	Important								X	X	X		X
	Secondaire	X	X				X	X					
Diagnostiquer	Essentiel		X	X		X	X		X	X	X	X	X
	Important	X			X								
	Secondaire							X					
Evaluer l'évolution	Essentiel				X			X	X	X	X		X
	Important												
	Secondaire	X	X	X		X	X					X	

Objectif et motivations de l'étude

Ce travail vise à répertorier les outils d'évaluation de la parole destinés à des enfants francophones d'âge préscolaire en analysant leurs caractéristiques psychométriques.

Si une démarche d'analyse psychométrique des outils d'évaluation est une condition importante à une pratique clinique de qualité, elle reste peu réaliste en raison des contraintes rencontrées par les professionnels de santé telles que le manque de temps ou encore le manque de connaissances académiques (Durieux et al., 2016). Ainsi, une enquête réalisée auprès de 129 orthophonistes francophones a montré un faible niveau de connaissances en psychométrie ainsi qu'un sentiment de compétence jugé insatisfaisant par plus de la moitié des répondants quant au processus de sélection d'un outil d'évaluation (dans ce même numéro, Cattini & Willems, 2024). L'accès à des ressources externes proposant des analyses critiques indépendantes constitue dès lors un facilitateur indispensable pour la mise en place d'une évaluation basée sur des données probantes.

Méthodologie

Recherche des outils francophones évaluant la parole des enfants d'âge préscolaire

Une recherche a été menée afin de repérer les outils francophones permettant d'évaluer la parole des enfants d'âge préscolaire suspectés de présenter ou présentant des difficultés de parole. Afin de répondre aux objectifs de recherche, nous avons analysé les propriétés d'outils

francophones évaluant les TSP. Une recherche a été réalisée sur les sites web des éditeurs (c'est-à-dire, De Boeck, ECPA, Gerip, Happyneuron, Hogrefe, Mot-à-mot, Orthoédition), deux bases de données bibliographiques (Medline ; PsycInfo) et les archives de revues francophones spécialisées (c.-à-d., ANAE et Glossa).

En complément de cette démarche, chaque auteur principal des tâches de performances a été contacté afin d'obtenir des références issues de la littérature grise concernant la validation des outils et le nom des batteries de performances ont été utilisé pour réaliser une recherche complémentaire dans les deux bases de données bibliographiques afin de trouver des études indépendantes (c.à.d., études réalisées indépendamment de l'étude de validation initiale). Le détail de la méthodologie de recherche (i.e. équations de recherche, date, processus de sélection) est consultable dans Cattini et al. (in prep). Les articles ou manuels de tests devaient répondre aux critères suivants : 1) les articles devait être en français ou en anglais ; 2) porter sur la validation d'un test (c.-à-d., batterie de performances, tâche de performances, questionnaire, grille d'observation) évaluant un ou des processus du traitement de la parole ; 3) concerner des enfants francophones âgés de 3 à 6 ans ; 3) porter sur un test toujours édité ou accessible sur demande aux auteurs ; et 4) ne pas être dans une phase de développement initiale.

Grille d'extraction et d'évaluation des caractéristiques psychométriques des outils

Afin de réaliser l'extraction des données concernant les qualités psychométriques des outils d'évaluation, la grille d'extraction du projet Tool2Care a été utilisée. L'extraction des données est composée de six volets : 1) la description générale de l'outil ; 2) les modalités de l'outil ; 3) la description factorielle de l'outil ; 4) la description générale des normes ; 5) la description des échantillons de référence ; et 6) les données psychométriques calculées sur les données normatives.

Une fois ces données de l'outil extraites à partir des différentes sources d'information, une grille d'analyse synthétique est complétée afin de calculer un score sur la base des données ainsi extraites pour chaque objectif d'évaluation (c-à-d, comprendre, détecter, diagnostiquer et/ou évaluer l'évolution). Les critères sont divisés en trois catégories principales : 1) les preuves de validité ; 2) les preuves de fidélité ; 3) les informations concernant les mesures de tendance centrale. Chaque qualité psychométrique est pondérée selon son importance au regard de l'objectif de l'évaluation en trois niveaux : essentiel, important et secondaire. Par exemple, la présence de preuves de validité théorique est essentielle afin de comprendre et diagnostiquer un patient tandis qu'elles sont secondaires pour estimer l'évolution ou encore

détecter les difficultés (voir Tableau 2). Enfin, il est à noter que pour ce qui est de la fidélité, le score tient compte de la qualité des indices (par exemple, une fidélité test-retest égal ou supérieur à 0.90 obtient un score maximal tandis qu'une fidélité test-retest strictement inférieure à 0.70 obtient un score de 0) alors que pour la validité le score est basé sur l'absence ou la présence de la qualité (c-à-d, 0 ou 1). Pour davantage d'informations concernant les grilles d'extraction et d'évaluation, nous renvoyons le lecteur à l'article de Burnay et al. (2024) dans ce même numéro.

Quelques adaptations mineures ont toutefois été réalisées dans le cadre de cette étude. Tout d'abord, l'évaluation de la validité théorique est une tâche particulièrement complexe qui dépend des données de la recherche sur le sujet, de la présence d'un consensus entre les experts du domaine ou encore des connaissances du juge. Au vu de notre sujet de recherche, nous avons jugé pertinent de différencier l'absence ou la présence de validité théorique sur deux niveaux : la présence de données/modèles théoriques justifiant le choix des épreuves et le contrôle de variables psycholinguistiques pour la sélection des items. Conformément au processus d'évaluation de la grille initiale, les preuves de validité sont jugées selon leur présence ou absence et, non, selon leur adéquation avec les données actuelles de la littérature. Par ailleurs, pour les batteries de performances, nous avons pris en compte les données de validation pour chaque sous-épreuves évaluant spécifiquement la parole. Les résultats présentés dans cet article peuvent donc légèrement différer d'une analyse tenant compte de la batterie dans son ensemble. Après une comparaison qualitative des résultats, ces différences sont toutefois mineures.

L'extraction des données a été réalisée par la première auteure (JC) et a été vérifiée par un chercheur spécialisé en psychométrie. En cas de doute, des concertations ont eu lieu afin d'harmoniser les procédures.

