



Faculté de Lettres et Sciences humaines
Institut des Sciences Logopédiques

L'identification d'un trouble développemental du langage oral chez les enfants bilingues français-portugais

Thèse présentée à l'Université de Neuchâtel pour l'obtention du titre
de docteur ès sciences humaines et sociales, discipline logopédie

Salomé SCHWOB

Sous la direction de :

Katrin SKORUPPA (professeure ordinaire, Université de Neuchâtel)

Membres du jury :

Madame Katrin SKORUPPA, professeure ordinaire, Université de Neuchâtel

Madame Elin THORDARDOTTIR, professeure, Université de McGill

Madame Anja BLECHSCHMIDT professeure,

Fachhochschule Nordwestschweiz

Monsieur Marco HESSELS, professeur ordinaire, Université de Genève

Monsieur Philippe PREVOST, professeur, Université de Tours

Soutenu le 21 octobre 2021

IMPRIMATUR

La Faculté des lettres et sciences humaines de l'Université de Neuchâtel, sur les rapports de Mme Katrin Skoruppa, directrice de thèse, professeure, Université de Neuchâtel ; Mme Anja Blechschmidt, professeure, Fachhochschule Nordwestschweiz ; M. Marco Hessels, professeur, Université de Genève ; M. Philippe Prévost, professeur, Université de Tours (France) ; Mme Elin Thordardottir, professeure, McGill University, Montréal (Canada) autorise l'impression de la thèse présentée par Mme Salomé Schwob en laissant à l'autrice la responsabilité des opinions énoncées.

Neuchâtel, le 21 octobre 2021

p.o. Elin Thordardottir
Le doyen
Matthieu Honegger

Remerciements

Pour commencer, mes remerciements vont à ma directrice de thèse, Katrin Skoruppa. Dès notre première rencontre, tu m'as accordé une grande confiance en acceptant que je travaille à tes côtés. Avec ton expérience internationale dans le monde de la recherche, tu as toujours été de bons conseils et tu n'as pas compté les heures pour m'aiguiller et commenter mes essais.

Je remercie également chaleureusement les membres de mon jury : Elin Thordardottir, Anja Bleschschmidt, Marco Hessels et Philippe Prévost. Merci d'avoir investi du temps pour lire ma thèse et d'y avoir apporté vos commentaires lors de mon colloque de thèse.

Mes remerciements vont également à toutes les familles et participants ayant accepté de prendre part à ma recherche, de m'avoir accueillie chez eux à plusieurs reprises. Merci aux logopédistes, enseignants et centres scolaires et parascolaires de m'avoir apporté votre aide dans la recherche de participants. Je remercie également les étudiants ayant contribué à l'analyse des résultats ou encore au recrutement.

Mes sincères remerciements vont également à mon assistante-étudiante, Patricia, qui m'a accompagnée partout en Suisse romande pour réaliser les épreuves en portugais. Merci aux personnes ayant acceptées de prêter leur voix pour mes tâches d'évaluation dynamique, Klara et Damien. Merci à mes relectrices, en particulier, Letizia, Maëlle, Mellie, Elodie et Nuzhat. Je remercie également mes collègues de Pam7 et collègues logopédistes pour votre soutien et vos conseils lors de la rédaction de cette thèse.

Finalement, je tiens à remercier ma famille, mes amis, mes parents et bien sûr Damien, sans qui tout ce travail n'aurait pas pu être possible. Merci pour votre soutien !

Résumé

Cette thèse s'inscrit dans le cadre des études actuelles portant sur la problématique de l'évaluation du langage oral des enfants bilingues dans un contexte logopédique. Pour diverses raisons, il est actuellement difficile d'identifier un trouble développemental du langage oral chez l'enfant bilingue (Armon-Lotem, 2012). Nous pouvons mentionner, de façon non exhaustive, qu'il est complexe de faire la distinction entre un trouble développemental du langage et des difficultés transitoires dues au manque d'exposition aux langues (Hasson et al., 2013). Plusieurs auteurs se sont intéressés à divers outils ou modes d'évaluation permettant de pallier ce problème (ex. Chiat, 2015 ; Thordardottir, 2015a). Toutefois, les études n'ont encore que peu testé l'efficacité de ces méthodes auprès d'enfants bilingues et se sont intéressées principalement à des combinaisons de langues fréquentes telles que l'anglais et l'espagnol aux États-Unis, laissant de côté des combinaisons moins courantes et rendant urgent la recherche à leur sujet dans des pays où elles sont majoritaires. De plus, la plupart des études se sont penchées sur des propositions statiques en observant les difficultés des enfants à un moment précis, sans analyser leur potentiel d'apprentissage (Camilleri et Law, 2007). En effet, peu de recherche se sont intéressées à l'utilisation d'une évaluation dynamique au sein du contexte logopédique (voir Orellana et al., 2019).

Au moyen de cinq articles, cette thèse de doctorat avait pour but de faire état des problématiques, des outils disponibles et a également eu comme objectif de créer de nouvelles tâches d'évaluation dynamique. Elle souhaitait ensuite tester et comparer l'efficacité des outils existants et nouvellement créés auprès d'une population monolingue et bilingue français-portugais âgée de 5 ;0 à 7 ;11 ans vivant en Suisse romande. Septante-deux enfants avec et sans trouble développemental du langage ont pris part à cette recherche et ont participé à trois sessions d'évaluation. Les enfants ont été assignés à quatre groupes (avec et sans trouble développemental du langage- monolingue et bilingue) en se basant sur une classification initiale

portant notamment sur l'évaluation clinique des logopédistes sur le terrain. Pour préciser cette classification, des sous-tests standardisés statiques en français et en portugais ainsi qu'une tâche *cross-linguistique* (Chiat, 2015) ont été proposés aux enfants. Leurs résultats ont été comparés aux normes monolingues de référence dans les tests et nous avons également testé l'adaptation de seuils de pathologie en fonction de leur taux d'exposition langagier (Thordardottir, 2015a) calculé via un questionnaire parental (basé sur Tuller, 2015). Nous nous sommes longuement penchées sur les tâches de répétition de non-mots fréquemment relevées dans la littérature comme prédictrices de trouble développemental du langage (Estes et al., 2007). Leur utilité clinique a été examinée à travers une méta-analyse et dans une note de recherche s'intéressant à diverses répétitions de non-mots (langues-spécifiques et *cross-linguistique*). En plus de ces tests statiques, nous avons voué un intérêt particulier à l'évaluation dynamique qui est encore peu étudiée et utilisée dans le contexte logopédique (Orellana et al., 2019). Deux situations d'évaluation dynamique consistant en l'apprentissage de nouveaux mots (substantifs et verbes) et d'une nouvelle règle morphologique ont été créées et proposées aux enfants. Elles se distinguaient en termes d'activité (jeu autonome sur ordinateur et lecture d'histoires avec interaction) ainsi que de procédure d'aide apportée par l'adulte (avec plus ou moins de médiation).

Nos résultats mettent en exergue de façon théorique et empirique l'intérêt des tâches de répétition de non-mots pour évaluer les enfants monolingues et bilingues. La tâche de répétition de non-mots en français s'est montrée la plus discriminante de nos sous-tests statiques. Les autres sous-tests en français et en portugais n'ont pas montré d'intérêt dans l'évaluation langagière de nos enfants bilingues, ne montrant que peu de différences entre les enfants avec et sans trouble développemental du langage même lorsqu'une adaptation de seuils de pathologie en fonction du taux d'exposition était opérée (Thordardottir, 2015a). Certaines questions adressées aux parents ont également montré leur intérêt diagnostique (âge de la phase

combinatoire et inquiétude parentale), bien qu'il s'agisse de post-prédictions. À la fin de l'analyse de nos sous-tests statiques, nous découvrons que douze enfants n'ont pas répondu comme attendu selon notre classification initiale : des enfants avec un trouble développemental n'échouaient dans aucune/une langue et des enfants avec un développement langagier présumé typique ont éprouvé des difficultés. Ainsi, ce sont uniquement quarante-neuf enfants qui ont été inclus dans nos analyses finales portant sur l'évaluation dynamique (à noter que onze autres enfants ont été exclus car ils ne répondaient pas aux critères généraux de l'étude). Les situations d'évaluation dynamique, notamment la tâche interactive de lecture d'histoires incluant des indices graduels de la part de l'expérimentateur, s'est montrée discriminante avec de hauts pourcentages de sensibilité (91 %) et de spécificité (85 %) (selon Plante et Vance, 1994). Grâce à l'analyse de précision diagnostique de nos situations d'évaluation dynamique, cinq enfants pouvaient être assignés différemment aux sous-groupes intermédiaires. Bien que notre objectif de diagnostic ait permis d'améliorer la classification de nos participants, des doutes subsistent quant à notre dernière classification que nous avons considérée comme complémentaire, reflétant ainsi le processus complexe d'identification des enfants bilingues avec un trouble développemental du langage.

Au terme de cette thèse, nous avons conclu que la répétition de non-mots en français et la situation d'évaluation dynamique interactive de lecture d'histoires étaient les tâches testées les plus précises et remplissaient, au moins en partie, notre objectif initial d'identification du trouble développemental du langage oral chez une population monolingue et bilingue français-portugais. Nous pensons également que les autres tâches dans les deux langues de l'enfant, qui n'ont montré que peu d'intérêt diagnostique dans notre thèse, sont tout de même importantes pour avoir davantage de connaissances sur ce qu'ils maîtrisent dans leurs langues dans une visée d'intervention. Cette recherche nous a donc permis d'enrichir les connaissances théoriques portant sur une combinaison de langues encore peu étudiée et fréquente en Suisse romande.

Nous avons essayé de fournir des preuves théoriques et empiriques quant à l'utilisation de certaines tâches et modes d'évaluation prometteurs dans le contexte clinique d'évaluation logopédique des enfants bilingues.

Mots clés : évaluation logopédique, trouble développemental du langage, bilinguisme.

Summary

This thesis is part of the current research on the issue of oral language assessment of bilingual children in a speech and language pathology context. For various reasons, it is currently difficult to identify a developmental language disorder in bilingual children (Armon-Lotem, 2012). We can mention, in a non-exhaustive way, that it is complex to distinguish between a developmental language disorder and temporary difficulties due to lack of language exposure (Hasson et al., 2013). Several authors have focused on various tools or modes of assessment to overcome this problem (e.g., Chiat, 2015; Thordardottir, 2015a). However, studies have done only a limited amount of testing of the effectiveness of these methods with bilingual children and have focused mainly on frequent language combinations such as English and Spanish in the United States, leaving out less common combinations and making research on them urgent in countries where they are in the majority. Moreover, most studies have focused on static propositions by observing children's difficulties at a specific point in time, without analyzing their learning potential (Camilleri & Law, 2007). Indeed, few research has focused on the use of dynamic assessment within the speech and language pathology context (see Orellana et al., 2019).

Through five articles, this doctoral thesis aimed to report on the issues, the available tools, and also had the objective of creating new dynamic assessment tasks. It then wanted to test and compare the effectiveness of the existing and newly created tools with a monolingual and bilingual French-Portuguese population aged 5;0 to 7;11 years old living in French-speaking Switzerland. Seventy-two children with and without developmental language disorders participated in this research and took part in three assessment sessions. The children were assigned to four groups (with and without developmental language impairment, monolingual and bilingual) based on an initial classification including clinical assessment by trained speech-language pathologists. To refine this classification, children were administered

static standardized subtests in French and Portuguese and a cross-linguistic task (Chiat, 2015). Their results were compared to monolingual norms of reference in the tests, and we also tested the adaptation of pathology thresholds according to their rate of language exposure (Thordardottir, 2015a) calculated via a parental questionnaire (based on Tuller, 2015). We looked extensively at non-word repetition tasks frequently identified in the literature as predictors of developmental language impairment (Estes et al., 2007). Their clinical utility was examined through a meta-analysis and in a research note examining a variety of non-word repetitions (language-specific and crosslinguistic). In addition to these static tests, we have devoted particular interest to dynamic assessment, which is still little studied and used in the speech and language pathology context (Orellana et al., 2019). Two dynamic assessment situations consisting of learning new words (nouns and verbs) and a new morphological rule were created and proposed to children. They differed in terms of activity (autonomous computer game and interactive story reading) as well as in the procedure of help provided by the adult (with more or less mediation).

