

**Standardisation et validation prédictive de la BATEDA,
un outil de dépistage de retards langagiers pour enfants
francophones camerounais d'âge préscolaire**

Aurélie Takam^{*}, Victor Boucher^{*}, Natacha Trudeau^{}**

^{*} Laboratoire de sciences phonétiques, Université de Montréal

aurelie.takamtaguemne@mail.mcgill.ca

victor.boucher@umontreal.ca

^{**} École d'orthophonie et d'audiologie, Université de Montréal

natacha.trudeau@umontreal.ca

auteure de correspondance :

Child language development and disorders Lab. , McGill University

aurelie.takamtaguemne@mail.mcgill.ca

Résumé :

Nous présentons le premier outil normalisé de dépistage des retards de langage destiné aux enfants francophones d'âge préscolaire des pays africains comme le Cameroun. Il faut noter que ces pays se caractérisent par l'absence de services publics de prise en charge orthophonique. A partir d'un échantillon de 159 enfants camerounais francophones de 4 et 5 ans, nous avons normalisé la batterie de tests suivant la cote z et les rangs en centiles. Les tests présentent une bonne fiabilité inter-juge, une cohérence interne d'environ 96% (coefficient $\alpha = 0,958$) et une validité prédictive de la réussite scolaire des enfants ($r = 0,728$).

Mots clés : test de langage, évaluation du langage, dépistage, psychométrie du langage, retard de langage, développement du langage et éducation primaire.

Standardization and predictive validation of BATEDA, a tool for assessing language delays in French speaking preschool children from Cameroon

Summary:

We present the first standard language-assessment tool for French speaking preschool children in African countries like Cameroon. These countries are characterized by the absence of public services in speech and language disorders. From a sample of 159 French-speaking Cameroon children aged 4 and 5 years, we normalized the language tests using z -scores and percentile ranks. The test battery has a good inter-judge reliability, an internal consistency of about 96% ($\alpha = .958$ Cronbach's) and a predictive validity of academic success of children ($r = 0.728$).

Key words: language test, language assessment, language screening, language psychometry, language delay, language development and primary education.

----- INTRODUCTION -----

La présente étude porte sur la standardisation du premier outil d'évaluation langagière répondant au besoin d'aide d'enfants camerounais (Afrique subsaharienne) d'âge préscolaire présentant des retards de langage oral. Il s'agit de la Batterie de tests de dépistage approfondi ou BATEDA, un outil dont la particularité réside dans son contexte d'utilisation (Takam et al., 2013). En effet, compte tenu de l'absence de services publics orthophoniques dans les pays subsahariens tels que le Cameroun, notre batterie de tests est destinée à l'usage des parents et des enseignants de la maternelle et du primaire. Trois conditions majeures ont sous-tendu cette orientation.

Premièrement, il y a l'inexistence d'un environnement propice à la prise en charge orthophonique publique au Cameroun. Nous pouvons principalement noter le faible nombre d'orthophonistes dans ce pays, l'absence de services publics d'orthophonie et la méconnaissance des difficultés langagières souvent assimilées à des déficiences mentales par les familles et la marginalisation des personnes aux prises avec ces difficultés (Takam, 2006). Cependant, cet état de fait occulte la situation réelle de ces difficultés dans la société. Le *DSM-IV* (2003) rapporte une fourchette de 3% à 7% d'enfants d'âge préscolaire et scolaire présentant des retards de langage. Au Cameroun, Takam (2006) a rapporté environ 6% d'enfant d'âge préscolaire et scolaire ayant des troubles de l'articulation dévoilant un retard de parole. En suivant ces différents taux de prévalence, nous pouvons estimer qu'environ 20000 de ces enfants (en moyenne) présenteraient des difficultés sous forme de retards de langage. En effet, l'Institut national de la statistique du Cameroun¹ estime qu'environ 20% de la population globale évaluée à plus de 19 millions sont constitués d'enfants d'âge préscolaire (4 à 5 ans) et scolaire (6 à 11 ans). Par ailleurs, le taux de scolarité primaire est d'environ 80% contre un taux de redoublement d'environ 24%.

Deuxièmement, compte tenu de leur fonction sociale d'éducateurs, les parents et les enseignants pourraient intervenir de façon bénéfique dans les cas de difficultés sans étiologie connue, ni condition pathologique associée. Ils apparaissent pour ainsi dire comme étant des intervenants potentiels. Enfin, la société camerounaise ne dispose d'aucune information pour prévenir ces problèmes, ni d'outil d'évaluation adapté qui permettrait de les repérer et de cibler une stratégie d'amélioration des niveaux de compétence verbale affectés. La nécessité d'un outil spécifique et l'originalité de l'étude tiennent donc au contexte camerounais qui diffère substantiellement du contexte occidental.

Il existe plusieurs outils d'évaluation destinés aux enfants d'âge préscolaire (4 et 5 ans) et couvrant divers aspects du langage. Cependant, leur inapplicabilité à la population cible tient à deux aspects majeurs : la situation sociolinguistique et le contexte culturel. Au plan sociolinguistique, ces tests s'adressent généralement à des enfants monolingues, ayant une langue indo-européenne comme langue de communication. En revanche, les enfants camerounais évoluent dans des environnements fortement plurilingues où la maîtrise de la langue véhiculaire varie suivant la zone géographique (urbaine *versus* rurale). Ce contexte pose le problème de niveau de connaissance variable de la langue véhiculaire. Au plan culturel, l'environnement quotidien des enfants en occident est marqué par une conscience sociale de la nécessité de prendre en charge les difficultés langagières. Cela implique une

¹ Voir <http://www.statistics-cameroon.org/> Tableau 4.04 : Répartition en pourcentage de la population par groupe d'âge

attitude préventive à travers l'attention familiale, scolaire et publique, et une organisation des services de soins de santé orientée vers une approche orthophonique. Ainsi, les tests sont généralement conçus pour être appliqués par des professionnels en orthophonie. Leur application implique donc plusieurs prérequis, limitant de ce fait leur accessibilité dans le contexte camerounais.

Compte tenu de ce contexte, l'usage de la batterie de tests vise de façon primordiale le dépistage des retards de langage et la détermination des structures problématiques (Takam et al., 2013). La BATEDA étant conçue pour être administrée par des personnes non spécialistes en la matière, il était impératif de standardiser la procédure d'évaluation et d'interprétation des résultats. C'est l'objectif général de notre étude.

