

Évaluation de la perception des sons de parole chez les populations pédiatriques : réflexion sur les épreuves existantes

Geneviève MELONI *, Hélène LOEVENBRUCK **, Anne VILAIN ***, Estelle GILLET-PERRET ****, Andrea MACLEOD *****

* Univ. Grenoble Alpes, LPNC et Grenoble INP, GIPSA-lab, 38000 Grenoble, France et Univ. de Montréal, Ecole d'orthophonie et d'audiologie, Montréal, Québec, Canada

** Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

*** Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

**** Centre Référent des Troubles du Langage et des Apprentissages, CHU Grenoble Alpes, 38000 Grenoble, France

***** Department of Communication Sciences & Disorders, Faculty of Rehabilitation Medicine, University of Alberta, Alberta, Canada

Auteure de correspondance :

genevieve.meloni@univ-grenoble-alpes.fr

Résumé :

Les enfants avec un Trouble du Développement des Sons de Parole (TDSP) présentent des difficultés dans l'acquisition des sons de la langue et une faible intelligibilité. Cela peut engendrer des difficultés scolaires et socio-économiques qui peuvent persister jusqu'à l'âge adulte. La prévalence des TDSP étant prédominante dans la patientèle des orthophonistes (Dodd, 2014), il est nécessaire de disposer de protocoles d'évaluation fiables. L'enjeu de cet article est de réaliser un inventaire des outils disponibles dans la clinique française pour évaluer la perception de la parole dans la population pédiatrique et de proposer une comparaison de ces outils en termes de caractéristiques psychométriques et de qualités propres à l'évaluation de la perception. L'objectif est de donner des clés aux orthophonistes pour le choix de leurs outils d'évaluation dans le cadre d'une pratique fondée sur les données probantes. Dans cette étude, nous avons recensé dix tâches utilisées auprès des enfants francophones. Nous avons examiné deux aspects de ces tests grâce à des grilles d'analyse. Premièrement, nous avons analysé le versant psychométrique selon une grille adaptée de Leclercq et Veys (2014) et Lafay et Cattini (2018). Puis, nous avons évalué les caractéristiques d'une tâche d'évaluation de la perception avec une grille reprenant certains critères de Locke (1980a). Les résultats démontrent que, même si les tâches sont adaptées aux âges des enfants et rapides à administrer, très peu d'études justifient le choix de leurs cibles ou établissent un lien entre les habiletés de perception et de production. Également, les tests ne satisfont pas les critères psychométriques, notamment ceux de sensibilité et spécificité. En effet, rares sont les tests qui incluent une population pathologique dans leur processus de standardisation. Cette étude encourage la pratique fondée sur les données probantes et propose de guider les orthophonistes dans leurs choix d'outils diagnostics.

Mots clés : Trouble du développement des sons de parole, évaluation, perception de parole, enfants francophones.

Speech sound perception assessment in paediatric population: considerations on existing tasks**Summary:**

Children with Speech Sound Disorders (SSD) have difficulties in the acquisition of speech sounds, and poor speech intelligibility. These difficulties have negative academic and socio-economic repercussions, often persisting until adulthood. There is a need for reliable assessment protocols, since the prevalence of speech sound disorders speech-language pathologists' caseloads is high (Dodd, 2014). The focus of this article is an inventory of French clinical tools for speech perception assessment in pediatric population and a comparison of these tools in terms of psychometrics and inherent characteristics of an efficient perception task. The goal is to provide speech-language pathologists with a framework for the selection of assessment protocols in an Evidence-Based Practice approach. In this study, we inventoried ten main perception tasks used for French-speaking children. We examined two aspects of these tests by using analysis grids. First, we analyzed the psychometric side with a grid adapted from Leclercq & Veys (2014) and Lafay & Cattini (2018). Second, we evaluated the characteristics of a perception task with a criterion grid adapted from Locke (1980a). The results showed that, even if the tests are adapted to children of different ages and quick to administer, very few include a justification for the selected targets or establish a link between perception and production abilities. In addition, these tests do not satisfy psychometric standards, especially sensibility

and specificity. In fact, only a few tests include a pathologic population in their standardization process. This study promotes Evidence-Based Practice approach and offers to guide clinicians in the selection of diagnostic tests.

Keywords: Speech Sound Disorders, Assessment, Speech perception, French-speaking children.

----- INTRODUCTION -----

Les Troubles du Développement des Sons de Parole (TDSP, en anglais Speech Sound Disorders) correspondent à un délai d'acquisition des sons de la langue qui entraîne une intelligibilité réduite (Baker & McLeod, 2011). Ce délai d'acquisition recouvre une combinaison de déficits à différents niveaux : des difficultés au niveau linguistique/cognitif avec une altération des représentations des sons et des règles phonotactiques, des difficultés de planification des mouvements orophonatoires, et/ou des difficultés du traitement perceptif (Dodd & McInstosh, 2008 ; International Expert Panel on Multilingual Children's Speech, 2012).

Les TDSP intéressent la clinique orthophonique pour plusieurs raisons. D'une part, ils représentent le trouble de la communication le plus fréquent dans la population pédiatrique (Dodd, 2014) ainsi que le trouble le plus représenté au sein des cabinets d'orthophonie (Broomfield & Dodd, 2004). D'autre part, les conséquences à long terme d'un TDSP peuvent être majeures et perdurer jusqu'à l'âge adulte, gênant notamment l'intégration dans le monde du travail (Felsenfeld, Broen et McGue, 1994 ; Pennington & Bishop, 2009).

Face à cette réalité, diagnostiquer les TDSP et établir le profil phonologique de l'enfant sont des enjeux essentiels pour son devenir. Pour être fiable, l'évaluation orthophonique doit s'appuyer sur des épreuves étalonnées sur la population générale (Haute Autorité de Santé (ANS), 2001). Cette démarche d'utilisation de tests standardisés s'intègre plus largement dans une volonté de fonder les pratiques des orthophonistes sur des données probantes (*Evidence-Based Practice (EBP)*, en anglais). En effet, les praticiens sont fortement encouragés à établir leurs décisions cliniques à partir de données issues de la recherche scientifique, de leur expérience de thérapeute et du profil singulier du patient (Dollaghan, 2007). Or, certains auteurs montrent que le choix des tests d'évaluation n'est généralement pas guidé par les qualités psychométriques de ceux-ci, le seul critère positivement corrélé à la préférence des outils d'évaluation étant la récence de la date de publication de l'outil (Betz et al., 2013). Face à ce constat, l'objectif de cette étude est d'offrir des clés aux cliniciens pour le choix des tests employés lors du diagnostic des enfants présentant un TDSP, dans le domaine particulier des épreuves évaluant les capacités perceptives des enfants. Nous abordons l'axe perceptif car, en comparaison au versant production, celui-ci bénéficie de moins d'attention de la part des chercheurs, ce qui a des répercussions en clinique. McLeod et Baker (2014) soulignent que seulement 15% des orthophonistes évaluent systématiquement les habiletés perceptives des enfants avec un TDSP, alors même que le nombre d'études démontrant d'un lien entre la production et perception de la parole ne cesse de grandir (MacAllister Byun & Tiede, 2017). Dans ce contexte, il apparaît comme essentiel de disposer de tâches efficaces pour l'évaluation de la perception de la parole.

1. Les capacités perceptives des enfants présentant un trouble du développement des sons de parole

Les Troubles du Développement des Sons de Parole constituent un déficit très fréquent chez les enfants en dessous de cinq ans (McLeod & Harrison, 2009) qui se caractérise principalement par un manque d'intelligibilité. Le terme de TDSP est un terme parapluie car il englobe des profils de sévérité et d'étiologie très différents (Nathan, 2001, Shriberg et al., 2010, Dodd, 2014, Bowen, 2015). Malgré le manque de consensus sur la classification des TDSP (Waring & Knight, 2013), nous retenons trois catégories principales :

- (1) Le trouble/retard phonologique constant qui est marqué par la présence de processus

- phonologiques caractéristiques des enfants plus jeunes, ou atypiques (Dodd, 2014),
- (2) Le trouble d'articulation qui correspond à une distorsion ou une substitution systématique d'un phonème ou d'une catégorie de phonèmes (Dodd, 2014),
 - (3) La dyspraxie verbale qui est considérée comme liée à un déficit du contrôle moteur de la parole (déficit au niveau de la planification et de la programmation des mouvements de la parole) (ASHA, 2007, Dodd, 2014).

La première question qui nous intéresse est de savoir si les enfants avec un TDSP démontrent des difficultés de perception de la parole. Peu d'études ont été menées et leurs résultats ne sont pas concordants comme le montre la méta-analyse de Hearnshaw et al., 2019. Plusieurs raisons interviennent dans cette absence de consensus (Hearnshaw et al., 2018 ; Hearnshaw et al., 2019). Le premier argument invoqué est celui du changement d'orientation de la recherche en phonologie qui, depuis les années 1970, a mis l'accent sur les processus phonologiques, écartant des domaines d'intérêt l'analyse du versant perceptif. Un second argument est celui de la diversité des méthodologies d'évaluation des habiletés perceptives. En effet, il a été suggéré qu'il serait nécessaire de placer l'enfant dans des conditions adverses (haut-parleurs, bruit environnant, multiples locuteurs, stimuli synthétiques) pour faire émerger les difficultés de perception de la parole (Rvachew, 2015). Ces déficits ne seraient pas démontrés dans des conditions d'écoute idéales (cabinet d'orthophonie, en situation duelle, avec peu de bruit environnant). C'est le cas, par exemple, des résultats de l'étude de Locke (1980b) qui propose une tâche de jugement de lexicalité à 131 enfants âgés de 3;1 à 9;11 ans. Il est demandé à l'enfant de juger si l'orthophoniste dit le mot correctement (le mot existe et est bien prononcé) ou non (le mot est prononcé avec une substitution). Les items de la tâche sont construits, de façon spécifique, pour chaque enfant, à partir des substitutions effectuées au cours d'une tâche antérieure de dénomination (procédure *Speech Production-Perception Task*, Locke 1980a et b). Les stimuli sont présentés en face à face et à voix haute par l'expérimentateur. Les résultats démontrent que deux tiers des enfants testés discriminent correctement les segments qu'ils produisent pourtant avec difficulté. Dans cette étude, la majorité des enfants ne présente pas de difficultés de perception.