Our results highlight theoretically and empirically the value of non-word repetition tasks to assess monolingual and bilingual children. The French repetition task proved to be the most discriminating of our static subtests. The other standardized subtests in French and Portuguese did not show any interest in the language assessment of our bilingual children, showing little difference between children with and without developmental language disorders even when pathology thresholds were adapted according to exposure rate (Thordardottir, 2015a). Some questions addressed to parents also showed diagnostic interest (age of combinatorial phase and parental concern), although these ones were post-predictions. At the end of the analysis of our subtests, we found that twelve children did not respond as expected according to our initial classification, i.e., children with developmental language disorder were failing in none/one language and children with presumed typical language development were struggling. Thus,

only forty-nine children were included in our final analyses of the dynamic assessment (note that an additional eleven children were excluded because they did not satisfy the general criteria of the study). The dynamic assessment situations, particularly the interactive story reading situation with graduated cues from the experimenter, proved to be discriminating with high percentages of sensitivity (91 %) and specificity (85 %) (according to Plante & Vance, 1994). Through the diagnostic accuracy analysis of our dynamic assessment situations, five children could be assigned differently to the intermediate subgroups. Although our diagnostic focus improved the classification of our participants, doubts remain about our final classification, which we considered complementary, reflecting the complex process of identifying bilingual children with developmental language impairment.

At the end of this thesis, we concluded that the French non-word repetition and the interactive dynamic assessment situation were the most accurate tasks tested and fulfilled, at least in part, our initial goal of identifying developmental oral language disorder in a monolingual and bilingual French-Portuguese population. We also believe that the other tasks in the child's two languages, which did not show diagnostic interest in our thesis, are still important to have more knowledge about what they know in their languages for intervention purposes. This research allowed us to enrich the theoretical knowledge about a language combination that is still little studied and frequent in French-speaking part of Switzerland. We tried to provide theoretical evidence for the use of some promising tasks in the clinical context of speech and language assessment of bilingual children.

Key words: speech and language assessment, developmental language disorder, bilingualism.

Table des matières

LISTE DES ABREVIATIONS	19
PREAMBULE	21
Contexte général de la thèse	21
Structure de la thèse	24
PARTIE I : REPERES THEORIQUES	25
1. INTRODUCTION THEORIQUE	27
1.1 Le trouble développemental du langage oral	27
1.1.1 Domaines pouvant être impactés par le TDL	31
1.2 Article : Le défi de l'évaluation langagière des enfants bilingues : état des lieux des ressources disponibles en logopédie	35
PARTIE II : PROBLEMATIQUE ET METHODE	79
2. PROBLEMATIQUE	81
2.1 Questionnement principal	86
2.2 Hypothèses concernant les sous-tests statiques langagiers	87
2.2.1 Modes d'évaluation	88
2.2.2 Tâches d'évaluation	90
2.3 Hypothèses à propos de l'évaluation dynamique	92
3. METHODE	97
3.1 Participants	97
3.1.1 Critères pour la constitution des sous-groupes	99
Statut linguistique	99
Statut clinique	102
Classifications utilisées.	102
3.1.2 Procédure de recrutement	107
3.2 Matériels	107
3.2.1 Sous-tests langagiers	108
Adaptation de seuils de pathologie	110
3.2.2 Questionnaire parental	110
3.2.3 Situations d'évaluation dynamique	111
3.3 Procédure de passation	113
3.4 Procédure de cotation et analyse	115

3.4.1	Accord inter-juge	116
3.5	Analyses statistiques	117
PARTIE III : L'ÉVALUATION DANS LES DEUX LANGUES DE L'ENFANT		119
4.	ARTICLE DE RECHERCHE	121
PARTIE IV : LA RÉPÉTITION DE NON-MOTS DANS L'ÉVALUATION LOGOPÉDIQUE DES ENFANTS MONOLINGUES ET BILINGUES		163
5.	ARTICLE DE RECHERCHE	165
6.	NOTE DE RECHERCHE	211
6.1	Informations additionnelles	249
PARTIE V : L'USAGE DE L'ÉVALUATION DYNAMIQUE AU SEIN DU BILAN LOGOPÉDIQUE		251
7.	ARTICLE DE RECHERCHE	253
7.1	Analyses additionnelles	322
7.2	Informations additionnelles	323
PARTIE VI : DISCUSSION ET CONCLUSION		325
8.	DISCUSSION	327
8.1	Synthèse de la thèse et implications cliniques	327
8.1.1	Sous-questions 1	327
8.1.2	Sous-questions 2	335
8.1.3	Synthèse au questionnaire	340
8.2	Limites et ouvertures	343
9.	CONCLUSION	351
BIBLIOGRAPHIE		353
ANNEXES		369
Annexe A : Création des items des tâches d'évaluation dynamique		371
Annexe B : Script et captures d'écran du jeu sur ordinateur		373
Annexe C : Histoires évaluation dynamique		379

Annexe D : Création des tâches d'évaluation	389
Annexe E : Questionnaire parental	391
Annexe F : Analyse des répétitions de non-mots	395
Annexe G : Analyse évaluation dynamique	399
Annexe H : Guide d'étayage	413

Liste des abréviations

DT : enfants avec développement langagier présumé typique

TDL : enfants avec un trouble développemental du langage oral

RNM : répétition de non-mots

Remarque préliminaire : les normes APA 7 (2019) ont été utilisées pour les citations, tableaux et figures.

Préambule

Cette thèse s'intègre dans le cadre de l'obtention du titre de docteur ès sciences humaines et sociales ayant pour discipline la logopédie. Elle s'intéresse au trouble développemental du langage oral (TDL) et à son évaluation dans un contexte logopédique monolingue et bilingue. Le TDL touche environ 5-10 % des enfants qu'ils soient monolingues ou bilingues sans cause apparente (Bishop et al., 2017 ; Norbury et al., 2016). Toutefois, en raison du nombre de plus en plus important d'enfants qui communique dans différentes langues (Grosjean, 2015), son identification est rendue complexe dans le contexte logopédique (Armon-Lotem, 2012). Il est difficile de faire la distinction entre des difficultés relevant d'un TDL et des difficultés passagères dues au manque d'exposition aux langues par l'enfant (Hasson et al., 2013). De plus, nous pouvons évoquer, de façon non exhaustive, que les cliniciens manquent d'outils normés pour les enfants bilingues et ne parlent souvent pas leur langue (Boerma et Blom, 2017). Ces raisons peuvent engendrer une sur- ou sous-identification des enfants avec un TDL dans la population bilingue (Camilleri et Law, 2007).

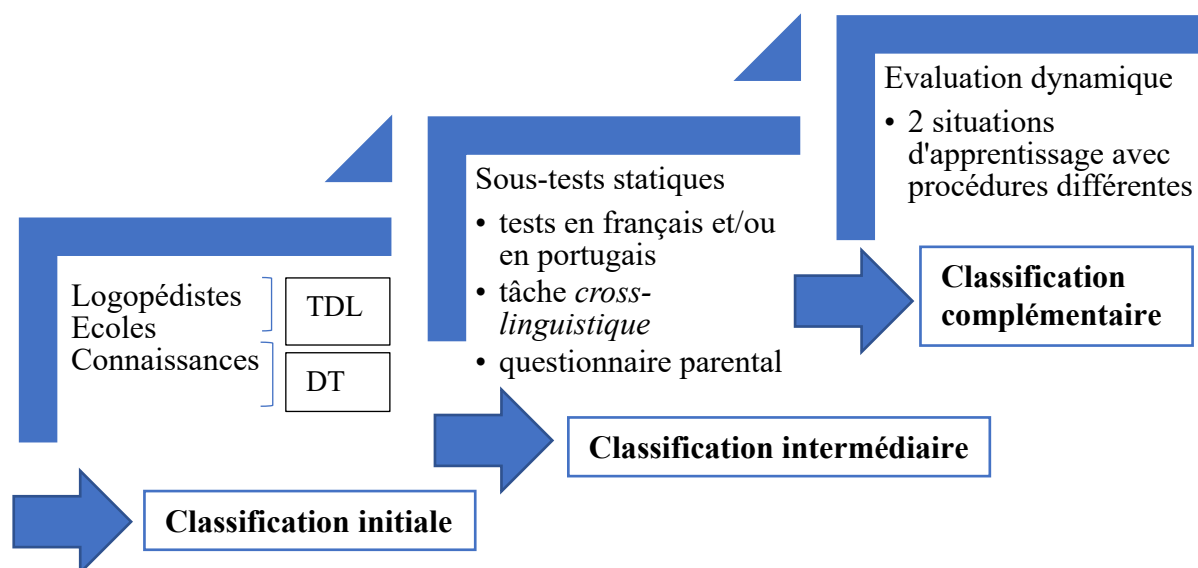
Contexte général de la thèse

Ce travail de doctorat propose de s'intéresser théoriquement à divers moyens et outils d'évaluation relevés dans la littérature scientifique pour évaluer les enfants bilingues. Les recherches et réflexions émanant de la théorie sont présentées dans un article (Schwob, 2020) apparaissant au point 1.2. Cette thèse souhaite aussi prêter une attention particulière à ces moyens et outils de façon empirique en poursuivant un objectif de diagnostic. Elle propose de tester et comparer l'efficacité de plusieurs outils d'évaluation dits classiques (ex. tests standardisés dans une et/ou les deux langues de l'enfant, répétition de non-mots, questionnaire parental), qui seront considérés comme des sous-tests statiques dans cette recherche, et innovants (situations d'évaluation dynamique) ainsi que différents modes d'évaluation (dans

une ou les deux langues de l'enfant, faisant usage de non-mots) auprès d'une population d'enfants d'âge scolaire partageant une combinaison de langues très fréquente en Suisse romande, le français et le portugais (Office Fédéral de la Statistique, 2020) et largement représentée en logopédie (Volpin et al., 2020). En testant l'efficacité des tâches et modes d'évaluation, cette recherche essaye au moyen des calculs de précision diagnostique, d'améliorer la classification initiale et imparfaite des participants à la thèse. Plus précisément, la Figure 1 ci-dessous permet d'apprécier le cheminement de la thèse ayant pour objectif le diagnostic du TDL.

Figure 1

De la classification initiale à l'identification des enfants avec un TDL



Note. Abréviations : TDL = enfants avec un trouble développemental du langage ; DT = enfant avec un développement langagier présumé typique.

Nous allons bien évidemment détailler les tâches utilisées et mentionnés dans la Figure 1 dans le chapitre traitant de la méthode (3). Toutefois, il convient de mentionner rapidement ici que les enfants de notre recherche ont été initialement considérés comme ayant un TDL ou un développement langagier présumé typique (DT) et nous tenterons grâce à des sous-tests statiques de proposer une classification intermédiaire puis en faisant usage de nos situations d'évaluation dynamique de proposer une classification complémentaire la plus précise possible. Bien que notre visée principale fût la mise au point de deux situations d'évaluation dynamique, nous avons également exploité les résultats issus de nos sous-tests statiques langagiers. Les résultats issus de l'ensemble de la recherche ont été valorisés par quatre articles empiriques, un étant publié (chapitre 5), deux étant acceptés pour publication (se situant aux points 4 et 6) et un étant soumis à une revue internationale (voir point 7). Le souhait de cette thèse de doctorat est de fournir des pistes cliniques aux logopédistes pour une identification plus précise d'un TDL chez cette population bilingue, cette étape étant fondamentale pour la mise en place d'un suivi logopédique adapté. Elle vise également à élargir le panorama des recherches portant sur des combinaisons de langues moins fréquentes que celles incluant l'anglais.

Cette recherche a été animée par différents raisons scientifiques (que nous avons évoqué brièvement ci-dessus et qui seront développées dans la Partie I de ce travail) et personnelles. Étant moi-même logopédiste depuis quelques années, je suis amenée à rencontrer, tous les jours, des patients qui communiquent dans des langues que je ne maîtrise pas. Il a été rapidement frustrant pour moi de ne pas pouvoir identifier précisément si ces enfants présentaient un TDL, raison pour laquelle je me suis intéressée et plongée dans la littérature scientifique de ce domaine. Ayant eu l'honneur d'intégrer l'équipe de recherche de Katrin Skoruppa à l'Université de Neuchâtel, j'ai ainsi développé mes questionnements durant mes quatre années de doctorat.