----- PRESENTATION DE LA BATEDA -----

Le tableau 1 ci-dessous résume les caractéristiques des épreuves qui constituent la BATEDA en indiquant le nombre d'items et la durée de l'évaluation, la nature de la tâche de l'enfant et de l'examineur. Takam et al. (2013) présentent l'outil de façon détaillée. Le test visuel est une adaptation du test standard « *Lea Symbols : Échelle d'acuité visuelle* ». Il évalue les deux yeux à la fois à l'aide d'un tableau situé à trois mètres de l'enfant et présentant quatre images sur dix lignes à des grandeurs décroissantes. Il y figure en plus quatre cartes représentant chacune des images que doit identifier l'enfant. Le test auditif détermine la capacité auditive de chaque oreille à l'aide d'un CD générant trois fréquences pulsées de 500, 1000 et 2000 Hz. Ces trois sons varient progressivement en amplitude de 0 à 80 dB SPL (mesurée à la sortie d'écouteurs intra-oraux) sur 80 secondes, ce qui permet d'établir une perte de l'ordre de 40 dB SPL sur les fréquences cibles. Une telle perte correspond à une surdité moyenne se rapportant à une difficulté à discriminer des sons de la parole à proximité et même dans un milieu silencieux. Les tests de parole déterminent respectivement les capacités discriminatoires et articulatoires des sons en moins de huit minutes. La détermination des sons cibles est faite à partir d'une étude antérieure décrivant les troubles articulatoires au Cameroun (Takam, 2006). Pour les épreuves langagières, nous avons construit les items en nous basant sur les particularités du langage oral entre quatre et six ans, les caractéristiques des retards de langage oral et les attentes langagières telles que définies par le système éducatif camerounais (Ministère de l'Éducation Nationale, 1998). Par ailleurs, nous nous sommes inspirés du format des tests nord-américains pour la présentation des épreuves. Ainsi, le *Test Of Language Development – Primary* (Newcomer, Hammill, 1982) a offert le format des tests morphologique et syntaxique respectivement à travers les sous-tests *Grammatical completion* et *Grammatical understanding*. Les tests de vocabulaire sont inspirés respectivement du *Peabody Picture Vocabulary Test* (Dunn, Dunn, 1981) et de l'*Expressive One-Word Picture Vocabulary Test* (Gardner, 1990). Enfin, le test pragmatique est inspiré d'un exercice utilisé dans les programmes scolaires au Cameroun.

| <i>Modalités et composantes</i> | <i>Tests et durée</i> | <i>Nature de la tâche de l'enfant</i> | <i>Réponses à relever</i> | <i>Support</i> |
|---------------------------------|---|---|---------------------------------|-------------------|
| <i>Vision</i> | <i>Lea Symbols : Échelle d'acuité visuelle – 5 minutes</i> | Montrer l'image correspondant à celle pointée | Cocher la ligne | Visuel |
| <i>Audition</i> | Fréquences pulsées de 500, 1000 et 2000 Hz – 2 minutes | Lever la main à l'écoute d'un son | Relever le temps du chronomètre | Auditif |
| <i>Articulation</i> | <i>Test de l'articulation des sons (TAS), 28 items - 2,29 minutes</i> | Répéter le son entendu | Écrire le son répété | Auditif |
| <i>Discrimination</i> | <i>Test de reconnaissance de sons (TRS), 37 items - 5,07 minutes</i> | Discriminer les paires de mots pareils et différents | Cocher la réponse de l'enfant | Auditif |
| <i>Morphologie</i> | <i>Test de dérivation des mots (TDM), 25 items - 3,58 minutes</i> <i>Test d'accord des mots (TAM), 25 items - 3,49 minutes</i> | Compléter la phrase entendue en trouvant le lexème attendu | Écrire le lexème | Auditif |
| <i>Vocabulaire</i> | <i>Test de vocabulaire réceptif (TVR), 50 items - 4,53 minutes</i> | Montrer l'image correspondant au lexème entendu | Cocher la case correspondante | Auditif et visuel |
| | <i>Test de vocabulaire expressif (TVE), 50 items - 5,49 minutes</i> | Nommer l'image | Écrire le lexème | Auditif et visuel |
| <i>Syntaxe</i> | <i>Test de compréhension des phrases (TCP), 25 items - 7,11 minutes</i> | Montrer l'image correspondant à la phrase entendue | Cocher la case correspondante | Auditif et visuel |
| <i>Pragmatique</i> | <i>Test de compréhension du discours (TCD), 1 histoire, 10 questions - 6,49 minutes</i> | Écouter une histoire et répondre aux questions de compréhension | Cocher la case correspondante | Auditif et visuel |

Tableau 1. Tests des modalités et des composantes langagières de la *BATEDA*.

----- PARAMETRES PSYCHOMETRIQUES DE STANDARDISATION D'UN TEST D'ÉVALUATION DU LANGAGE -----

Les paramètres psychométriques qui caractérisent tout outil d'évaluation standard portent sur la normalisation des scores, la fiabilité et la validité de l'outil (Anastasi, Urbina, 1997 ; McCauley, 2001 ; Shipley, McAfee, 2009). La normalisation des scores permet de déterminer la population cible et de quantifier ses performances langagières. Cette population se définit par rapport aux critères tels que l'âge, le sexe, la zone résidentielle et le niveau socioéconomique. Après avoir administré les tests à un échantillon représentatif de cette population, la quantification du langage se fait à partir de la conversion statistique des résultats bruts aux tests en scores standards afin d'obtenir la norme de référence pour la population cible. C'est à partir de la distribution de ces scores normés (écart-type et rang en centile) que le seuil de réussite aux tests est déterminé.

Les principales mesures permettant d'évaluer l'efficacité d'un test sont la fiabilité inter-juge, la consistance interne et le test re-test. La fiabilité inter-juge détermine le degré d'objectivité

de la procédure d'administration et de cotation. En effet, lorsqu'un test est correctement administré, les résultats ne doivent pas varier selon la personne qui l'utilise. La consistance interne permet d'évaluer l'homogénéité du contenu des items qui constituent le test, rendant ainsi compte de la représentativité du test par rapport à l'habileté évaluée. Enfin, le test re-test mesure la consistance du test dans le temps. Il s'agit de s'assurer que les résultats d'un enfant reflètent réellement ses habiletés. Pour ce faire, l'enfant est réévalué quelques semaines après sa première évaluation afin de comparer ses deux résultats. En considérant les différentes mesures de fiabilité, Wasserman et Bracken (2003) suggèrent qu'une excellente mesure de consistance interne devrait être supérieure à 0,80 alors qu'un coefficient de 0,75 et plus serait excellent pour la fiabilité interjuges. Selon Anastasi et Urbina (1997), des coefficients se situant entre 0,80 et 0,90 seraient l'expression d'un très bon degré de fidélité d'un outil psychométrique. De plus, l'Administration for children and families (2011) situe le seuil acceptable entre 0,65 et 0,70.