D'autres travaux, au contraire, démontrent des difficultés perceptives chez les enfants avec un TDSP. Edwards et al. (1999) soumettent des enfants typiques et des enfants avec un TDSP, âgés de 3;8 à 5;4 ans, à une tâche de désignation d'images en présentant des mots tronqués (paradigme de « gating ») ou diffusés dans différentes conditions de bruit. Comparés aux enfants typiques, les enfants avec un TDSP ont davantage de difficultés à reconnaître les mots présentés dans des conditions de bruit ou dont le signal acoustique a été tronqué. D'autres recherches se sont concentrées sur la perception par les enfants avec un TDSP de la variabilité entre les locuteurs. Hearnshaw et al., en 2018, ont mené une étude auprès de 12 enfants avec un TDSP et de 13 enfants tout-venant, entre 4 et 5 ans, en utilisant le logiciel SAILS (Rvachew et al., 2004). Le programme SAILS est un logiciel informatique qui teste le jugement de lexicalité. Il est demandé à l'enfant de juger si le mot qu'il entend est correct ou non. Les mots du logiciel ont été enregistrés auprès d'adultes typiques, d'enfants typiques et d'enfants ayant un TDSP. De ce fait, certains des mots, enregistrés auprès d'enfants avec un TDSP, comportent des distorsions ou des substitutions sur les consonnes en position initiale et finale de mots. Les résultats de Hearnshaw et al. (2018) montrent que les enfants avec un TDSP présentent des scores de jugement perceptif correct significativement moins élevés que les enfants tout-venant et que les scores globaux de production sont corrélés à ceux de perception.

Ces difficultés de perception observées chez les enfants avec un TDSP semblent être associées à des spécificités sur le plan neuroanatomique et neurofonctionnel. Certaines recherches

mettent en lien les déficits comportementaux avec des différences d'anatomie ou d'activation cérébrale. Des études de neuroimagerie fonctionnelle récentes révèlent des patrons d'activation cérébrale particuliers chez les enfants avec un TDSP lors du traitement de mots ou de pseudomots, avec, comparés aux enfants contrôles, une activation accrue des régions dorsales de perception de la parole et une moindre activation des régions ventrales (e.g. Preston et al., 2012). Sur le plan neuroanatomique, l'étude de Preston et al. (2014) compare les structures cérébrales à partir d'imagerie par résonance magnétique (IRM) chez un groupe de 54 enfants tout-venant et de 23 enfants présentant des erreurs de parole résiduelles (sous-type de TDSP caractérisé par des erreurs sur un ou quelques phonèmes, séquelles d'un TDSP plus important). Les sujets de leur étude sont âgés de 8 à 11 ans. Leurs résultats mettent notamment en lumière un excès de matière grise au niveau des gyri temporaux supérieurs bilatéraux (gyrus de Heschl). Il a été montré que les régions temporales supérieures sont impliquées notamment dans le traitement acoustico-phonétique (Hickok et al., 2011). Considérant cela, Preston et al. (2014) indiquent que : « *a possible underlying mechanism associated with residual SSEs might involve reduced synaptic pruning of regions that are responsible for fine-grained phonetic perception [...]*¹ » (p. 30). En d'autres termes, au cours du développement cérébral des enfants avec un TDSP, l'élagage synaptique aurait été moins efficient dans les régions temporales. Ce manque d'affinage des réseaux spécifiques au traitement de la parole pourrait être un facteur sous-jacent aux difficultés de perception de parole observées chez les enfants avec un TDSP.

Une seconde question est celle de savoir dans quelle mesure le trouble de perception est responsable des difficultés de production. Pour examiner cette question, Jamieson et Rvachew (1992) ont élaboré une tâche qui permet un entraînement à l'identification de contrastes phonétiques entre consonnes fricatives, en partant de contrastes très marqués puis en manipulant le contenu spectral des consonnes de façon à obtenir des contrastes de plus en plus fins. Cet entraînement a été proposé à six enfants entre cinq et sept ans avec un TDSP pendant 20 sessions de pratique de six minutes. Leurs résultats montrent une amélioration significative des habiletés de perception, qui se transfère pour trois enfants aux habiletés en production de parole. Ce lien perception/production est également présent chez Barrozo et al., (2016) qui démontrent une corrélation entre la sévérité du TDSP et la présence d'un déficit de traitement auditif central chez des enfants entre sept et dix ans. Les sujets présentant à la fois un TDSP et un déficit du traitement auditif central ont de moins bonnes performances aux tâches de dénomination d'images et de jugement de rimes. L'ensemble de ces études indiquerait donc que les enfants avec un TDSP, ou du moins une partie d'entre eux, démontrent un déficit au plan de la perception de parole et que ce déficit pourrait être relié à leurs erreurs en production (Hearnshaw et al., 2019).

2. Réflexion sur les caractéristiques psychométriques des tests

Dans le cadre d'une démarche en lien avec une pratique fondée sur des données probantes, (*Evidence-Based Practice*), les praticiens sont encouragés à orienter leurs décisions cliniques en fonction de preuves. Ces preuves doivent être récoltées à partir de trois sources : l'expérience professionnelle, la situation particulière du patient et les données issues de la recherche scientifique. Dans le cadre d'une démarche diagnostique, il convient de répartir les patients selon différentes classifications. Ces classifications des troubles peuvent être basées sur l'étiologie (classification de Shriberg et al., 2010), sur les caractéristiques linguistiques (classification de Dodd 2014) ou sur les processus psycholinguistiques (Stackhouse et al.,

¹« un mécanisme sous-jacent possiblement associé aux erreurs résiduelles pourrait impliquer un élagage synaptique réduit des régions responsables de la perception fine des contrastes phonétiques [...] » (notre traduction)

2007). Il est donc nécessaire d'utiliser un outil qui permette de classer de façon optimale le patient, c'est-à-dire d'identifier si le patient présente un trouble et de quel type de trouble il s'agit (Dollaghan, 2004). C'est ce que l'on appelle le pouvoir diagnostic (Leclercq & Veys, 2014). Pour détenir un certain pouvoir diagnostique, les tests doivent concentrer cinq types de caractéristiques psychométriques : la fidélité, la validité, la sensibilité/spécificité, la standardisation et la normalisation (Lafay & Cattini, 2018).

La fidélité (ou fiabilité) peut se définir comme la constance, la stabilité du test (Belot & Tricot, 2001). Il y a plusieurs façons d'envisager la fidélité d'un test. Une première façon est de considérer la stabilité du test dans le temps. En effet, les résultats du test doivent être similaires s'ils sont administrés plusieurs fois et doivent aboutir au même diagnostic (Belot & Tricot, 2001). Il s'agit de la fidélité temporelle (ou fidélité test-retest). Pour évaluer la qualité de la fidélité test-retest de l'épreuve, on calcule un coefficient de corrélation : un indice de 0 indique qu'il n'y a pas de relation entre les scores des différentes administrations du test, et donc que le score est dû au hasard, et un indice de 1 indique une corrélation parfaite, le score de la première passation prédisant le score de la seconde (Betz et al., 2013). On considère qu'un indice de 0,90 est recommandé pour les tests diagnostiques (Salvia et al., 2010). La deuxième façon d'envisager la fidélité est celle de la fidélité inter-juges. Ainsi, le test doit donner des résultats comparables avec deux expérimentateurs différents. La stabilité du test, indépendamment de la personne qui l'a administré et corrigé, garantit des résultats les plus objectifs possibles (Lafay & Cattini, 2018). L'indice reconnu dans la littérature pour évaluer la fidélité inter-juges est un Kappa de Cohen, qui doit être au minimum de 0,60. Enfin, une troisième façon d'envisager la fidélité se fait sous l'angle de la cohérence des items. Ceux-ci doivent être appropriés pour évaluer l'aptitude testée et doivent réagir de la même façon. De ce fait, le choix des items doit être justifié par une analyse approfondie. Certains concepteurs de tests utilisent l'indice Alpha de Cronbach, qui doit être de 0,70 pour considérer que les items sont cohérents (Leclercq & Veys, 2014).

