

# L'évaluation critériée statique formelle en orthophonie

Julie CATTINI \*, Guillaume DUBOISDINDIEN \*\*

\* Orthophoniste, Lëtzebuerg, Luxembourg

\*\* Orthophoniste, France

\*\* Chercheur, Ph.D, Université de Lille, UMR 8163 – STL – Savoirs Textes Langage, CNRS, Lille, France. <https://orcid.org/0000-0002-7068-9881>

## **Auteurs de correspondance :**

Julie Cattini, courriel : [juliecattini@hotmail.com](mailto:juliecattini@hotmail.com)

Guillaume Duboisdindien, courriel : [duboisdindien@hotmail.com](mailto:duboisdindien@hotmail.com)

**Note des auteurs :** Les auteurs ont contribué, de façon concertée, décisive et équitable à la création de ce manuscrit. Il s'agit d'une œuvre de collaboration indivisible, dans laquelle la contribution de chaque auteur ne peut être isolée. Les auteurs déclarent n'avoir aucun conflit d'intérêts, financiers ou autres.

## Résumé :

Les outils d'évaluation normés statiques sont plébiscités par les orthophonistes pour poser un diagnostic et déterminer le profil atypique ou pathologique d'un patient. Les outils d'évaluation critériée sont, quant à eux, de bons candidats pour compléter les résultats de première ligne issus des approches évaluatives normées et pour évaluer les patients ne pouvant être comparés à une population dite normocognitive. A travers un état de l'art, cet article défend l'importance d'une approche d'évaluation critériée en orthophonie.

Tout d'abord, une définition générale de l'évaluation critériée est proposée sur la base d'une taxonomie des différentes évaluations possibles pour les compétences langagières. Ce premier volet est ensuite illustré par plusieurs outils d'évaluation critériée employés en orthophonie et en orthopédagogie afin de dégager une liste de huit caractéristiques méthodologiques. Enfin, nous proposons de délimiter les principes méthodologiques et les enjeux cliniques ou pédagogiques de cette évaluation encore mal renseignée dans le domaine de l'orthophonie.

Les conclusions et perspectives proposées ouvrent la voie pour permettre aux cliniciens d'élaborer un programme de rééducation ou encore observer l'évolution d'un patient par rapport à lui-même dans une approche fonctionnelle et systémique.

**Mots-clefs :** évaluation critériée, orthophonie, communication, apprentissage

## Static criterion-referenced assessment in Speech and Language Pathology

### Summary:

Static norm-referenced assessments have the speech and language pathologists' preference to establish a diagnosis and determine a patient's unusual or pathological profile. However, criterion-referenced assessment tools are relevant to complement the results of standardized evaluative approaches performed in the first instance and to assess patients who cannot be compared to an average population. This article argues for the importance of a criterion-referenced assessment approach in speech-language pathology through state of art.

First, a general definition is suggested further to recent studies focusing on assessments and their taxonomies. This first section is then illustrated by several criterion-referenced assessment tools used in speech-language pathology and remedial education to draw up a list of eight methodological characteristics. Finally, we propose to define this assessment's methodological principles and clinical or pedagogical issues, which still need to be better understood in speech therapy. The proposed conclusions and perspectives allow clinicians to develop a rehabilitation program or observe the evolution of a patient to himself in a functional and systemic approach.

**Keywords:** criterion-referenced assessment, speech language pathology, communication, learning

## -----INTRODUCTION-----

Depuis plusieurs années, la recherche sur l'évaluation clinique en orthophonie a diversifié ses objets d'étude. D'une part, elle tend à proposer une meilleure compréhension psychométrique et statistique des données collectées à l'aide d'évaluations normées traditionnelles (Lafay & Cattini, 2018 ; Leclercq & Veys, 2014) et des interprétations qu'il sera possible d'en extraire (McCauley, 1996). D'autre part, elle dégage des stratégies alternatives d'évaluation afin d'affiner le tableau clinique du patient (Dockrell & Marshall, 2015 ; Hasson & Joffe, 2007 ; Laing & Kamhi, 2003). De telles approches ont une incidence significative sur l'action clinique que l'orthophoniste pourra mobiliser (Laveault & Grégoire, 2014 ; Mottier Lopez & Laveault, 2014). Toutefois, il est notoire qu'en clinique les tests normés sont devenus une composante omniprésente des protocoles d'évaluation (Kerr et al., 2003). Pour rappel, les tests normés sont préconisés pour définir la présence ou l'absence d'un trouble. Ils n'ont pas pour vocation de fournir un pronostic, de développer des objectifs thérapeutiques et de vérifier l'efficacité de la thérapie proposée (Lefebvre & Trudeau, 2005 ; McCauley & Swisher, 1984 ; Piérart, 2006). Ce dernier point n'est pas une vision partagée par la majorité des cliniciens (Huang et al., 1997 ; Kerr et al., 2003 ; Rondal & Seron, 2000). La prise de décision clinique étant fondée sur l'interprétation des mesures, Kerr et al. (2003) avertissent que le manque de connaissances peut entraîner des erreurs de jugement et affaiblir la crédibilité du professionnel de santé.

Dans cet article, nous portons une attention particulière à l'évaluation dite *critériée*. Au cours de notre projet de recherche visant à proposer des outils méthodologiques afin de construire des mesures critériées en clinique, il s'est rapidement avéré que, les termes se rapportant à l'évaluation critériée changent selon les différents auteurs et les disciplines qui s'y sont consacrés. McCauley (1996) recense les termes anglophones suivants bien qu'aucune distinction définitoire ne soit réellement statuée : *criterion-referenced assessment*, *domain-referenced testing*, *objectives-referenced testing*, *competency-based testing* ou *mastery testing*. Par ailleurs, la définition, varie de fait, selon la discipline, le propos des auteurs, le type d'épreuves proposées ou encore la compétence visée.

En 1994, dans le domaine des Sciences de l'éducation, Grégoire évoque l'évaluation critériée de cette manière : « *Suivant celle-ci, les performances ne sont plus évaluées en référence à celles des autres enfants mais en référence aux objectifs hiérarchisés du curriculum d'apprentissage.* »

En 2019, dans le domaine des Sciences du langage, Fabiano-Smith, étudiant les outils d'évaluation des troubles des sons de la parole chez l'enfant, évoque la mesure critériée en ces termes : « *L'évaluation critériée s'appuie sur des critères publiés dans des études sur le développement phonologique de l'enfant à des fins de comparaison.* »

Notre recherche dans la littérature, des années 70 à aujourd'hui, a fourni des définitions diverses de l'évaluation critériée alors que, comme le spécifie Grégoire (1994), « (...) *construire des tests critériés est beaucoup plus complexe que de construire des tests normés.* »

Notre objectif sera donc de proposer un état de l'art de l'évaluation critériée afin de pouvoir dégager les bases définitoires et les caractéristiques de ce type d'épreuves tant pour les concepteurs de test que pour le clinicien dans sa pratique quotidienne. Si les données de la littérature appuient la nécessité de changer les pratiques évaluatives, il est primordial de définir les autres approches disponibles et de démontrer leur valeur ajoutée pour le patient.

## APPROCHE DEFINITOIRE DE L'ÉVALUATION CRITERIEE

En 2019, à partir d'une étude Delphi<sup>1</sup>, Denman et al. ont proposé une première taxonomie des moyens d'évaluation langagière chez l'enfant d'âge scolaire sur la base de la littérature contemporaine. Selon ces auteurs, la forme de l'évaluation, thème central de cet article, peut se décomposer en quatre catégories de caractérisation (Figure 1) : 1. standardisé vs non standardisé ; 2. normée vs critériée vs descriptive ; 3. statique<sup>2</sup> vs dynamique<sup>3</sup> ; 4. décontextualisée vs contextualisée vs ciblée.