En ce qui concerne la validité d'un test langagier, elle renvoie à sa capacité à bien évaluer les compétences langagières ciblées. Cette mesure est un moyen de minimiser le nombre de faux-positifs et de faux-négatifs identifiés par un outil. La principale mesure de validité d'un test, la validité prédictive (en anglais *predictive validity*), vise à établir la preuve que les résultats au test donnent une estimation d'une compétence ultérieure. Il y a aussi la validité de contenu (en anglais *content validity*) qui détermine dans quelle mesure les items choisis sont représentatifs de la compétence à évaluer, la validité de construction (en anglais *construct validity*) qui reflète la solidité empirique de la conceptualisation du langage adoptée pour l'élaboration des tests. Notons enfin les mesures de sensibilité et de spécificité qui évaluent la capacité d'un test à poser un diagnostic précis. Pour une bonne validité prédictive, Glascoe et al. (1992) situent le score préférable autour de 70%. Cependant, le seuil acceptable serait de 30% (Beaucage, Bonnier Viger, 1996). Enfin, pour plusieurs outils de dépistage et d'évaluation langagière, ce score se situerait entre 30% et 50% (Aylward, Stancin, 2008).

----- OBJECTIFS SPECIFIQUES -----

De façon spécifique, la présente étude de standardisation de la BATEDA vise à appliquer les paramètres psychométriques que sont la normalisation des scores, la fiabilité interjuge, la cohérence interne et la validité prédictive. Il est important de noter que les critères de construction des items tests résumés ci-dessus (lire *Présentation de la BATEDA*) et présentés dans Takam et al. (2013) confèrent une validité de contenu à l'outil. Nous avons également effectué des mesures de validité de construit, mais les résultats seront présentés dans une étude subséquente. Cependant, nous n'avons pas procédé aux mesures de fiabilité test re-test. Il en est de même pour les mesures de sensibilité et de spécificité qui feront l'objet d'une enquête ultérieure.

----- METHODE -----

1. Echantillon

La collecte des données s'est faite dans sept écoles maternelles et primaires à Yaoundé, la capitale politique du Cameroun. Comme plusieurs grandes villes africaines, Yaoundé est un milieu urbain qui se caractérise par une grande diversité linguistique. Il s'agit donc d'une ville fortement plurilingue, ce qui au fil du temps, a favorisé l'usage du français comme principale

langue de communication. Il faut dire que le français ayant le statut de langue officielle, la scolarisation a contribué à faire de lui une des langues premières des enfants dans les zones urbaines. Dans notre étude, nous avons catégorisé les enfants afin de ne considérer que ceux dont le français était la principale langue de communication. Ainsi, nous avons constitué un échantillon « par quota » de 159 enfants afin de représenter les principales catégories socioéconomiques.

Trois écoles privées ont permis d'avoir accès aux enfants de milieu aisé (36 enfants, soit 22,6% de l'échantillon), et quatre autres écoles (des écoles publiques) ont mis à disposition les enfants de milieu socioéconomique moyen à pauvre (123 enfants, 77,4% de l'échantillon). La distribution de cet échantillon par tranches d'âge est équilibrée avec 79 enfants de 4,00 ans à 4,11 ans (soit 49,7% de l'échantillon) et 80 enfants de 5,00 ans à 5,11 ans (soit 50,3% de l'échantillon). Nous avons également veillé à l'équilibre de l'échantillon par sexe : 80 filles (50,3% de l'échantillon) et 79 garçons (49,7% de l'échantillon). Pour la validation prédictive, la collecte des notes scolaires s'est faite auprès de 32 des 159 enfants échantillonnés. La répartition de ce sous-échantillon suivant le niveau socioéconomique est de 11 enfants de milieu aisé contre 21 enfants de milieu moyen à pauvre. Il s'agit de 27 enfants de quatre ans et de 5 enfants de cinq ans.

2. Procédure de collecte et d'analyse des données

Deux étudiants gradués de niveau maîtrise spécialisés en linguistique appliquée et la chercheuse principale ont participé à la collecte des données. Leur rôle était d'administrer la BATEDA aux enfants pris individuellement. Les enfants étaient assignés à chaque examinateur de façon aléatoire par les enseignants. La collecte des réponses des enfants s'est faite à l'aide des fiches de réponses faisant partie de l'outil. Les données de normalisation se présentent sous forme de scores bruts par enfant pour chaque item et pour chaque test. L'évaluation s'est déroulée le jour dans une salle bien éclairée et à l'abri de tout dérangement. L'enfant était confortablement assis sur une chaise placée à trois mètres du mur le plus éclairé de la salle et sur lequel on avait affiché le tableau d'images du test visuel. Le reste du matériel était disposé sur la table placée devant l'enfant et l'examineur était assis à la gauche de l'enfant tel qu'illustré sur la figure 1 ci-dessous. Le matériel servant à effectuer l'évaluation comprend deux disques compacts (le disque CD 1 intitulé « *Test de l'audition* » et le disque CD 2 intitulé « *Batterie de tests de dépistage approfondi* »), un équipement visuel (le tableau d'images *LEA chart* et les quatre cartes *LEA symbol cards* pour le test *Lea Symbols: Échelle d'acuité visuelle*, et le catalogue d'images), des fournitures (crayon, punaises, ruban à mesurer en mètres, le livret de fiches des réponses et le livret de grilles d'évaluation) et un équipement électronique (un *walkman calibré* muni de deux piles AA fonctionnelles, une paire d'*inserts auditifs* munis de mousses jaunes ou grises connectée au walkman à l'aide d'un adaptateur et d'une fiche Y à deux têtes).