Résultats

La recherche d'outils a permis de répertorier 32 outils permettant d'évaluer la parole des enfants d'âge préscolaire. Nous comptabilisons 7 batteries de performances³ et 25 tâches de performances⁴. Soixante-deux références ont été consultées afin de compiler les données

³ Une batterie de performances désigne un ensemble de tests utilisés conjointement afin de tenir compte de plusieurs aspects des aptitudes d'une personne.

⁴ Une tâche de performances désigne une épreuve visant à évaluer le niveau de performances d'un individu pour une compétence précise.

disponibles. Les références utilisées pour chaque outil sont librement téléchargeables sur la plateforme ORBI : [XXXXXX](#).

Caractéristiques psychométriques des outils

Echantillons de référence

Les batteries de performance répertoriées possèdent des échantillons de référence âgés de 2;3 à 11;11 ans provenant de la Belgique, de la France ou de la Suisse. Les échantillons des tâches de performances sont âgés de 1;8 à 10;3 ans et proviennent de Belgique, du Canada, des Etats-Unis, de France ou de Suisse. L'effectif par groupe est généralement inférieur à 100 avec une étendue de 20 à 147 pour les batteries de performances et de 12 à 75 pour les tâches de performances. Le détail des caractéristiques des outils d'évaluation en termes d'échantillon de référence⁵ et de données de validation sont librement téléchargeables sur la plateforme ORBI en suivant le lien fourni préalablement.

Batteries de performances

Le relevé synthétique des données de validité et de fidélité des sept batteries de performances analysées est disponible en Annexe A (Tableau A1). Aucun test ne dispose à l'heure actuelle d'information nécessaire pour l'ensemble des domaines.

Données sur la validité. La validité théorique est la mieux représentée dans le manuel des outils avec la mention de modèles théoriques liés au contenu de l'évaluation pour quatre sur sept des batteries et des preuves d'un contrôle des variables psycholinguistiques des items pour trois d'entre elles. Les autres indices de validité (c.à.d, validité prédictive, validité de la structure interne, relations avec d'autres variables) ne sont mentionnés dans aucun manuel des batteries. Seule la validité discriminante est évaluée pour l'épreuve de répétition de logatomes de l'Exalang 5-8 par le biais d'une étude indépendante.

Données sur la fidélité. Des estimations de la consistance interne sont précisées dans le manuel de deux batteries (c.à.d, Evalo 2-6 - Coquet et al., 2009 ; ISADYLE - Piérart et al., 2010) ainsi de la fidélité test-retest (c.à.d, Exalang 3-6 - Helloin & Thibault, 2006 ; Exalang 5-8 - Helloin et al., 2010). La fidélité interjuge est indiquée dans le manuel de l'Evalo 2-6 et une étude indépendante permet de donner des précisions pour l'épreuve de répétition de logatomes de l'Exalang 5-8.

⁵ Les échantillons rapportés sont ceux des enfants au développement typique pour l'établissement des normes, excepté pour les épreuves de Masson et al. (2017) qui évaluent le pouvoir discriminant des épreuves afin de différencier les enfants présentant un trouble phonologique ou une dyspraxie verbale

Sensibilité aux changements. Des données partielles sont disponibles dans le manuel de l'ISADYLE et des études indépendantes fournissent des informations pour trois épreuves des Exalang 3-6 et 5-8.

Distribution des scores. Cinq manuels sur les sept batteries de performances précisent si les données sont normales ou fournissent des données corrigées. La moitié des outils permettent de calculer l'erreur standard de mesure mais un seul (c.à.d, ISADYLE) fournit cette donnée de manière transparente pour chaque épreuve dans le manuel.

Tâches de performances

Le relevé synthétique des preuves de validité et de fidélité des 25 tâches de performances analysées sont disponibles en Annexe A (Tableau A2).

Données sur la validité. La validité théorique est la mieux représentée avec un contrôle des variables psycholinguistiques des items pour 23 des épreuves d'évaluation sur 24 et des justifications pour le choix du type d'épreuves pour 21 d'entre eux. Seule l'ELDP (Macchi et al., 2012) propose une analyse en composante principale afin de fournir des données concernant la structure interne. Trois des tâches de performances possèdent des preuves de validité convergente (c.à.d, DRAP - Niederberger et al., 2021 ; répétition de non-mots de Courcy - Thordardottir et al., 2011 ; répétition de non-mots de Poncelet - Poncelet & Van der Linden, 2003). Des preuves de validité prédictive ne sont disponibles pour aucun outil. Enfin, les preuves de validité comprennent des données concernant le pouvoir discriminant **pour 10 outils sur 24 : neuf huit** pour discriminer les enfants présentant un trouble phonologique (TP) de ceux présentant une dyspraxie verbale (DV), deux pour distinguer des enfants au développement typique (DT) d'enfants présentant un Trouble Développementale du Langage (TDL) et un seul outil de dépistage fournit des données de sensibilité et spécificité pour discriminer les enfants DT des enfants présentant un TSP (c.à.d, ESPP – MacLeod et al., 2014).

Données sur la fidélité. Les données concernant la fidélité sont sous-représentées. La fidélité interjuges est documentée pour huit des tâches de performances sur 24. Les preuves de consistance interne et de fidélité test-retest sont rares avec respectivement deux (c.à.d, DRAP, ELDP) et un outil (c.à.d, répétition de non-mots de Poncelet) concernés.

Sensibilité aux changements. Neuf outils fournissent des informations concernant la sensibilité aux changements afin de fournir des preuves essentielles pour mieux caractériser l'évolution d'un patient.

Distribution des scores. Seuls trois outils précisent si les données sont normales ou fournissent des données corrigées (c.à.d, ELDP ; épreuve des logatomes de Ferrand - Ferrand, 1978 ; EULALIES version courte - Warnier et al., 2022) et dix tâches de performances fournissent des données suffisantes pour calculer l'erreur standard de mesure.

Analyse critique des outils selon l'objectif de l'évaluation

Tout type d'outils confondu, les scores obtenus à la grille d'évaluation sont faibles avec une moyenne de 3,09 ($\pm 1,14$) sur un score total de 10 points. L'étendue des scores est de 1 à 6 et les résultats sont comparables quel que soit l'objectif premier de l'outil d'évaluation (comprendre $M = 2,79$, $ET = 1,3$; détection $M = 3$, $ET = 1,1$; diagnostiquer $M = 3,4$, $ET = 1$; évolution, $M = 3$, $ET = 1$).