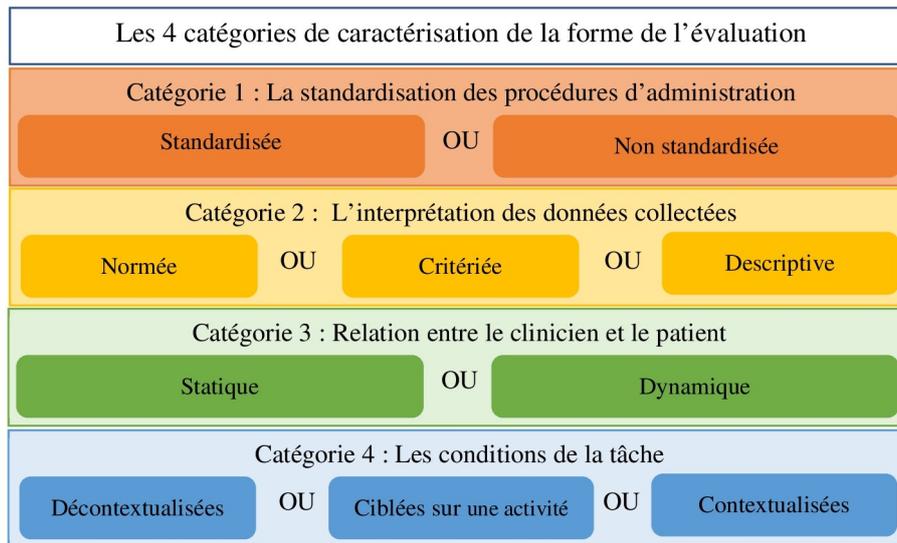


Figure 1. Les 4 catégories de caractérisation de la forme de l'évaluation d'après Denman et al. (2019).

L'interprétation des données collectées d'une épreuve évaluative peut être de trois natures : normée, critériée ou descriptive. L'évaluation critériée est définie par ces auteurs comme « une comparaison des performances d'un enfant par rapport à un niveau ou à un critère prédéterminé (c'est-à-dire les compétences attendues en fonction de l'âge de l'enfant, de son niveau scolaire ou du programme d'études) ». Afin de compléter cette définition, les auteurs utilisent une approche contrastive en opposant l'évaluation normée (i.e. évaluation qui compare quantitativement les résultats d'un patient à ceux d'un échantillon de pairs appariés qui ont accompli la même tâche) et l'évaluation descriptive (i.e. évaluation conçue pour fournir des données descriptives ou qualitatives sur les capacités du patient) à l'évaluation critériée. Cette

<sup>1</sup> La méthode Delphi vise à recueillir par l'intermédiaire de questionnaires l'avis justifié d'un panel d'experts issus de différents domaines en vue d'établir un consensus scientifique sur un thème : le premier avis est ouvert (en phase 1), le ou les suivants sont fermés (en phase 2 et plus). Dans cette procédure rétroactive, la confrontation des experts est évitée et leur anonymat préservé. Les résultats sont ensuite discutés et justifiés phase après phase jusqu'à l'obtention d'une convergence aussi forte que possible des réponses (Ekionea et al., 2011 ; Okoli & Pawlowski, 2004 ; Paliwoda, 1983).

<sup>2</sup> L'évaluation comprend des procédures conçues pour mesurer les performances à un moment donné dans des conditions fixes (i.e. sans faire varier le contexte et en ne fournissant aucun étayage au patient). Elle fournit une indication des connaissances actuelles, sans fournir d'informations sur le potentiel d'apprentissage.

<sup>3</sup> L'évaluation comprend des procédures conçues pour évaluer les performances d'un enfant dans des contextes variés et/ou avec des aides thérapeutiques différentes. Celles-ci permettent de décrire le potentiel d'apprentissage et d'identifier les aides et les techniques d'enseignement les plus appropriées. La relation interactive entre le patient et l'évaluateur est un facteur primordial de cette forme d'évaluation.

approche par opposition défendue de manière récurrente par de nombreux auteurs (Brown & Hudson, 2002 ; Burton, 2006 ; Klecan-Aker & Colson, 2009 ; McCauley, 1996) est insatisfaisante voire réductrice car elle peut éclipser les constituants communs de ces trois modes d'interprétation si les autres caractéristiques de la forme de l'évaluation ne sont pas spécifiées.

Pour éclaircir ce point, prenons l'exemple d'une évaluation du discours narratif chez l'enfant d'âge scolaire. Premièrement, l'évaluation pourrait être réalisée en utilisant l'épreuve de récit provenant du test français Exalang 5-8 (Thibault et al., 2010) pour situer les résultats du patient par rapport à la norme de référence. Cette épreuve consiste pour l'essentiel à présenter une séquence animée que l'enfant devra restituer oralement après le visionnage. Une fois le corpus transcrit, l'analyse se fait sur base de critères généraux et subjectifs : 4 points sont accordés pour la cohérence (niveau macrostructurel) et 7 points sont attribués pour la cohésion (niveau microstructurel). Les scores sont ensuite interprétés au regard des mesures de tendance centrale (i.e. interprétation normée). Lors d'une interprétation critériée, le même support d'évaluation pourrait être exploité afin d'analyser la macrostructure et la microstructure du discours de l'enfant dans son récit par l'intermédiaire du même corpus oral transcrit. L'interprétation des résultats se confronterait alors aux compétences attendues en fonction de l'âge de l'enfant selon les données disponibles dans la littérature telles que les stades développementaux proposés par Applebee (1978) (voir Schelstraete, 2011). Nous convenons de fait que l'épreuve est identique dans ses procédures d'administration (i.e. standardisée), la relation entre le clinicien et le patient (i.e. statique) et les conditions de la tâche (i.e. contextualisée) ; néanmoins, les données recueillies et les interprétations réalisées sont de natures diverses (normée vs critériée) et permettent, ainsi, d'atteindre des objectifs différents.

Notons qu'une épreuve peut donc être critériée tout en variant sur les autres niveaux de caractérisation. Ainsi, elle peut être standardisée ou non, elle peut aussi être statique ou dynamique, etc. (voir Figure 2).

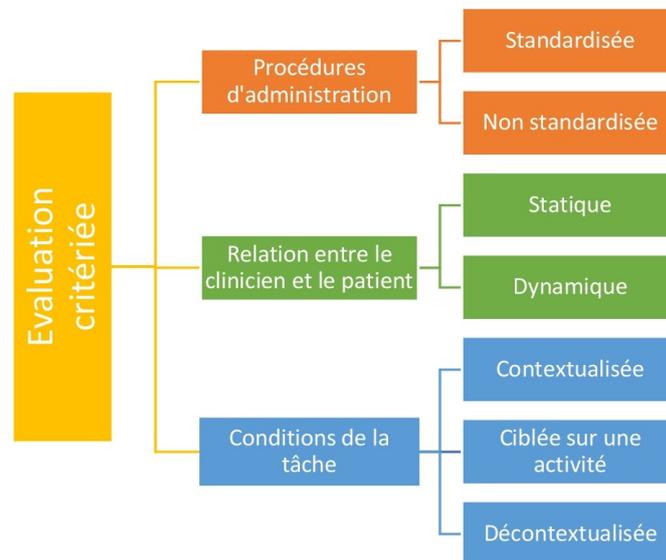


Figure 2. Les aspects formels de l'évaluation critériée sur base de la taxonomie de Denman et al. (2019)

Dans la littérature du domaine, le terme « évaluation critériée » regroupe des concepts très variés qu'il est nécessaire de définir avec soin. Etant donné que les évaluations dynamiques (normalisées, critériées ou descriptives) sont une forme d'évaluation requérant une

méthodologie différente, nous axerons sur les évaluations critériées statiques formelles<sup>4</sup>. Nous optons pour une approche par l'exemplification en présentant des outils d'évaluation utilisables en orthophonie afin de tirer des recommandations générales et plurifonctionnelles.