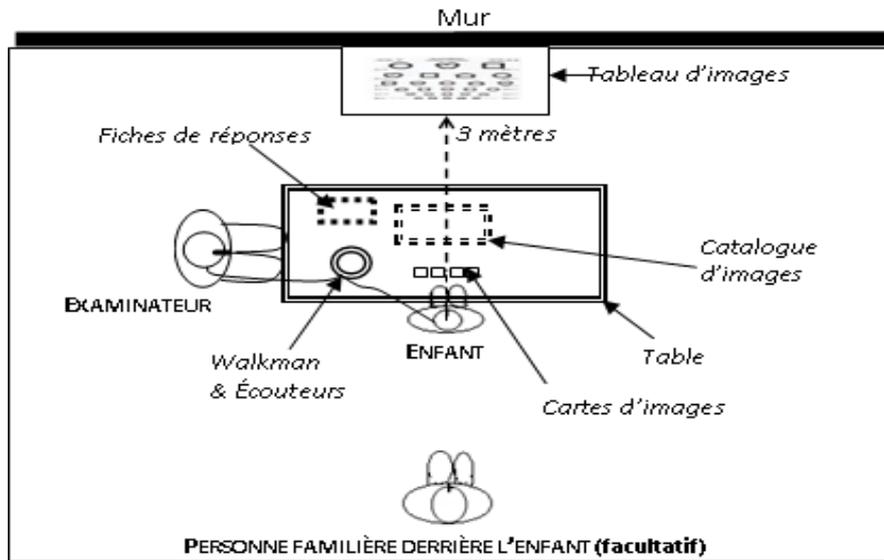


Figure 1. Procédure d'évaluation.

L'évaluation a commencé par les tests visuel et auditif qui sont apparus comme des mesures sélectives ; tout enfant ne réussissant pas à ces tests était exclu de l'échantillon. Le seuil de réussite était de 20/32 pour la vision. Quant à l'audition, l'identification d'une perte était déduite si la réponse de l'enfant dépassait 40 secondes sur le chronomètre du *Walkman*. Puis, ont suivi sans interruption les tests de parole et de langage tous enregistrés sur un disque compact (le CD 2). La durée totale de l'évaluation était d'environ une heure et aucun des enfants évalués n'a manifesté de signe de fatigue. En ce qui concerne la cotation des réponses des enfants aux tests de parole et de langage, on attribuait 1 point pour une réponse juste et 0 point pour une réponse fautive ou un silence. Il faut noter que pour que les résultats d'un test soient valides, il fallait avoir soumis tous les items-tests à l'enfant. L'addition de tous les points accumulés a permis d'obtenir le score brut de chaque enfant par test.

Pour la validation prédictive, la collecte des notes scolaires s'est déroulée douze mois après la collecte des données de normalisation de la BATEDA. Il s'est agi de recueillir les notes des enfants dans les six matières les plus déterminantes de sa réussite scolaire et qui portent sur le français écrit et le calcul. Comme on peut le lire sur le tableau 2 ci-dessous, les notes reflètent précisément les performances des enfants en dictée, en réponses aux questions de compréhension de texte, en rédaction de texte, en lecture à voix haute, en mathématiques et en résolution de problèmes. Étant donné les taux de corrélation très élevés entre ces notes et la moyenne finale (significatifs au niveau 0,0001 et variant entre 0,8 et 0,9), nous n'avons utilisé que la moyenne globale des notes dans ces six matières pour étudier la validité prédictive des tests. L'analyse statistique s'est faite à l'aide de SPSS (version 17.0).

| <i>Domaine</i> | <i>Matières</i> | <i>Description</i> | <i>Échelle de notation</i> | <i>r</i> |
|----------------|------------------|---|----------------------------|----------|
| Français | <i>Dictée</i> | Reproduction écrite d'un texte lu à haute voix. | 10 | ,893 |
| | <i>Questions</i> | Questions de compréhension du texte de la dictée. | 10 | ,863 |
| | <i>Rédaction</i> | Production écrite d'un texte sur un fait suggéré. | 10 | ,902 |
| | <i>Lecture</i> | Lecture à haute voix d'un texte. | 10 | ,803 |
| Calcul | <i>Maths</i> | Exercices de calcul mathématiques : addition, soustraction, multiplication et division. | 10 | ,901 |
| | <i>Problème</i> | Résolution de problèmes impliquant des opérations mathématiques. | 10 | ,831 |

Tableau 2. Choix des notes scolaires pour les analyses de validation prédictive de la batterie de tests (r = corrélation avec la moyenne finale selon l'échelle des notes).

----- RESULTATS -----

1. Mesures de fiabilité

La cohérence interne : nous avons examiné la statistique l'*alpha de Cronbach* (ou α) afin de voir s'il y a une corrélation entre les items testant les modalités de parole d'une part et les composantes langagières d'autre part. Cette mesure permet généralement de savoir si les tests évaluent correctement les habiletés cibles. Le tableau 3 qui suit récapitule les résultats obtenus. Les tests des modalités de parole ont un taux de fiabilité interne d'environ 86% ($\alpha = 0,862$). Ce taux est de 95% pour les tests langagiers ($\alpha = 0,954$) et d'environ 96% pour l'ensemble des items de la BATEDA ($\alpha = 0,958$).

| <i>Batteries de tests</i> | <i>Nombre d'items</i> | <i>Moyenne</i> | <i>Erreur standard de la moyenne</i> | <i>Écart-type</i> | <i>Alpha de Cronbach</i> |
|------------------------------|-----------------------|----------------|--------------------------------------|-------------------|--------------------------|
| <i>Modalités^a</i> | 65 | 42,29 | 1,174 | 14,798 | ,862 |
| <i>Langage^a</i> | 184 | 92,28 | 2,036 | 25,669 | ,954 |
| <i>BATEDA^a</i> | 273 | 143,70 | 3,456 | 43,576 | ,958 |

^a Scores cumulés : tests des modalités (*articulation des sons et reconnaissance des sons*) ; tests des composantes langagières (*dérivation des mots, accord des mots, vocabulaire réceptif, vocabulaire expressif, compréhension de phrases et compréhension de discours*) ; BATEDA (tous les tests)

Tableau 3. Mesures de fiabilité de cohérence des tests (*Alpha de Cronbach*).

La fiabilité interjuge : étant donné que trois examinateurs ont administré les tests auprès de groupes d'enfants différents, nous avons comparé la moyenne de groupe de ces trois juges afin d'évaluer l'objectivité de la procédure d'évaluation. Cette comparaison a été effectuée à l'aide du test non-paramétrique de *Kruskal-Wallis* pour plusieurs échantillons indépendants parce que la distribution des scores par juge était anormale. Par ailleurs, la variation du

nombre d'enfants par examinateur se justifie par la constitution aléatoire des groupes. De ce fait, le premier examinateur a évalué 66 enfants, le deuxième, 57 enfants et le troisième, 36 enfants. Les résultats que présente le tableau 4 ci-dessous indiquent que la BATEDA a une fiabilité inter-juge de 0,911 au niveau 0,05. Cette mesure indique que la procédure d'évaluation est objective et que les examinateurs n'ont pas eu d'impact sur les résultats de l'évaluation. Ceci reflète le fait que tous les tests étaient enregistrés sur un CD, ce qui n'a laissé aucune marge de manœuvre aux examinateurs lors de l'évaluation des enfants.

| Variables | Examineurs | | | Khi-deux | Indice de fiabilité | Niveau de signification |
|-----------|------------|----|----|----------|---------------------|-------------------------|
| | 1 | 2 | 3 | | | |
| BATEDA | 66 | 57 | 36 | ,187 | ,911 | 0.05 |

Tableau 4. Mesures de fiabilité inter-juge (Khi-deux de *Kruskal-Wallis*).