La validité est en général définie comme le fait que le test mesure effectivement ce qu'il est censé mesurer. La compétence que le test est censé mesurer est appelée le construit (Gaul Bouchard et al., 2009). Là encore, cette notion peut se mesurer avec des indices différents. Le premier indice concerne la validité du contenu (ou validité théorique). Il est indiqué que le test doit se référer à un modèle théorique validé et argumenté. Le test doit ainsi être validé dans le cadre de ce modèle théorique (Leclercq & Veys, 2014). Un deuxième indice fait référence à la validité concourante (ou concomitante ou concordante). Il s'agit de pouvoir comparer les résultats au test en question avec les résultats obtenus à d'autres tâches du même type. Les concepteurs des tests doivent pouvoir établir des corrélations avec d'autres tâches du même acabit (Leclercq & Veys, 2014). Un troisième indice correspond à l'indice de validité prédictive. Ainsi, les résultats du test doivent être corrélés avec le fonctionnement dans la vie quotidienne, par exemple, les résultats scolaires (Leclercq & Veys, 2014). Un quatrième indice a trait à la validité de construit, c'est-à-dire au fait que la tâche doit démontrer qu'elle évalue bien le construit. Si le construit est clairement défini, il est possible de comprendre comment la tâche a été construite et comment les items ont été choisis (Gaul Bouchard et al., 2009). Dans ce contexte, la tâche doit démontrer que tous les items mesurent le même construit, et, par conséquent, la tâche doit démontrer que les items sont reliés statistiquement (validité factorielle, Lafay & Cattini, 2018). Également, la tâche doit démontrer que son score est corrélé avec les caractéristiques majeures des individus : sexe, âge, quotient intellectuel, etc., (validité en lien avec les caractéristiques de l'individu, Lafay & Cattini, 2018).

La sensibilité d'un test correspond à son degré de précision pour déterminer si un enfant présente une pathologie. La spécificité concerne le degré de précision pour déterminer qu'un enfant ne présente pas de pathologie. Une façon de garantir qu'un test permet effectivement de diagnostiquer une pathologie est d'inclure dans l'échantillon une partie de population présentant la pathologie en question (Leclercq & Veys, 2014). Cela permet de valider que les enfants porteurs de troubles sont bien identifiés par le test.

La standardisation peut se définir comme l'uniformisation des conditions de passation et de cotation des épreuves (Lafay & Cattini, 2018). Le test doit apporter une description de la qualification de l'expérimentateur, une description des consignes de passation et une description des modalités de cotation. Ces trois paramètres doivent être précisément décrits pour « limiter la subjectivité, ainsi que les erreurs de mesure ou les biais d'interprétation » (Lafay & Cattini, 2018, p. 130).

Enfin, la normalisation fait référence aux scores qu'un échantillon de la population a obtenus au test (Lafay & Cattini, 2018). Plusieurs critères doivent être pris en compte pour que cet échantillon soit pris en référence. Premièrement, l'échantillon doit être décrit avec précision et il doit être représentatif de la population vue en clinique. Deuxièmement, cet échantillon doit être suffisamment grand. Certains auteurs recommandent un échantillon de 100 sujets par tranche d'âge (Leclercq & Veys, 2014). Enfin, pour déterminer le seuil de pathologie, le test doit faire référence aux tendances centrales (moyenne et écart-type). Toutefois, pour que ces tendances soient valables, il faut que la population soit répartie selon une courbe gaussienne. Si la distribution n'est pas normale, le test doit indiquer des percentiles (Leclercq et Veys, 2014).

3. Réflexion sur les qualités propres à l'évaluation de la perception

En plus des critères psychométriques qui s'appliquent à l'ensemble des outils d'évaluation, il nous a semblé intéressant de réfléchir à des critères qui s'appliqueraient spécifiquement aux tâches évaluant le traitement perceptif.

Locke (1980a) réalise un état des lieux des tests de la clinique anglophone et présente différents critères pour évaluer la qualité du test dans sa capacité à donner des informations sur les habiletés perceptives pertinentes pour la clinique. Locke (1980a et b), repris en partie par Rvachew (2015), définit huit critères que les tâches d'évaluation de la perception doivent contenir :

- 1) Le test devrait examiner la perception de sons en fonction des erreurs en production. Selon Locke (1980a), il est essentiel d'orienter l'évaluation du versant perceptif en fonction des caractéristiques des productions de l'enfant. Selon l'auteur, l'erreur en production proviendrait, dans les représentations phonologiques de l'enfant, d'une agglomération de deux catégories phonémiques en une seule. Par exemple, l'enfant qui produit /s/ pour /ʃ/ posséderait dans ses représentations phonologiques une seule catégorie pour les deux phonèmes. C'est cette substitution qu'il faut examiner au cours de l'évaluation du versant perceptif.
- 2) Le test devrait examiner les phonèmes dans les mêmes contextes phonétiques pour les tâches de production et de perception.

- 3) Dans une tâche de discrimination entre deux catégories, le test devrait permettre une comparaison des performances de l'enfant sur des contrastes contrôlés présentant les mêmes écarts acoustiques que les contrastes cibles. Selon l'auteur, il faut d'une part tester si l'enfant perçoit les deux catégories phonémiques et d'autre part, si l'enfant peut percevoir la faible distance acoustique qui sépare les deux catégories. En d'autres termes, il s'agit de différencier si ce sont deux sons en particulier que l'enfant ne distingue pas parce qu'il ne dispose que d'une seule catégorie de son, ou bien si le problème se situe au niveau de la perception de la faible distance acoustique. Prenons l'exemple de l'enfant qui produit indifféremment /s/ et /ʃ/, il faudrait que le test propose également un item de contrôle, qui aurait la même distance acoustique (déterminé à partir d'un arbre de confusion auditive). Cet item contrôle pourrait être /f/.
- 4) Les items du test devraient être prononcés par des adultes pour évaluer si l'enfant parvient à associer la forme adulte avec sa propre représentation interne. De ce fait, Locke (1980a) ne fait pas écouter les propres productions des enfants pour évaluer la perception.
- 5) Le test devrait présenter les items plusieurs fois, pour permettre de s'assurer que la réponse correcte n'est pas due au hasard.
- 6) Le test devrait permettre de différencier les erreurs en lien avec un déficit perceptif et les erreurs en lien avec un déficit attentionnel. Pour cela, proposer plusieurs essais permet également d'avoir une information quantitative et de limiter l'impact des erreurs d'inattention temporaires.
- 7) Le test devrait être court et requérir une réponse facile de l'enfant, en termes de capacités conceptuelles, sans mettre en jeu des compétences cognitives trop importantes.
- 8) Le test devrait évaluer la direction de la substitution. Là encore, si l'enfant perçoit /s/ pour /ʃ/, il serait important de savoir si l'enfant est capable de percevoir la différence dans la direction /ʃ/ vers /s/. Dans ce contexte, l'enfant accepterait que le /ʃ/ soit remplacé par un /s/ mais pas l'inverse : le /s/ serait bien identifié et ne pourrait pas être perçu comme un /ʃ/.

La mise en place de critères pour évaluer la qualité des tâches de perception est un apport majeur dans la clinique. Cela offre aux orthophonistes des repères dans le choix des tests qui seront proposés aux enfants. Un des intérêts des critères proposés par Locke (1980a) est de considérer l'interaction production-perception et d'envisager les représentations phonologiques comme un ensemble de connaissances, à la fois sur le versant moteur mais aussi sur le versant perceptif. Ce type de position abonde dans le sens d'une conception multidimensionnelle de la phonologie, dans laquelle le développement de la parole implique l'accroissement progressif intégré des connaissances articulatoires, perceptives et symboliques (Munson et al., 2005).

En revanche, il semble que, pour Locke (1980a), les représentations phonologiques s'organisent de façon linéaire. Ainsi, si l'enfant produit /ipopopam/ pour hippopotame, cela correspond pour l'auteur à une substitution du /t/ par le /p/ et non pas à une harmonisation au niveau de la structure syllabique. Or, dans cet exemple, on peut soutenir l'hypothèse que ce sont les autres phonèmes du mot qui altèrent la production du /t/. Si on raisonne de cette façon, on se place dans une conception multilinéaire de la phonologie car la structure syllabique peut influencer le niveau segmental (Bérubé et al., 2015). La prise en charge d'un enfant qui ferait ce type d'erreurs devrait donc cibler l'amélioration du contrôle intersyllabique, plutôt que de se limiter

à l'entraînement segmental. L'actualisation des critères proposés par Locke (1980a) devrait donc intégrer cette conception non linéaire de la phonologie.

D'autres auteurs se sont intéressés aux différents biais qui sont à contrôler lors d'une tâche de perception (Stackhouse et al., 2007). Il s'agit de facteurs confondants qui peuvent influencer le résultat de la tâche en dehors d'un déficit perceptif. Premièrement, la notion de « même/différent », qui est nécessaire pour réussir une tâche de discrimination dans laquelle il est demandé de juger si les deux stimuli présentés diffèrent ou non, ne serait acquise qu'à partir de 4 ans. Avant cet âge, les enfants ne pourraient pas porter un jugement de ce type sur les stimuli perçus. Un second biais concerne la mémoire auditivo-verbale. En effet, lorsqu'on demande à l'enfant de comparer deux stimuli, celui-ci doit conserver les informations phonémiques dans la boucle phonologique pour pouvoir faire le traitement comparatif. Si la boucle phonologique ne permet pas de maintenir en mémoire ces informations, la tâche de comparaison sera difficilement réalisée. Enfin, il faut prendre en compte la variable attentionnelle. En effet, ce type de tâche nécessite de mobiliser une attention soutenue.

----- OBJECTIFS -----

La présente étude a pour objectif de recenser et de comparer les principaux outils d'évaluation de la perception, à destination des enfants présentant un Trouble du Développement des Sons de Parole, et utilisés en clinique orthophonique en France. La comparaison des outils d'évaluation vise à déterminer leur pertinence pour diagnostiquer un déficit du versant perceptif chez l'enfant avec un TDSP. Cette comparaison se fait au travers d'une grille d'analyse qui présente l'originalité de disposer de deux axes. Le premier axe envisage les caractéristiques psychométriques des tests (selon Lafay & Cattini, 2018 ; Leclercq & Veys, 2014) et le second s'intéresse aux caractéristiques spécifiques d'un test de la perception de la parole. L'enjeu de cette étude est de donner des arguments aux orthophonistes au moment du choix de leurs outils d'évaluation.