## ILLUSTRATIONS D'ÉVALUATIONS CRITÉRIÉES

Nous répertorions dans cette section 4 outils appartenant à différents domaines de l'orthophonie pouvant être catégorisées comme des évaluations critériées. Effectivement, les auteurs de ces tests proposent une ou des épreuves d'évaluation dont les performances du sujet sont comparées à un niveau ou à un critère prédéterminé.

Afin de tenter de dégager des caractéristiques des évaluations critériées, nous analysons chaque outil en caractérisant : a) la population visée, b) les objectifs d'évaluation, c) la validité théorique, d) la méthodologie, e) la standardisation et f) les éventuelles qualités psychométriques documentées.

### 1. En sciences psychologiques – Evaluation des jalons du Comportement verbal et Programmes d'Intervention

Dans le souhait d'une intervention auprès d'un enfant présentant un Trouble du Spectre Autistique (TSA), une évaluation complète permettant d'évaluer les compétences verbales et non verbales afin d'établir un projet thérapeutique précis est primordiale (American Speech-Language-Hearing Association, 2020). L'Évaluation des jalons du Comportement verbal et Programmes d'Intervention (ou *Verbal Behavior Milestones Assessment and Placement Program* en anglais – VB-mapp) est un outil d'évaluation critériée, élaboré aux Etats-Unis par Sundberg en 2008, qui souhaite répondre à cette demande. Depuis 2017, la forme révisée de 2014 a été traduite et elle est disponible en français (Sundberg, 2017).

#### a. Population visée

Jeunes enfants présentant un TSA ou un autre trouble du développement.

#### b. Objectifs

La fonction de l'outil est de fournir un programme d'évaluation du langage appuyé sur les comportements communicationnels. Il permettra ainsi de proposer une intervention adaptée selon le niveau de développement de l'enfant et les obstacles présents. Il peut également être utilisé pour observer l'évolution du patient.

---

<sup>4</sup> Une évaluation formelle est une évaluation élaborée par un groupe d'experts et qui se base sur des données normées ou critériées. La standardisation de l'outil, quant à lui, peut varier. C'est pour cette raison que nous distinguons « formel » de « standardisé ».

### c. Validité théorique

La VB-mapp rassemble les procédures et la méthodologie d'enseignement de l'*Applied Behavioral Analysis* avec l'analyse des comportements verbaux de Skinner. Toutefois, la méthode de sélection et d'analyse des données externes n'est pas explicitée. A titre d'exemple, les jalons (i.e. repères développementaux) ont été conçus par « *une moyenne issue de cinquante profils de développement obtenus à partir d'une grande variété d'études* » et « *une variété d'ouvrages servant de guides sur le développement de l'enfant a été consultée* ». (Sundberg, 2017, p.17)

### d. Méthodologie

La VB-mapp se décompose en cinq éléments : 1) l'évaluation des jalons qui fournit un échantillon des compétences de langage et des compétences associées de 0 à 48 mois ; 2) l'évaluation des obstacles qui comprend 24 obstacles rencontrés par les enfants présentant un TSA ; 3) l'évaluation de l'orientation afin de guider les choix du milieu éducatif ; 4) l'analyse de tâches et compétences du soutien du VB-mapp qui est une évaluation plus spécifique des jalons et 5) le guide de développement du curriculum et de rédaction des objectifs du Projet Educatif Individualisé qui fournit des suggestions pour sélectionner les objectifs.

### e. Passation et cotation

Les conditions de passation sont laissées au libre choix de l'évaluateur : le manuel ne précise pas de consignes, ne fournit pas un matériel propre à l'évaluation et n'indique pas d'information sur la temporalité ni sur la séquentialité à appliquer. En revanche, la cotation est précisée avec une notation pour chaque compétence en trois niveaux : acquis (1 point) ; en émergence (0,5 point) ou déficitaire (0 point).

### f. Qualités psychométriques

Deux études ont évalué la validité et la fidélité de la VB-mapp. Au niveau de la validité concomitante, Dixon *et al.*, (2015) ont montré une forte corrélation entre le VB-mapp et le *Promoting the Emergence of Advanced Knowledge Relational Training System* (PEAK) également utilisé pour évaluer les répertoires verbaux des TSA. Plus récemment, Montallana *et al.* (2019) ont évalué la fidélité inter-juges de la VB-mapp. Ils montrent une fidélité satisfaisante pour l'évaluation des jalons, mais une forte variation selon les domaines, et une faible fidélité en ce qui concerne l'évaluation des obstacles.

## 2. En sciences de l'Éducation - Echelle d'évaluation de la production écrite

La production écrite est essentielle à la réussite scolaire, socio-professionnelle et à de bonnes compétences communicationnelles. Cependant, l'incapacité à produire un langage soutenu, précis et adapté à l'écrit reste une faiblesse omniprésente pour de nombreux élèves. L'évaluation initiale proposée doit permettre de dresser un profil des compétences développementales et de définir des mesures d'actions appropriées (Dockrell *et al.*, 2014). Dans ce cadre, en Angleterre, Dunsmuir *et al.* (2015) ont développé une grille d'analyse critériée de production écrite : la *Writing Assessment Measure* (WAM).

**a. Population visée**

Enfants du CE2 à la 6<sup>ème</sup>.

**b. Objectifs**

L'objectif principal est d'identifier les besoins spécifiques des élèves par rapport à leur développement en écriture et contribuer à la planification de l'enseignement ultérieur de façon ciblée.

**c. Validité théorique**

Les éléments objectifs du National Curriculum (2010) établi au Royaume-Uni ont été utilisés pour définir les critères de performance en matière d'expression écrite. Cela permet de s'assurer de la pertinence des données au regard du contexte d'apprentissage et des socles de compétences à atteindre des élèves ainsi que de mieux éclairer la pratique et l'apprentissage.

**d. Méthodologie**

La WAM est basée sur la structure et le format de la *Wechsler Objective Language Dimensions* (WOLD – Weschler, 1996) et des descripteurs du National Curriculum (2010). Elle se décompose en sept domaines : graphisme, orthographe, ponctuation, structure de phrase et grammaire, vocabulaire, macrostructure et idées. Quatre énoncés descriptifs avec des critères de compétence clairement définis et sans ambiguïté sont proposés dans chaque domaine.

**e. Passation et cotation**

La tâche consiste en une rédaction de textes narratifs en 15 minutes à partir d'une ébauche écrite. Ensuite, les productions sont notées dans 7 domaines principaux d'expression écrite sur une échelle de 4 points ; le niveau 4 devant être atteint par la majorité des élèves de 6<sup>ème</sup>.

**f. Qualités psychométriques**

Dans leur article, Dunsmuir *et al.* (2015) ont évalué la validité de construit (âge), la validité concomitante (comparaison avec le subtest *Written Expression* de la *Wechsler Objective Reading Dimensions*), la cohérence interne, la fidélité interjuges et la fidélité temporelle. Sur la base des résultats de cette étude, la WMA a démontré qu'elle est valide, stable dans le temps et peut être notée de manière fidèle par différents évaluateurs. Notons toutefois que la validité concomitante a été évaluée avec le test ayant servi de base à la création de la WAM, une bonne corrélation n'est donc pas surprenante.