2. Normalisation

Distribution et standardisation des tests : Le test de *Levene* a révélé l'homogénéité de la variance pour toutes les variables et l'exploration des données n'a détecté aucune valeur extrême. Le tableau 5 ci-dessous présente les statistiques descriptives des tests en indiquant les mesures de tendance centrale (moyenne, médiane et centile) et les mesures de dispersion (écart-type, score minimum et score maximum). De façon générale, les moyennes des tests langagiers sont presque identiques aux valeurs médianes, ce qui est une propriété de la distribution normale des données.

| | Variables | Moy. | Méd. | ESM | E.-T. | Min. | Max. | Centile 10 | Centile 95 |
|---------|-----------|-------|------|------|-------|------|------|------------|------------|
| Parole | TAS | 23,88 | 25 | ,275 | 3,468 | 11 | 28 | 19,00 | 28,00 |
| | TRS | 26,61 | 28 | ,673 | 7,060 | 10 | 37 | 16,10 | 36,00 |
| Langage | TDM | 8,85 | 9 | ,431 | 5,440 | 0 | 21 | 2,00 | 18,00 |
| | TAM | 11,48 | 11 | ,394 | 4,967 | 1 | 22 | 5,00 | 19,00 |
| | TVR | 31,14 | 31 | ,622 | 7,847 | 14 | 48 | 21,00 | 44,00 |
| | TVE | 23,36 | 23 | ,546 | 6,879 | 3 | 42 | 16,00 | 35,00 |
| | TCP | 11,62 | 12 | ,222 | 2,801 | 5 | 19 | 8,00 | 17,00 |
| | TCD | 5,82 | 6 | ,155 | 1,960 | 1 | 10 | 3,00 | 9,00 |

N = échantillon ; Moy. = moyenne ; Méd. = médiane ; ESM = erreur standard de la moyenne ; E.-T. = écart-type ; Min. = score minimum ; Max. = score maximum.

TAS = Test de l'articulation des sons ; TRS = Test de reconnaissance des sons ; TDM = Test de dérivation des mots ; TAM = Test de l'accord des mots ; TVR = Tests de vocabulaire réceptif ; TVE = Test du vocabulaire expressif ; TCP = Test de compréhension de phrases ; TCD = Test de compréhension de discours.

Tableau 5. Statistiques descriptives des variables de normalisation (*N*=159).

Cependant, tel n'est pas le cas pour les tests de parole ; le *TAS* et le *TRS* ne sont pas normalement distribués comme l'illustrent les histogrammes A et B de la figure 2 ci-dessous. Ils indiquent des asymétries reflétant un plafonnement des scores pour ces deux tests. Ce plafonnement apparaît à travers les scores minimums assez élevés (respectivement 11 sur 28 et 10 sur 37). Toutefois, il faut noter que ces asymétries étaient prévisibles dans la mesure où l'échantillon d'étude était constitué d'enfants dont l'âge prédisposait à la maîtrise de la majorité des sons. Quant aux tests de langage, les histogrammes C à H de la figure 2 ci-dessous indiquent la normalité de la distribution de tous les scores. Suite à cette vérification des données, nous avons standardisé les scores à partir de la cote *z* et des rangs en centile. Chaque score brut correspond à un score standard et le seuil inférieur de réussite à un test correspond au 10ème rang en centile, c'est-à-dire qu'il se situe entre 1,25 et 1,50 écart-type en dessous de la moyenne (McCauley, 2001).

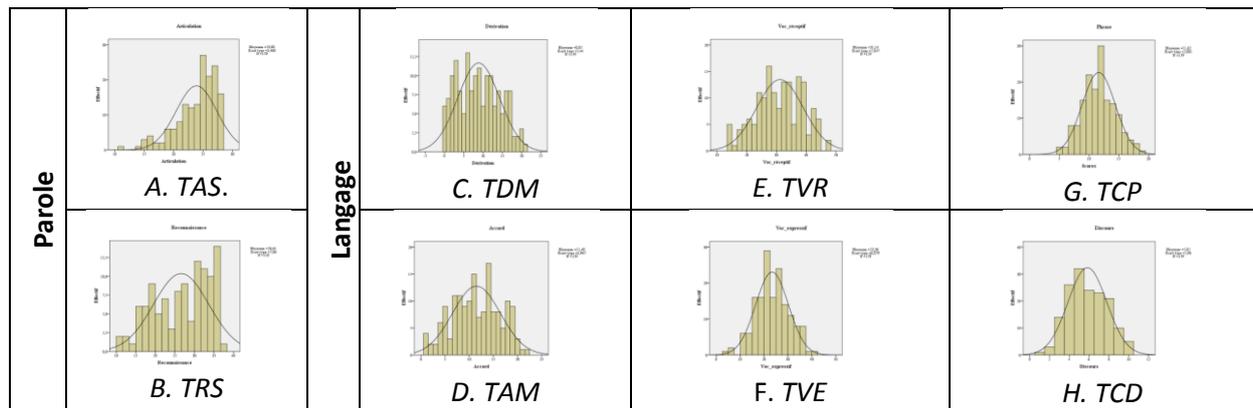


Figure 2. Histogrammes de distribution des scores aux tests (n = 159).

TAS = Test de l'articulation des sons ; TRS = Test de reconnaissance des sons ; TDM = Test de dérivation des mots ; TAM = Test de l'accord des mots ; TVR = Tests de vocabulaire réceptif ; TVE = Test du vocabulaire expressif ; TCP = Test de compréhension de phrases ; TCD = Test de compréhension de discours.