Batteries de performances

Les résultats de l'analyse critique des sous-épreuves issues des batteries de performances selon l'objectif de l'évaluation sont disponibles dans le Tableau 3.

La majorité des batteries de performances (6/7) ciblent plus d'un objectif d'évaluation.

Toutefois, aucune batterie de performances n'obtient un score supérieur à la moitié (> 5) quel que soit cet objectif. Le score maximal est de 5 sur 10 pour le dépistage et le diagnostic pour l'épreuve de répétition de logatomes de l'Exalang 5-8. Quatre batteries de performances sur sept (57%) obtiennent un résultat faible (≤ 3) pour un des objectifs d'évaluation fixés. Les qualités psychométriques nécessaires pour répondre aux standards de qualité par rapport aux objectifs fixés par les auteurs pour chaque outil ne sont pas rencontrées. Prenons l'exemple de la détection où la sensibilité est une caractéristique essentielle, celle-ci n'a été évaluée que dans une batterie de performances sur les quatre visant cet objectif. Par ailleurs, les données de sensibilité ont été fournies par la réalisation d'une étude indépendante pour une seule des épreuves de la batterie et n'est pas le fruit de la validation de l'outil par les auteurs.

Tableau 3. Evaluation de la qualité psychométrique des batteries de performances

Nom de l'outil	Objectif(s) de l'outil selon les auteurs	Score – Grille d'évaluation			
		Comprendre (/10)	Détecter (/10)	Diagnostiquer (/10)	Evaluer l'évolution (/10)
Diabolab 3 évaluation	Comprendre et évaluer l'évolution	1	0	1	0
ELOLA	Diagnostiquer	3	1	2	2
Evalo 2-6	Comprendre, détecter et diagnostiquer	4	4	4	4
Exalang 3-6	Détecter, diagnostiquer et évaluer l'évolution du patient	2	3	3	3

Exalang 5-8	Comprendre, détecter, diagnostiquer et évaluer l'évolution	2	5	5	4
ISADYLE	Comprendre, détecter, diagnostiquer et évaluer l'évolution	4	3	4	3
NEPSY-II	Comprendre et diagnostiquer	1	2	2	0

Légende. Les cases grisées correspondent aux objectifs ciblés par les concepteurs de l'outil. Diabolab 3 évaluation (Menin-Sicard & Sicard, 2006) ; ELOLA : Batterie d'évaluation du langage oral de l'enfant aphasique (De Agostini et al., 1998) ; Evalo 2-6 : Batterie d'évaluation du langage de l'enfant de 2 à 6,3 ans (Coquet et al., 2009) ; Exalang 3-6 : Examen des troubles du langage et de la parole 3-6 (Helloin & Thibault, 2006) ; Exalang 5-8 : Examen des troubles du langage et de la parole 5-8 (Helloin et al., 2010) ; ISADYLE : ISADYLE : Instruments pour le Screening et l'Approfondissement des Dysfonctionnements du Langage (Piérart et al., 2010) ; NEPSY-II : Bilan neuropsychologique de l'enfant - 2nde édition (Korkman & Kirk et al., 2012)

Tâches de performances

Les résultats de l'analyse critique des tâches de performances selon l'objectif de l'évaluation sont disponibles dans le Tableau 4.

La majorité des tâches de performances (22/25) ciblent un seul objectif d'évaluation. Une épreuve (c-à-d, la répétition de non-mots de Poncelet) obtient un score supérieur à la moitié (6/10) par rapport à l'objectif initial visé par l'outil. Les épreuves visant à comprendre les difficultés obtiennent un score allant de 2 à 4 sur 10. Elles se caractérisent par ailleurs par des scores très faibles pour les autres objectifs d'évaluation. Concernant la détection des difficultés, le DRAP spécialement conçu pour être rapide (i.e. 3 à 5 minutes de passation) démontre le score le plus haut avec un score de 4 sur 10. Toutefois, aucune donnée n'est disponible concernant la sensibilité de l'outil, caractéristique essentielle pour s'assurer que l'outil permet une bonne détection. Les résultats sont comparables au niveau du diagnostic avec une majorité d'outils (72%) qui obtient un score de 3 sur 10 avec l'épreuve de répétition de non-mots de Poncelet qui fournit un score égal à 5. Toutefois, les données de sensibilité et de spécificité concernent spécifiquement la discrimination des enfants présentant un TP et une DV. Enfin, une seule épreuve vise à évaluer l'effet d'une intervention : l'ICS. Cependant, aucune donnée francophone n'est encore disponible pour ce questionnaire hétéro-rapporté. Aucune tâche de performances ne répond donc aux standards psychométriques recommandés selon le ou les objectifs d'évaluation ciblés par les auteurs.

Tableau 4. Evaluation de la qualité psychométrique des tâches de performances selon l'objectif de l'évaluation qualifié par les auteurs

Nom de l'outil	Score GRID			
	Comprendre (/10)	Détecter (/10)	Diagnostiquer (/10)	Evaluer l'évolution (/10)
Comprendre				
Décision lexicale de Maillart	2	0	1	2

Nom de l'outil	Score GRID			
	Comprendre (/10)	Détecter (/10)	Diagnostiquer (/10)	Evaluer l'évolution (/10)
Discrimination auditive de Maillart	2	0	1	2
ISSEF	2	0	1	0
TFP	4	2	3	5
TIPLO	2	0	1	0
TPF – Standard	2	0	1	0
TPF – Manitoba	4	2	3	3
Détecter				
CL-NWR	4	3	5	5
DRAP	6	4	4	6
Epreuve des logatomes de Ferrand	1	2	2	0
ESPP	4	3	5	4
TDFP	4	2	3	5
TPFC-D	4	1	3	3
Diagnostiquer				
Dénomination d'images de Martínez Perez	2	2	3	0
Dénomination rapide de Masson	2	2	3	0
Discrimination phonémique de Martínez Perez	2	2	3	0
Epreuve de diadococinésies de Martínez Perez	2	2 0	3 1	0
ELDP	3	2	3	1
Histoire de Thomas	2	2	3	0
Jugement phonologique de Martínez Perez	2	2	3	0
Répétition de non-mots de Courcy	3	3	4	0
Répétition d'images de Martínez Perez	2	2	3	0
Comprendre et diagnostiquer				
Eulalies – version courte	4	3	4	4
Répétition de non-mots de Poncelet	5	5	6	4
Détecter et évaluer l'évolution du patient				
ICS	3	1	2	2