**g. Commentaires**

Puisqu'elle a été élaborée en référence aux objectifs de réussite du National Curriculum et qu'elle utilise un système de notation analytique, la WAM peut être considérée comme utile pour identifier des objectifs spécifiques pertinents pour les enfants qui éprouvent des difficultés d'écriture.

**3. En Sciences du Langage et en linguistique clinique - Inventaire des structures syllabiques chez l'enfant francophone**

La dyspraxie verbale est un trouble des sons de la parole dont les caractéristiques diagnostiques sont difficiles à identifier. C'est dans cette perspective que l'Inventaire des Structures Syllabiques chez l'Enfant Francophone (ISSEF – Aubry et al., 2019) a été conçue au Québec par trois orthophonistes et diffusée en accès libre.

#### **a. Population visée**

Enfants présentant une (suspicion de) dyspraxie verbale.

#### **b. Objectifs**

Selon les auteures, l'objectif de cet outil est d'évaluer spécifiquement et de manière exhaustive la production des différentes structures syllabiques en français du Québec, par le biais d'une tâche de dénomination. Il permet également de faciliter l'identification d'objectifs pour le plan de traitement.

#### **c. Validité théorique**

Sa conception théorique s'appuie sur les liens entre les indicateurs diagnostiques de la dyspraxie verbale référencés dans la littérature et la structure syllabique : a) simplification des structures syllabiques, b) augmentation des erreurs avec l'allongement des séquences de sons, c) erreurs de type complexification, d) tâtonnement et maintien de la structure syllabique avec des transformations de phonèmes. Toutefois, ces indicateurs conviennent aux enfants anglophones alors que le système phonologique du français est différent de celui de l'anglais en ce qui concerne les aspects suprasegmentaux (Brosseau-Lapr e et al., 2018).

#### **d. M ethodologie**

La s election des 197 items a  et e r ealis ee selon quatre param etres : 1) la repr esentativit e des diff erentes structures syllabiques, 2) la repr esentativit e des diff erentes s equences de sons possibles en franco-qu eb ecois, 3) le caract ere imageable des mots cibl es et 4) le niveau de vocabulaire attendu chez des enfants de 3  a 6 ans. La t ache consiste en une d enomination d'images avec une hi erarchisation d'indices si l'enfant m econna it le mot. L'analyse des productions de l'enfant peut  etre r ealis ee  a plusieurs niveaux : a) patrons de transformation phonologique, b) analyse par structures syllabiques, c) analyse descriptive de la production des mots polysyllabiques, d) inventaire phon etique, e) inventaire des structures syllabiques et f) inventaire des patrons d'erreur produits par l'enfant.

#### **e. Qualit es psychom etriques**

La validation du contenu a  et e r ealis ee en deux volets : 1) aupr es de cinq orthophonistes avec des enfants  ag es de 3  a 6 ans ; 2) aupr es de 74 enfants pour contr oler les capacit es de d enomination sur les items choisis et noter les indices fournis. Selon les informations obtenues aupr es des auteures, un article est actuellement sous presse concernant ce point. Concernant les proc edures d'administration, l'outil est standardis e au niveau des consignes de passation et de cotation. N eanmoins, il se veut flexible car les sections  eval uees sont au libre choix de l' evaluateur.

## f. Commentaires

Le manuel est exhaustif concernant l'analyse des productions de l'enfant. En revanche, il manque des informations concernant l'élaboration de l'outil. Aucune donnée n'est fournie concernant les bases de données utilisées pour respecter les critères de sélection des items cités tels que le niveau de vocabulaire des enfants de 3 à 6 ans ou encore l'imageabilité. Nous notons également un manque de clarté concernant les appuis théoriques et, tout particulièrement, la langue sur laquelle se fonde l'outil.

## 4. En Sciences du Langage et en linguistique clinique - Échelle de Communication Verbale de Bordeaux

Les tests psycholinguistiques normatifs traditionnels développés en orthophonie et en linguistique clinique fournissent habituellement des informations fondamentales sur les capacités cognitivo-linguistiques des individus chez qui est suspecté un trouble acquis du langage et de la communication. Cependant, la performance à ces tests ne permet pas de prédire aisément dans quelle mesure, et avec quelle efficacité, ces personnes communiquent dans la vie quotidienne. *L'Échelle de Communication Verbale de Bordeaux* (ECVB – Darrigrand & Mazaux, 2000) est une évaluation critériée qui s'emploie à caractériser ces phénomènes.

### a. Population visée

Personnes aphasiques adultes ou présentant un trouble majeur de la communication verbale.

### b. Objectifs

Cet outil permet d'évaluer la perception de la personne aphasique sur la fonctionnalité de sa communication et sa qualité de vie afin d'établir les objectifs de la rééducation orthophonique et d'en évaluer les progrès.

### c. Validité théorique

Cette échelle repose sur un inventaire de quelques échelles fonctionnelles de la communication telles que le *Communication Activities of Daily Living 2* (Holland et al., 1999) et le *Amsterdam–Nijmegen Everyday Language Test* (Blomert et al., 1994) avec toutefois une dimension écologique jusqu'alors négligée dans la plupart de ces évaluations qui privilégiaient des démarches par jeux de rôle et des entretiens dirigés.

### d. Méthodologie

Les aspects de la communication fonctionnelle sont regroupés sous les items suivants : expression des besoins et des intentions, conversation, utilisation du téléphone, communication dans les relations sociales, lecture, écriture, stratégies de compensation, désir et manière de communiquer, ainsi que le calcul mental et la gestion du budget.

### e. Passation et cotation

L'administration de l'ECVB est simple et brève, sa notation est rapide et, selon les auteurs, facile à interpréter pour les personnes aphasiques. Par ailleurs, les utilisateurs sont invités à proposer l'échelle à un intime qui partage la vie quotidienne de la personne afin de contourner

des difficultés de compréhension. Les réponses sont recueillies à l'aide d'une échelle de Likert à quatre niveaux (jamais, quelquefois, souvent, toujours).

#### f. Qualités psychométriques

L'ECVB a été validée auprès de 127 sujets adultes dont l'aphasie était d'origine vasculaire dans 93 % des cas, et le délai moyen entre la lésion vasculaire et l'entretien est de 4,5 ans. Selon les auteurs (Darrigrand & Mazaux, 2000), la *fidélité inter-juges* de cette échelle est considérée comme satisfaisante. Sa *validité de construit* (i.e. *validité factorielle* et sensibilité aux *facteurs démographiques*), sa *validité concomitante* (comparaison avec l'ASRS, l'ECVB et le TLC) et *prédictive*, et sa *cohérence interne* sont bonnes.

#### g. Commentaires

L'ECVB présente des limites importantes pour les personnes présentant un trouble de la compréhension. En effet, la formulation des questions notamment plus longue, peut nuire au traitement cognitif, et à leur intégration.

Comme nous avons pu le voir dans les exemples précédents, l'évaluation critériée a le potentiel d'évaluer le patient ou l'élève au plus proche de son niveau de performance, de ses capacités. La construction d'une évaluation permettant une interprétation optimale, contrôle les différents aspects de la tâche afin de permettre une prise d'information exhaustive (i.e. évaluation formative).

Seuls deux outils sur les quatre étudiés étaient définis comme des évaluations dites critériées. Pourtant l'élaboration d'une évaluation avec un mode d'interprétation critériée diffère de l'évaluation normée (Laveault & Grégoire, 2014). Cette absence de clarté questionne sur le cadre méthodologique engagé par les auteurs pour élaborer leur outil.