----- MÉTHODOLOGIE -----

1. Sélection des outils d'évaluation recensés

Le choix des outils d'évaluation intégrés dans cette étude s'est appuyé sur une recherche systématique à partir des sites internet des maisons d'édition qui fournissent du matériel en orthophonie (Ortho Édition, Mot à Mot Éditions, HAPPY Neuron, Orthopratic, Pearson, De Boeck Supérieur, les Editions Passe-Temps) et des espaces de revente d'outils en orthophonie (Espace Orthophonie, Unithèque). Sur chaque site internet, nous avons orienté les recherches afin de cibler les outils d'évaluation du langage oral. Si cela n'a pas été possible, nous avons passé en revue l'ensemble du matériel présenté sur le site. Pour chaque item, nous avons lu la présentation de l'outil pour déterminer si le test pouvait être intégré dans notre étude.

Les critères d'inclusion d'un test dans l'étude étaient :

- Présence d'un test ou d'un sous-test d'évaluation de la perception de parole
- Disponible pour des enfants de 3 à 11 ans
- Présence d'un étalonnage
- Outil d'évaluation ou de diagnostic
- Outil en français

Les critères de non inclusion d'un test dans cette étude étaient :

- Absence de test ou de sous-test d'évaluation de la perception de la parole
- Outil ciblant les enfants après 11 ans et les adultes (uniquement)
- Outil sans étalonnage
- Outil destiné au dépistage
- Outil destiné à la rééducation
- Livre théorique

Notre recension a permis de collecter 159 outils dont 33 doublons. 152 outils ont été exclus car ils ne remplissaient pas les critères d'inclusion. Six outils d'évaluation ont donc été conservés dans l'étude.

À ce matériel, nous avons ajouté quatre outils que nous connaissions par notre expérience de clinicienne et qui sont fréquemment utilisés. Ces outils ne sont pas ressortis dans notre revue systématique car deux d'entre eux ne sont plus édités (BEPL-A et EDP 4-8) et deux autres ne sont pas des produits commerciaux (ELDP et TIPLO).

La liste finale des outils inclus dans cette étude est disponible dans le tableau 1.

Tableau 1

Description des outils d'évaluation de l'étude

Nom du test ou de la batterie	Auteurs	Date	Âge	Etalon-nage	Type de tâche proposée
BEPL-A	Chevrie-Muller, Simon, Le Normand et Fournier	1997	2 ans 9 mois à 4 ans 3 mois	288 enfants	Jugement de lexicalité
CLEA	Pasquet, Parbeau-Gueno et Bour	2014	2 ans 6 mois à 14 ans 11 mois	1039	Jugement de lexicalité
EDP 4-8	Autesserre, Deltour et Lacert	1988	4 à 8 ans	364 enfants	Décodage phonologique
ELDP	Macchi, Descours, Girard, Guitton, Morel, Timmermans et Boidein	2012	5 à 11 ans	274 enfants	Décodage phonologique
EVALEO	Launay, Maeder, Roustit et Touzin	2018	6 à 15 ans	1524	Décodage phonologique
EVALO	Coquet, Ferrand et Roustit	2009	2 à 6 ans	880 enfants	Jugement de lexicalité Décodage phonologique
Exalang 3-6	Helloin et Thibault	2006	2 ans 8 mois à 5 ans 10 mois	468 enfants	Décodage phonologique

Exalang 5-8	Thibault, Helloin et Croteau	2010	5 à 8 ans	377 enfants	Décodage phonologique
ISADYLE	Piérart, Comblain, Grégoire et Mousty	2010	3 à 12 ans	1144 enfants	Jugement de lexicalité
TIPLO	Parisse et Soubeyrand	2002	1 an 6 mois à 2 ans 9 mois	19 enfants	Jugement de lexicalité

2. Grille d'analyse des qualités psychométriques

Notre grille d'analyse des qualités psychométriques (tableau 2) prend principalement appui sur celle présentée par Lafay et Cattini (2018). La structuration et les indices proposés par ces auteurs nous ont paru adaptés à la comparaison de tests orthophoniques. La grille proposée s'inspire également de celles de Leclercq et Veys (2014) et Gaul Bouchard et al. (2009). Les grilles de Lafay et Cattini (2018), Leclercq et Veys (2014) et Gaul Bouchard et al. (2009) sont elles-mêmes issues et traduites de McCauley et Swisher (1984).

Tableau 2*Grille des critères psychométriques (adaptée de Lafay et Cattini, 2018)*

Critères		Définition / mots clés	Questions / Indices
Standardisation		Uniformisation des conditions de passation	<ul style="list-style-type: none"> • La qualification de l'examineur est-elle définie ? • Les consignes de passation sont-elles décrites avec précision ? Les aides et rétroactions à apporter à l'enfant ? • La cotation est-elle décrite avec précision ?
	Validité de contenu	Référence à une théorie étayée scientifiquement	<ul style="list-style-type: none"> • Le modèle théorique est-il décrit ? L'outil fait-il référence à des articles scientifiques ?
Validité	Validité de critère	Validité concurrente	<ul style="list-style-type: none"> • Les résultats du test sont corrélés avec les résultats à un autre test du même acabit
		Validité prédictive	<ul style="list-style-type: none"> • Les résultats du test sont corrélés avec des activités de la vie quotidienne • Les résultats sont-ils corrélés aux résultats scolaires ?
	Validité de construit	Validité / caractéristiques indiv.	<ul style="list-style-type: none"> • Les résultats du test sont corrélés avec les caractéristiques biologiques, démographiques et cognitives des sujets • Effet de l'âge, du sexe, du statut socio-économique, du QI, de la latéralité, du pays ?
Validité factorielle		<ul style="list-style-type: none"> • Les différents items ou sous-tests qui mesurent la même compétence doivent être reliés • Le test montre-t-il une corrélation entre les différentes épreuves ? en référence au cadre théorique 	
Pouvoir discriminant	Spécificité	Capacité du test à identifier les personnes saines	<ul style="list-style-type: none"> • L'étalonnage du test comporte-t-il une population pathologique ?
	Sensibilité	Capacité du test à identifier les personnes avec une pathologie	<ul style="list-style-type: none"> • Référence à des pourcentages de vrais/faux positifs négatifs ?
Fidélité	Test-retest	Stabilité temporelle	<ul style="list-style-type: none"> • Coefficient de corrélation de 0,90
	Inter-juge	Stabilité entre les expérimentateurs	<ul style="list-style-type: none"> • Kappa de Cohen de 0,60
	Cohérence interne	Cohérence des items	<ul style="list-style-type: none"> • Alpha de Cronbach de 0,70
Normalisation	Description de l'échantillon	Description précise de la population + représentativité	<ul style="list-style-type: none"> • Y a-t-il une description précise de l'échantillon (âge, niveau socio-économique, langue, variété dialectale) • La population est-elle cohérente avec la clinique ?
	Taille		<ul style="list-style-type: none"> • 100 enfants par tranche d'âge
	Mesure de tendance centrale	Présence de moyennes et écart-type	<ul style="list-style-type: none"> • Les moyennes et écart-types sont-ils mentionnés ? • La répartition est-elle gaussienne ? Sinon, percentiles ?

3. Grilles d'analyse des qualités propres à un test de perception de la parole

La comparaison des outils d'évaluation, en plus de prendre appui sur les critères psychométriques, se fonde sur des éléments qui sont décrits dans la littérature comme pertinents pour évaluer spécifiquement la perception (Locke, 1980 a et b ; Stackhouse et al., 2007, Rvachew, 2015). Les critères que nous avons choisis de retenir pour le versant spécifique à l'évaluation de la perception sont présentés dans le tableau 3.

Tableau 3

Grille des critères d'une tâche d'évaluation de la perception

Critères	Questions / Indices
Lien perception-production	<ul style="list-style-type: none"> • Est-ce que le test permet de comparer les habiletés de production et de perception ? • Est-ce que ce lien production-perception a été pris en compte au cours de l'étalonnage ?
Justification des cibles choisies	<ul style="list-style-type: none"> • Est-ce que les bilans justifient les cibles linguistiques proposées ? • Quelles sont les justifications (en lien avec les observations des erreurs fréquentes chez les enfants porteurs de TDSP, et/ou avec la distribution fréquentielle des segments de la langue) ?
Evaluation des différents niveaux de structuration de la phonologie (du trait phonémique à la structure du mot)	<ul style="list-style-type: none"> • Est-ce que le test propose des items cibles avec des altérations à différents niveaux (au niveau du trait phonémique, au niveau du phonème -consonne et voyelle-, au niveau de la syllabe, au niveau de la structure syllabique du mot).
Répétition	<ul style="list-style-type: none"> • Le test propose-t-il plusieurs tentatives pour chaque item (entre autres, pour identifier un éventuel trouble des capacités attentionnelles) ?
Nombre d'items	<ul style="list-style-type: none"> • Le test justifie-t-il le nombre d'items proposés ? • Dans le cadre des tâches de discrimination phonémique, le test justifie-t-il le nombre d'items similaires et d'items différents ?
Lecture labiale	<ul style="list-style-type: none"> • Le test contrôle-t-il la présence ou l'absence de la lecture labiale (stimuli produits en direct par l'orthophoniste ou présentés en audio-visuel) ?
Durée de passation	<ul style="list-style-type: none"> • Le test a-t-il une durée de passation limitée (permettant une passation fonctionnelle pour la clinique quotidienne et permettant de préserver les ressources attentionnelles de l'enfant pour la poursuite de l'évaluation) ?
Âges d'application	<ul style="list-style-type: none"> • Le test indique-t-il l'âge d'application ? <p>Cet âge d'application est-il en adéquation avec les compétences cognitives et attentionnelles du public visé ?</p>
Biais cognitif	<ul style="list-style-type: none"> • Le test contrôle-t-il les biais cognitifs possibles (maîtrise de la notion de similaire/différent, mémoire auditivo-verbale, capacités attentionnelles)