Structure de la thèse

Ce travail comporte six Parties. La Partie I comporte un chapitre théorique présentant le TDL et est suivi d'un article publié (Schwob, 2020) qui se focalise sur son évaluation au sein d'une population bilingue afin de comprendre l'état de l'art sous-tendant cette thèse de doctorat et en dressant un inventaire des outils disponibles. La Partie II exposera la problématique, les questions de recherche et hypothèses théoriques de la thèse. Les éléments méthodologiques concernant notre population de recherche et notre procédure seront également présentés dans ce chapitre. Les Parties III et IV évoqueront l'usage de nos sous-tests langagiers ainsi que leurs analyses. La Partie III portera sur l'utilisation de tests standardisés dans les deux langues de l'enfant et se concentrera uniquement sur les enfants bilingues de notre cohorte. Nous aborderons ici plusieurs modes d'évaluation et outils disponibles en clinique pour réaliser une évaluation langagière des enfants bilingues. La Partie IV s'intéressera exclusivement à l'usage de la répétition de non-mots dans le bilan langagier des enfants monolingues et bilingues. Celle-ci débutera par une revue systématique de la littérature et méta-analyse (chapitre 5, Schwob et al., 2021) puis une note de recherche s'intéressant à plusieurs types de répétition de non-mots (langues-spécifiques et *cross-linguistique*) sera également présentée (chapitre 6). La Partie V se focalisera sur une méthode d'évaluation innovante et prometteuse : l'évaluation dynamique. Un article de recherche portant sur deux situations d'évaluation dynamique que nous avons créées sera présenté (chapitre 7). Enfin, la Partie VI proposera une synthèse de la thèse de doctorat et exposera des éléments de discussion en lien avec la pratique clinique. Nous terminerons ensuite par une conclusion. Les références théoriques seront présentées à la fin de chaque article et la bibliographique générale se tiendra à la fin de la thèse.

Partie I : Repères théoriques

1. Introduction théorique

Pour mieux comprendre le domaine de cette thèse de doctorat, cette introduction théorique débutera par un sous-chapitre portant sur le TDL. Nous exposerons l'historique de cette pathologie et faisant mention de quelques précisions terminologiques. Nous évoquerons aussi les causes possibles du TDL et finirons par présenter les domaines pouvant être impactés par celui-ci. Seules quelques spécificités inhérentes au bilinguisme seront brièvement évoquées dans cette section, car à la suite de celle-ci, un article publié (point 1.2) permettra d'apprécier plus spécifiquement le TDL dans le contexte du bilinguisme et de faire un état des lieux des outils disponibles sur le terrain pour effectuer une évaluation logopédique des enfants bilingues. Ces aspects sont essentiels pour saisir la problématique de cette thèse.

1.1 Le trouble développemental du langage oral

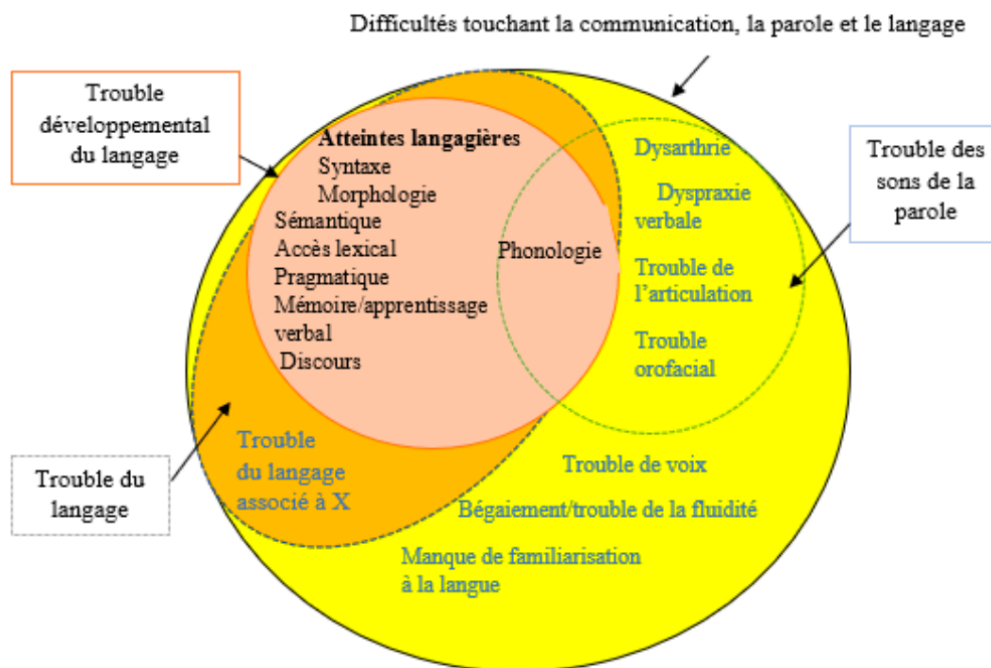
La première description de trouble du langage chez les enfants est datée du début du XIX^{ème} siècle (pour un historique voir de Montfort Supple et Söderpalm, 2010). Il a été rapidement associé aux pathologies adultes (ex. aphasie) et a intéressé de nombreuses disciplines, dont la médecine qui a fourni les premières observations et hypothèses explicatives (voir de Weck et Marro, 2010). Le champ de la logopédie et la profession, telle que nous la connaissons maintenant, se sont largement développés au XX^{ème} siècle (de Montfort Supple et Söderpalm, 2010 ; de Weck et Marro, 2010).

Des terminologies différentes en fonction des champs disciplinaires, des époques et des langues de recherche (ex. anglais- français) ont été utilisées. La première description d'un trouble langagier chez l'enfant a été nommée aphasie congénitale (de Montfort Supple et Söderpalm, 2010) puis a évolué au fil des années et des siècles en passant par aphasie de développement (voir Paul et Norbury, 2012), dysphasie et trouble spécifique ou primaire du langage (voir Breault et al., 2019). Dans le monde francophone, de Weck et Marro (2010) font

état, dans leur ouvrage, du terme de trouble du développement du langage. En plus des nombreuses terminologies employées, la notion de « spécificité » du trouble du langage a été fréquemment remise en question (Breault et al., 2019). Par exemple, le système international de classification biomédicale, DSM-5, (American Psychiatric Association, 2013) qui répertorie le trouble du langage dans la catégorie « troubles de la communication », estime que des performances intellectuelles en dessous de la norme sont possibles et que la présence d'un trouble du spectre autistique n'est pas un critère d'exclusion à un trouble du langage (Avenet et al., 2016). Ainsi, pour choisir des appellations et des critères communs aux troubles du langage, un projet international nommé CATALISE a décidé par consensus auprès d'un large panel d'experts dans le domaine (Bishop et al., 2017). Le terme de trouble développemental du langage (TDL par la suite) a été choisi quand il s'agit d'un trouble persistant impactant divers domaines langagiers, ayant des impacts fonctionnels dans la vie des patients et faisant référence à une étiologie incertaine (Breault et al., 2019). Le terme de difficultés de langage est privilégié quand il renvoie à des difficultés transitoires comme cela peut être le cas dans une situation de bilinguisme lorsque l'enfant n'est pas suffisamment exposé à la langue de la scolarisation (Breault et al., 2019). Enfin, le terme de trouble du langage associé à « x » est utilisé lorsque le trouble du langage est relié à une condition biomédicale connue (ex. trouble du spectre de l'autisme ; Breault et al., 2019). La Figure 2 présentée ci-dessous met en lumière les spécificités du TDL et permet aussi d'apprécier visuellement ses frontières avec le trouble des sons et de la parole, le trouble du langage associé à une pathologie neuro-développementale reconnue et les difficultés liées au manque d'exposition à la langue de scolarisation.

Figure 2

Le trouble développemental du langage oral, consensus CATALISE



Note. Traduction de Breault et al. (2019) et tiré de l'article de Bishop et al. (2017).

Pour pouvoir attribuer le diagnostic de TDL, l'enfant doit échouer à des tests langagiers standardisés mais ce critère n'est pas suffisant (Bishop et al., 2017). En effet, l'écart à la norme pour définir un score déficitaire se situe généralement entre -1 et -2 écart-type (Thordardottir et al., 2011) mais cet écart n'est pas unanimement défini dans la littérature. De plus, ces écarts à la norme sont encore plus problématiques lorsqu'il s'agit d'enfants bilingues puisqu'ils ne peuvent pas être comparés à la population monolingue de référence dans les tests (Thordardottir, 2015a). Il s'agit alors d'utiliser des outils d'évaluation ne se concentrant pas essentiellement sur les incapacités de l'enfant avec un TDL ; cela peut-être le cas de l'évaluation dynamique qui s'intéresse à la capacité d'apprentissage de l'enfant (Camilleri et Law, 2007 ; nous y reviendrons dans la suite de cette thèse et évoquerons également d'autres pistes possibles). Il est aussi nécessaire de prendre en considération l'impact fonctionnel du

TDL dans la vie de l'enfant mais ici encore peu d'outils valides se sont développés et ne sont que peu utilisés sur le terrain (Breault et al., 2019).

Les causes de TDL sont difficilement définissables et semblent multifactorielles (Leonard, 2014). Il a été reconnu que certains facteurs génétiques sont présents dans l'étiologie du TDL (Rice, 2013), la présence de difficultés langagières dans l'entourage familial constitue ainsi un facteur de risque (Bishop et al., 2017) de la même façon qu'être un garçon ou avoir des parents peu formés ou scolarisés (Rudolph, 2017). Durant longtemps, le TDL ou plus anciennement le trouble spécifique du langage devait se présenter en l'absence d'autres problématiques (ex. déficience intellectuelle), toutefois, comme mentionné précédemment, ces critères d'exclusion variaient et n'ont fait l'objet d'aucun consensus. C'est la raison pour laquelle le projet CATALISE a décidé de ne faire usage d'aucun critère d'exclusion pour définir les troubles du langage (Bishop et al., 2017 ; Breault et al., 2019). Il est notamment reconnu que d'autres troubles développementaux peuvent être fréquemment associés au TDL (ex. troubles spécifiques des apprentissages, Bishop et al., 2017). De plus, comme évoqué auparavant, des conditions biomédicales peuvent être assimilées aux troubles du langage, c'est notamment le cas dans des pathologies développementales plus larges (voir Figure 2, trouble du langage associé à « x »).

Ainsi, selon des études récentes, le TDL touche environ 5-10 % des enfants monolingues et bilingues (Bishop et al., 2017 ; Kohnert, 2010 ; Norbury et al., 2016 ; Tomblin et al., 1996). Le TDL se manifeste dans les différentes langues acquises par l'enfant (Kohnert, 2010 ; Paradis et al., 2011). Certains déficits du TDL ont été largement documentés mais semblent également caractéristiques d'autres troubles du développement. C'est le cas notamment d'un déficit en mémoire de travail (Breault et al., 2019). Même si des difficultés sont observables dans le domaine de la mémoire de travail, il convient alors d'affiner ses hypothèses avant de conclure à la présence d'un TDL. Ainsi, une analyse de différents domaines langagiers et des différentes

modalités langagières est nécessaire pour attester de la présence d'un TDL. Étant donné que les difficultés rencontrées sont belles et bien hétérogènes (Thomas et al., 2019), Bishop et al. (2017) recommandent ainsi d'évaluer les domaines de la phonologie, de la syntaxe, du lexique, de la pragmatique, du discours et de la mémoire verbale lors d'une évaluation langagière orale des enfants. Toutefois, les auteurs du consensus n'ont pas réussi à se mettre d'accord sur des sous-type de TDL (Bishop et al., 2017) alors que dans le passé, d'autres auteurs, tels que Van der Lely (2005) avait proposé des sous-classifications de TDL avec un domaine langagier plus sévèrement atteint (ex. « TDL grammatical »). En effet, plusieurs études longitudinales ont attesté la persistance des difficultés chez les enfants avec un TDL, toutefois, elles relèvent que le profil des difficultés semble varier durant l'enfance et que les enfants ne présentent pas un sous-type de TDL de façon durable (Conti-Ramsden et Botting, 1999 ; Tomblin et al., 2003). Le TDL semble donc dynamique (Conti-Ramsden et Botting, 1999) avec des difficultés phonologiques, lexicales et syntaxiques dans la petite enfance et s'apparentant à des difficultés pragmatiques lorsque les enfants sont plus âgés (Paul et Norbury, 2012). Nous avons choisi, dans ce qui suit, de présenter les différents domaines pouvant être impactés dans le TDL.