## LES CARACTERISTIQUES DES EVALUATIONS CRITERIEES STATIQUES FORMELLES

Sur la base des illustrations présentées dans la section précédente, nous sommes en mesure d'extraire et de proposer un certain nombre de caractéristiques méthodologiques communes pour constituer une **évaluation critériée**.

### 1. Le principe de validité théorique

La littérature internationale est une source importante de preuves scientifiques. Elle permet de baser la méthodologie de l'évaluation critériée, sur des modèles<sup>5</sup> et des données externes interdisciplinaires. Comme le précisent Laing et Kamhi (2003), la qualité des mesures collectées par une évaluation critériée est dépendante de la qualité des données empiriques (e.g.

---

<sup>5</sup> Un modèle est développé dans le prolongement d'une théorie, dont il est en fait une projection intellectuelle et consensuelle par un groupe de scientifiques en vue de répondre à un problème ouvert. Cependant son champ d'actions est basé sur des domaines d'application plus réduits. De fait, le modèle, comme outil de description, de représentation et d'analyse, constitue donc un moyen de découvrir de nouvelles relations, d'établir de nouveaux faits, d'énoncer de nouvelles hypothèses, de définir des actions d'intervention, de corriger certaines erreurs ou insuffisances par rapport à la perception et à la compréhension d'une partie du réel obtenue en évaluation normative, et enfin de prévoir, si cela est possible, le cours des événements (Willett, 1996).

repères développementaux) ou théoriques (e.g. modèles connexionnistes) sur lesquelles elles se fondent.

## 2. Le principe de spécification des objectifs

Afin de s'assurer de la validité de contenu du test, il est nécessaire de réaliser une définition opérationnelle du construit que le test souhaite mesurer par la référence à un modèle théorique, la construction d'un tableau de spécifications<sup>6</sup>, l'analyse de contenu d'entretiens et/ou l'observation directe de comportement (Laveault & Grégoire, 2014). Cette spécification des objectifs en complément de la validité théorique permet de s'assurer d'une bonne validité de contenu de l'outil.

## 3. Le principe de standardisation terminologique

L'évaluation critériée étant orientée sur des critères fixes dédiés à analyser une entité, il est préférable de s'engager à utiliser une terminologie reconnue et standardisée. Pour cela, il est nécessaire de se référer à un ensemble structuré de termes d'un domaine ou d'un champ d'activité spécialisé reconnu. En effet, il a l'avantage d'être accompagné le plus souvent de définitions sur ces termes et donc sur ce domaine ou champ d'activité.

## 4. Le principe de standardisation de la tâche

Il est nécessaire que l'évaluation critériée comporte des procédures permettant de définir et de contrôler les critères quantitatifs et qualitatifs de la tâche (i.e. aspects de présentation, condition de réalisation, niveaux d'exigence attendus) par le biais d'une standardisation de l'outil (Laveault & Grégoire, 2014). La standardisation permet de créer un environnement de passation toujours identique au sein de la tâche, de faire en sorte que les conditions de départ du test soient les mêmes pour tous et en tout temps. Elle permet également d'assurer la fidélité du test.

## 5. Le principe d'exhaustivité

L'évaluation critériée, contrairement à l'évaluation normée, concerne un domaine spécifique afin d'en couvrir le contenu (McCauley, 1996). Cela nécessite une sélection exhaustive des items ou des critères préétablis pour l'analyse du niveau de performance, lors de la réalisation d'une tâche.

## 6. Le principe de délimitation des besoins de l'individu

Le but fondamental de l'évaluation critériée est de comprendre la raison des difficultés rencontrées par le patient et de planifier l'intervention. En contexte de soin, selon Gordon et al. (2000), le diagnostic est le raisonnement clinique par lequel le praticien détermine l'affection dont souffre le patient, et qui permet de proposer une action clinique cohérente. Il repose sur la recherche des causes (i.e. étiologie) et des effets (i.e. symptômes) de l'affection afin de constituer un tableau clinique (Audétat et al., 2011 ; Elstein et al., 1978). Ensuite, des solutions thérapeutiques peuvent être proposées au patient et à son entourage. Afin de remplir ces

---

<sup>6</sup> Tableau servant à planifier la récolte d'informations dans le cadre d'une procédure d'évaluation. L'objectif est que l'instrument utilisé permette d'appréhender, dans des proportions adéquates, les différents aspects que l'on se propose d'investiguer ou d'étudier, et ceux-là seulement.

objectifs, il est donc primordial de ne pas seulement comparer le patient aux autres mais d'offrir une interprétation individualisée.

## 7. Le principe de consultation décisionnelle d'éthique du soin et du diagnostic

L'évaluation critériée tient une place avantageuse en offrant une évaluation réaliste des causes et des effets de la gêne rapportée en accord avec les valeurs du patient. Elle ne prend pas le parti de comparer un individu à un autre individu considéré comme « normo-cognitif » mais de le situer par rapport à une situation clinique donnée.

## 8. Le principe d'interprétation des données récoltées

L'évaluation critériée, par définition, permet d'interpréter les résultats du patient sur base de critères d'analyse ou de performance préétablis. Il s'agit d'un processus dépendant de la qualité des données externes fournies, du niveau d'expertise du clinicien (i.e. expertise scientifique, connaissance du contexte, raisonnement clinique et compétences interpersonnelles) et du niveau de granularité des données récoltées. L'analyse et l'interprétation des résultats constituent l'étape qui permet la comparaison quantitative ou qualitative des différentes solutions envisagées sur une base rationnelle. Il est donc essentiel que le clinicien se fonde sur une approche systématique et rigoureuse. Parmi les tâches accomplies par l'orthophoniste, cette étape est sans doute celle qui fait le plus appel à ses connaissances techniques et à sa rigueur scientifique. Rappelons que la validité des résultats d'un test est dépendante tant de la construction de l'outil que des interprétations réalisées par l'utilisateur (Laveault & Grégoire, 2014).

## ----- DISCUSSION -----

La limite entre ce que les performances d'un sujet donnent à voir en évaluation normative et ce que l'expertise du clinicien couplée aux besoins concrets du patient permettent de caractériser de manière empirique, est une limite fondamentale à laquelle des auteurs issus de différentes disciplines ont tenté de répondre par le prisme de l'évaluation critériée (Alderson & Wall, 1993 ; Brown & Hudson, 2002).

Dans le cadre d'une approche hypothético-déductive, basée sur des critères préétablis et dépendante de l'interprétation des données cliniques recueillies, les résultats du patient ne sont plus comparés par rapport aux résultats d'une population de référence mais par rapport à lui-même dans une situation circonscrite à un contexte et à un observable.

De fait, l'évaluation critériée a l'avantage d'atteindre des objectifs multiples : décrire et comprendre les difficultés du sujet (Laveault & Grégoire, 2014 ; McCauley & Swisher, 1984), faciliter la mise en œuvre d'objectifs (Justice et al., 2006) et de moyens thérapeutiques (McCauley, 1996), planifier l'intervention (Laveault & Grégoire, 2014), évaluer de manière continue et fréquente des progrès (Justice et al., 2006 ; Proger & Mann, 1973) et enfin, identifier d'éventuels transferts de compétences / habiletés (Martinez et al., 2015).

Une telle approche soulève de nombreuses réflexions qui mériteraient d'être approfondies en recherche fondamentale interdisciplinaire et en recherche appliquée à la clinique et à la pédagogie.