4. Cotation

La cotation des différents outils d'évaluation en fonction de la grille globale d'analyse (les deux axes) a été réalisée par une orthophoniste diplômée. Cette cotation s'appuie sur les manuels de passation fournis avec les batteries de tests. Dans le cadre de TIPL0, nous nous sommes fondées sur l'article de Parisse et Soubeyrand (2002) en plus du manuel accompagnant le test. Dans le cadre de EVALO, nous avons également pris appui sur « L'ouvrage de référence » de Coquet et al. (2009). Pour la cotation, nous avons opté pour un système de points :

- 2 points sont attribués lorsque le critère est complètement rempli, c'est-à-dire que le manuel de test ou les articles mentionnent le critère et disposent des qualités suffisantes pour le remplir,
- aucun point n'est attribué lorsque le critère n'est pas rempli ou n'est pas mentionné dans le manuel de test ou les articles,
- 1 point est attribué lorsque le critère est mentionné mais la description donnée par le manuel ne permet pas de juger de la validation du critère.

En annexe se trouve un tableau indiquant en précision les seuils décidés pour attribuer les points.

----- RÉSULTATS -----

Tableau 4

Synthèse des résultats obtenus par les sept tâches comparées

	ISADYLE	TIPLO	EVALO	EVALO	BEPL-A	EDP 4-8	ELDP	EVALE O	Exalang 3-6	Exalang 5-8	CLE A	Total (22)
Nom de l'épreuve	Gnosies auditives	Epreuve testant la compréhension et la perception phonétique	Epreuve "petit" Gnosies auditivo- verbales	Epreuve "grand" : Jugement "pareil" versus "pas pareil"	Gnosies auditivo- phonétique s	Epreuve de discrimina- tion phono- logique	Epreuve Lilloise de Discrimina- tion Phono- logique	Discri- mina- tion phono- logique	Gnosies auditivo - verbales	Similarité - Dissembla nces		
Type de tâche	Appariemen t mot / image	Apparieme nt mot / image	Apparieme nt mot / image	Discrimina- tion phono- logique	Apparieme nt mot / image	Discrimina- tion phono- logique	Discrimina- tion phono- logique	Discri- mina- tion phono- logique	Discri- mina- tion phono- logique	Discrimi- nation phono- logique		
Standardisation	2	2	2	2	2	2	2	2	2	2	2	22
Validité de contenu	2	2	2	2	2	2	2	2	1	2		
Validité concurrente	0	0	0	0	0	0	0		1	1		
Validité prédictive	0	0	0	0	0	0	0		1	1		
Validité en relation avec les caractéristiques individuelles	2	0	0	0	0	2	2		0	0	2	
Validité factorielle	2	1	0	0	0	2	2		1	1	2	
Spécificité/sensibilité	0	0	2	2	0	2	1	2	1	2		
Test-retest	1	0	0	0	0	0	0		2	2		
Inter-juge	0	0	1	1	0	0	0		0	2		
Cohérence interne	2	0	0	0	0	0	0		1	1		
Description de l'échantillon	2	1	2	2	2	2	2	2	2	2	2	
Taille	1	1	0	0	2	1	1	2	1	2	1	
Mesure de tendance centrale	1	0	0	0	0	2	2		2	2	2	
Sous-total (39)	15	7	9	9	8	15	14		15	20		77

Lien perception/production	2	0	1	0	0	2	0	0	0	0	
Justification des cibles choisies	1	2	0	0	0	2	1	2	0	0	
Différents niveaux phonologiques	0	2	2	1	1	0	2	0	2	2	2
Tentatives multiples	0	0	0	0	0	0	2	0	0	0	
Nombre d'items	0	0	0	0	0	1	0	0	0	0	
Lecture labiale	1	1	0	0	0	1	1	1	1	1	
Durée de passation	2	2	2	2	2	2	2	2	2	2	
Âges d'application	2	2	2	2	2	2	2	2	2	2	
Biais cognitif	2	1	2	2	2	2	2	0	2	2	
Sous-total (27)	10	10	9	7	7	12	12		9	9	67
Total (66)	25	17	18	16	15	27	26		24	29	144

1. Critères psychométriques

Notre premier axe de comparaison des épreuves évaluant la perception a trait aux caractéristiques psychométriques. Certains critères sont validés par tous les tests, comme celui de la validité théorique (20/20). Pour construire un sous-test dans une batterie, les auteurs réfèrent à divers modèles psycholinguistiques (modèle de Maillart et al. (2004) pour ISADYLE ou modèle de Coquet (2006) pour EVALO). Les épreuves ne rentrant pas dans le cadre d'une batterie complète telles que EDP 4-8 ou ELDP font davantage référence à une définition théorique de la compétence qu'elles entendent tester.

Le critère de standardisation est également bien respecté par les différentes épreuves (20/20). Les consignes à donner aux enfants sont systématiquement décrites ainsi que la méthodologie de la cotation.

Enfin, le critère de description de l'échantillon d'étalonnage est validé par les différentes épreuves (13/14). Les principales caractéristiques rapportées sont l'âge, le sexe, la variété dialectale du français ou la région géographique du domicile, la catégorie socio-culturelle des parents/adultes responsables et le type de garde ou de scolarisation.

D'autres critères sont moins mentionnés dans les manuels des épreuves. C'est le cas de la taille de l'échantillon de population par tranche d'âge. Le critère indiqué dans la littérature est celui d'au minimum 100 enfants par tranche d'âge (Gaul Bouchard et al., 2009 ; Leclercq & Veys, 2014). A ce critère, nous avons ajouté celui de la régularité des tranches d'âge, c'est-à-dire que, si le critère de groupe d'âge défini par le test est de 6 mois, il faut que celui-ci soit respecté pour tout l'étalonnage. Seules la BEPL-A et EVALEO valident ces deux critères. Les autres tests ne présentent pas suffisamment d'enfants dans leur étalonnage (pour l'ELDP, l'étalonnage se fait sur 45 enfants par groupe d'âge, pour l'EDP 4-8, en moyenne 40 enfants par groupe d'âge et pour TIPLO, 15 enfants par groupe d'âge) ou bien les tests élargissent leurs tranches d'âge pour atteindre les 100 enfants par groupe (ISADYLE propose des groupes de 6 mois d'âge pour les plus jeunes enfants et de 1 an pour les plus âgés, alors que le test démontre une progression en fonction de l'âge même chez les enfants plus âgés). Concernant les mesures de tendances centrales, là encore le critère n'est que peu validé par les tests (5/20). L'ensemble des épreuves proposent des mesures en moyenne avec des écart-types. Ce type de mesure de tendances centrales nécessite une répartition normale des scores de la population d'étalonnage. C'est le cas pour EDP 4-8 et pour certains groupes d'âge de ELDP et ISADYLE. Cependant, dans certains cas, la répartition de la population n'est pas normale. Face à cela, certains bilans (ELDP et ISADYLE) proposent une mesure en percentile mais d'autres préfèrent conserver les écart-types (BEPL-A). Dans le cadre de EVALO, aucune indication n'est donnée sur la répartition des scores de la population. Finalement, EVALEO propose un test de normalité pour une grande majorité des épreuves proposées mais la discrimination phonologique n'en fait pas partie : les informations sur la répartition de la population pour ce subtest ne sont pas disponibles. Concernant la spécificité et la sensibilité, qui sont pourtant les critères essentiels d'un outil de diagnostic, seulement trois outils (EVALO, ELDP et EDP 4-8) tentent d'intégrer le critère (7/14). Pour cela, les tests incluent une population porteuse de pathologie du langage (EVALO) ou en difficulté (ELDP qui, cependant, exclut les enfants avec un diagnostic de trouble spécifique du langage) ou de population à risque (EDP 4-8). En revanche, aucun test ne propose un calcul des vrais ou faux positifs/négatifs à partir de cette population en difficulté. Cela nous apparaît comme une limite essentielle de l'utilisation de ces tests qui ont pourtant un objectif diagnostic.

Concernant les critères de validité de construit, il est demandé aux évaluations de décrire comment les caractéristiques individuelles des sujets (validité en relation avec les caractéristiques individuelles) et les caractéristiques des items soumis (validité factorielle) interagissent avec la mesure du construit. Si le construit est bien défini, il est possible d'élaborer des hypothèses sur la façon dont l'âge ou les caractéristiques linguistiques, par exemple, vont influencer la mesure du construit. Pour la validité en lien avec les caractéristiques individuelles des sujets (6/20), il n'y a qu'ISADYLE, EDP 4-8 et ELDP qui apportent des informations, notamment sur la façon dont l'âge est corrélé avec les scores à la tâche. EVALO, TIPLO et BEPL-A ne donnent pas d'indication. Concernant la façon dont les caractéristiques des items influencent les scores (7/20), très peu d'informations sont disponibles. Les scores ou les pourcentages de réussite en fonction des items sont disponibles pour TIPLO, ISADYLE, EDP 4-8 et ELDP et à partir de ces informations, il serait possible de déduire des informations sur la façon dont les contrastes ou les structures syllabiques testées interagissent avec les scores des enfants. Cette analyse des caractéristiques des items est en lien avec le critère de cohérence interne qui évalue si les différents items sont cohérents, c'est-à-dire s'ils réagissent dans la même direction. La cohérence interne n'est validée que par le test ISADYLE qui rapporte un alpha de Cronbach de 0,78.