1.1.1 Domaines pouvant être impactés par le TDL

La phonologie est fréquemment impactée dans le TDL chez les enfants monolingues et bilingues (Armon-Lotem et de Jong, 2015). La présence d'erreurs de parole persistantes, l'inintelligibilité des enfants ainsi que des difficultés de conscience phonologique sont fréquemment relevées chez les enfants avec un TDL (Bishop et al., 2017). Selon le consensus CATALISE (Bishop et al., 2017), des difficultés phonologiques ou articulatoires isolées avant cinq ans s'apparentent à un trouble des sons et de la parole (dans la Figure 2) et supposent un bon pronostic d'évolution. À contrario, des difficultés phonologiques persistantes après cinq ans avec d'autres atteintes langagières sont associées à un TDL. Chez les enfants bilingues, les

compétences phonologiques peuvent parfois être retardées en raison d'un transfert inter linguistique typique, lorsque certains sons ne sont, par exemple, pas partagés par les deux systèmes phonologiques de l'enfant (voir la revue systématique d'Hambly et al., 2013). Il convient ainsi de se renseigner suffisamment sur l'acquisition phonologique des deux langues et sur le taux d'exposition aux langues de l'enfant avant de pouvoir conclure à un TDL (Hambly et al., 2013). Les capacités de précision phonologique sont fréquemment évaluées par une tâche de répétition de non-mots (RNM) ayant montré son potentiel dans l'évaluation des enfants monolingues (Coady et Evans, 2008 ; Estes et al., 2007) et bilingues (Girbau et Schwartz, 2007 ; Schwob et al., 2021, article présent au chapitre 5). Toutefois, pour évaluer précisément la phonologie, il convient de proposer une tâche de RNM ne variant pas en termes de longueur syllabique, pour réduire l'effet de la mémoire à court-terme verbale, mais proposant plusieurs niveaux de complexité phonologique (structures syllabiques présentes) (Dos Santos et Ferré, 2018). Nous reviendrons sur la RNM dans la suite cette thèse (Partie IV). Puisque des difficultés phonologiques sont présentes dans divers troubles, tels que dans le trouble des sons et de la parole ou le TDL, voir Figure 2, il semble indispensable d'évaluer conjointement les autres domaines langagiers que nous allons évoquer ci-dessous.

Les domaines de la syntaxe et de la morphologie sont fréquemment relevés comme impactés dans le TDL (Van der Lely, 2005). L'usage adapté de la morphologie grammaticale est complexe pour les enfants avec un TDL acquérant différentes langues (Leonard, 2014). L'omission ou la substitution des morphèmes grammaticaux sont fréquemment observées chez les enfants avec un TDL. Toutefois, ces éléments sont également relevés dans le développement langagier typique des enfants bilingues, puisque certains enfants parlent des langues ne possédant pas ces indications morphologiques (Kohnert et Danahy, 2007) et commettent donc des erreurs en acquérant des langues les nécessitant (Armon-Lotem et De Jong, 2015). D'autres sources de difficultés sont relevées quel que soit le statut linguistique des enfants avec un TDL,

c'est, par exemple le cas des difficultés dans la compréhension des monèmes interrogatifs (voir Schulz, 2015).

Au niveau lexical, les enfants avec un TDL présentent souvent un manque du mot (Leonard, 2014) car le mot est en mémoire mais le chemin pour y accéder est inapproprié ou inefficace, ou encore les connexions entre les mots ne sont pas suffisamment riches. Leur stock lexical semble également être réduit (McGregor et al., 2013) et leur capacité d'apprentissage de nouveaux mots restreinte (ex. Gray, 2003). La diversité lexicale est pauvre notamment pour les verbes ; les enfants avec un TDL feraient fréquemment usage des mêmes verbes généraux dans leurs productions (Kambanaros et Grohmann, 2015). L'évaluation lexicale chez les enfants bilingues est particulièrement complexe car il est possible qu'ils possèdent un stock lexical plus restreint que les enfants monolingues mais ce n'est pas le cas lorsque les stocks totaux dans les deux langues de l'enfant sont pris en considération (Thordardottir et al., 2006). De plus, la quantité d'exposition aux langues influence fortement l'acquisition dans le domaine lexical avec des différences pour le vocabulaire réceptif et expressif (Thordardottir, 2011).

Les difficultés pragmatiques des enfants avec un TDL affectent la compréhension et la production de leurs messages en contexte (Bishop et al., 2017). La gestion des tours de parole et des pannes conversationnelles sont fréquemment relevées comme problématiques (voir de Weck et Marro, 2010). Les enfants avec un TDL sont souvent décrits comme des partenaires de communication peu performants (Hardley et Rice, 1991). Ces difficultés sont caractéristiques des troubles du spectre de l'autisme mais sont également présentes lorsque les enfants ne présentent pas cette pathologie développementale (Bishop et al., 2017). Particulièrement dans l'évaluation du domaine pragmatique, il convient de faire attention aux différences culturelles.

Les enfants avec un TDL semblent rencontrer des difficultés dans la production et la compréhension des discours, bien que celles-ci dépendent largement du genre discursif (de

Weck et Marro, 2010). Les auteurs s'accordent pour dire que des difficultés quant à la microstructure (référant à la complexité du langage utilisé, voir Rezzonico et al., 2016) et la macrostructure narrative (structure du récit) sont impactées chez les enfants avec un TDL (Gagarina et al., 2015 ; Rezzonico et al., 2016). Les difficultés de macrostructure semblent communes à tous les enfants monolingues et bilingues avec un TDL alors que les compétences de microstructure semblent davantage liées à l'exposition reçue (Boerma et al., 2016).

Des difficultés à retenir des informations verbales sont régulièrement notées chez les enfants avec un TDL mais il est souvent difficile de dire si cela est une cause du TDL ou une conséquence de celui-ci (Vance, 2008). Les difficultés liées à la mémoire à court-terme verbale se reflètent dans une tâche de RNM. En effet, plus les items sont longs, plus les enfants avec un TDL ont de la difficulté à retenir l'information verbale. Toutefois, il semble que les enfants avec un développement langagier typique (DT) soient également sensibles à la longueur des items (Marini et al., 2017).

Les difficultés des enfants avec un TDL peuvent donc toucher l'un ou plusieurs de ces domaines de façon dynamique durant l'enfance (Conti-Ramsden et Botting, 1999). L'hétérogénéité des difficultés doit donc être prise en considération dans l'évaluation logopédique en testant plusieurs de ces domaines et en prêtant une attention particulière à la situation de bilinguisme afin de ne pas interpréter par exemple des difficultés transitoires typiques dans le développement bilingue à des difficultés s'apparentant à un TDL. Nous allons porter une attention particulière à la situation de bilinguisme dans le sous-chapitre qui suit.

**1.2 Article : Le défi de l'évaluation langagière des enfants bilingues :
état des lieux des ressources disponibles en logopédie**

Schwob, S.

État : Publié

Schwob, S (2020). Le défi de l'évaluation langagière des enfants bilingues : état des lieux des ressources disponibles en logopédie. *Travaux neuchâtelois de linguistique*, 72(2), 63-87. http://www.unine.ch/files/live/sites/tranel/files/Tranel/72/63-87_Schwob.pdf

Abstract

Approximately 7% of monolingual and bilingual children have a developmental language disorder (Kohnert, 2010). These children need to be assessed and diagnosed and offered appropriate care. Speech and language therapists are responsible for assessing and managing the language of monolingual and bilingual children. The language assessment of bilingual children is a challenge for speech and language therapists (Boerma & Blom, 2017; Estienne & Van der Linden, 2014; Tuller & al., 2018) because it is currently hard to distinguish between difficulties related to a developmental disorder and a possible delay due to the bilingual situation as both populations present similar errors. (Armon-Lotem, 2012). The lack resources and adapted assessment tools enhances the challenges in differentiating the two populations (Kehoe, 2009). The purpose of this paper is therefore to provide an overview of current knowledge about bilingual language development and developmental language disorder, to show the speech and language therapist's difficulties with bilingual language assessments, and to draw up a list of current resources available to them. Several tools will be described and research attesting to their effectiveness in differentiating between bilingual children with and without language disorders will be presented. We will then discuss their relevance to practice using a clinical vignette.

Introduction théorique

La mission des logopédistes est de pouvoir recevoir et évaluer les capacités et difficultés langagières des enfants de divers horizons linguistiques et culturels. Une évaluation de qualité permet ensuite de leur proposer, s'il elle s'avère nécessaire, une prise en charge adaptée. A l'heure de la mondialisation et de la diversification des flux migratoires, de plus en plus d'enfants parlant plusieurs langues sont signalés en service de logopédie. Nous verrons à travers cette introduction théorique comment sont définis les troubles développementaux du langage et ce que constitue une évaluation logopédique. Nous nous pencherons également sur l'évaluation langagière des enfants bilingues¹ qui représente actuellement un des défis les plus importants de la pratique actuelle des logopédistes².

Les troubles développementaux du langage et leur évaluation

La plupart des enfants développent leur langage de manière typique, sans difficultés. Toutefois, environ 10% des enfants âgés entre 4 et 5 ans présentent des troubles du langage (Norbury et al., 2016). Chez les trois quarts environ, il s'agit de 'troubles développementaux du langage' (TDL : terminologie ayant fait l'objet d'un consensus international, voir Bishop et al., 2017; Breault et al., 2019) qui apparaissent en l'absence de conditions biomédicales sous-jacentes ou de déficience intellectuelle attestée (Norbury et al., 2016). Malgré des décennies de recherche, les causes de ces troubles sont difficilement définissables mais semblent multifactorielles, incluant des facteurs génétiques (Leonard, 2014 ; Rice, 2013).

¹ Dans ce document, nous avons privilégié l'utilisation du terme de bilinguisme pour faire référence à l'acquisition de plusieurs langues.

² Les divers exemples de pratique logopédique sont expliqués en prenant en considération le cas de la Suisse. Nous nous référons, entre autres, à la vignette clinique suisse de l'ouvrage de Law et al. (2019). Des différences en fonction des régions et des pays peuvent notamment être recensées mais ne seront pas traitées dans ce document.

Le terme de logopédie est utilisé comme synonyme d'orthophonie mais sera privilégié dans le document.

Le diagnostic de TDL peut être formulé en prenant en considération un décalage par rapport à la norme (Estienne et Vander Linden, 2014). En effet, les scores des enfants avec un TDL aux évaluations standardisées réalisées par le logopédiste doivent se situer en dessous de la moyenne par rapport aux autres enfants du même âge (Leonard, 2014). Généralement, les critères de déviation se situent entre -1 et -2 écart-types (ET) par rapport à la norme de l'âge en question (Thordardottir et al., 2011). Le TDL se caractérise également par sa persistance. C'est pour cette raison que le diagnostic est généralement posé vers 5-6 ans lorsque l'évolution reste lente et que les difficultés persistent malgré la prise en charge logopédique (Leclercq et Leroy, 2012). Nous relevons une grande hétérogénéité des profils des enfants avec un TDL (Thomas et al., 2019) avec le versant réceptif et/ou expressif qui est plus ou moins touché. Les domaines langagiers qui sont impactés diffèrent également en fonction des enfants. En effet, les difficultés peuvent se situer au niveau phonologique, lexical, syntaxique, discursif et/ou pragmatique de façon isolée ou combinée (de Weck et Marro, 2010).

L'évaluation logopédique des enfants bilingues

L'évaluation logopédique des enfants bilingues diffère quelque peu de l'évaluation des enfants monolingues. Pour mieux en saisir les raisons, nous allons évoquer brièvement comment se définit le bilinguisme en Suisse et comment se développe le langage des enfants bilingues. La présence et le profil de difficultés langagières chez une population bilingue seront également mentionnés. Puis, nous allons exposer les raisons pour lesquelles l'évaluation logopédique bilingue est actuellement problématique en clinique.

Le développement du langage des enfants bilingues

Le monde et, par conséquent, la Suisse est multilingue et multiculturelle (OFS, 2018 ; Williams et McLeod, 2012). En effet, plus de deux tiers des Suisses utilisent fréquemment une deuxième langue (OFS, 2018). Ce bilinguisme est dû à divers facteurs tels que la migration religieuse, économique et/ou politique (Grosjean, 1993). Les langues les plus utilisées en Suisse sont les quatre langues nationales mais également d'autres langues telles que l'anglais et le portugais. En raison du multilinguisme de plus en plus fréquent dans les différentes régions du monde, les recherches au sujet du bilinguisme ont largement été développées ces vingt dernières années (de Houwer, 2009). L'hétérogénéité des individus parlant plusieurs langues rend complexe la définition du bilinguisme. Dans ce papier, nous adoptons celle de Grosjean (2015) qui est actuellement la plus utilisée en clinique et en recherche francophone. L'auteur définit le bilinguisme comme 'la maîtrise d'au moins une compétence linguistique dans une autre langue' (2015 : 16). L'âge d'acquisition des langues est important puisqu'il permet de distinguer deux types de bilinguisme, appelés simultané ou successif. L'enfant bilingue simultané apprend ces deux langues simultanément depuis la naissance ou avant l'âge de 3 ans (Paradis, 2007). A contrario, l'enfant bilingue successif acquiert sa deuxième langue après l'âge de 3 ans. En fonction des âges et/ou des contextes d'apprentissage, une des deux langues peut être considérée comme dominante et l'autre comme minoritaire (Paradis, 2007 ; Kohnert, 2010).