3. Validation prédictive

Nous avons examiné, au moyen de mesures de corrélations, la capacité de la batterie de tests de langage (BATEL) à prédire les performances scolaires des enfants sur six matières scolaires. Le tableau 6 ci-dessous présente les statistiques descriptives des variables utilisées pour les analyses, soient les notes scolaires obtenues par 32 des enfants dans l'échantillon. Les données ne présentent aucune valeur extrême, et toutes les moyennes se rapprochent des valeurs médianes. Nous avons uniquement considéré la moyenne des notes scolaires retenues (*Note_moy*) qui présente une distribution normale (voir figure 3). Pour ce qui est des tests langagiers, nous les avons regroupés par composantes (morphologie, vocabulaire, syntaxe et discours) et les histogrammes de la figure 3 ci-dessous présentent également des distributions normales. De plus, nous avons procédé à l'évaluation de la linéarité par un examen visuel des diagrammes de dispersion ; aucune distribution ne s'écartait de l'hypothèse de linéarité.

| | <i>Variables</i> | <i>Moy.</i> | <i>Méd.</i> | <i>ESM</i> | <i>E.-T.</i> | <i>Min.</i> | <i>Max.</i> | <i>Centile 5</i> | <i>Centile 95</i> |
|--------------|-------------------------------|-------------|-------------|------------|--------------|-------------|-------------|------------------|-------------------|
| | Note_moy. ¹ | 7,05 | 7,20 | ,42 | 2,42 | 1,33 | 10,90 | 1,59 | 10,83 |
| Tests | Morpho | 22,91 | 24,50 | 1,594 | 9,01 | 6,00 | 39,00 | 7,30 | 37,05 |
| | Lexique | 59,00 | 61,50 | 2,718 | 15,37 | 18,00 | 82,00 | 30,35 | 81,35 |
| | Syntaxe | 11,19 | 11,50 | ,501 | 2,83 | 5,00 | 18,00 | 6,30 | 16,05 |
| | Discours | 5,78 | 6,00 | ,421 | 2,37 | 2,00 | 10,00 | 2,00 | 10,00 |
| | BATEL ² | 98,88 | 99,00 | 4,56 | 25,78 | 42,00 | 149,00 | 53,05 | 141,85 |

¹ Moyenne des six notes scolaires. ² Score global de la *Batterie de tests de langage*.

Tableau 6. Statistiques descriptives des notes scolaires et des tests ($n = 32$).

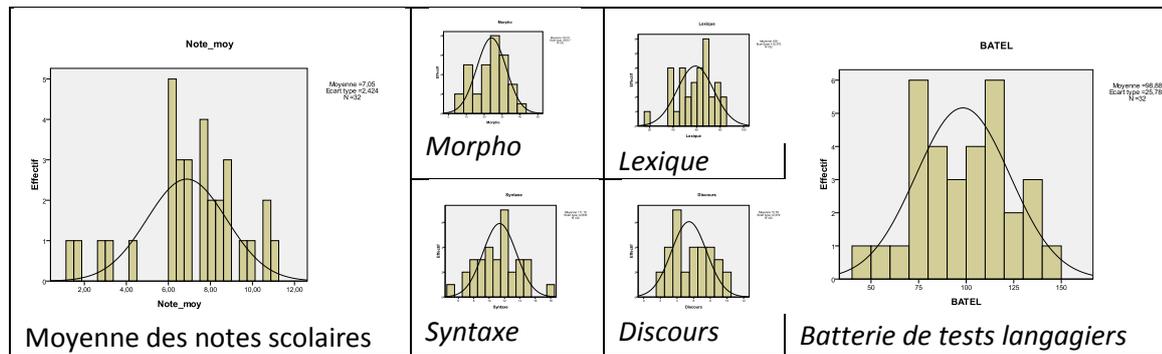


Figure 3. Histogrammes de distribution des scores aux tests langagiers ($n = 32$)

Mesures de corrélation : nous avons effectué des analyses de corrélation non paramétriques en évaluant le coefficient rho de Spearman (r) afin de comparer les performances d'un même individu dans chacun des tests à sa note scolaire moyenne et de vérifier leur degré de concordance. Le tableau 7 ci-dessous résume les coefficients de corrélations obtenus qui sont significatifs au niveau 0,01. De façon générale, les scores globaux aux tests de langage (BATEL) présentent une excellente validité prédictive du succès scolaire des enfants ($r = 0,728$), indiquant que les connaissances en langage oral contribuent de façon marquée à l'apprentissage du langage écrit, mais aussi à la maîtrise des exercices de mathématiques. En effet, tout comme le français écrit, l'enseignement des mathématiques se fait en français. Ainsi, il est logique qu'un enfant ayant des difficultés à comprendre les mots et les phrases soit incapable de résoudre des problèmes impliquant des calculs mathématiques. La meilleure validité prédictive revient aux tests de vocabulaire ($r = 0,743$). Cependant, les autres tests sont tout aussi importants pour une évaluation approfondie du langage compte tenu de la variabilité du profil des difficultés langagières chez les enfants. On sait qu'un enfant peut très bien être performant en vocabulaire et présenter des difficultés sur le plan morphologique, syntaxique et/ou discursif. À cet effet, une stratégie qui consisterait à limiter l'évaluation au vocabulaire pourrait mener à la présence de faux-négatifs lors du dépistage. Pour éviter de tels résultats, il est important d'évaluer toutes les composantes langagières.

| <i>Tests</i> | r |
|--------------|--------|
| Lexique | ,743** |
| Morpho | ,532** |
| Discours | ,514** |
| Syntaxe | ,482** |
| BATEL | ,728** |

**La corrélation est significative au niveau 0.01 (bilatéral).

Tableau 7. Corrélations (r) entre les notes scolaires et les scores aux tests (scores z).

----- DISCUSSION DES RESULTATS -----

1. Mesures de fiabilité

En ce qui concerne la cohérence interne des tests, nous avons remarqué que l'*alpha de Cronbach* variait selon le nombre d'items : plus il y avait d'items, plus grand était *alpha*. Ceci s'explique par le fait que cette statistique est sensible au nombre d'items (Urbina, 2004). En outre, en comparant le taux de fiabilité des tests de langage ($\alpha = 0,954$) à celui de la BATEDA ($\alpha = 0,958$), nous pouvons noter qu'au-delà de 150 items, l'indice de cohérence interne ne varie pas de façon marquée. En considérant qu'une mesure de consistance interne devrait de préférence être supérieure à 0,80, nous pouvons en déduire que la BATEDA a une bonne cohérence interne (Wasserman, Bracken, 2003). Cette fiabilité interne des tests est également confortée par la mesure de l'erreur standard de la moyenne (ou ESM) de chaque test (en anglais *standard error of measurement* ou *SEM*) qui est inférieure au tiers de chaque écart-type. Généralement, un taux d'erreur supérieur à cette fraction de l'écart-type constitue une limite de la fiabilité du test (McCauley, 2001).