Certains critères sont très peu respectés par les protocoles de test. Les critères de validité concurrente ou de validité prédictive ne sont jamais abordés dans les manuels (0/20). De fait, il est donc difficile d'objectiver que les résultats obtenus par le test ont une valeur dans la vie quotidienne ou bien sont cohérents avec un autre test diagnostic. De la même façon, la fidélité test-retest (1/14) n'est abordée que par ISADYLE. Enfin, la fidélité inter-juges (4/20) n'est que vaguement abordée alors que ces tests ont pour objectif d'être proposés à l'ensemble des orthophonistes et que de nombreux expérimentateurs ont participé à l'étalonnage.

2. Qualités propres à l'évaluation de la perception

Parmi les critères inhérents à l'évaluation des habiletés perceptives chez l'enfant, certains sont remplis par chacune des sept épreuves que nous avons comparées. Ces critères sont l'adéquation de la tâche à l'âge des enfants (20/20) et la rapidité du protocole de passation (20/20). Les tests s'appliquent à rendre la tâche accessible à leur population cible, en utilisant des renforçateurs (petites saynètes vidéo pour TIPLO), en adaptant leur support visuel (utilisation de photos pour ISADYLE), en proposant des types de protocoles différents (EVALO) ou des listes avec des longueurs d'items différentes (ELDP). Les critères d'adaptation à l'âge et de durée de passation, fortement liés à la prise en compte des biais cognitifs inhérents aux tâches, sont également bien respectés par les tests de notre comparaison (19/20). Les biais cognitifs font l'objet d'une importante littérature qui décrit l'intérêt et les limites des différents protocoles d'évaluation pour les enfants (Locke, 1980a, Stackhouse et al., 2007, Macchi et al., 2012). La tâche d'appariement mot/image est la plus compréhensible par des enfants très jeunes et elle permet une réponse non verbale. C'est pour cela que l'épreuve TIPLO peut être proposée à partir de 2 ans et demi. Cependant, elle fait intervenir la variable de la fréquence lexicale et du niveau de vocabulaire de l'enfant. Si l'enfant ne connaît pas le mot qui lui est proposé, il ne pourra pas désigner la bonne image, et cela, même s'il a bien décodé les phonèmes. Les tests mettent en place des stratégies pour contrôler ces biais, comme par exemple, faire dénommer les items et en expliquer les concepts avant de proposer la tâche en perception pour s'assurer que l'enfant connaît les items. Pour les tâches de discrimination de mots ou de pseudomots, bien qu'elles soient décrites comme les tâches les plus pures pour tester le décodage phonémique, elles sont soumises à la capacité de réaliser un jugement de similitude ainsi qu'à la mémoire de travail et aux capacités attentionnelles. Dans la littérature, il est indiqué qu'on ne peut pas proposer une

tâche de discrimination phonologique avant 4 ans (Stackhouse et al., 2007). La majorité des épreuves de notre inventaire respecte cette limite d'âge : les épreuves de discrimination commencent respectivement à 4 ans pour l'EDP et EVALO et à 5 ans pour l'ELDP. En plus de cela, les épreuves proposent des items d'entraînement pour s'assurer que l'enfant comprend la notion de « même » et de « différent ». Seul Exalang 3-6 propose une tâche avant 4 ans mais celle-ci consiste en une épreuve de discrimination de mots et non de pseudomots, tâche qui est moins exigeante cognitivement.

Les autres critères en revanche sont moins respectés par les épreuves de notre inventaire. Le choix des cibles n'est que rarement justifié (8/20). Pour EVALEO, ISADYLE, EVALO et BEPL-A, la justification tient dans le fait de tester, de façon non systématique et/ou exhaustive, la perception des traits consonantiques de voisement, de lieu d'articulation ou de mode. Pour l'EDP 4-8, les auteurs ont construit les pseudomots en se fondant sur la fréquence de distribution des segments en français et ont choisi les items les plus pertinents en fonction d'une étude pilote auprès de 20 adultes et 60 enfants. Également, pour TIPLO, le choix des cibles a pris appui sur les données de la littérature concernant les étapes du développement des capacités perceptives. Quant au nombre d'items proposés aux enfants, six tâches sur sept omettent d'indiquer de quelle façon ce seuil a été établi (1/20). Seul EDP 4-8 justifie le nombre d'items proposés par son étude pilote. Également, il n'est pas possible de savoir à partir de combien d'essais il est possible d'avoir un bon aperçu des habiletés de perception chez un enfant. De la même façon, la grande majorité des épreuves ne propose qu'une seule tentative par item (2/20). Seul ELDP propose de tester les cibles, dans le même contexte phonétique mais dans un item différent, une fois dans la liste à vitesse normale et une seconde fois dans la liste à vitesse rapide.

Concernant le lien entre la perception et la production, seuls ISADYLE et EDP 4-8 présentent des données étalonnées sur cette relation (5/20). Les consignes de EVALO indiquent de faire nommer les items pour noter s'il y a des substitutions phonémiques ou d'autres patrons d'erreurs. Comme nous l'avons expliqué plus haut, le lien perception-production nous apparaît pourtant essentiel dans le cadre d'une évaluation du traitement de la parole.

Pour le critère de la prise en compte des différents niveaux phonologiques (8/20), TIPLO, EVALO et ELDP adoptent une perspective multilinéaire, testant à la fois les traits phonologiques mais également les voyelles et les structures syllabiques.

Enfin, aucune épreuve ne teste spécifiquement l'apport de la lecture labiale en proposant une condition avec lecture labiale et une autre condition sans lecture labiale. Dans le meilleur des cas, la lecture labiale est prise en compte, soit elle est absente car les stimuli sont enregistrés à l'audio seul, soit elle est présente car il est spécifié que l'orthophoniste doit prononcer les items sans cacher ses lèvres.

----- CONCLUSION -----

L'objectif de cette étude était de proposer une comparaison des principales épreuves disponibles actuellement et évaluant la perception de la parole pour les enfants présentant un Trouble du Développement des Sons de Parole. Cette comparaison a pris appui à la fois sur des critères psychométriques mais également sur des critères décrits dans la littérature comme pertinents pour l'évaluation spécifique des habiletés de perception. Ce travail s'intègre dans une démarche de transfert aux orthophonistes de clés pour argumenter la sélection de leurs outils d'évaluation et pour intégrer une pratique fondée sur les données probantes. Notre état des lieux met en lumière qu'aucune tâche actuellement disponible en clinique ne respecte l'ensemble des critères

psychométriques ni les critères spécifiques à une évaluation de la perception. Cela sous-entend qu'aucune tâche ne permet d'évaluer efficacement la perception, ce qui a des conséquences dans la caractérisation des troubles et pour l'intervention clinique.

Concernant les critères spécifiques aux tâches de perception, l'adaptation à l'âge des enfants, la courte durée de la passation et la prise en compte des biais cognitifs sont intégrés dans la construction des épreuves. Cependant, les modalités d'analyse des résultats de l'enfant restent globales. Les résultats donnés par les tests permettent essentiellement de situer l'enfant face à une population d'étalonnage. Plus rarement, les résultats permettent de faire un lien avec le versant production (seulement deux protocoles envisagent ce lien) ou d'envisager le niveau phonologique auquel se situent les difficultés de perception (seulement trois épreuves évaluent à la fois les traits segmentaux et la structure syllabique). Également, la justification théorique du choix des cibles n'est pas systématique. Enfin, aucune épreuve, excepté ELPD qui propose un versant de son épreuve avec un débit de parole rapide, ne met au défi le système perceptif dans un contexte de bruit ou de signal dégradé. Or, il semblerait que les déficits perceptifs des enfants avec un TDSP soient davantage mis en évidence dans des conditions adverses de perception (parole accélérée, bruit parasite, multilocuteurs).

D'un point de vue psychométrique, là encore, aucun outil ne rassemble les critères reconnus pour une pratique de diagnostic. La standardisation et la validité théorique sont présentes dans tous les manuels mais la corrélation avec d'autres épreuves ou avec des indices de la vie quotidienne (tels que le niveau scolaire) est manquante. De la même façon, les mesures de tendances centrales sont systématiquement données en moyenne et écart-type alors que parfois la répartition de la population n'est pas normale. Ces manques ne permettent pas d'établir un diagnostic assuré du déficit.

Notre conclusion suit les observations de Leclercq et Veys (2014) et de Lafay et Cattini (2018) selon lesquelles les concepteurs de tests doivent développer des outils reposant sur les connaissances actuelles des déficits et sur un réel effort de validation des tâches. Ces outils devraient permettre d'identifier les forces et les faiblesses de chaque enfant présentant un TDSP, de façon à proposer une intervention ciblée pour l'enfant. Notre conclusion soutient et encourage également l'évaluation des habiletés perceptives de tous les enfants présentant un TDSP dans la mesure où des études actuelles confirment qu'au moins une partie d'entre eux aurait des difficultés avec la perception de la parole (Hearnshaw et al., 2019).

----- BIBLIOGRAPHIE -----

American Speech-Language-Hearing Association (ASHA) (2007). Childhood apraxia of speech (Position statement). <https://www.asha.org/policy/PS2007-00277/>

Autesserre, D., Deltour, J.-J. et Lacert, P. (1988). *EDP 4-8. Epreuve de Discrimination phonémique pour enfants de 4 à 8 ans*. EAP.