Légende. Les cases grisées correspondent aux objectifs ciblés par les concepteurs de l'outil. CL-NWR : Crosslinguistic Nonword Repetition Test (Schwob & Skoruppa, 2022) ; Décision lexicale de Maillart (Maillart & Schelstraete, 2004) ; Dénomination d'images de Martínez Perez (Masson, 2017) ; Dénomination rapide de Masson (Masson, 2017) ; Discrimination auditive de Maillart (Maillart & Schelstraete, 2004) ; Discrimination phonémique de Martínez Perez (Masson, 2017) ; DRAP : Dépistage Rapide de l'Articulation et de la Parole (Niederberger et al., 2021) ; Epreuve de diadococinésies de Martínez Perez (Masson, 2017) ; ICS : Échelle d'intelligibilité en contexte (McLeod et al., 2012) ; ELDP : Épreuve Lilloise de Discrimination Phonologique (Macchi et al., 2012) ; ESPP : Évaluation sommaire de la phonologie chez les enfants d'âge préscolaire (MacLeod et al., 2014) ; Epreuve des logatomes de Ferrand (Ferrand, 1978) ; Eulalies – version courte (Warnier et al., 2022) ; Histoire de Thomas (Masson, 2017) ; ISSEF : Inventaire des Structures Syllabiques chez l'Enfant Francophone (Aubry et al., 2021) ; Jugement phonologique de Martínez Perez (Masson, 2017) ; Répétition de non-mots de Courcy (Thordardottir et al., 2011) ; Répétition de non-mots de Poncelet (Poncelet & Van der Linden, 2003) ; Répétition d'images de Martínez Perez

(Masson, 2017) ; TDFP : Test de Dépistage Francophone de Phonologie (Rvachew et al., 2012) ; TFP : Test Francophone de Phonologie (Paul & Rvachew, 2008) ; TIPL0 : Test Informatisé pour la Perception du Langage Oral (Parijsse & Soubeyrand, 2003) ; TPF : Test de Phonologie du Français (Bérubé et al., 2013) ; TPFC-D : Test de Phonologie du Français Canadien-Dépistage (Bérubé et al., 2013).

Discussion

L'analyse et l'évaluation critique des outils francophones permettant d'évaluer la parole sont cohérentes avec les résultats obtenus dans les études antérieures que cela soit en cognition mathématique (Cattini & Lafay, 2024), en langage oral (Leclercq & Veys, 2014 ; Macchi et al., 2023), en langage écrit (Gaul-Bouchard et al., 2009) ou en perception de la parole (Meloni et al., 2022). Si nous comparons les qualités psychométriques de batteries et des tâches de performances, nous pouvons constater que les batteries, généralement éditées par des maisons d'édition, fournissent très peu de données concernant la validité de l'outil mais documentent davantage la fidélité. Toutefois, les données concernant cette dernière sont souvent fournies par le biais d'études indépendantes complémentaires et non dans le manuel du test. Les tâches de performances, mises en libre accès par des chercheurs et dont l'objectif premier n'était pas forcément d'être reprises en clinique, sont accompagnées quant à elles davantage de données concernant la validité avec un accent mis particulièrement sur une justification théorique de la sélection de l'épreuve et des items. En second lieu, les données concernant la sensibilité et la spécificité sont les plus représentées. Le taux de preuves disponibles au niveau de la fidélité est comparable entre les batteries de performances (53%) et les tâches de performances (50%) même si le type de preuves diffère légèrement (voir Tableaux A1 et A2). Cette étude souligne la nécessité de renforcer de la qualité psychométrique des outils existants en complétant les données manquantes pour les outils existants.

Enfin, seules la présence ou l'absence d'éléments de validité théorique ont été pris en compte dans cette analyse. Une analyse plus fine de la validité théorique envisagée via les liens entre les outils d'évaluation de la parole et une démarche d'évaluation psycholinguistique font l'objet d'une autre publication (Cattini et al., in prep).

Recommandations cliniques

A la lumière des données disponibles et en attendant de disposer d'outils plus satisfaisants au niveau psychométrique, les recommandations suivantes peuvent être proposées.

Afin de détecter un TSP, l'orthophoniste pourra recourir au test de Dépistage Rapide de l'Articulation et de la Parole (DRAP) ou encore à l'Echelle d'Intelligibilité en Contexte (ICS)

tout en restant très prudent en cas de score négatif au regard de l'absence de données concernant la sensibilité de ces outils. Lors d'une évaluation de première ligne, l'orthophoniste pourrait sélectionner l'Epreuve Sommaire de la Phonologie pour les enfants d'âge Préscolaire (ESPP) qui montre une sensibilité de 80%.

Afin de poser un diagnostic de TSP, l'orthophoniste se trouve face à une difficulté majeure : aucun outil n'a évalué le pouvoir discriminant pour différencier les enfants DT et TSP, excepté l'ESPP conçu initialement pour de la détection. Nous recommandons de réaliser 1) une évaluation informelle pour collecter des données concernant l'intelligibilité par les différentes personnes de l'entourage ; 2) une tâche de dénomination d'images afin d'objectiver un écart de performances pour la production de la parole par rapport aux compétences attendues pour l'âge (par exemple, Isadyle, Evalo 2-6) et 3) une tâche de discrimination auditive avec l'Epreuve Lilloise de Discrimination Phonologique (ELDP) démontrant des preuves de structure interne et de validité théorique. Afin de réaliser le diagnostic différentiel de dyspraxie verbale, la tâche de dénomination et de répétition d'images, ~~un calcul de la vitesse articulatoire et de la durée des pauses lors de la production des séries diadococinésiques de Martinez Perez (Masson, 2017)~~ ainsi qu'une évaluation de la stabilité de la parole avec l'Histoire de Thomas (Masson, 2017) s'avèrent la meilleure option sur base des données actuelles.