Le développement linguistique des enfants bilingues et monolingues présente des similitudes et des différences quantitatives et qualitatives (Werker et al., 2009). Les différences peuvent s'expliquer par la présence de divers facteurs d'influence du développement langagier bilingue (Paradis, 2019). Nous allons nous y intéresser dans les lignes qui suivent. L'âge du début d'acquisition et, par conséquent, le type de bilinguisme est un facteur d'influence important (de Almeida et al., 2016). Les recherches au sujet des étapes développementales des enfants bilingues successifs sont actuellement minoritaires car elles dépendent de l'âge du début

d'acquisition de la langue et les données sont hétérogènes. Nous pouvons mentionner que la quantité et la qualité du langage adressé à l'enfant bilingue est primordiale pour le développement harmonieux de ses deux langues (Paradis, 2019 ; Vygotski, 1934/1985). La typologie et la proximité des langues de l'enfant peuvent aussi faire différer le développement linguistique et la quantité d'exposition nécessaire à son apprentissage (Paradis, 2019). De plus, les différences entre les enfants bilingues peuvent également varier en fonction du domaine linguistique. Par exemple, Thordardottir (2015a) relate que le développement morphosyntaxique chez les enfants bilingues est influencé par la quantité d'exposition, mais à moindre mesure que pour le vocabulaire (Thordardottir, 2011). En effet, les connaissances lexicales de l'enfant bilingue sont souvent partagées entre ses deux langues (Altman et al., 2018). En ce qui concerne le domaine phonologique, l'âge d'exposition aux langues semble avoir une influence importante notamment en ce qui concerne la précision phonologique et les capacités de discrimination (Paradis, 2007 ; Ruiz-Felter et al., 2016). Des interactions sont possibles entre les différents systèmes phonologiques de l'enfant (Fabiano-Smith et Goldstein, 2010). La modalité linguistique semble aussi jouer un rôle dans le développement linguistique des enfants bilingues. En effet, l'écart entre le lexique réceptif et expressif peut être plus important chez les enfants bilingues que chez les enfants monolingues (Keller et al., 2015). Le statut social des langues ainsi que le parcours migratoire de l'enfant peuvent également avoir une incidence sur son développement langagier (Simon et al., 2010). Enfin, les facteurs psychoaffectifs tels que la motivation, le style d'apprentissage et la confiance en soi de l'enfant peuvent également avoir un impact (Paradis et al., 2011).

Le trouble développemental du langage chez les enfants bilingues

Il n'existe pas plus de risque à présenter un TDL chez les enfants bilingues que chez les enfants monolingues. Le chiffre de prévalence est identique à la population monolingue, soit

d'environ 7% (Kohnert, 2010). De même, il n'existe pas d'avantages ou d'inconvénients à apprendre une deuxième langue en ce qui concerne l'apparition d'un TDL (Kohnert, 2010 ; Leonard, 2014). Le TDL se manifeste dans les différentes langues de l'enfant (Kohnert, 2010 ; Paradis et al., 2011). Comme pour les étapes d'acquisition, les profils de difficultés sont relativement similaires chez les enfants bilingues et monolingues. Leonard (2014) relève que le domaine morphosyntaxique et, plus particulièrement, la morphologie grammaticale, est souvent affectée par les TDL quelles que soient les langues de l'enfant. Toutefois, chaque langue possède ses spécificités linguistiques et les types d'erreurs commises semblent être différentes en fonction des langues (Leonard, 2014). En français, certaines recherches dont celles de Paradis et Crago (2001) ont trouvé un nombre plus important d'erreurs morphosyntaxiques chez les enfants monolingues avec un TDL que chez ceux ayant un développement typique (DT) âgés de 7 ans. Toutefois, d'autres études telles que celle de Thordardottir et Namazi (2007) n'observent pas particulièrement d'erreurs morphosyntaxiques chez les enfants de 4 ans monolingues avec un TDL. Ils relèvent plutôt un décalage chronologique avec des utilisations se référant à des enfants avec un DT plus jeunes. Ils notent également l'existence d'erreurs présentes seulement chez les enfants avec un TDL, soit des erreurs sur les infinitifs et sur les participes passés dans des phrases sans sujet. Il semble donc qu'il existe des différences inter-langues mais également beaucoup de différences dues aux choix méthodologiques des études (âge des participants, type de tâche, méthode de cotation et condition de passation, etc.).

Actuellement, en clinique logopédique, il est difficile pour les logopédistes de faire la différence entre des difficultés relevant d'un TDL et des difficultés temporaires liées à la situation de bilinguisme (Camilleri et Law, 2007). En effet, certains types d'erreurs sont similaires dans les deux cas de figure (Armon-Lotem, 2012). Par exemple, des erreurs dans la morphologie verbale et nominale (Armon-Lotem, 2012) ainsi que des difficultés dans la microstructure narrative sont recensées dans les deux situations (Boerma et al., 2016). Dans le contexte d'apprentissage

bilingue, ces erreurs sont dues à l'input reçu (jugé encore insuffisant par exemple) et disparaissent ainsi au gré des expositions langagières et des feedbacks donnés alors qu'elles perdurent en présence d'un réel TDL.

L'évaluation logopédique des enfants bilingues devrait donc, dans l'idéal, tenir compte de ces aspects et observer si les difficultés de l'enfant sont présentes ou non dans ses deux langues (Kehoe, 2009). Toutefois, les évaluations langagières des enfants bilingues représentent actuellement un défi pour les logopédistes pour différentes raisons. Certaines langues, même européennes, ne disposent d'aucun ou de peu d'outils standardisés. De plus, nous relevons que certains tests sont normés pour une population monolingue spécifique et sont utilisés par des professionnels d'autres pays par manque d'outils. Par exemple en francophonie, il existe un large éventail de tests français utilisés également par des professionnels d'autres pays francophones (la Suisse par exemple), malgré les importantes différences régionales et socio-culturelles, ce qui conduit à une évaluation biaisée (Grégoire, 2006). Il arrive également que les logopédistes ne parlent pas la langue de l'enfant (Williams et McLeod, 2012) et que pour des raisons économiques et administratives, le recours à des interprètes de qualité ne soit pas aisé (Boerma et Blom, 2017). Enfin, la majorité des tests existants ne sont pas normés pour une population bilingue. En effet, même si le logopédiste est capable d'en faire l'usage, il réalise une évaluation avec des outils normés pour des enfants monolingues. Quelques exceptions de tests normés pour des enfants bilingues existent ayant une combinaison de langues incluant l'anglais et l'espagnol (ex. test 'BESA' de Peña et al., 2018). La conception de test pour une population bilingue n'est pas aisée puisqu'il faut prendre en considération que le développement bilingue dépend de nombreux facteurs (présentés ci-dessus) et qu'il est donc difficile d'établir des normes avec une population autant hétérogène (Boerma et Blom, 2017). Pour ces diverses raisons, les logopédistes utilisent fréquemment des tests ayant des normes monolingues dans la langue qu'ils maîtrisent (Estienne et Vander Linden, 2014). Ces tests

monolingues sont sensibles aux différences culturelles et linguistiques et au degré d'exposition à la langue ce qui rend difficilement interprétable leurs résultats pour une population bilingue. En effet, il arrive fréquemment que les difficultés des enfants bilingues aux épreuves ne soient pas imputables à de réelles difficultés langagières mais plutôt à un manque d'exposition à la langue (Hasson et Botting, 2010). Les réelles capacités et difficultés langagières des enfants bilingues ne ressortent donc pas à travers ces épreuves. Les logopédistes se voient donc confrontés à des erreurs de diagnostics, soit en sur-diagnostiquant les enfants bilingues ou encore en attribuant toutes leurs difficultés à leur background linguistique bilingue (Kohnert, 2010).

Objectifs de l'article

Suite à cette introduction théorique, nous relevons que le pourcentage d'enfants parlant plusieurs langues croît en raison de la mondialisation et des flux migratoires. Les enfants bilingues présentant des difficultés langagières sont orientés en service de logopédie et les logopédistes ont pour mission de déceler si leurs difficultés sont de l'ordre d'un TDL ou de difficultés passagères dues à la situation de bilinguisme. Toutefois, un des problèmes majeurs des logopédistes en Suisse et dans le monde est actuellement de pouvoir réaliser une évaluation langagière des enfants bilingues qui soit précise et fiable. L'objectif de ce papier est alors de dresser un panorama des outils à disposition des cliniciens pour réaliser une évaluation langagière des enfants bilingues en limitant les erreurs de diagnostics. Nous évoquerons également les résultats de plusieurs recherches ayant utilisé un ou plusieurs de ces outils. Il faut noter que certains de ces instruments sont déjà largement utilisés en clinique et d'autres sont encore théoriques et méritent ainsi d'être testés sur le terrain. Nous terminerons par présenter brièvement une vignette clinique inspirée de cas réels afin d'illustrer la procédure et l'application de certains outils présentés dans l'article.

Les outils à disposition des cliniciens

Plusieurs instruments qui vont être présentés ont été créés dans le cadre de l'action COST intitulée 'Language Impairment in a Multilingual Society : Linguistic Patterns and the Road to Assessment' (Armon-Lotem et al., 2015) et portent le nom de LITMUS (Language Impairment Testing in Multilingual Settings). Il s'agit notamment d'une répétition de non-mots (Chiat, 2015), d'une répétition de phrases (Marinis et Armon-Lotem, 2015), d'un questionnaire parental (Tuller, 2015) et de tâches narratives (Gagarina et al., 2015). Nous verrons que ces outils LITMUS, largement utilisés en recherche, ne sont actuellement pas normés, mais peuvent tout de même présenter un intérêt en clinique à des fins d'évaluation complémentaire notamment. Ils peuvent être demandés aux auteurs à des fins de recherche.

Évaluation dans les deux langues

La façon recommandée pour évaluer l'enfant bilingue est de réaliser une évaluation dans les deux langues de l'enfant (ASHA, 2004 ; Kehoe, 2009 ; Peña et al., 2016) ce qui permet de réduire le taux de diagnostic erroné. Cependant, les résultats aux épreuves restent difficiles à interpréter puisque la plupart des tests possèdent des normes monolingues (McLeod et al., 2017; Scheidnes et Tuller, 2016). Scheidnes et Tuller (2016) ont utilisé des tests standardisés avec normes monolingues pour évaluer les capacités langagières des enfants bilingues successifs anglais-français âgés entre 6 et 12 ans vivant en France. Les auteurs observent que l'utilisation de telles évaluations est inadéquate car elle ne met pas en lumière les réelles capacités des enfants testés. Ces derniers obtiennent des scores déficitaires dans plusieurs sous-tests des deux langues. Les scores déficitaires aux sous-tests de la langue première, considérée comme dominante dans cette recherche, peuvent être attribués à une acquisition langagière non aboutie ou mise en suspens pour investir l'autre langue (Anderson, 2012). Les auteurs notent toutefois l'importance de porter son attention sur la langue première des enfants bilingues successifs pour

éviter de se focaliser sur la langue seconde et ainsi avoir un panorama langagier des deux langues. Cependant, il est nécessaire d'interpréter avec précaution les résultats obtenus aux tests avec normes monolingues (Scheidnes et Tuller, 2016). Certains auteurs mentionnent même que les résultats aux tests monolingues devraient être interprétés qualitativement plutôt que quantitativement (Lefebvre, 2014). Un outil conçu en France portant le nom 'ELAL d'Avicenne' (évaluation langagière des allophones et primo-arrivants) (Rezzoug et al., 2018) permet d'obtenir des profils de langues utilisées par l'enfant et non des scores de déviation par rapport à la norme. Cet outil permet d'évaluer la langue première des enfants bilingues successifs. Il a été validé auprès d'une population d'enfants dont la langue première était l'arabe, le soninké ou le tamoul. L'évaluation se réalise par un interprète, guidé par le clinicien. L'évaluation de la langue seconde est quant à elle réalisée à l'aide de tests francophones avec normes monolingues. La combinaison des scores des diverses épreuves des différentes langues permet de mettre en exergue des profils de langues. Le clinicien peut ainsi connaître si le langage se développe de façon harmonieuse ou si seule une des deux langues est investie. Enfin, un profil de vulnérabilité langagière signale si des difficultés sont présentes dans les deux langues de l'enfant (Rezzoug et al., 2018).