Pour ce qui est de l'évaluation de la fiabilité inter-juge effectuée dans cette étude, il est important de rappeler que l'outil est destiné à l'usage de personnes n'ayant pas nécessairement une formation professionnelle en psychométrie du langage. Cette mesure a permis de confirmer l'objectivité de la procédure d'administration de la batterie avec un coefficient de 0,911. Nous pouvons donc retenir que la BATEDA a une fiabilité inter-juge significative à $p < 0,05$ (Wasserman, Bracken, 2003). Cette fiabilité se justifie par le fait que les examinateurs étaient contraints par la présentation audio de tous les tests (tests des modalités et du langage). En effet, pour tous les tests, les tâches et les items étaient enregistrés sur un même CD et l'examineur ne pouvait intervenir ni dans l'ordre de passation ni dans la durée d'administration des tests. Il devait être attentif, écouter et suivre le test en même temps que l'enfant, s'abstenir de parler, de presser ou d'encourager l'enfant de quelque façon que ce soit, arrêter immédiatement le test si l'enfant exprimait un malaise ou ne souhaitait plus continuer. Par ailleurs, l'indice de fiabilité obtenu implique aussi que les juges ont coté les réponses de façon comparable et donc que les instructions à cet égard étaient claires. Pendant l'évaluation, l'examineur se servait de la *Fiche de réponses* uniquement pour relever les répliques de l'enfant soit en cochant une case soit en écrivant un court énoncé (un son ou un mot). Il n'y avait aucune référence aux réponses attendues et la cotation des réponses ne se faisait qu'à la fin de l'administration de tous les tests à l'aide de la *Grille d'évaluation* conçue à cet effet.

2. Normalisation

Nos analyses de normalisation ont permis de standardiser les scores de chaque test. Chaque score brut correspond donc à un score standard qui permet de déterminer si le score d'un enfant est supérieur ou égal au seuil inférieur de réussite du test. Des neuf tests normalisés, deux tests ont présenté des distributions asymétriques, soient le *Test de l'articulation des sons* (TAS) et le *Test de reconnaissance de sons* (TRS). Ces asymétries, qui ont résulté du plafonnement des scores supérieurs à la moyenne, s'expliquent par le fait qu'autour de quatre ans, l'enfant n'ayant aucun déficit pouvant affecter son développement maîtrise déjà la majorité des sons de sa langue (Kent, 1994 ; Shipley, McAfee, 2009). Ceci ne constitue donc pas une entorse à la normalisation des scores obtenus.

3. Validité prédictive

Les coefficients de corrélation obtenus permettent de situer la validité prédictive des tests de langage construits à environ 73%, ce qui est largement supérieur au seuil suggéré par la littérature qui se situerait à 30% (Beaucage, Bonnier Viger, 1996). Ces résultats permettent donc d'affirmer que la batterie de tests langagiers a une bonne validité prédictive (Glascoe et al., 1992). De plus, les résultats obtenus corroborent ceux de la littérature qui indiquent des liens pertinents entre les composantes langagières testées à l'oral et les performances en langage écrit (Gerrits, De Bree, 2009 ; Plaza, 2003 ; Plaza, Cohen, 2004 ; Wood et al., 2005). Rappelons que, dans cette étude, les performances en français écrit étaient reflétées par quatre matières portant sur la connaissance de l'orthographe, la grammaire et la compréhension de textes en français (*Dictée, Questions, Rédaction et Lecture*). De plus, nos résultats vont au-delà en montrant que les connaissances langagières peuvent aussi permettre de prédire les performances en calculs mathématiques. Cette corrélation s'explique par le fait que l'exécution des exercices mathématiques nécessite la compréhension de consignes qui les accompagnent. De ce fait, l'enfant doit recourir à des unités lexicales spécifiant le type d'activité ou d'opération à effectuer. Ainsi par exemple, lorsqu'on demande à un enfant de tracer des figures planes, non seulement il doit connaître les noms des formes en question, mais il doit aussi être en mesure de saisir les notions abstraites spécifiant la nature du travail avant de pouvoir accomplir la tâche correspondante. En somme, l'analyse des données de validation prédictive de la BATEDA suggère que les connaissances langagières contribuent positivement à la réussite aux épreuves de français écrit et de mathématiques chez les enfants du primaire.

4. Retombées de l'étude sur le plan clinique

La démarche d'élaboration et de normalisation de la BATEDA visait à pallier un manque de ressource chronique dans un contexte particulier à l'Afrique sub-saharienne. Les qualités psychométriques de l'outil étant confirmées, la batterie pourra être utilisée dans ce contexte pour identifier les enfants qui pourraient bénéficier d'une intervention visant à améliorer leurs habiletés langagières. De toute évidence, des interventions adaptées au contexte de ce milieu devront être élaborées et mises à l'essai. Par ailleurs, étant donné que la BATEDA teste les habiletés en français d'enfants de 4 et 5 ans évoluant dans un environnement multilingue et une culture camerounaise, il pourrait s'avérer un outil intéressant pour les pays africains comparables au Cameroun, que ce soit les pays frontaliers ou ceux de la sous-région de l'Afrique de l'ouest francophone. De plus, les orthophonistes travaillant avec des populations

similaires pourraient en faire usage, soit dans d'autres contextes africains, soit en occident auprès d'enfants nouvellement arrivés en Europe ou en Amérique, par exemple. Toutefois, il est important de noter qu'une telle application ne pourrait se faire qu'à condition que l'enfant ait évolué dès l'âge de deux ans au plus dans un environnement urbain francophone comparable à celui des villes camerounaises. Cette restriction est d'autant plus importante que suivant les étapes du développement psycholinguistique, on sait que le développement du langage est très marqué chez l'enfant entre 18 et 24 mois avec l'apparition de combinaisons lexicales et l'expansion du vocabulaire. De plus, il doit être âgé de quatre ans au moins et de cinq ans onze mois au plus, et démontrer une acuité visuelle et auditive normale après avoir passé les tests visuel et auditif.