Après avoir identifié les difficultés, il est recommandé de poursuivre l'évaluation afin de comprendre plus finement le profil du patient et déterminer la prise en charge la plus appropriée. Dans ce cadre, l'évaluation psycholinguistique vise à identifier les processus déficitaires et préversés du traitement de la parole tant au niveau de la réception que de la production (Terband et al., 2019). Cette analyse basée sur un modèle du traitement de la parole de type boîte-flèche (Stackhouse & Wells, 1997 ; Terband et al., 2019 ; Meloni, 2022) nécessite d'évaluer chaque processus de traitement perceptif ou de production de la parole afin de poser des hypothèses sur la nature des difficultés du patient et, ainsi, individualiser la rééducation. A l'heure actuelle, il n'existe aucune batterie de performances permettant d'évaluer l'ensemble des processus de la parole sur la base d'un modèle théorique actuel. En outre, les épreuves ayant pour objectif premier de comprendre les difficultés du patient présentent généralement le niveau de preuve le plus faible concernant leurs caractéristiques psychométriques. L'orthophoniste doit donc utiliser les épreuves existantes en tenant compte de leurs limites ou en construire. Par exemple, il n'existe aucune épreuve francophone évaluant la stimulabilité de sons. Ce type d'épreuve a pourtant de nombreux avantages : 1) la

stimulabilité des sons est un bon prédicteur de l'effet de l'intervention ; 2) des études ont montré son potentiel pour mesurer l'évolution du patient à la suite d'une intervention et 3) la modalité dynamique permet de sélectionner des cibles stimulables pour des patients, par exemple, rapidement frustrés en cas d'échec (Glaspey & Stoel-Gammon, 2007 ; Powell & Miccio, 1996). Le développement d'une échelle francophone de stimulabilité serait donc une réelle plus-value. Concernant les tâches de dénomination d'images permettant de réaliser une analyse descriptive détaillée de la production de la parole, nous pouvons actuellement recommander 1) l'Inventaire des Structures Syllabiques de l'Enfant Francophone (ISSEF) pour une analyse structurelle des erreurs de parole et 2) le Test de Phonologie du Français (TFP ; Bérubé et al. 2023) si une analyse phonologique non linéaire s'avère plus pertinente au regard de la situation clinique.

Pour terminer, peu d'outils permettent d'évaluer l'évolution du patient. Si la plupart des auteurs des batteries de performances évoquent cet objectif dans leur manuel associé, aucune d'entre elles ne démontrent une bonne sensibilité aux changements ou encore des preuves de validité prédictive. Afin d'évaluer l'évolution dans le temps d'un enfant d'âge préscolaire, faute de mieux, nous recommandons d'utiliser le Test Francophone de Phonologie (Paul & Rvachew, 2008) ou la version courte d'Eulalies (Warnier et al., 2022) qui sont deux tâches de dénomination d'images ayant montré une taille d'effet forte de sensibilité aux changements (respectivement $d = 1,8$ et $\eta^2 = 0,30 - 0,41$).

Limites

Cette étude rencontre plusieurs limites. Tout d'abord, il n'a pas toujours été possible d'échanger avec les auteurs afin d'obtenir toutes les informations complémentaires. Deuxièmement, la recherche des études et l'extraction des données n'ont été réalisées que par la première auteure ce qui n'a pas rendu possible le calcul d'un accord inter-juges. Troisièmement, il est important de garder en tête que la pondération des critères de validité théorique est basée sur la présence ou l'absence de preuves. Le lecteur doit donc être vigilant quant aux appuis théoriques mis en avant par les auteurs. Pour terminer, les recommandations cliniques formulées ne sont pas prescriptives et doivent être prises avec du recul. Ces recommandations sont dégagées indépendamment de la situation clinique rencontrée et permettent uniquement au clinicien de prendre en compte le facteur psychométrique dans son processus décisionnel. De plus, il est important de rappeler qu'un outil obtenant une note faible n'est pas forcément un outil n'étant pas valide sur le plan clinique mais il peut s'agir

d'un outil n'ayant pas fait l'objet d'études de validation. Prenons l'exemple de Diabolab évaluation, cet outil est basé sur une réflexion rigoureuse sur le plan des critères d'analyse et de la conception des épreuves. Il offre la possibilité de réaliser des analyses poussées partir d'échantillon de parole court que nul autre outil n'offre. Néanmoins, la validité et la fidélité de cette démarche doivent encore être démontrées.

Conclusion

Les résultats de cette étude permettent de fournir un répertoire et une analyse critique détaillée des outils d'évaluation de la parole destinés aux enfants francophones d'âge préscolaire. Sa particularité est d'intégrer des outils souvent méconnus du grand public car ils ont été conçus par des chercheurs en étant peu ou pas diffusés. Au vu de l'absence de différences entre les tâches issues de ces deux cadres de conception différents, cela ouvre de nouvelles perspectives intéressantes pour de nombreux orthophonistes. Afin de prendre des décisions éclairées, l'orthophoniste pourra se référer aux tableaux et au matériel supplémentaire mis à disposition en accès libre afin de soutenir sa sélection d'un ou plusieurs outils selon les objectifs de son évaluation. Toutefois, au regard des résultats obtenus, il peut être rapidement décourageant de faire face à un manque majeur d'outils validés que cela soit pour proposer une évaluation de première ligne que pour comprendre finement le profil du patient. Du point de vue de la recherche, il est primordial de réaliser des études évaluant les qualités psychométriques des outils disponibles et de proposer des outils évaluant des domaines non investigués jusqu'ici tel que la stimulabilité à l'aide d'une évaluation dynamique (Glaspey & Stoehl-Gammon, 2007). En attendant, comme souvent, l'orthophoniste devra faire preuve de créativité et de méthodes afin de répondre aux besoins de ses patients avec des outils « maison » comme cela est le cas depuis de nombreuses années (Limbrick et al., 2013).