Anaya et al. (2018) se sont penchés sur la façon d'analyser les productions des enfants dans leur deux langues à un test de lexique expressif (les auteurs ont utilisé la tâche de dénomination du test bilingue 'BESA' de Peña et al., 2018). L'objectif des auteurs était de déterminer le type d'analyse le plus sensible et spécifique pour analyser les productions langagières des enfants d'âge scolaire bilingues espagnol-anglais et également de différencier adéquatement les enfants avec et sans TDL. Les auteurs relatent qu'il est recommandé d'utiliser ce qu'ils nomment l'*analyse conceptuelle* (= '*conceptual scoring*') qui consiste à compter le nombre d'items connu par les enfants dans leurs différentes langues. Ce procédé permet de réduire les différences lexicales souvent obtenues par les enfants bilingues et monolingues aux

tâches de dénomination. Toutefois, les auteurs notent qu'il est préférable d'effectuer une administration de la tâche par langue à des moments différents et ensuite de comptabiliser le nombre total d'items produits correctement, c'est ce qu'ils appellent l'*analyse conceptuelle croisée* (= '*across-test conceptual scoring*'). En effet, lorsque les deux langues sont testées simultanément, il n'est pas aisé pour les enfants bilingues d'accéder aux deux langues et de faire preuve, entre autres, de capacité d'inhibition. Les résultats sont plus élevés en utilisant la procédure croisée tant pour le groupe avec un DT et avec un TDL.

Adaptation du seuil de pathologie

Thordardottir (2015b) propose que grâce aux données récoltées concernant les langues et les taux d'exposition de l'enfant (ex. via un questionnaire parental, voir 2.3), le logopédiste puisse adapter les critères de déviation à la norme des tests monolingues. Dans une situation monolingue, Thordardottir (2015b) suggère d'utiliser le critère de déviation de la norme de -1,25 ET. Dans la situation de bilinguisme, plusieurs cas de figures peuvent se présenter. Si l'enfant possède un bilinguisme équilibré (50% d'exposition à une langue et 50% d'exposition dans l'autre), une évaluation monolingue dans une des deux langues de l'enfant peut se réaliser en fixant le seuil de pathologie entre -1,75 et -2 ET. Si l'enfant possède une langue dominante avec un taux de 75% d'exposition et de 25% d'exposition dans sa langue minoritaire, une évaluation dans la langue dominante est ainsi privilégiée avec une adaptation des seuils entre -1,5 et -1,75 ET. Toutefois, si l'évaluation ne peut être effectuée que dans la langue minoritaire ayant 25% d'exposition, une adaptation des seuils va s'opérer entre -2,25 et -2,5 ET. Malgré l'adaptation des seuils, il faut être attentif au fait que ce type d'évaluation ne prend en considération que les capacités et les difficultés de l'enfant dans une de ses deux langues et peut ainsi conduire à des erreurs de diagnostic (Hasson et Botting, 2010).

Utilisation des questionnaires parentaux

L'utilisation d'un questionnaire parental semble être un outil prometteur et complémentaire dans le contexte de l'évaluation des enfants bilingues. Il fournit des informations sur la qualité et la quantité d'input reçu dans les langues, sur le développement langagier et permet de relever la présence de facteurs de risque d'un TDL (Boerma et Blom, 2017), tels que l'apparition tardive des premiers mots et de la combinatoire, des difficultés langagières dans l'entourage familial ainsi que des capacités langagières considérées comme faibles par les parents (Tuller, 2015). Boerma et Blom (2017) s'intéressent à deux facteurs du questionnaire PabiQ (Tuller, 2015), soit l'apparition tardive des premiers mots ainsi que l'inquiétude parentale quant au développement langagier d'enfants monolingues hollandais et bilingues hollandais-turc/marocains âgés de 5-6 ans. Elles observent que l'apparition tardive des premiers mots constitue bel et bien un facteur de risque important de développer un TDL. Toutefois concernant l'inquiétude parentale, elles observent que les parents des enfants bilingues se questionnent régulièrement sur le développement langagier de leur enfant et ils semblent parfois inquiets alors qu'aucune difficulté n'est attestée. Elles notent ainsi que dans leur recherche, l'inquiétude parentale semble être un indicateur de trouble plus robuste chez une population de parents d'enfants monolingues que bilingues. Tuller et al. (2015) observent que les réponses parentales au questionnaire PabiQ sont en accord avec les résultats aux tests standardisés d'enfants bilingues français-arabes âgés entre 6 et 12 ans avec un DT et avec un TDL. En effet, les parents semblent estimer les compétences de leurs enfants de façon relativement fiable. Les auteurs relatent également que les données récoltées concernant les facteurs de risque ont renforcé la validité des résultats obtenus aux tests standardisés.

D'autres types de questionnaire parental peuvent être utilisés. Il s'agit d'inventaires de la communication tels que les 'MacArthur communicative development inventories' de Fenson et collaborateurs (2007) que les parents doivent compléter afin d'obtenir un aperçu des gestes,

du lexique et des phrases compris et produits par l'enfant. Ces outils sont intéressants sur le plan clinique et existent dans 60 langues (O'Toole et al., 2017). Toutefois, ils ne s'étendent pas au-delà de 36 mois, ce qui permet ainsi de ne déceler que les parleurs tardifs (O'Toole et al., 2017). D'autres inventaires parentaux existent, nous pouvons citer, entre autres, le 'DLPF' (= nouvel outil pour l'évaluation du développement du langage de production en français de Bassano et al., 2005) qui s'adresse aux parents d'enfants parlant le français et qui a l'avantage de s'étendre jusqu'à la quatrième année de vie. O'Toole et al. (2017) ont utilisé les questionnaires 'McArthur' pour tester le vocabulaire connu par des enfants bilingues âgés entre 24 et 36 mois parlant différentes combinaisons de langue. Les auteurs observent que les enfants obtiennent des meilleurs scores au questionnaire testant leur langue première que leur langue seconde, ce qui rejoint les résultats des enfants bilingues aux tests standardisés avec normes monolingues (Boerma et Blom, 2017). Hansen et al. (2019) observent plusieurs corrélations entre les observations faites via le questionnaire parental et les scores aux évaluations standardisées. O'Toole et al. (2017) nous signalent qu'il n'est pas toujours aisé pour certains parents de milieux défavorisés de répondre de façon fidèle aux questions posées. De Houwer et al. (2005) recommandent ainsi que plusieurs adultes effectuent cette évaluation des compétences afin d'obtenir une plus grande fiabilité dans les résultats relevés.

Utilisation des tâches de répétition

Pour éviter les biais liés à l'évaluation des enfants bilingues, il existe aussi la possibilité d'utiliser des tâches qui ne sont pas ou peu affectées par les connaissances préalables ou l'expérience de l'enfant (Campbell et al., 1997). L'échec aux tâches de répétition de phrases et de non-mots³ est actuellement reconnu comme étant un marqueur de TDL chez une population

³ Nous allons utiliser de manière synonyme les termes de 'non-mots' et de 'logatomes' tous deux étant considérés comme du matériel verbal dépourvu de lexicalité.

monolingue de différentes langues (Chiat, 2015 ; Conti-Ramsden, 2003). La présence d'erreurs dans la répétition de non-mots reflète des difficultés en mémoire à court-terme verbale et de précision phonologique (Ferré et Dos Santos, 2015 ; Li'el et al., 2019). La répétition de phrases est, quant à elle, considérée comme une tâche complexe impliquant des représentations linguistiques et nécessitant des capacités mémorielles (Vang Christensen Rikke, 2019).

La tâche de répétition de non-mots a l'avantage de pouvoir être utilisée avec des enfants dès le plus jeune âge (Guiberson et Rodríguez, 2015). L'utilisation de logatomes permet de s'assurer qu'aucun des enfants n'ait déjà rencontré ces items et ainsi d'éviter les biais linguistiques (Chiat, 2015 ; Kapantzoglou et al., 2011). Toutefois, bien que les résultats à la tâche de répétition de non-mots soient moins affectés par les connaissances lexicales et syntaxiques préalables de l'enfant, les scores sont, tout de même, affectés par les spécificités de sa langue. Un enfant est donc plus à même de répéter des non-mots qui possèdent les mêmes caractéristiques phonologiques de sa propre langue. Utiliser une répétition de non-mots créée en respectant les règles phonotactiques d'une langue de l'enfant fournit ainsi des résultats controversés dans la littérature scientifique. En effet, plusieurs études (Boerma et al., 2015 ; Gutiérrez-Clellen et Simon-Cerejido, 2010 ; Kohnert et al., 2006) obtiennent des résultats ne permettant pas de juger la répétition de non-mots comme discriminante pour les enfants bilingues avec un TDL. En effet, les scores obtenus par les enfants sont dépendants de leur expérience langagière dans la langue de la répétition de non-mots. Les enfants sont donc plus à même de répéter les non-mots de la tâche créée en se basant sur la langue pour laquelle ils ont été plus exposés. Kohnert et al. (2006) suggèrent ainsi d'utiliser deux répétitions de non-mots créées dans les différentes langues de l'enfant. Boerma et al., (2015) incitent à privilégier une répétition de non-mots quasi-universelle. D'autres auteurs (de Almeida et al., 2016 ; Thordardottir et Brandeker, 2013) obtiennent des résultats de spécificité et sensibilité élevés pouvant justifier la tâche de répétition de non-mots comme étant discriminante. La tâche

LITMUS francophone utilisée par de Almeida et al. (2016) fournit des scores de sensibilité et de spécificité élevés et presque équivalents entre les groupes d'enfants monolingues français et bilingues français-portugais d'âge scolaire. Les différences quant aux résultats obtenus par les recherches peuvent, en partie, s'expliquer par des différences méthodologiques quant à la création des tâches (nature des stimuli et méthode de cotation, etc.) (Boerma et Blom, 2017). Afin d'éviter les biais liés à l'exposition aux langues et à leurs propriétés, la création d'une tâche LITMUS de répétition de non-mots quasi-universelle a été proposée par Chiat (2015). Celle-ci comprend 16 items ayant été construits en évitant les variations segmentales et phonotactiques de plusieurs langues, et en se basant sur des phonèmes fréquents des langues du monde. Boerma et al. (2015), entre autres, ont évalué la valeur clinique de cette tâche auprès d'enfants monolingues hollandais et bilingues ayant le hollandais comme langue seconde âgés entre 5 et 6 ans et les résultats semblent prometteurs puisque le bilinguisme n'affecte pas les performances à cette tâche. En se basant sur la tâche de Chiat (2015), Dos Santos et Ferré (2018) ont développé une tâche de répétition de non-mots comprenant des items dépendants au français et des items non-dépendants. Les auteurs obtiennent des différences significatives entre les enfants bilingues français-anglais/arabe d'âge scolaire avec un DT et ceux avec un TDL. Les différences sont aussi relevées pour les groupes monolingues. Toutefois, la tâche semble légèrement plus sensible et spécifique pour la population monolingue de l'étude que pour la population bilingue.

En ce qui concerne la répétition de phrases, les scores à cette tâche peuvent également être influencés par différents facteurs, comme dans le cas de la répétition de non-mots. C'est le taux d'exposition langagière qui est le plus souvent mentionné comme facteur d'influence aux scores de la répétition de phrases (Scheidnes, 2020). Par exemple, si la tâche est réalisée dans la langue minoritaire de l'enfant, celui-ci sera susceptible d'obtenir des scores moindres (Marinis et Armon-Lotem, 2015). Toutefois, d'autres auteurs tels que Fleckstein et al. (2018)

n'obtiennent pas ce résultat. Ces auteurs ont utilisé la tâche LITMUS de répétition de phrases en français auprès d'une population monolingue français et bilingue français-anglais/arabe âgée entre 5 et 8 ans avec et sans TDL. Ils obtiennent des différences significatives entre les enfants avec un TDL et ceux avec un DT tant pour le groupe monolingue que bilingue. De plus, les auteurs ne trouvent aucune corrélation entre les facteurs de développement bilingue (ex. âge d'exposition, taux d'exposition, etc.) et les scores obtenus à la tâche. Les auteurs font donc mention de l'atout diagnostic de la tâche de répétition de phrases francophones.