## 1. Les données théoriques : une composante indispensable

Si nous prenons du recul par rapport aux caractéristiques d'une évaluation critériée, développées précédemment, il apparaît clairement que la qualité des données recueillies est dépendante de la validité théorique de l'outil. De ce fait, il apparaît nécessaire de s'appuyer sur des données externes pour la sélection des épreuves, l'élaboration des items et des critères d'analyse, la standardisation terminologique et, bien entendu, la définition opérationnelle du construit évalué. Idéalement, l'évaluation normée doit également répondre à ces standards de qualité. Pourtant il a été démontré que ce n'est pas le cas pour la majorité des outils à la disposition des orthophonistes francophones (Gaul Bouchard et al., 2009 ; Lafay & Cattini, 2018 ; Leclercq & Veys, 2014) alors que ceux-ci sont utilisés prioritairement pour la pose des diagnostics. Le manque d'appuis théoriques ne peut qu'entraver la construction d'outils diagnostiques ayant un pouvoir discriminant satisfaisant. Lors de la revue des différents outils critériés que nous avons précédemment détaillés, nous avons pu constater un manque de transparence méthodologique concernant la phase d'élaboration. Ne bénéficiant pas de normes pour comparer le patient à un autre individu, il est primordial de se saisir de données développementales ou de critères de performances clairement établis pour interpréter les résultats par le biais d'un raisonnement hypothético-déductif robuste. Si la qualité de l'outil est dépendante des preuves externes pour toutes les formes d'évaluation, elle est essentielle dans le cas d'une évaluation critériée. Une épreuve normée peut s'accommoder d'items statistiquement discriminants (i.e. capacité à distinguer aussi clairement et aussi finement que possible les individus en fonction du critère considéré) pour catégoriser les individus d'une population selon la présence d'un trouble ou non. Une épreuve critériée doit fournir des items permettant d'évaluer un niveau de performance. Cela nécessite une connaissance très précise des variables en jeu lors de la réalisation d'une tâche et la gradation précise des niveaux de difficulté selon l'âge, le niveau scolaire et/ou la pathologie de l'individu (Grégoire, 1994). Par conséquent, les universaux dégagés convergent tous vers une prise en compte des données externes, une explicitation de celles-ci au clinicien et une prise en compte du patient en tant qu'individu.

## 2. L'évaluation critériée statique au service des patients avec un profil complexe

Lors de l'évaluation, la présence d'atypies doit inviter le clinicien à poursuivre sa recherche d'éléments pathologiques ou du moins problématiques pour la qualité de vie du patient. Il est bien établi que les résultats des tests ne sont pas valides pour un patient qui ne correspond pas à la représentativité de la population de référence, même si le test est administré conformément aux instructions (Laing & Kamhi, 2003). Les profils de patients considérés comme complexes sont au cœur de cette réflexion et justifieraient un usage plus répandu des évaluations permettant des interprétations critériées. Dans ce cas de figure, ce que nous désignons par *profils complexes* se définit par des caractéristiques qui sortiraient des échantillons normatifs traditionnels chez un patient. Ces caractéristiques pourraient concerner un système social, culturel, étiologique ou identitaire même moins représentatif de ces échantillons. Plusieurs situations peuvent correspondre à cette condition. Elles sont finalement assez fréquentes dans la pratique orthophonique.

Par exemple, ce contexte prévaut particulièrement pour l'évaluation du langage et des habiletés psycholinguistiques des personnes bilingues (De Lamo White & Jin, 2011; Laing & Kamhi, 2003; Marrus & Hall, 2017). La littérature est riche d'études tentant de définir des outils d'évaluation alternatifs (Orellana et al., 2019) ou de déterminer les épreuves moins sensibles à

cette variable (Thordardottir & Brandeker, 2013). L'utilisation d'outils normés élaborés à partir d'une population monolingue peut conduire à un nombre important de faux positifs.

D'autres situations peuvent être identifiées au regard de la pathologie du patient rencontré. En situation clinique, les orthophonistes prennent souvent le parti de réaliser des épreuves normées qui se référeraient de manière purement spéculative - *mais faute de mieux* -, à un âge équivalent en fonction des performances obtenues par le patient à une épreuve (i.e. l'âge développemental) en complétant par une évaluation descriptive (McCauley & Swisher, 1984). Ce raisonnement suppose que les compétences se développeraient de manière indépendante les unes des autres et sans être sujettes aux facteurs environnementaux et cognitifs. Par conséquent, un même nombre d'erreurs représenterait un même profil d'un individu à l'autre.

Dans le cadre de la déficience intellectuelle notamment, il n'est pas rare que les cliniciens comparent les scores du patient par rapport à une norme présumée représentative à partir des résultats obtenus à un examen normatif de quotient intellectuel. Dans ce cas, l'extrapolation est double : le test cognitif permettrait de donner un âge correspondant au développement du patient dans sa globalité ; cet âge, dit représentatif des compétences du patient, permettrait de situer celui-ci par rapport aux normes attendues à ce niveau développemental.

Prenons également l'exemple des personnes atteintes de troubles cognitifs légers. Celles-ci peuvent arriver avec une plainte pour un manque du mot, des difficultés de planification discursive et des disfluences (Duboisdindien et al., 2018). Pour autant, les évaluations normalisées des habiletés psycholinguistiques des personnes avec un trouble cognitif léger ne sont pas assez spécifiques pour objectiver ces phénomènes. Faute d'outils discriminants, les patients sont renvoyés chez eux avec leur plainte. Toutes ces situations génèrent des contretemps dans la prise de décision clinique, lors de l'évaluation orthophonique

Une approche par évaluation critériée statique part d'une vision ontologique du tableau clinique qui est à l'étude, vers une vision critériologique. L'objectif étant de distinguer des formes qui permettraient d'objectiver des atypies grâce à un certain nombre de caractères dynamiques.

### **3. Vers un consensus terminologique pour la conception d'outils plus robustes**

Historiquement, l'évaluation critériée a été introduite dans le domaine des sciences de l'éducation et des sciences médicales depuis les années 1970 (Proger & Mann, 1973). Une telle interdisciplinarité dans des domaines à la fois différents dans leurs approches et complémentaires dans leur dimension éthique, rend complexe l'élaboration d'une définition universelle et exhaustive de l'évaluation critériée. Une autre difficulté réside dans l'hétérogénéité des usages que cet outil revêt ainsi que dans sa diversité de formes, de compétences visées et de supports d'évaluation (e.g. tâches de récit narratif, de résolution de problèmes, d'évaluation des compétences d'étudiants, manipulations de matériels ou examen oral/écrit). Pour autant, il est raisonnable d'affirmer l'existence d'un concept commun pour lequel on peut dégager au moins neuf caractéristiques explicites.

Le domaine de l'évaluation orthophonique est lui aussi traversé par des problématiques de consensus terminologique. Ce premier constat souligne l'importance d'adopter une terminologie commune, même si cela implique des changements d'habitude ou de pratique. Il serait pertinent à l'avenir que les concepteurs qui développent une évaluation – quelle qu'elle soit – emploient et vérifient l'adéquation du terme utilisé pour la qualifier. Ces systématismes

dans les appellations permettraient tant au niveau clinique qu'au niveau de la formation initiale et continue, d'utiliser les bons « outils » méthodologiques et de se référer aux données de la littérature afin d'avoir des résultats robustes. Définir un objectif clair et concis permettrait aux concepteurs de déterminer en amont les différentes caractérisations de l'outil (e.g. forme, conditions de la tâche). Ainsi, ils se référeraient aux standards méthodologiques pour assurer la robustesse de l'épreuve.