Bibliographie

- ASHA. (2020). Practice portal - Clinical topics : Speech Sound Disorders - Articulation and Phonology. Retrieved March 17, 2020, from <https://www.asha.org/practice-portal/clinical-topics/articulation-and-phonology/>
- Aubry, N., Bissonnette, I., Raymond, S., Perron, M., & Duchesne, L. (2021). Inventaire des structures syllabiques chez l'enfant francophone : un outil pour planifier l'intervention en dyspraxie verbale. *Revue Canadienne d'Orthophonie et d'Audiologie*, 45(3), 189–207.
- Bates, S., Titterington, J., & Child Speech Disorder Research Network, UK. A. I. (2021, Oct 15). *Good Practice Guidelines for the Analysis of Child Speech (2nd edition)*. Ulster University.
- Bérubé, D., Bernhardt, B. M., & Stemberger, J. P. (2013). Un Test de Phonologie du Français : construction et utilisation. *Revue Canadienne d'orthophonie et d'audiologie*, 37(1), 26–40.
- Burnay, J., Grégoire, J., Monseur, C., & Willems, S. (2024). Lutter contre les freins à l'Evidence-Based Assessment : Création d'une grille pour examiner la qualité des outils d'évaluation en psychologie, neuropsychologie et orthophonie. *ANAE*, XX(XX), XX–XX.
- Cattini, J., Duboisindien, G., & Maillart, C. (in prep). *L'évaluation de la parole chez les enfants d'âge préscolaire : inventaire des outils francophones disponibles*.
- Cattini, J., & Willems, S. (June 2024). Analyse des pratiques professionnelles et du sentiment de compétences des orthophonistes francophones pour une sélection raisonnée des tests normés. *ANAE: Approche Neuropsychologique des Apprentissages chez l'Enfant*, 190, 237-246.
- Cattini, J., & Lafay, A. (June 2024). Analyse critique des caractéristiques psychométriques des batteries de performances évaluant la cognition mathématique. *ANAE: Approche Neuropsychologique des Apprentissages chez l'Enfant*, 190, 280-290.
- Coquet, F., Ferrand, P., Roustit, J. (2009). EVALuation du développement du Langage Oral chez l'enfant de 2 ans 3 mois à 6 ans 3 mois. (EVALO 2-6). Isbergues : Ortho Edition.
- Daub, O., Cunningham, B. J., Bagatto, M. P., Johnson, A. M., Kwok, E. Y., Smyth, R. E., & Cardy, J. O. (2021). Adopting a Conceptual Validity Framework for Testing in Speech-Language Pathology. *American Journal of Speech-Language Pathology*, 30(4), 1894–1908. https://doi.org/10.1044/2021_AJSLP-20-00032
- de Agostini, M., Metz-Lutz, M., van Hout, A., Chavance, M., Deloche, G., Pavao-Martins, I., & Dellatolas, G. (1998). Batterie d'évaluation du langage oral de l'enfant aphasique (ELOLA) : standardisation française (4-12 ans). *Revue de Neuropsychologie*, 8(3), 319–367.
- Duboisindien, G., & Cattini, J. (sous presse). L'expertise et le raisonnement cliniques de l'orthophoniste dans la pratique évaluative. In Collège Français d'Orthophonie (Ed.), *Orthophonie : Identité professionnelle et expertise*. France : De Boeck Supérieur.

- Durieux, N., Pasleau, F., Piazza, A., Donneau, A.-F., Vandenput, S., & Maillart, C. (2016). Information behaviour of French-speaking speech-language therapists in Belgium: results of a questionnaire survey. *Health Information & Libraries Journal*, 33(1), 61–76. <https://doi.org/10.1111/hir.12118>
- Eadie, P., Morgan, A., Ukoumunne, O. C., Ttofari Eecen, K., Wake, M., & Reilly, S. (2015). Speech sound disorder at 4 years: prevalence, comorbidities, and predictors in a community cohort of children. *Developmental Medicine & Child Neurology*, 57(6), 578–584. <https://doi.org/10.1111/dmcn.12635>
- Ferrand, P. (1978). Etalonnage de l'épreuve de logatomes. *Revue de Laryngologie*, 99(7–8), 453–458.
- Gaul Bouchard, M.-E., Fitzpatrick, E. M., & Olds, J. (2009). Analyse psychométrique d'outils d'évaluation utilisés auprès des enfants francophones. *Revue Canadienne d'orthophonie et d'audiologie*, 33(3), 129–139.
- Glaspey, A., & Stoel-Gammon, C. (2007). A dynamic approach to phonological assessment. *Advances in Speech Language Pathology*, 9(4), 286–296. <https://doi.org/10.1080/14417040701435418>
- Piérart, B., Comblain, A., Grégoire, J., & Mousty, P. (2010). *ISADYLE : Batterie de tests*. De Boeck Supérieur.
- Helloin, M.-C., & Thibault, M.-P. (2006). *Exalang 3-6*. Campbell: Happy Neuron Pro.
- Korkman, M., & Kirk, U. (2012). *NEPSY-II - Bilan neuropsychologique de l'enfant - 2nde édition*. Canada : ECPA.
- 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), 127–144.
- Leclercq, 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(131).
- Limbrick, N., McCormack, J., & McLeod, S. (2013). Designs and decisions: The creation of informal measures for assessing speech production in children. *International Journal of Speech-Language Pathology*, 15(3), 296–311. <https://doi.org/10.3109/17549507.2013.770552>
- Macchi, L., Herman, F., Colli-Vaast, L., Merle, A., & Danchin, P. (2023). Propriétés psychométriques des tests francophones de langage oral chez l'enfant. *Études de Linguistique Appliquée : Revue de Didactologie Des Langues-Cultures*.
- Macchi, L., Descours, C., Girard, E., Guitton, E., Timmermans, N., & Boidein, F. (2012). *ELDP: Epreuve Lilloise de Discrimination Phonologique destinée aux enfants de 5 ans à 11;6 ans*.