----- CONCLUSION -----

A l'issue des analyses, il ressort que la batterie de tests construits présente les principales caractéristiques psychométriques des outils standards, à savoir, la normalisation des scores, la fiabilité et la validité des tests. Le *Test de l'articulation des sons* (TAS) et le *Test de reconnaissance de sons* (TRS) ont présenté des distributions asymétriques résultant du plafonnement des scores supérieurs à la moyenne. Toutefois, cette situation prévisible compte tenu de l'âge de l'échantillon d'enfants ne constitue pas un obstacle pour la normalisation des tests. En effet, autour de 4 ans, un enfant n'ayant aucun déficit pouvant affecter son développement maîtrise déjà la majorité des sons de sa langue. Par ailleurs, avec un indice de 0,911, les mesures de fiabilité inter-juge effectuées ont permis de vérifier l'objectivité de la procédure d'administration et de cotation des tests. La nécessité de ces mesures s'est justifiée par le fait que l'outil est destiné à l'usage de personnes n'ayant pas nécessairement une formation professionnelle en psychométrie du langage. Par ailleurs, les tests ont présenté une excellente cohérence interne ($\alpha = 0,958$), ce qui permet de conclure que les items évaluent correctement les compétences cibles. Cependant, dans cette étude, nous n'avons pas estimé la stabilité du test dans le temps qui se mesure par la fiabilité test-retest. Toutefois, compte tenu du caractère dynamique de tout test normalisé, cette mesure sera effectuée dans nos études ultérieures. Il en est de même pour ce qui est de l'évaluation de la sensibilité et de la spécificité de la BATEDA. En ce qui concerne la validation prédictive, l'étude s'est particulièrement intéressée aux composantes langagières en relation avec les tâches scolaires. Les résultats des analyses de corrélation obtenus corroborent ceux de la littérature qui indiquent des liens significatifs entre les composantes langagières testées et les performances en langage écrit. Les coefficients de corrélation obtenus situent cette validité à environ 73% ($r = 0,728$), ce qui confère à la batterie de tests langagiers construits une bonne validité prédictive. Enfin, la BATEDA est un outil standard permettant de pallier le manque de ressources adaptées au contexte camerounais. De plus, elle pourrait s'appliquer aux enfants des milieux urbains francophones comparables au contexte des villes camerounaises.

Remerciements

Les travaux de recherche effectués dans cette étude ont été réalisés grâce à une subvention du Centre de recherche pour le développement international (CRDI), Ottawa, Canada. Lucia Nzenge et Lucie Ngayambena ont participé à la collecte des données.

----- BIBLIOGRAPHIE -----

Administration for Children and Families (2011). *Resources for measuring services and outcomes in Head Start Programs serving infants and toddlers*. Washington, DC: Department of Health and Human Services. Consulté le 20.10.2014 de Administration for children and families: http://www.acf.hhs.gov/sites/default/files/opre/resources_for_measuring_services_and_outcomes.pdf

American Psychiatric Association (2003). *DSM-IV Text Revision. Manuel diagnostique et statistique des troubles mentaux*. Issy-les-Moulineaux : Masson.

Anastasi, A., Urbina, S. (1997). *Psychological testing (7th ed.)*. Upper Saddle River, NJ: Prentice Hall.

Aylward, G.P., Stancin, T. (2008). Screening and assessment tools: measurement and psychometric considerations. In M.L. Wolraich, D.D. Drotar, P.H. Dworkin, E.C. Perrin (Eds.), *Developmental-Behavioral Pediatrics: Evidence-Based Practice* (pp. 123-130). Philadelphie, PA: Mosby Elsevier.

Beaucage, C., Bonnier Viger, Y. (1996). *Épidémiologie appliquée. Une initiation à la lecture critique de la littérature en sciences de la santé*. Boucherville : Gaëtan Morin éditeur.

Dunn, L., Dunn, L.M. (1981). *Peabody Picture Vocabulary Test - Revised (PPVT-R)*. Minesota: American Guidance Services.

Gardner, M.F. (1990). *Expressive One-Word Picture Vocabulary Test (revised)*. Novato, CA: Academic Therapy Publications.

Gerrits, E., de Bree, E. (2009). Early language development of children at familial risk of dyslexia: speech perception and production. *Journal of Communication Disorders*, 42(3), 180-194. DOI: 10.1016/j.jcomdis.2008.10.004

Glascoc, F.P., Byrne, K.E., Ashford, L.G., Johnson, K.L., Chang, B., Stickland, B. (1992). Accuracy of the Denver-II in developmental screening. *Pediatrics*, 89(6), 1221-1225.

Kent, R.D. (1994). *Reference manual for communicative sciences and disorders: speech and language*. Austin, TX: Pro-Ed.

McCauley, R.J. (2001). *Assessment of Language Disorders in Children*. Mahwah, NJ, London: Lawrence Erlbaum Associates.

Ministère de l'Éducation Nationale (1998). *Programme officiel de l'enseignement primaire : Niveau I*. Yaoundé, Cameroun: Ministère de l'Éducation Nationale.

Newcomer, P., Hammill, D.D. (1982). *Test Of Language Development (TOLD)*. Austin, TX: Pro-Ed.

Plaza, M. (2003). The role of naming speed, phonological processing and morphological/syntactic skill in the reading and spelling performance of second-grade

children. *Current Psychology Letters [online]*, 1(10). Consulté le 20.10.2014 de CPL : <http://cpl.revues.org/index88.html>

Plaza, M., Cohen, H. (2004). Predictive influence of phonological processing, morphological/syntactic skill, and naming speed on spelling performance. *Brain and Cognition*, 55(2), 368-373. DOI: 10.1016/j.bandc.2004.02.076

Shipley, K.G., McAfee, J.G. (2009). *Assessment in speech-language pathology: A resource manual* (4th Edition.). New York: Delmare, Cengage Learning.

Takam, A. (2006). *Troubles de l'articulation (de la parole) chez l'enfant Camerounais de trois à huit ans parlant le ghomala' et le français*. Thèse de doctorat en linguistique appliquée. Yaoundé : Université de Yaoundé 1.

Takam, A. (2010). *Vers un outil d'évaluation du langage préscolaire et détecteur de prérequis pour l'éducation de base au Cameroun en cas de retard de langage*. Thèse de doctorat en linguistique. Montréal : Université de Montréal, faculté des arts et sciences. Consulté le 20.10.2014 de Papyrus, université de Montréal: https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/4851/Takam_Taguemne_Aurelie_2010.pdf?sequence=6

Takam, A., Boucher, V., Trudeau, N. (2013). Évaluation langagière en contexte subsaharien : Construction et pré-tests de la BATEDA pour enfants francophones camerounais d'âge préscolaire. *African Journal of Applied Linguistics*, 7, 123-138.

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: John Wiley & Sons, Inc.

Wasserman, J.D., Bracken, B.A. (2003). Psychometric characteristics of assessment procedures. In J.R. Graham, J.A. Naglieri (Eds.), *Handbook of psychology: assessment psychology* (Vol. 10, pp. 43-66). Hoboken, NJ: John Wiley & Sons, Inc.

Wood, F.B., Hill, D.F., Meyer, M.S., Flowers, D.L. (2005). Predictive assessment of reading. *Annals of Dyslexia*, 55(2), 193-216. DOI: 10.1007/s11881-005-0011-x