

MATÉRIEL SUPPLÉMENTAIRE

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*, 210, 141-156.

1. Questionnaires à destination des orthophonistes, à propos de la fréquence d'utilisation des tests.
« J'indique la fréquence à laquelle j'utilise chaque test, en pensant à un enfant auquel le test est applicable (ex. âge adapté). Pour cela, je pense à un cas d'application : un enfant suspecté de présenter des difficultés langagières orales, quelles qu'elles soient. Le patient auquel je pense doit avoir un âge compris dans l'intervalle d'application du test ou être dans un niveau scolaire d'application du test, et ce, pour chaque test. Par exemple, pour l'ELO, le niveau d'application est entre la PSM et le CM2, je peux donc penser à un enfant de CE1, avec des difficultés de langage oral. Je peux penser à des enfants de difficultés et d'âges différents pour chaque test. »

2. Fréquence moyenne d'utilisation des tests par les orthophonistes, dans des situations où les tests sont applicables.

Tests ou batteries	Fréquence moyenne d'utilisation
EVALO 2-6 (Coquet et al., 2009)	45,99 %
ELO (Khomsi, 2001)	40,96 %
EVALéo 6 - 15 (Maeder et al., 2018)	34,05 %
EXALang 11-15 (Helloin et al., 2009)	31,55 %
EXALang 8-11 (Thibault et al., 2012)	30,33 %
EXALang 5-8 (Thibault et al., 2010)	27,29 %
N-EEL (Chevrie-Muller & Plaza, 2001)	24,59 %
T.L.O.C.C. (Maurin, 2006)	23,54 %
EXALang 3-6 (Helloin & Thibault, 2006)	23,37 %
L2MA-2 (Chevrie-Muller et al., 2010)	22,42 %
E.CO.S.SE. (Lecocq, 1996)	18,80 %
Péléa (Boutard et al., 2011)	18,49 %
TCS (Maeder, 2006)	12,72 %
Kikou 3-8 (Boutard & Bouchet, 2009)	12,39 %
Grille de Hénin - étude de la motricité et des praxies oro-faciales (Hénin, 1981)	12,07 %
EVIP (Dunn et al., 1993)	10,45 %
PEES 3-8 (Boutard & Bouchet, 2009)	9,59 %
IFDC (Kern & Gayraud, 2010)	9,26 %
Children's Communication Checklist (Bishop & Baird, 2001; Maillart, 2003)	9,05 %
ELDP (Macchi et al., 2012)	7,43 %
ECSP (Guidetti & Tourrette, 2009)	6,36 %
BILO-2 (Khomsi et al., 2007)	5,92 %
BILO Petits (Khomsi & Khomsi, 2009)	5,81 %
CLéA (Pasquet et al., 2014)	5,66 %
BILO EC-2 (Khomsi & Khomsi, 2007)	5,45 %
BILO-3C (Khomsi et al., 2007)	5,19 %
Test des concepts de base Boehm – 3 (Boehm, 2009)	5,07 %
Épreuve de décision lexicale (Maillart et al., 2004)	4,18 %
BEPL-A (Chevrie-Muller et al., 1997)	4,13 %
THaPho (Ecalte, 2007)	3,93 %
Épreuve de discrimination (Maillart & Schelstraete, 2004)	3,90 %
BIMM (Gatignol et al., 2007)	3,80 %
ISADYLE (Piérart et al., 2010)	3,52 %
ELOLA (de Agostini et al., 1998)	3,42 %
Test de dépistage de dénomination et de désignation d'images (Kremin & Dellatolas, 1995)	3,38 %
BEPL-B (Chevrie-Muller et al., 1997)	3,27 %
DEDALE (Deltour & Hupkens, 2011)	2,77 %
DLPF (Bassano et al., 2005)	2,50 %

Note. Les tests en italiques ont fait l'objet d'une analyse psychométrique.

3. Échelle de cotation des critères psychométriques.