- Macleod, A. A. N., Sutton, A., Sylvestre, A., Thordardottir, E., & Trudeau, N. (2014). Outil de dépistage des troubles du développement des sons de la parole: Bases théoriques et données préliminaires. *Revue Canadienne d'orthophonie et d'audiologie*, 38(1), 40–57.
- Maillart, C., & Schelstraete, M.-A. (2004). L'évaluation des troubles phonologiques : illustration de la démarche linguistique par la présentation d'épreuves qualitatives. In *Les troubles du langage et du calcul chez l'enfant* (pp. 113–147).
- Masson, Pauline. *Etude de la spécificité d'épreuves langagières chez des enfants présentant une dyspraxie verbale versus un trouble phonologique*. Unpublished master's thesis, Université de Liège, Liège, Belgique, 2017. <https://matheo.uliege.be/handle/2268.2/2321>
- McCauley, R., & Swisher, L. (1984). Use and misuse of norm-referenced test in clinical assessment. *Journal of Speech and Hearing Disorders*, 49(4), 338–348. <https://doi.org/10.1044/jshd.4904.338>
- McLeod, S., Harrison, L. J., & McCormack, J. (2012). The intelligibility in context scale: Validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research*, 55(2), 648–655. [https://doi.org/10.1044/1092-4388\(2011/10-0130\)](https://doi.org/10.1044/1092-4388(2011/10-0130))
- Meloni, G., Loevenbruck, H., Vilain, A., Gillet-Perret, E., & MacLeod, A. (2022). Evaluation de la perception des sons de parole chez les populations pédiatriques: Réflexion sur les épreuves existantes. *Glossa*, 132, 1–27.
- Meloni, G. (2022). *Evaluation des troubles du développement des sons de parole dans le cadre d'un modèle psycholinguistique chez l'enfant francophone*. Thèse de doctorat, Université de Grenoble Alpes & Université de Montréal. <https://doi.org/1866/28078>
- Anne Menin-Sicard, Etienne Sicard. DIADOLAB 3 - Logiciel d'évaluation et de rééducation de la parole - Manuel d'utilisation. 2019. hal-02073314
- Niederberger, N., Winkler-Kehoe, M., & Bouchut, A.-L. (2021). DRAP : un test de Dépistage Rapide de l'Articulation et de la Parole pour les enfants de 3 à 6 ans. *Glossa*, 131, 33–58.
- Parisse, C., & Soubeyrand, E. (2003). Computerized test for oral language perception: TIPL0. *Journal of Multilingual Communication Disorders*, 1(1), 53–61. <https://doi.org/10.1080/1476967031000106522>
- Paul, M., & Rvachew, S. (2008). *Test Francophone de Phonologie*. Université McGill. https://www.mcgill.ca/child-phonology/files/child-phonology/TFP_for_website.pdf
- Poncellet, M., & van der Linden, M. (2003). L'évaluation du stock phonologique de la mémoire de travail : élaboration d'une épreuve de répétition de non mots pour population francophones. *Revue de Neuropsychologie*, 13(3), 377–407.
- Powell, T. W., & Miccio, A. W. (1996). Stimulability: A useful clinical tool. *Journal of Communication Disorders*, 29(4), 237–253. [https://doi.org/10.1016/0021-9924\(96\)00012-3](https://doi.org/10.1016/0021-9924(96)00012-3)
- Rvachew, S., Marquis, A., Brosseau-Lapré, F., Paul, M., Royle, P., & Gonnerman, L. M. (2012). Speech articulation performance of francophone children in the early school years: Norming of the Test de Dépistage Francophone de Phonologie. *Clinical*

Linguistics & Phonetics, 27(12), 950–968.
<https://doi.org/10.3109/02699206.2013.830149>

- Schwob, S., & Skoruppa, K. (2021). L'identification du trouble développemental du langage oral chez les enfants bilingues français-portugais : les ressources à disposition des orthophonistes. *Revue Canadienne d'orthophonie et d'audiologie*, 45(4), 251–266.
- Stackhouse, J., & Wells, B. (1997). *Children's speech and literacy difficulties: A psycholinguistic framework*. Wiley.
- Terband, H., Maassen, B., & Maas, E. (2019). A psycholinguistic framework for diagnosis and treatment planning of Developmental Speech Disorders. *Folia Phoniatrica et Logopaedica*, 71(5–6), 216–227. <https://doi.org/10.1159/000499426>
- Thibault, M.-P., Helloin, M.-C., & Croteau, B. (2010). *Exalang 5-8, Batterie Informatisée d'Examen du Langage Oral et Ecrit chez l'enfant de 5 à 8 ans version 2*. Mont-Saint-Aignan, France : Orthomotus.
- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, 54(2), 580–597. [https://doi.org/10.1044/1092-4388\(2010/09-0196\)](https://doi.org/10.1044/1092-4388(2010/09-0196))
- Warnier, M., Maillart, C., Rose, Y., & MacLeod, A. A. N. (2022). Exploring word production in three-year-old monolingual French-speaking children. *Clinical Linguistics & Phonetics*, 1–19. <https://doi.org/10.1080/02699206.2022.2092424>
- Youngstrom, E. A., van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24(4), 331–363. <https://doi.org/10.1111/cpsp.12207>

ANNEXE A – Relevé synthétique des preuves de validité et de fidélité

Tableau A1. Relevé synthétique des preuves de validité et de fidélité des batteries de performances

Nom de l'outil	Validité						Fidélité				Score	
	Théorique		Structure interne	Relations avec d'autres variables	Prédictive	Sensibilité et spécificité	Sensibilité au changement	Temporelle	Interjuge	Consistance interne	Distribution	Erreur standard de mesure
	Epreuves	Items										
Diabolab 3 évaluation	-	+	-	-	-	-	-	-	-	-	-	-
ELOLA	+	+	-	-	-	-	-	-	-	-	-	-
Evalo 2-6	+	-	-	-	-	-	-	≥ 0,80	≥ 0,80	+	+	+
Exalang 3-6	-	-	-	-	-	-	+/-	≥ 0,90	-	-	+	+
Exalang 5-8	-	-	-	-	-	+/-	+/-	≥ 0,80	+/-	-	+	+
ISADYLE	+	+	-	-	-	-	-	-	-	≥ 0,80	+	+
NEPSY-II	+	-	-	-	-	-	-	-	-	-	+	-
	4/7	3/7	0/7	0/7	0/7	1/7	2/7	2/7	2/7	2/7	5/7	4/7
	57%	43%	0%	0%	0%	14%	29%	29%	29%	29%	71%	57%