D'autres facteurs d'influence inhérents à la tâche sont aussi relevés comme pouvant faire varier les scores à la tâche tels que le lexique et les catégories de mots utilisées (Fitton et al., 2019). Il semble alors plus sûr de réaliser une évaluation au moyen de tâches de répétition de phrases dans les deux langues de l'enfant (Marinis et Armon-Lotem, 2015).

Utilisation des tâches narratives

L'évaluation langagière via une tâche narrative a l'avantage de pouvoir mettre en lumière un large éventail des capacités langagières du sujet en contexte (Gagarina et al., 2015). Cette tâche est riche au niveau clinique car elle fournit beaucoup de renseignements linguistiques, discursifs et pragmatiques au clinicien (Rezzonico et al., 2016). Les enfants monolingues avec un TDL présentent des difficultés pour raconter et comprendre une histoire (Tsimpli et al., 2016). Plus précisément, ils semblent avoir des difficultés dans la macrostructure qui réfère à l'organisation de la narration ainsi qu'à la pertinence des informations transmises (Boerma et Blom, 2017) ainsi que dans la microstructure comprenant les aspects linguistiques des narrations tels que la longueur moyenne des énoncés, la complexité syntaxique et le nombre de mots différents ayant été produits (Gagarina et al., 2015 ; Rezzonico et al., 2016). Dans le contexte de bilinguisme, les recherches suggèrent que les capacités narratives de macrostructure acquises dans la langue première peuvent être transférées à la langue seconde (Boerma et Blom,

2017). Par conséquent, des difficultés dans la macrostructure narrative semblent communes à tous les enfants monolingues et bilingues avec un TDL (Boerma et al., 2016). De même, les enfants bilingues avec un DT semblent ne pas avoir de difficultés dans la macrostructure et performant comme des enfants monolingues avec un DT. En comparaison aux capacités de macrostructure, celles liées à la microstructure seraient davantage influencées par l'input reçu (Boerma et al., 2016). Des difficultés dans la microstructure ne seraient donc pas présentes que chez les enfants avec un TDL mais seraient également observés chez des enfants bilingues n'ayant pas reçu assez d'input dans la langue testée. Govindarajan et Paradis (2019) relatent dans leur recherche⁴ que les enfants bilingues avec un DT qui ont plus d'expérience avec la langue testée et qui bénéficient d'un environnement langagier riche produisent des narrations de meilleure qualité. A contrario, l'exposition langagière n'aurait aucun impact sur les enfants avec un TDL. A noter qu'il existe des résultats contradictoires concernant l'utilisation de la macrostructure pour différencier les enfants bilingues avec un TDL et avec un DT. Certains auteurs ne trouvent pas de différences entre les enfants bilingues avec un TDL et avec un DT quant aux scores liés à la macrostructure (Altman et al., 2016). Les différences méthodologiques entre les études peuvent expliquer ces résultats divergents (profil des participants, tâches utilisées, méthode de cotation et condition de passation, etc.).

Un outil d'évaluation narrative pour les enfants bilingues a été conçu par Gagarina et al. (2012). Cet outil, nommé LITMUS MAIN (the Multilingual Assessment Instrument for Narratives) a été adapté dans plusieurs langues pour des enfants âgés entre 3 et 10 ans⁵. Il permet l'évaluation de plusieurs langues chez le même enfant dans les versants expressifs et réceptifs (Gagarina et al., 2012). Les tâches ont été créées en prenant en considération les aspects culturels des stimuli visuels. Des scores quant au rappel de l'histoire et de narration sont

⁴ Leur population était composée d'enfants bilingues ayant l'anglais comme langue seconde et étant âgés de 5 ans.

⁵ La version française s'intitule OMER (outil multilingue pour l'évaluation du récit) et a été adaptée par Haiden et Knapp (2012).

calculés. L'évaluation peut être réalisée dans chacune des deux langues en respectant un intervalle de 4 à 7 jours afin de réduire l'effet d'entraînement et de transfert (Gagarina et al., 2012). L'outil n'est pas encore normé à l'heure actuelle mais des recherches montrent son pouvoir discriminant chez une population monolingue et bilingue avec un TDL. Par exemple, Boerma et al. (2016) trouvent des scores moindres aux différents tâches du MAIN (tâches de génération d'histoire ainsi que de compréhension aux questions posées par l'expérimentateur) chez les enfants avec un TDL monolingues hollandais et bilingues avec un TDL ayant le hollandais comme langue seconde âgés entre 5 et 6 ans. Les enfants avec un TDL possèdent une faible compréhension narrative, produisent moins d'éléments discursifs. La compréhension et la production d'états mentaux liés aux personnages semblent très difficiles pour les enfants avec un TDL (Boerma et al., 2016), de même que l'expression des objectifs, des buts et des finalités de l'histoire (Altman et al., 2016). Dans l'étude de Boerma et al. (2016), les enfants bilingues performant comme les enfants monolingues et l'impact de l'input langagier n'a pas eu d'influence. Certains auteurs nous rendent attentifs au fait que lors d'une évaluation langagière, la tâche de narration ne doit pas être administrée seule (Boerma et Blom, 2017). En effet, en fonction de la/des cultures dont l'enfant est issu ainsi que de ses pratiques familiales (Sladkova et al., 2014), l'enfant peut être exposé à différents types de narration et/ou être focalisé majoritairement sur certaines parties des histoires (ex. insistance sur les aspects chronologiques, sur les personnages, etc.) (Barra et McCabe, 2013). Ainsi, ses compétences narratives vont en être influencées. De plus, si la narration est utilisée à des fins d'évaluation, les enfants provenant de cultures qui racontent des histoires de façon similaire que la culture de la société seront favorisés, créant un biais culturel (Gutiérrez-Clellen et Quinn, 1993).

Évaluation dynamique

Une autre alternative pour l'évaluation des enfants bilingues est la méthode d'évaluation dynamique. Le but de celle-ci est de connaître le potentiel d'apprentissage de l'enfant et non son niveau d'habileté dans une seule situation. Concrètement, l'évaluation dynamique consiste à faire apprendre des nouveaux mots, de nouvelles structures syntaxiques et narratives aux enfants (Kehoe, 2009). L'utilisation d'un langage inventé permet de s'assurer qu'aucun des enfants n'ait déjà rencontré les items, et, par conséquent, d'éviter les biais linguistiques (Kapantzoglou et al., 2012). Ce type d'études portent fréquemment le nom de *fast-mapping* qui réfère aux premières étapes d'un nouvel apprentissage. Les recherches ont mis en avant des différences quant à la capacité d'apprentissage des enfants avec et sans TDL. En effet, les enfants monolingues avec un TDL nécessitent de plus d'exposition à un mot pour l'acquérir que des enfants avec un DT de même âge, mais du même nombre d'exposition que des enfants avec un DT plus jeunes (Gray, 2003; Oetting et al., 1995). Kapantzoglou et al. (2012) se sont, quant à eux, intéressés aux capacités d'apprentissage des enfants d'âge préscolaire bilingues ayant l'espagnol comme langue dominante. Les auteurs observent également des difficultés d'apprentissage de nouveaux mots par les enfants bilingues avec un TDL malgré la présence d'un support apporté par l'adulte. En effet, les enfants avec un TDL ont besoin de plus d'exposition pour retenir les mots en comparaison avec des pairs avec un DT de même âge. D'autres auteurs se sont intéressés à de nouveaux apprentissages dans le domaine de la (morpho)syntaxe. Anderson (2001) a mis au point une tâche pour des enfants monolingues espagnols âgés entre 3 et 6 ans qui consiste en l'apprentissage d'une nouvelle règle morphologique. Il observe que les enfants avec un TDL ont plus de peine à la comprendre et à la produire versus des enfants avec un DT de même âge et plus jeunes. Pour différencier les enfants avec et sans TDL bilingues ayant l'espagnol comme langue première âgés entre 4 et 6 ans, Roseberry et Connell (1991) ont évalué leur capacité à apprendre une règle

morphosyntaxique. Ils ont utilisé un suffixe qui permet de nommer ‘une partie d’un objet’. La règle est d’abord expliquée aux enfants et ils doivent ensuite dénommer des images en utilisant adéquatement la règle enseignée. Le groupe d’enfants avec un DT obtient de meilleurs résultats pour apprendre la règle morphosyntaxique que les enfants bilingues avec un TDL.

Plusieurs recherches se sont aussi intéressées au domaine discursif. Les biais linguistiques ne peuvent être éradiqués dans ce domaine mais c’est le potentiel d’évolution de l’enfant qui est observé. Dans ce domaine, l’évaluation dynamique est utilisée pour aider et évaluer l’enfant dans l’élaboration de ses compétences. C’est ainsi une vision interactionniste qui est, la plupart du temps, privilégiée en s’inspirant de la théorie de Vygotski (1934/1985). L’intérêt de la narration est que chez tous les enfants, des capacités peuvent être renforcées (lexicales, syntaxiques, en lien avec la cohésion ou la planification narrative, etc.). Dans cette lignée, l’évaluation dynamique comprend trois phases : le prétest, la médiation ou enseignement et le posttest (Peña et al., 2014). Le potentiel d’apprentissage est donc pris en considération (Gutiérrez-Clellen et Peña, 2001). L’évaluateur compare ainsi les performances de l’enfant au moment du prétest et du posttest et observe également à quel point la médiation a nécessité des efforts de sa part. L’enfant qui peut faire usage des recommandations et modèles fournis lors du moment de médiation possède des stratégies cognitives et semble avoir des capacités langagières dans la norme. En effet, les enfants présentant un TDL ont des difficultés persistantes lors du posttest, montrent moins de gains de la situation de médiation (Peña et al., 2014) et nécessitent plus d’étayage de la part de l’adulte (Petersen et al., 2017). L’évaluation dynamique nécessite du temps lors du bilan. Par exemple, l’étude de Peña et al. (2014) s’est déroulée sur trois sessions d’environ une heure. Certains auteurs considèrent alors que les premières séances de la prise en charge logopédique peuvent parfois être encore considérées comme de l’évaluation et non comme de l’intervention dans le but d’observer le potentiel de

changement chez l'enfant (Hasson et Joffe, 2007). De plus, les aspects n'ayant pas évolué après la phase de médiation peuvent constituer des objectifs lors de la thérapie (Petersen et al., 2017).

Vignette clinique

Afin d'illustrer l'utilisation des différents outils disponibles dans le contexte d'évaluation logopédique bilingue, nous allons exposer le bilan logopédique de Vanessa, fillette de 5 ; 2 ans bilingue français-portugais, vivant en Suisse romande et qui est actuellement en 2^{ème} année primaire. Notre choix s'est porté sur cette enfant hypothétique en sachant que le bilinguisme français-portugais est le plus fréquent en Suisse romande (OFS, 2018) et dans les services de logopédie (Volpin et al., 2017).

Contexte logopédique

Sara est une logopédiste monolingue francophone. Elle rencontre beaucoup de familles issues de différentes cultures et fait parfois appel à des interprètes lorsque la famille ne parvient pas à communiquer ou comprendre en français. Pour avoir une représentation des langues parlées par les enfants en consultation, elle utilise souvent les fiches langues disponibles sur le site du CNRS (<https://lgidf.cnrs.fr/fiches-langues>) afin de connaître les caractéristiques des langues en comparaison à la langue française. Sara est donc consciente que certaines erreurs transitoires peuvent être dues à la situation de bilinguisme et ne relèvent pas nécessairement de la pathologie. Le français et le portugais sont deux langues romanes qui partagent beaucoup de similarités (Kihm, 2016). Toutefois, Sara reste vigilante dans ses interprétations.

Premier entretien et anamnèse

La famille de Vanessa est portugaise ; ses parents sont à l'aise avec le français et le portugais. Sara les interroge au sujet de son développement langagier. Elle s'appuie sur le

questionnaire parental PabiQ (Tuller, 2015) mais l'utilise oralement lors de l'anamnèse qu'elle effectue (Kehoe, 2009). La famille de Vanessa parle le français et le portugais à la maison. Vanessa est une enfant bilingue simultanée puisqu'elle a été exposée au portugais dès sa naissance par son entourage familial et au français au moment de l'entrée en crèche dès la première année de vie. Les premiers mots sont apparus précocement vers 10 mois. Par contre, la combinaison de deux mots est apparue plus tardivement vers 30 mois et des difficultés ont ensuite été observées. Les parents remarquent, dès le plus jeune âge, peu de diversité lexicale et des erreurs dans l'ordre des mots. Vanessa a fréquenté la crèche 3 jours et demi par semaine avant sa scolarisation. Depuis le début de sa scolarité, son taux d'exposition au français a augmenté. La mère de Vanessa mentionne même qu'elle est plus à l'aise en français. Le pourcentage d'exposition est estimé à 50% pour le français et 50% pour le portugais. Les parents sont inquiets concernant le développement langagier de Vanessa. Des difficultés sont observées dans les deux langues et un décalage par rapport aux enfants de l'entourage familial est noté. Les parents relèvent actuellement des difficultés quant à l'apprentissage de nouveaux mots et des productions de phrases incorrectes. La compréhension semble meilleure que la production selon les parents. Le frère aîné de Vanessa présente un TDL et est scolarisé dans une école spécialisée. L'enseignante de Vanessa est également inquiète, elle relève peu de progrès depuis l'entrée à l'école.