Critères	0	1	2	3
Validité de contenu	Non indiquée	Analyse insuffisante des items ¹ , quel que soit le nombre d'items	Analyse partielle des items ² , nombre d'items par épreuve < 20	Analyse complète des items ³ , nombre d'items par épreuve ≥ 20
Objectifs du test	Non indiqués	<i>Pour les batteries</i> : objectif général + précision des objectifs pour moins de 50 % des domaines étudiés ou des épreuves <i>Pour les tests à épreuve unique</i> : objectif très général, flou et peu spécifique	<i>Pour les batteries</i> : objectif général + objectifs pour 50 à 99 % des domaines étudiés et des épreuves	<i>Pour les batteries</i> : objectif général + objectifs pour tous les domaines étudiés et toutes les épreuves <i>Pour les tests à épreuve unique</i> : objectif clair, précis
Validité de construit	Non indiquée	Indication de références ou de modèles théoriques, sans analyse statistique	Indication de références ou de modèles théoriques, avec une analyse statistique incomplète ⁴ , ne portant que sur une partie des résultats ou sur un sous-groupe de moins de 100 sujets	Indication de références ou de modèles théoriques, avec une analyse statistique complète ⁵ sur un sous-groupe de plus de 100 sujets
Validité concourante (seuil : $r = .40$)	Non indiquée	Indiquée, mais sans preuve chiffrée	Indiquée, mais inférieure au seuil ou sur un sous-groupe de moins de 100 sujets	Supérieure ou égale au seuil et sur un sous-groupe de plus de 100 sujets
Validité prédictive (seuil : $r = .30$)				
Fidélité test-retest (seuil : $r = .80$)				
Fidélité inter-juges (seuil : $r = .80$)				
Cohérence interne (seuil : $r = .80$)				

¹ Analyse insuffisante : mention des critères de choix des items (ex. complexité, structure, fréquence) sans preuve chiffrée ou sans l'indication des pourcentages de réussite des items, sans que cela conduise à une modification du choix des items, c'est-à-dire à une version définitive et améliorée de l'épreuve.

² Analyse partielle des items : analyse statistique classique des items avec un seul indice renseigné : l'indice de difficulté ou de discrimination.

³ Analyse partielle des items : analyse statistique classique des items avec deux indices renseignés : l'indice de difficulté et de discrimination.

⁴ Exemples d'analyses statistiques incomplètes : ANOVA mettant en évidence des différences de scores liées à l'âge des sujets.

⁵ Exemples d'analyse statistique complète : analyse factorielle, analyse multi-traités multi-méthodes.

Critères	0	1	2	3
Instructions d'utilisation (administration, cotation, interprétation)	Non indiquées	Clair et précis pour seulement un type d'instructions parmi les 3 suivantes : administration, cotation, interprétation (cette dernière peut comprendre des analyses de cas)	Clair et précis pour 2 types d'instructions	Clair et précis pour les 3 types d'instructions
Qualifications de l'examineur	Non indiquées	Terme générique et flou (ex. praticien, professionnel, clinicien)	Liste lacunaire de professionnels, sans lien avec l'objectif ou les conditions de passation du test ⁶	Indication précise en lien avec l'objectif du test : orthophoniste, ou liste précise de professionnels (dont l'orthophoniste)
Erreur type de mesure	Non indiquée	Pas de tableau des erreurs type de mesure (ETM), ni de tableau des intervalles de confiance (IC), mais présence de données permettant à l'utilisateur de calculer les ETM et donc les IC ⁷	Présence d'un tableau des ETM, mais pas de tableau des IC	Présence de tableaux des IC
Description de l'échantillon d'étalonnage	Non indiquée	Pas de preuve de représentativité de l'échantillon, indications chiffrées à propos d'1 ou 2 critère(s) descriptif(s) ⁸	Pas de preuve de représentativité de l'échantillon, indications chiffrées à propos de 3 ou 4 critères descriptifs	Preuves de représentativité de l'échantillon, indications chiffrées à propos d'au moins 4 critères descriptifs
Taille de l'échantillon	Non indiquée	Moins de 50 participants par sous-groupe	Entre 50 et 99 participants par sous-groupe	100 participants ou plus par sous-groupe
Normes	Aucun indicateur de tendance centrale	Indicateur de tendance centrale mentionné, mais non adapté à la distribution de la population contrôle ⁹	Indicateur de tendance centrale adapté à la distribution de la population contrôle, mais mention confuse ou imprécise ou mal placée dans le manuel ou le logiciel	Indicateur de tendance centrale adapté à la distribution de la population contrôle et mention claire, précise ¹⁰ et bien placée dans le manuel ou le logiciel
Sensibilité et spécificité (seuil : $r = .80$)	Non évoquées	Évocation de l'une ou l'autre, sans preuve chiffrée	Indication de taux de sensibilité / spécificité inférieurs au seuil	Indication de taux de sensibilité / spécificité, égaux ou supérieurs au seuil

⁶ Par exemple, le test peut être utilisé dans une école, sans que les enseignants soient mentionnés comme potentiels utilisateurs.