Légende. Diabolab 3 évaluation (Menin-Sicard & Sicard, 2006) ; ELOLA : Batterie d'évaluation du langage oral de l'enfant aphasique (De Agostini et al., 1998) ; Evalo 2-6 : Batterie d'évaluation du langage de l'enfant de 2 à 6,3 ans (Coquet et al., 2009) ; Exalang 3-6 : Examen des troubles du langage et de la parole 3-6 (Helloin & Thibault, 2006) ; Exalang 5-8 : Examen des troubles du langage et de la parole 5-8 (Helloin et al., 2010) ; ISADYLE : ISADYLE : Instruments pour le Screening et l'Approfondissement des Dysfonctionnements du Langage (Piérart et al., 2010) ; NEPSY-II : Bilan neuropsychologique de l'enfant - 2nde édition (Korkman & Kirk et al., 2012)

Tableau A2. Relevé synthétique des preuves de validité et de fidélité des tâches de performances

Nom de l'outil	Validité						Fidélité				Score	
	Théorique		Structure interne	Relations avec d'autres variables	Prédictive	Sensibilité et spécificité	Sensibilité au changement	Test-retest	Inter-juges	Consistance interne	Distribution	Erreur standard de mesure
	Epreuves	Items										
CL-NWR	+	+	-	-	-	+	+	-	≥ 0,90	-	-	+
Décision lexicale de Maillart	+	+	-	-	-	-	+	-	-	-	-	-
Dénomination d'images de Martinez Perez	+	+	-	-	-	+	-	-	-	-	-	-
Dénomination rapide de Masson	+	+	-	-	-	+	-	-	-	-	-	-

Nom de l'outil	Validité						Fidélité			Score		
	Théorique		Structure interne	Relations avec d'autres variables	Prédictive	Sensibilité et spécificité	Sensibilité au changement	Test-retest	Inter-juges	Consistance interne	Distribution	Erreur standard de mesure
	Epreuves	Items										
Discrimination auditive de Maillart	+	+	-	-	-	-	+	-	-	-	-	-
Discrimination phonémique de Martínez Perez	+	+	-	-	-	+	-	-	-	-	-	-
DRAP	-	+	-	+	-	-	+	-	≥ 0,80	≥ 0,80	-	+
Epreuve de diadococinésies de Martínez Perez	+	+	-	-	-	+-	-	-	-	-	-	-
ICS	+	+	-	-	-	-	-	-	NA	-	-	-
ELDP	+	+	+	-	-	-	-	-	-	≥ 0,70	+	+
Epreuve des logatomes de Ferrand	-	-	-	-	-	-	-	-	-	-	+	-
ESPP	+	+	-	-	-	+	+	-	≥ 0,80	-	-	+
Eulalies – version courte	+	+	-	-	-	-	+	-	≥ 0,80	-	+	+
Histoire de Thomas	+	+	-	-	-	+	-	-	-	-	-	-
ISSEF	+	+	-	-	-	-	-	-	-	-	-	-
Jugement phonologique de Martínez Perez	+	+	-	-	-	+	-	-	-	-	-	-
Répétition de non-mots de Courcy	+	+	-	+	-	+	-	-	-	-	-	-
Répétition de non-mots de Poncelet	+	+	-	+	-	+	+	≥ 0,80	-	-	-	+
Répétition d'images de Martínez Perez	+	+	-	-	-	+	-	-	-	-	-	-
TDFP	+	+	-	-	-	-	+	-	≥ 0,90	-	-	+
TFP	+	+	-	-	-	-	+	-	≥ 0,90	-	-	+
TIPLO	-	+	-	-	-	-	-	-	-	-	-	-
TPF - Standard	+	+	-	-	-	-	-	-	-	-	-	-
TPF – Manitoba	+	+	-	-	-	-	-	-	≥ 0,90	-	-	+

Nom de l'outil	Validité					Fidélité				Score		
	Théorique		Structure interne	Relations avec d'autres variables	Prédictive	Sensibilité et spécificité	Sensibilité au changement	Test-retest	Inter-juges	Consistance interne	Distribution	Erreur standard de mesure
	Epreuves	Items										
TPFC-D	+	+	-	-	-	-	-	-	≥ 0,80	-	-	+
	21/24	23/24	1/25	3/25	0/25	11/25 10/25	9/25	1/25	8/24	2/25	3/25	10/25
	88%	96%	4%	12%	0%	44% 40%	36%	4%	33%	8%	12%	40%

Légende. NA : Non applicable. CL-NWR : Crosslinguistic Nonword Repetition Test (Schwob & Skoruppa, 2022) ; Décision lexicale de Maillart (Maillart & Schelstraete, 2004) ; Dénomination d'images de Martinez Perez (Masson, 2017) ; Dénomination rapide de Masson (Masson, 2017) ; Discrimination auditive de Maillart (Maillart & Schelstraete, 2004) ; Discrimination phonémique de Martinez Perez (Masson, 2017) ; DRAP : Dépistage Rapide de l'Articulation et de la Parole (Niederberger et al., 2021) ; Epreuve de diadococinésies de Martinez Perez (Masson, 2017) ; ICS : Échelle d'intelligibilité en contexte (McLeod et al., 2012) ; ELDP : Épreuve Lilloise de Discrimination Phonologique (Macchi et al., 2012) ; ESPP : Évaluation sommaire de la phonologie chez les enfants d'âge préscolaire (MacLeod et al., 2014) ; Epreuve des logatomes de Ferrand (Ferrand, 1978) ; Eulalies – version courte (Warnier et al., 2022) ; Histoire de Thomas (Masson, 2017) ; ISSEF : Inventaire des Structures Syllabiques chez l'Enfant Francophone (Aubry et al., 2021) ; Jugement phonologique de Martinez Perez (Masson, 2017) ; Répétition de non-mots de Courcy (Thordardottir et al., 2011) ; Répétition de non-mots de Poncelet (Poncelet & Van der Linden, 2003) ; Répétition d'images de Martinez Perez (Masson, 2017) ; TDFP : Test de Dépistage Francophone de Phonologie (Rvachew et al., 2012) ; TFP : Test Francophone de Phonologie (Paul & Rvachew, 2008) ; TIPLO : Test Informatisé pour la Perception du Langage Oral (Parisse & Soubeyrand, 2003) ; TPF : Test de Phonologie du Français (Bérubé et al., 2013) ; TPFC-D : Test de Phonologie du Français Canadien-Dépistage (Bérubé et al., 2013)