Baker, E. et McLeod, S. (2011). Evidence-based practice for children with speech sound disorders: Part 1 narrative review. *Language, Speech, and Hearing Services in Schools*, 42(2), 102-139. [https://doi.org/10.1044/0161-1461\(2010/09-0075\)](https://doi.org/10.1044/0161-1461(2010/09-0075))

Barrozo, T. F, Pagan-Neves, L. O., Vilela, N., Carvallo, R. M. M. et Wertzner, H. F. (2016). The influence of (central) auditory processing disorder in speech sound disorders. *Brazilian Journal of Otorhinolaryngology*, 82(1), 56-64. <https://doi.org/10.1016/j.bjorl.2015.01.008>

- Belot, C. et Tricot, M. (2001). *Les tests en orthophonie. Tome 1 : langage oral, langage écrit, enfants, adolescents*. OrthoEdition.
- Betz, S., Eickhoff, J. et Sullivan, S. (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))
- Bérubé, D., Bernhardt, B. M. et Stemberger (2015). A test of Canadian French phonology: Construction and use. *Canadian Journal of Speech-Language Pathology and Audiology, 39*(1), 61-100. <https://www.cjslpa.ca/detail.php?ID=1172&lang=en>
- Bowen, C. (2015). *Children's speech sound disorders* (2e ed.). Wiley Blackwell.
- Broomfield, J. et Dodd, B. (2004). The nature of referred subtypes of primary speech disability. *Child Language Teaching and Therapy, 20*(2), 135-151. <https://doi.org/10.1191/0265659004ct267oa>
- Chevrie-Muller, C., Simon, A.-M., Le Normand, M.-T. et Fournier, S. (1997). *Batterie d'Évaluation PsychoLinguistique. BEPL-A et BEPL-A « forme courte » de 2 ans 9 mois à 4 ans 3 mois*. Pearson Canada Assessment.
- Coquet, F., Roustit, J. et Ferrand, P. (2009). *ÉVALO 2-6 - l'ouvrage de référence*. OrthoEdition.
- Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports, 1*(3), 189-196. <https://doi.org/10.1007/s40474-014-0017-3>
- Dodd, B. et McIntosh, B. (2008). The input processing, cognitive linguistic and oro-motor skills of children with speech difficulty, *International Journal of Speech-Language Pathology, 10*(3), 169-178. <https://doi.org/10.1080/14417040701682076>
- Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Brookes Publishing.
- Dollaghan, C. A. (2004). Evidence-based practice in communication disorders: What do we know, and when do we know it? *Journal of Communication Disorders, 37*(5), 391-400. <https://doi.org/10.1016/j.jcomdis.2004.04.002>
- Edwards, J., Fourakis, M., Beckman, M. et Fox, R. (1999). Characterizing knowledge deficits in phonological disorders. *Journal of Speech, Language, and Hearing Research, 42*(1), 169-186. <https://doi.org/10.1044/jslhr.4201.169>
- Felsenfeld, S., Broen, P. A. et McGue, M. (1994). A 28-year follow-up of adults with a history of moderate phonological disorder: Educational and occupational results. *Journal of Speech and Hearing Research, 37*(6), 1341-1353. <https://doi.org/10.1044/jshr.3706.1341>
- Gaul Bouchard, M.-E., Fitzpatrick, E. M. et 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. <https://www.cjslpa.ca/detail.php?ID=984&lang=fr>

Haute Autorité de Santé (2001). *L'orthophonie dans les troubles spécifiques du développement du langage oral chez l'enfant de 3 à 6 ans*, Service des recommandations et références professionnelles

https://www.has-sante.fr/jcms/c_271995/fr/l-orthophonie-dans-les-troubles-specifiques-du-developpement-du-langage-oral-chez-l-enfant-de-3-a-6-

Hearnshaw, S., Baker, E. et Munro, N. (2019). Speech perception skills of children with speech sound disorders: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, 62(10), 3771-3789. https://doi.org/10.1044/2019_JSLHR-S-18-0519

Hearnshaw, S., Baker, E. et Munro, N. (2018). The speech perception skills of children with and without speech sound disorder. *Journal of Communication Disorders*, 71, 61-71. <https://doi.org/10.1016/j.jcomdis.2017.12.004>

Helloin, M.-C. et Thibault, M.-P. (2006). *EXALang 3-6. Batterie d'examen des fonctions langagières chez l'enfant de 3 à 6 ans*. Happyneuron.

Hickok, G., Houde, J. et Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron Perspective*, 69(3), 407-422. <https://doi.org/10.1016/j.neuron.2011.01.019>

International expert panel on multilingual children's speech. (2012). *Multilingual children with speech sound disorders: Position paper*. Bathurst, Australia: Research Institute for Professional Practice, Learning and Education (RIPPLE), Charles Sturt University. <http://www.csu.edu.au/research/multilingual-speech/position-paper>

Jamieson, D., G. et Rvachew, S. (1992). Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology*, 16(3), 201-210. https://www.cjslpa.ca/files/1992_JSLPA_Vol_16/No_03_177-250/Jamieson_Rvachew_JSLPA_1992.pdf

Lafay, A. et Cattini, J. (2018). Analyse psychométrique des outils d'évaluation mathématique utilisés auprès des enfants francophones, *Canadian Journal of Speech-Language Pathology and Audiology*, 42(2), 127-144. <https://www.cjslpa.ca/detail.php?lang=fr&ID=1232>

Launay, L., Maeder, C., Roustit, J. et Touzin, M. (2018). *Évaluation du langage écrit et du langage oral de 6 à 15 ans (EVALéo 6-15)*. OrthoEdition.

Leclercq, A.-L. et Veys, E. (2014). Réflexions sur le choix des tests standardisés lors du diagnostic de dysphasie, *A.N.A.E.*, 131, 374-382.

Locke, J. (1980a). The inference of speech perception in the phonologically disordered child. Part I: A rationale, some criteria, the conventional tests, *Journal of Speech and Hearing Disorders*, 45(4), 431-444. <https://doi.org/10.1044/jshd.4504.431>

Locke, J. (1980b). The inference of speech perception in the phonologically disordered child. Part II: Some clinically novel procedures, their use, some findings. *Journal of Speech and Hearing Disorders*, 45(4), 445-468. <https://doi.org/10.1044/jshd.4504.445>

- MacAllister Byun, T. et Tiede, M. (2017). Perception-production relations in later development of American English rhotics, *PLoS ONE*, 12(2), e0172022. <https://doi.org/10.1371/journal.pone.0172022> accès libre
- Macchi, L., Vansteene, C., Timmermans, N. et Boidein, F. (2013). Epreuve Lilloise de Discrimination Phonologique (ELDP) : Présentation et illustration par deux études de cas clinique, *Les cahiers de l'ASELF*, 10(3), 3-20.
- Macchi, L., Descours, C., Girard, E., Guitton, E., Morel, C., Timmermans, N. et Boidein, F. (2012). *ELDP. Epreuve Lilloise de Discrimination Phonologique destinée aux enfants de 5 ans à 11;6 ans*. <https://medecine.univ-lille.fr/orthophonie/reseau-professionnel>
- Maillart, C., Schelstraete, M.-A. et Hupet, M. (2004). Phonological representations in children with SLI: A study of French. *Journal of Speech, Language and Hearing Research*, 47(1), 187-198. [https://doi.org/10.1044/1092-4388\(2004/016\)](https://doi.org/10.1044/1092-4388(2004/016))
- McCauley, R. J. et Swisher, S. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorder*, 49(1), 34-42. <https://doi.org/10.1044/jshd.4901.34>
- McLeod, S. et Baker, E. (2014). Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics and Phonetics*, 28(7-8), 508–531. <http://dx.doi.org/10.3109/02699206.2014.926994>.
- McLeod, S. et Harrison, L. J. (2009). Epidemiology of speech and language impairment in a nationally representative sample of 4- to 5-year-old children. *Journal of Speech, Language, and Hearing Research*, 52(5), 1213 – 1229. [https://doi.org/10.1044/1092-4388\(2009/08-0085\)](https://doi.org/10.1044/1092-4388(2009/08-0085))
- Munson, B., Edwards, J. et Beckman, M. (2005). Phonological knowledge in typical and atypical speech and language development: nature, assessment, and treatment. *Topics in Language Disorders*, 25(3), 190-206. <https://doi.org/10.1097/00011363-200507000-00003>
- Nathan, E. (2001). *The development of speech processing skills in children with and without speech difficulties*. Thèse de doctorat “Philosophy in Human Communication Science”. University College London. <https://discovery.ucl.ac.uk/id/eprint/1349803/1/366979.pdf>
- Parisse, C. et Soubeyrand, E. (2002). Un test informatisé pour la perception du langage oral : TIPLo. *Glossa*, 79, 6-21. Test téléchargeable sur <https://www.ortolang.fr/market/tools/tiplo>
- Pasquet, F., Parbeau-Guéno, A. et Bourg, E. (2014). *CLÉA. Communiquer, Lire et Écrire pour Apprendre*. ECPA.
- Pennington, B. et Bishop, D. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, 60, 283-306. <https://doi.org/10.1146/annurev.psych.60.110707.163548>
- Piérart, B., Comblain, A., Grégoire, J. et Mousty, P. (2010). *Batterie ISADYLE*. De Boeck Solal.