Donner un statut souverain à l'évaluation normée parce qu'elle a cette capacité à évaluer un large spectre de compétences au regard d'un échantillon représentatif de la population, est non pertinent voire contre-productif pour l'évaluation normée elle-même et évidemment pour caractériser un fonctionnement. Enfin, nous espérons qu'en respectant les caractéristiques propres à l'évaluation critériée statique, plus d'orthophonistes seront susceptibles de monter une évaluation de ce type pour évaluer un observable problématique pour leur patient (voir Limbrick et al. (2013) pour une discussion sur le sujet)-en effectuant des recherches de données externes pour offrir un maximum de validité à leur outil.

## ----- CONCLUSION -----

Depuis les années 1990, nous observons un regain d'intérêt pour les démarches hypothético-déductives telles qu'elles ont été originellement décrites par Whewell (1837, 1840) dans le domaine des sciences de l'éducation. Le renforcement intellectuel manifeste des recherches cliniques orientées en Evidence Based Medicine ainsi que le besoin de rationaliser les pratiques professionnelles et leur accueil dans de nombreux domaines de la société (Kaufman & Kaufman, 2018) ont favorisé l'attrait des praticiens et des acteurs à l'interface de la recherche clinique.

L'évaluation critériée s'inscrit dans cette démarche. En effet, l'évaluation critériée statique implique une évaluation consciencieuse des compétences, des symptômes et des stratégies d'un individu par rapport à lui-même au regard d'un observable délimité. Cette évaluation consciencieuse permet d'objectiver des comportements déficitaires et/ou stratégiques (e.g. emploi d'occurrences spécifiques en phonologie ou en morphosyntaxe, usage de stratégies opératoires en résolution de problèmes). Elle peut être intégrée dans un continuum dans lequel le clinicien emploie tout aussi bien une évaluation normée statique qu'une évaluation critériée dynamique. Premier avantage, elle présente l'opportunité de placer les connaissances du clinicien/de l'évaluateur et les demandes du patient/élève au cœur d'un dispositif de critères basés sur des données externes. Second avantage et sans prétendre à l'exhaustivité, elle permet une approche raisonnée de l'évaluation d'un individu à l'aune de ses ressources et dans la perspective à long terme d'offrir tous les moyens de résilience et de réadaptation possibles pour le patient/élève face aux situations problématiques.

L'évaluation orthophonique se définit par un ensemble de mesures et de collectes de données cliniques permettant de répondre à la plainte rapportée par le patient et/ou son entourage. L'interprétation critériée représente un potentiel important pour répondre à la démarche d'évaluation continue, individualisée et fonctionnelle afin de proposer au patient les meilleurs soins possibles au regard des données de la littérature, de son expertise, des caractéristiques du patient et du système dans lequel chacun évolue.

## ----- BIBLIOGRAPHIE -----

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129. <https://doi.org/10.1093/applin/14.2.115>
- Applebee, A. N. (1978). *The Child's Concept of Story: Ages 2 to 17*. University of Chicago Press.
- ASHA. (n.d.). *Practice portal - Clinical topics: Speech Sound Disorders - Articulation and Phonology*. Retrieved March 17, 2020, from <https://www.asha.org/practice-portal/clinical-topics/articulation-and-phonology/>
- Aubry, N., Bissonnette, I., & Raymond, S. (2019). *Inventaire des structures syllabiques chez l'enfant francophone (ISSEF)*. <https://www.ciussc-capitalesnationale.gouv.qc.ca/inventaire-des-structures-syllabiques-chez-lenfant-francophone-issef>
- Audétat, M.-C., Faguy, A., Jacques, A., Blais, J.-G., & Charlin, B. (2011). Étude exploratoire des perceptions et pratiques de médecins cliniciens enseignants engagés dans une démarche de diagnostic et de remédiation des lacunes du raisonnement clinique. *Pédagogie Médicale*, 12(1), 7-16. <https://doi.org/10.1051/pmed/2011014>
- Blomert, L., Kean, M. L., Koster, C., & Schokker, J. (1994). Amsterdam-Nijmegen everyday language test: construction, reliability and validity. *Aphasiology*, 8(4), 381-407. <https://doi.org/10.1080/02687039408248666>
- Brousseau-Lapré, F., Rvachew, S., Macleod, A. A. N., Findlay, K., Bérubé, D., Bernhardt, B., & Findlay, K. (2018). Une vue d'ensemble : les données probantes sur le développement phonologique des enfants francophones canadiens. *Revue Canadienne d'orthophonie et d'audiologie*, 42(1), 1-19.
- Brown, J. D., & Hudson, T. (2002). *Criterion-Referenced Language Testing*. Cambridge University Press.
- Burton, K. (2006). Designing criterion-referenced assessment. *Journal of Learning Design*, 1(2), 73-82.
- Darrigrand, B., & Mazaux, J. M. (2000). *Echelle de communication verbale de Bordeaux (E.C.V.B.) : manuel d'utilisation*. Ortho-Edition.
- De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders*, 46(6), 613-627. <https://doi.org/10.1111/j.1460-6984.2011.00049.x>
- Denman, D., Kim, J.-H., Munro, N., Speyer, R., & Cordier, R. (2019). Describing language assessments for school-aged children: A Delphi study. *International Journal of Speech-Language Pathology*, 1-11. <https://doi.org/10.1080/17549507.2018.1552716>
- Dixon, M. R., Belisle, J., Stanley, C., Rowsey, K., Daar, J. H., & Szekely, S. (2015). Toward a behavior analysis of complex language for children with autism: Evaluating the relationship

- between PEAK and the VB-MAPP. *Journal of Developmental and Physical Disabilities*, 27(2), 223-233. <https://doi.org/10.1007/s10882-014-9410-4>
- Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2014). Assessing children's writing products: The role of curriculum based measures. *British Educational Research Journal*, 41(4), 575-595. <https://doi.org/10.1002/berj.3162>
- Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: assessing language skills in young children. *Child and Adolescent Mental Health*, 20(2), 116-125. <https://doi.org/10.1111/camh.12072>
- Duboisdindien, G., Grandin, C., Boutet, D., & Lacheret-Dujour, A. (2018). A Multimodal corpus to check on pragmatic competence for Mild Cognitive Impaired aging people. *Corpus*, 19(19). <https://doi.org/10.4000/corpus.4295>
- Dunsmuir, S., Kyriacou, M., Batuwitige, S., Hinson, E., Ingram, V., & O'Sullivan, S. (2015). An evaluation of the Writing Assessment Measure (WAM) for children's narrative writing. *Assessing Writing*, 23, 1-18. <https://doi.org/10.1016/j.asw.2014.08.001>
- Ekionea, J. B., Bernard, P., & Plaisent, M. (2011). Consensus par la méthode Delphi sur les concepts clés des capacités organisationnelles spécifiques de la gestion des connaissances. *Recherches Qualitatives*, 29(3), 168-192.
- Elstein, A., Shulman, L., & Sprafka, S. (1978). *Medical problem solving: An analysis of clinical reasoning*. Harvard University Press.
- Fabiano-Smith, L. (2019). Standardized Tests and the Diagnosis of Speech Sound Disorders. *Perspectives of the ASHA Special Interest Groups*, 4(1), 58-66. [https://doi.org/10.1044/2018\\_PERS-SIG1-2018-0018](https://doi.org/10.1044/2018_PERS-SIG1-2018-0018)
- Gaul Bouchard, M.-E., Fitzpatrick, E. M., & Olds, J. (2009). Analyse psychométrique d'outils d'évaluation utilisés auprès des enfants francophones. *Revue Canadienne d'orthophonie et d'audiologie*, 33(3), 129-139.
- Grégoire, J. (1994). Le diagnostic des troubles de l'acquisition de la lecture. In J. Grégoire & B. Piérart (Eds.), *Évaluer les troubles de la lecture : Les nouveaux modèles théoriques et leurs implications diagnostiques*. (pp. 33-48). De Boeck Supérieur.
- Gordon, J., Hazlett, C., Ten Cate, O., Mann, K., Kilminster, S., Prince, K., ... & Newble, D. (2000). Strategic planning in medical education: enhancing the learning environment for students in clinical settings. *Medical education*, 34(10), 841-850. <https://doi.org/10.1046/j.1365-2923.2000.00759.x>
- Hasson, N., & Joffe, V. (2007). The case for dynamic assessment in speech and language therapy. *Child Language Teaching and Therapy*, 23(1), 9-25. <https://doi.org/10.1177/0265659007072142>
- Holland, A., Frattali, C., & Fromm, D. (1999). *Communication activities of daily living (2nd ed.)*. Pro-Ed.