⁷ La connaissance du coefficient de fidélité test-retest et des écarts-types permettent de calculer les erreurs-types de mesure (ETM) et donc les indices de confiance (IC) selon la formule : $IC = [\text{score brut} - \text{ETM seuil} ; \text{score brut} + \text{ETM seuil}]$.

⁸ Critères considérés : âge, sexe, origine géographique, catégorie socioprofessionnelle ou niveau socioculturel des parents.

⁹ Ex. Les moyennes et *ET* sont indiqués, mais sans analyse de la distribution des données contrôles, ou avec des éléments montrant une distribution non gaussienne pour au moins une note, et l'utilisation des percentiles n'est pas recommandée en cas de distribution non gaussienne.

¹⁰ Ex. pour les percentiles : indication des P2-3 ou du P5.

4. Scores des 15 critères psychométriques pour les 27 tests de langage oral.

Tests	Cohérence interne	Échantillon décrit, voire représentatif	Erreur type de mesure	Fidélité inter-juges	Fidélité test-retest	Instructions d'utilisation	Normes	Objectifs du test	Qualifications de l'examinateur	Sensibilité Spécificité	Taille de l'échantillon	Validité concourante	Validité de construit	Validité de contenu	Validité prédictive
BILO EC-2	0	2	0	0	0	3	2	3	1	0	3	2	1	1	0
BILO-2	0	2	0	0	0	3	2	3	1	0	3	0	2	1	0
BILO-3C	0	2	0	0	0	3	2	3	1	0	2	0	2	1	0
BILO-Petits	0	2	0	0	0	3	2	3	1	0	3	0	1	1	0
Boehm-3	2	2	2	0	2	3	3	3	3	1	2	3	1	1	0
CCC	2	1	0	2	0	3	1	3	1	1	1	1	1	0	0
CLEA	0	2	0	0	0	3	3	3	1	0	1	2	2	1	0
ECOSSE	0	2	0	0	0	3	1	3	2	0	3	0	2	3	0
ECSP	0	1	0	0	0	3	1	3	1	1	1	2	2	1	0
ELDP	0	2	0	0	0	2	3	3	3	0	1	0	1	1	0
ELO	0	1	0	0	0	3	2	3	0	1	2	0	1	1	0
EVALéo	2	3	0	1	1	3	2	3	2	0	3	0	1	1	0
EVALO 2-6	2	2	0	1	0	3	1	3	3	1	2	0	2	2	0
EVIP	2	2	3	0	2	3	1	3	3	0	2	2	2	3	1
EXALang 11-15	1	2	3	1	1	3	1	3	3	1	2	1	2	0	1
EXALang 3-6	0	2	0	0	2	2	1	3	3	1	2	1	1	1	1
EXALang 5-8	1	1	0	1	1	3	1	3	3	1	1	1	1	0	1
EXALang 8-11	1	2	3	1	1	3	1	3	3	1	2	1	2	0	1
Hénin	1	0	0	0	1	2	1	3	0	0	2	0	0	2	0
IFDC	1	2	0	0	1	2	3	3	3	0	1	1	2	2	1
Kikou 3-8	0	2	0	0	0	1	3	3	3	1	2	1	1	1	0
L2MA-2	0	2	0	0	0	3	1	3	3	0	1	0	1	1	0
N-EEL	0	2	0	0	2	3	1	3	1	1	3	1	1	1	1
PEES 3-8	0	2	0	0	0	1	3	3	1	1	2	1	2	1	0
Pélea	0	2	0	0	0	1	1	3	1	1	1	0	2	1	0
T.L.O.C.C.	2	2	0	0	0	3	1	3	1	1	3	3	3	2	0
TCS	2	3	0	0	0	3	1	3	3	0	2	2	2	1	0