- Preston, J. L., Felsenfeld, S., Frost, S. J., Mencl, W. E., Fulbright, R. K., Grigorenko, E. L., Landi, N., Seki, A. et Pugh, K. R. (2012). Functional brain activation differences in school-age children with speech sound errors: Speech and print processing. *Journal of Speech, Language and Hearing Research*, 55(4), 1068–1082. [https://doi.org/10.1044/1092-4388\(2011/11-0056\)](https://doi.org/10.1044/1092-4388(2011/11-0056))
- Preston, J. L., Molfese, P. J., Mencl, W. E., Frost, S. J., Hoeft, F., Fulbright, R. K., Landi, N., Grigorenko, E. L., Seki, A., Felsenfeld, S. et Pugh, K. R.. (2014). Structural brain differences in school-age children with residual speech sound errors. *Brain and Language*, 128(1), 25–33. <https://doi.org/10.1016/j.bandl.2013.11.001>
- Rvachew, S. (2015) Perceptually based interventions. Dans C. Bowen (dir), *Children's speech sound disorders* (2e ed., p. 188-191). Wiley Blackwell.
- Rvachew, S., Nowak, M. et Cloutier, G. (2004) Effect of phonemic perception training on the speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech-Language Pathology*, 13(3), 250-263. [https://doi.org/10.1044/1058-0360\(2004/026\)](https://doi.org/10.1044/1058-0360(2004/026))
- Salvia, J., Ysseldyke, J. et Bolt, S. (2010). *Assessment in special and inclusive education* (11e ed.). Wadsworth.
- Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeney, J. L., Potter, N. L., Scheer-Cohen, A. R., Strand, E. A., Tilkens, C. M. et Wilson, D. L. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics*, 24(10), 795-824. <https://doi.org/10.3109/02699206.2010.503006>
- Stackhouse, J., Vance, M., Pascoe, M. et Wells, B. (2007). *Compendium of auditory and speech tasks: Children's speech and literacy difficulties 4*. Wiley Blackwell.
- Thibault, M.-P., Helloin, M.-C. et Croteau, B. (2010). *EXALang 5-8. Batterie d'examen des fonctions langagières chez l'enfant de 5 à 8 ans*. Happyneuron.
- Waring, R. et Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language & Communication Disorders*, 48(1), 25-40. <https://doi.org/10.1111/j.1460-6984.2012.00195.x>

----- ANNEXE -----

Tableau 5*Seuils de décision d'attribution des tests*

Critère	Cotation
Lien perception-production	<ul style="list-style-type: none"> • 2 points : le test met en relation l'épreuve de perception avec une épreuve de production et propose une analyse quantifiée de la comparaison des deux versants. • 1 point : le test indique qu'utiliser les items de la tâche de perception en production est une source d'information clinique • 0 point : le test n'aborde pas la production
Justification des cibles	<ul style="list-style-type: none"> • 2 points : le test justifie les cibles par des études développementales de la perception chez des enfants tout-venant, des enfants avec un TDSP ou avec la fréquence de distribution des sons de la langue. • 1 point : le test justifie les cibles par la description d'une étude pilote. • 0 point : le test ne justifie pas les cibles.
Evaluation des différents niveaux phonologiques	<ul style="list-style-type: none"> • 2 points : le test propose d'évaluer la perception sur les traits segmentaux (consonantique et vocalique) et sur la structure des syllabes • 1 point : le test propose d'évaluer la perception sur les traits consonantiques et la structure syllabique. • 0 point : le test propose d'évaluer la perception sur les traits consonantiques.
Tentative multiple	<ul style="list-style-type: none"> • 2 points : le test propose plusieurs fois le même item à percevoir ou le test propose la même cible dans le même contexte phonétique mais dans un item différent. • 1 point : le test propose plusieurs fois la même cible dans des contextes phonétiques différents. • 0 point : le test ne propose pas plusieurs tentatives.
Nombre d'items	<ul style="list-style-type: none"> • 2 points : le test justifie la quantité d'items proposés par la littérature par une analyse statistique des items. • 1 point : le test justifie la quantité d'items pour une étude pilote et le retrait d'items non pertinents. • 0 point : le test ne justifie pas la quantité d'items proposés.
Lecture labiale	<ul style="list-style-type: none"> • 2 points : la lecture labiale est contrôlée et la condition avec versus sans lecture labiale est étalonnée • 1 point : la lecture labiale est contrôlée parce que les stimuli sont enregistrés (audio, sans lecture labiale) ou parce que le test spécifie une production par l'expérimentateur (avec lecture labiale) • 0 point : la lecture labiale n'est pas contrôlée
Durée de passation	<ul style="list-style-type: none"> • 2 points : la durée de passation est courte et il y a une justification indiquant que les enfants supportent d'être testés pendant cette durée • 1 point : la durée de passation est courte sans justification • 0 point : la durée de passation est longue
Âges d'application	<ul style="list-style-type: none"> • 2 points : la tâche est adaptée à l'âge des enfants, le manuel démontre une réflexion de l'adaptation (type de tâche, renforçateurs) • 1 point : la tâche est adaptée mais le manuel ne démontre pas de réflexion. • 0 point : la tâche n'est pas adaptée selon la littérature.
Biais cognitifs	<ul style="list-style-type: none"> • 2 points : Le manuel du test avertit des différents biais et met en place un protocole qui les limite. • 1 point : Le manuel du test avertit des différents biais mais ne met rien en place. • 0 point : Le manuel n'indique pas de biais cognitifs.
Standardisation	<ul style="list-style-type: none"> • 2 points : La qualification de l'expérimentateur, l'objectif du test, les consignes et la cotation sont indiqués avec des précisions (exemple de cotation). • 1 point : Un de ces quatre points manque. • 0 point : Plus d'un de ces quatre points sont manquants.
Validité de contenu	<ul style="list-style-type: none"> • 2 points : Le manuel décrit le modèle théorique et/ou la définition auxquels il se rapporte en intégrant des références de la littérature scientifique.

	<ul style="list-style-type: none"> • 1 point : Le manuel décrit le modèle théorique et/ou la définition auxquels il se rapporte sans référence à la littérature scientifique. • 0 point : Le manuel ne décrit pas de modèle théorique.
Validité concurrente	<ul style="list-style-type: none"> • 2 points : Le manuel donne un indice de corrélation entre les résultats au test et les résultats à un autre test de même acabit. • 1 point : Le manuel aborde la notion avec un indice différent. • 0 point : Le manuel n'aborde pas la notion.
Validité prédictive	<ul style="list-style-type: none"> • 2 points : Le manuel donne un indice de corrélation entre les résultats au test et les résultats à une activité de la vie quotidienne ou scolaire testant la perception. • 1 point : Le manuel aborde la notion avec un indice différent. • 0 point : Le manuel n'aborde pas la notion.
Validité en relation avec les caractéristiques individuelles	<ul style="list-style-type: none"> • 2 points : Le manuel donne un indice statistique de l'effet des caractéristiques principales des individus, notamment les caractéristiques indiquées dans la description de l'échantillon (sexe, âge, catégories socio-culturelles, etc.). • 1 point : Le manuel aborde la notion avec un indice différent. • 0 point : Le manuel n'aborde pas la notion.
Validité factorielle	<ul style="list-style-type: none"> • 2 points : Le manuel analyse les effets des facteurs de construction des items (effet de la modification des traits phonologiques, de la structure syllabique, etc.) • 1 point : Le manuel indique des pourcentages de réussite de chaque item • 0 point : Le manuel n'aborde pas la notion.
Spécificité-Sensibilité	<ul style="list-style-type: none"> • 2 points : La population d'étalonnage intègre une population porteuse de pathologie diagnostiquée. • 1 point : La population d'étalonnage intègre une population tout-venant recrutée largement (le suivi en orthophonie n'est pas un critère d'exclusion). • 0 point : La population d'étalonnage n'inclut pas de population avec une pathologie.
Effet test-retest	<ul style="list-style-type: none"> • 2 points : Le manuel indique un coefficient de corrélation d'au minimum .90. • 1 point : Le manuel aborde la notion avec un indice différent. • 0 point : Le manuel n'aborde pas la notion.
Inter-juge	<ul style="list-style-type: none"> • 2 points : Le manuel indique un Kappa de Cohen de 0,60 • 1 points : Le manuel aborde la notion avec un indice différent. • 0 point : Le manuel n'aborde pas la notion.
Cohérence interne	<ul style="list-style-type: none"> • 2 points : Le manuel indique un Alpha de Cronbach d'au moins 0,70. • 1 point : Le manuel aborde la notion avec un indice différent comme le pourcentage de réussite par item et la description d'une étude pilote préliminaire. • 0 point : Le manuel n'aborde pas la notion.
Description de l'échantillon	<ul style="list-style-type: none"> • 2 points : L'échantillon est décrit en termes des caractéristiques intéressantes pour l'évaluation du construit. • 1 point : L'échantillon est décrit avec peu de détails. • 0 point : L'échantillon n'est pas décrit.
Taille de l'échantillon	<ul style="list-style-type: none"> • 2 points : L'échantillon comprend aux alentours de 100 sujets par tranche d'âge et les tranches d'âge sont régulières. • 1 point : L'échantillon comprend moins de 100 sujets par tranche d'âge. • 0 point : La taille de l'échantillon n'est pas décrite.
Mesure de tendance centrale	<ul style="list-style-type: none"> • 2 points : La répartition de la population est normale et les mesures sont données en moyennes et écart-types et pour certains sous-groupes de l'étalonnage, la population n'est pas normale et les mesures sont données en percentiles. • 1 point : La répartition n'est pas normale et les mesures uniquement données en percentiles. • 0 point : Il n'y a pas d'indication sur la répartition normale ou la répartition n'est pas normale mais les mesures sont données en moyenne et écart-types.