- Huang, R.-J., Hopkins, J., & Nippold, M. A. (1997). Satisfaction with standardized language testing: A survey of speech-language pathologists. *Language, Speech, and Hearing Services in Schools*, 28(1), 12-29. <https://doi.org/http://dx.doi.org.proxy.library.vanderbilt.edu/10.1044/0161-1461.2801.12>
- Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology*, 15(2), 177-191. [https://doi.org/10.1044/1058-0360\(2006/017\)](https://doi.org/10.1044/1058-0360(2006/017))
- Kaufman, A. B., & Kaufman, J. C. (2018). *Pseudoscience: The conspiracy against science*. MIT Press.
- Kerr, M. A., Guildford, E., & Bird, E. K. (2003). Standardized language test use: a canadian survey. *Revue Canadienne d'orthophonie et d'audiologie*, 27(1), 10-28. [https://cjslpa.ca/files/2003\\_JSLPA\\_Vol\\_27/No\\_01\\_1-92/Kerr\\_Guildford\\_Kay-Raining\\_Bird\\_JSLPA\\_2003.pdf](https://cjslpa.ca/files/2003_JSLPA_Vol_27/No_01_1-92/Kerr_Guildford_Kay-Raining_Bird_JSLPA_2003.pdf)
- Klecan-Aker, J. S., & Colson, K. (2009). Criterion-referenced assessment for language organization: An example of evidence-based practice. *Forum on Public Policy*.
- Lafay, A., & Cattini, J. (2018). Analyse psychométrique des outils d'évaluation mathématique utilisés auprès des enfants francophones. *Revue Canadienne d'orthophonie et d'audiologie*, 42(2), 127-144. [https://cjslpa.ca/files/2018\\_CJSLPA\\_Vol\\_42/No\\_02/CJSLPA\\_Vol\\_42\\_No\\_2\\_2018\\_MS\\_1129.pdf](https://cjslpa.ca/files/2018_CJSLPA_Vol_42/No_02/CJSLPA_Vol_42_No_2_2018_MS_1129.pdf)
- Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools*, 34(1), 44-55. [https://doi.org/10.1044/0161-1461\(2003/005\)](https://doi.org/10.1044/0161-1461(2003/005))
- Laveault, D., & Grégoire, J. (2014). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. De Boeck Supérieur.
- Leclercq, A.-L., & Veys, E. (2014). Réflexions sur le choix de tests standardisés lors du diagnostic de dysphasie. *ANAE : Approche Neuropsychologique des Apprentissages chez l'Enfant*, 26(131), 374-382.
- Lefebvre, P., & Trudeau, N. (2005). L'orthophoniste et les tests normalisés. *Fréquences*, 17(2), 17-20.
- Limbrick, N., McCormack, J., & McLeod, S. (2013). Designs and decisions: The creation of informal measures for assessing speech production in children. *International Journal of Speech-Language Pathology*, 15(3), 296-311. <https://doi.org/10.3109/17549507.2013.770552>
- Marrus, N., & Hall, L. (2017). Intellectual disability and language disorder. *Child and Adolescent Psychiatric Clinics of North America*, 26(3), 539-554. <https://doi.org/10.1016/j.chc.2017.03.001>

- Martinez, T., Dor, O., & Maillart, C. (2015). Préciser, argumenter et évaluer, les objectifs thérapeutiques pour améliorer la prise en charge orthophonique. *Rééducation Orthophonique*, 261, 63-89.
- McCauley, R. (1996). Familiar strangers : Criterion-referenced measures in communication disorders. *Language, Speech and Hearing Services in Schools*, 27, 122-131. <https://doi.org/10.1044/0161-1461.2702.122>
- McCauley, R., & Swisher, L. (1984). Use and misuse of norm-referenced test in clinical assessment. *Journal of Speech and Hearing Disorders*, 49(4), 338-348. <https://doi.org/10.1044/jshd.4904.338>
- Montallana, K. L., Gard, B. M., Lotfizadeh, A. D., & Poling, A. (2019). Inter-rater agreement for the milestones and barriers assessments of the Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP). *Journal of Autism and Developmental Disorders*, 49(5), 2015-2023. <https://doi.org/10.1007/s10803-019-03879-4>
- Mottier Lopez, L., & Laveault, D. (2014). L'évaluation des apprentissages en contexte scolaire. *Mesure et Évaluation En Éducation*, 31(3), 5-34. <https://doi.org/10.7202/1024962ar>
- Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, 42(1), 15-29. <https://doi.org/10.1016/j.im.2003.11.002>
- Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal of Speech-Language Pathology*, 28(3), 1298-1317. [https://doi.org/10.1044/2019\\_AJSLP-18-0202](https://doi.org/10.1044/2019_AJSLP-18-0202)
- Paliwoda, S. J. (1983). Predicting the future using Delphi. *Management Decision*, 21(1), 31-38. <https://doi.org/10.1108/eb001309>
- Piérart, B. (2006). Nature et fonctions des bilans. In F. Estenne & B Piérart (Eds.), *Les bilans de langage et de voix : fondements théoriques et pratiques* (pp. 7-13). Masson.
- Proger, B. B., & Mann, L. (1973). Criterion-reference measurement. *Journal of Learning Disabilities*, 6(2), 73-84. <https://doi.org/10.1177/002221947300600203>
- Rondal, J. A., & Seron, X. (2000). *Troubles du langage : bases théoriques, diagnostic et rééducation*. Mardaga.
- Schelstraete, M. A. (2011). *Traitements du langage oral chez l'enfant : Symptômes et interventions*. Elsevier Health Sciences France.
- Sundberg, M. (2017). *VB-MAPP : Évaluation du comportement verbal et programme d'intervention*. De Boeck Supérieur.
- Sundberg, M. L. (2008). *VB-MAPP Verbal Behavior Milestones Assessment and Placement Program: A language and social skills assessment program for children with autism or other developmental disabilities: Guide*. AVB Press.

- Thibault, M.-P., Helloin, M.-C., & Croteau, B. (2010). *Exalang 5-8 : Batterie informatisée pour l'examen du langage oral et écrit chez l'enfant de 5 à 8 ans* (No. 2). Orthomotus.
- Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, 46(1), 1-16. <https://doi.org/10.1016/j.jcomdis.2012.08.002>
- Wechsler, D. (1996). Wechsler Objective Language Dimensions. Kent, United Kingdom, *The Psychological Corporation*.
- Whewell, W. (1837). *History of the inductive sciences*. Parker.
- Whewell, W. (1840). *The philosophy of the inductive sciences: founded upon their history*. Parker.
- Willett, G. (1996). Paradigme, théorie, modèle, schéma : qu'est-ce donc ? *Communication et Organisation*, 10. <https://doi.org/10.4000/communicationorganisation.1873>