Plusieurs facteurs de risque sont donc relatés en faveur de la présence d'un TDL : difficultés dans l'entourage familial, inquiétude parentale, combinaison tardive de deux mots et compétences langagières considérées comme faibles par les parents (Tuller, 2015).

Évaluation logopédique

Puisque Sara ne parle et ne comprend pas le portugais, elle décide d'utiliser des outils monolingues francophones⁶ et d'adapter les seuils de pathologie en fonction du taux d'exposition aux langues de Vanessa. Cette dernière possède un taux d'exposition équilibré et Sara adapte ainsi les seuils de pathologie entre -1,75 et -2 ET (versus -1,25 ET pour des enfants monolingues) (Thordardottir, 2015b). Sara choisit également d'utiliser l'analyse conceptuelle des stocks lexicaux expressifs et réceptifs et de la compréhension orale de Vanessa avec l'aide des parents pour le portugais (Anaya et al., 2018). Elle utilise également deux tâches de répétition (non-mots et phrases) qui semblent moins sujettes à l'expérience préalable de l'enfant (Campbell et al., 1997). Enfin, elle termine son évaluation par une tâche narrative prometteuse pour les enfants bilingues (MAIN de Gagarina et al., 2012).

Phonologie

Des erreurs de parole sont présentes dans le discours spontané de Vanessa (ex.: [tɛnɛt]⁷ pour 'tablette'; [pakoler] pour 'parascolaire'). La tâche de répétition de logatomes fournit des scores à -4,39 ET. Ce sont surtout des erreurs de type substitution qui sont observées (ex. : [wadʃ] pour 'ladon' ; [rɛ̃poni] pour 'rinpoli', etc). La longueur est un critère de complexité (plus de difficultés à répéter les non-mots longs). Le score à cette tâche est clairement en dessous de la norme.

Selon la fiche langue du portugais (Kihm, 2016), la phonologie du français sur laquelle est basée la tâche ne devrait pas poser de problème aux lusophones.

⁶ Les tests francophones Exalang 3-6 (Helloin et Thibault, 2006) et Evalo 2-6 (Coquet et al., 2009) ont été utilisés.

⁷ Pour les productions de Vanessa, nous avons privilégié une transcription phonétique.

Lexique

Les lexiques actifs ont été évalués par une tâche de dénomination qui a été réalisée en prenant en considération les deux langues de Vanessa à des moments différents avec l'aide des parents (Anaya et al., 2018). Les scores totaux (verbes et substantifs réunis dans les deux langues) se situent à -2,6 ET. Vanessa possède peu de stock lexical quelle que soit la langue. L'ébauche phonémique n'est pas facilitatrice. Elle utilise parfois des circonlocutions à la place de l'item (ex. : [met *fi*] pour 'jupe' ; [couper pain] pour 'couteau'). Nous relevons aussi des paraphrasies sémantiques (ex. : [mouche] pour 'oiseau').

Les lexiques passifs sont évalués par une tâche de désignation en appliquant l'analyse conceptuelle (Anaya et al., 2018). Les mots se rapportant aux objets quotidiens sont connus en portugais et les mots utilisés à l'école sont connus en français. Sara observe donc une répartition des stocks lexicaux de Vanessa en fonction des domaines et des langues (Thordardottir, 2011). Les scores totaux pour les deux langues se situent à -0,5 ET n'étant pas dans la zone pathologique.

Syntaxe et morphosyntaxe

Ce domaine semble particulièrement difficile pour Vanessa. A une tâche de répétition de phrases, des scores à -4,29 ET sont relevés. Bien souvent, le sens des phrases est préservé lors de la répétition mais l'ordre des mots et les items attendus ne sont pas produits (ex. : la phrase 'la petite fille marche dans le jardin' est répétée [une *fi* aller jadẽ]).

Une tâche de compléments de phrases fournit des scores à -5,29 ET. Dans cette tâche, la compréhension du début de la phrase n'est pas toujours aisée par Vanessa puisque ses productions ne sont pas adaptées sémantiquement et syntaxiquement (ex : 'Le singe va écrire avec ... [dessin] '). Beaucoup de phrases agrammatiques sont aussi observées dans son discours.

D'autres informations sont relevées via une tâche de production de phrases et lors de productions spontanées (données qualitatives) : une fréquente absence de verbe (ex. [*tit* fille un doudou] pour 'la petite fille berce le doudou'), et l'ordre canonique n'est pas toujours respecté ([la pomme moi *mãze*] pour 'je mange la pomme'). Vanessa utilise le 'moi' mais pas encore le pronom personnel 'je'.

En prenant comme référence Kihm (2016), Sara sait que la structure canonique est identique en français et en portugais. D'ailleurs, la maman relève des difficultés dans l'ordre des mots également en portugais dès le plus jeune âge.

Une tâche de compréhension morphosyntaxique a été réalisée en prenant en compte les 2 langues de Vanessa avec l'aide des parents. En effet, Sara s'est inspirée du concept de l'analyse conceptuelle (Anaya et al., 2018) pour traiter ce domaine langagier. En prenant en compte les 2 langues, Vanessa obtient des scores à -1,3 ET (alors qu'en prenant que le français, les scores se situent à -2,6 ET), elle ne se situe donc pas dans la zone pathologique si nous appliquons une adaptation de seuils en fonction de ses taux d'exposition aux langues. Au niveau analytique, nous notons une meilleure compréhension syntaxique en portugais. Toutefois, nous notons que la longueur est un critère de complexité pour Vanessa. ; plus il y a d'éléments à traiter dans la phrase, plus elle échoue.

Discours

En raison des difficultés langagières dans les différentes sphères du langage, le discours produits par Vanessa à la tâche MAIN⁸ est difficilement compréhensible pour autrui. Vanessa parle beaucoup malgré ses difficultés mais peu d'éléments concernent l'histoire (scores de 6 sur 17). Aucune production d'états mentaux liés aux personnages n'est relevée et les buts et les finalités de l'histoire ne sont pas énoncés. Aux questions posées par Sara, Vanessa a de la

⁸ Utilisation de la version francophone nommée OMER (Haiden et Knapp, 2012).

difficulté à répondre ; les états internes des personnages ne semblent pas saisis et les réponses sont souvent peu en lien avec la narration (scores de 3 sur 10).

Discussion

Nous relevons que Vanessa présente des difficultés dans les différentes sphères du langage oral. L'ensemble des domaines expressifs est impacté malgré une adaptation des seuils de pathologie en fonction du taux d'exposition aux langues (Thordardottir, 2015b). En effet, en essayant de prendre en considération les deux langues de Vanessa, nous observons des difficultés phonologiques, lexicales, (morpho)syntaxiques et discursives qui semblent similaires à celles relevées dans les recherches portant sur le TDL en français (Paradis et Crago, 2010). Nous relevons également que la longueur est un critère de complexité dans les différentes tâches pour Vanessa nous faisant supposer que la mémoire à court-terme verbale soit impactée, comme l'étant fréquemment dans le cas des TDL (Ferré et Dos Santos, 2015 ; Li'el et al., 2019). Vanessa s'exprime volontiers malgré ses difficultés qui rendent son discours peu compréhensible pour son interlocuteur. Le versant réceptif semble meilleur lorsque nous prenons en considération les deux langues de Vanessa (Anaya et al., 2018). Les tâches de répétition qui semblent moins affectées par l'expérience préalable de l'enfant (non-mots et phrases) sont clairement échouées par Vanessa. En discutant avec les parents, nous notons également plusieurs facteurs de risque de TDL (cf. premier entretien et anamnèse ; Tuller et al., 2015 ; Tuller, 2015).

Les diverses observations sont réalisées en essayant de prendre en considération les deux langues de Vanessa ce qui laisse penser que les observations réalisées attestent de réelles difficultés langagières. Afin de pouvoir observer les capacités d'apprentissage de Vanessa, Sara propose une prise en charge logopédique sur la base de ces observations mais effectuera une évaluation dite dynamique au début de la prise en charge. Elle pourra ainsi évaluer son potentiel

d'apprentissage (Hasson et Joffe, 2007), en vue de pouvoir clairement déterminer si les difficultés observées émanent d'un véritable TDL.

Discussion générale et conclusion

Plusieurs outils ont été présentés qui montrent leur intérêt pour distinguer les enfants bilingues avec et sans TDL, y compris les enfants qui parlent des langues non partagées par les logopédistes. Ces outils ont été utilisés dans plusieurs recherches mettant en lumière parfois des résultats différents en raison de la grande hétérogénéité méthodologique entre les études (âge, langues, types de bilinguisme des participants, tâches et méthode de cotation utilisées, etc.). Certains de ces outils ont également pu être illustrés dans la vignette clinique de cet article. Toutefois, il s'agit ici de rappeler que les résultats issus de ces tâches doivent être interprétés avec précaution puisqu'ils émanent presque toujours d'épreuves normées pour des enfants monolingues (Scheidnes et Tuller, 2016). Même les tâches peu ou moins biaisées par l'expérience préalable de l'enfant peuvent parfois être sujettes au biais d'exposition (Boerma et al., 2015) et il semble donc préférable de les administrer dans les deux langues de l'enfant ou alors, dans le cas de la répétition de non-mots, d'avoir recours à une tâche quasi-universelle (Chiat, 2015). Parmi les outils proposés, nous supposons que c'est l'évaluation des deux langues qui peut être considérée comme la plus périlleuse en Suisse romande, puisque les langues maîtrisées par les logopédistes ne sont que peu similaires à celles des enfants (Volpin et al., 2017). De plus, l'utilisation d'une évaluation dite dynamique semble rare en clinique car elle est souvent couteuse en temps. Toutefois, comme évoqué dans la vignette clinique, il y a la possibilité de poursuivre l'évaluation au début de la prise en charge afin d'estimer le potentiel d'apprentissage de l'enfant (Hasson et Joffe, 2007).

Nous voyons au terme de notre illustration clinique que c'est l'utilisation d'outils de façon complémentaire qui nous a permis de comprendre au mieux les difficultés langagières de

la patiente et d'obtenir son profil langagier. Plusieurs auteurs (Boerma et Blom, 2017 ; Li'el et al., 2019 ; Paradis et al., 2013) nous montrent également l'intérêt de combiner plusieurs outils tels que le questionnaire parental, les répétitions de non-mots et phrases et les tâches narratives pour augmenter le pouvoir discriminant de l'évaluation. Par exemple, Boerma et Blom (2017) obtiennent une augmentation de la valeur discriminante en associant la tâche de répétition de non-mots quasi-universelle (Chiat, 2015) et une tâche narrative (MAIN de Gargarina et al., 2012) mais seulement chez le groupe d'enfants bilingues hollandais-turc/marocain. Paradis et al. (2013) s'étant intéressés aux apprenants anglophones possédant des langues premières diverses âgés de 5 ans avec et sans TDL, obtiennent une amélioration de la validité discriminante en associant plusieurs tests langagiers anglophones. Cette combinaison permettant de détecter 92% des enfants avec un TDL.

À noter que la vignette clinique présentée concerne un type de bilinguisme (simultané et équilibré) et une combinaison de langue fréquente en Suisse romande et ne se veut donc pas exhaustive sur le plan de la pratique logopédique en Suisse romande. De plus, la collaboration avec les parents de la patiente semblait aisée et a permis ainsi l'utilisation de certaines tâches nécessitant leur aide (ex. questionnaire parental et analyses conceptuelles). Toutefois, nous sommes conscients que ce n'est pas toujours possible d'agir ainsi.

Les tâches présentées au fil de cet article ont une réelle utilité clinique et la recherche à leur sujet est encore fondamentale afin d'apporter des nouveaux savoirs et pour tenter de normer ces instruments auprès d'une population bilingue malgré sa forte hétérogénéité. Une évaluation logopédique des enfants bilingues comprenant une combinaison des différents instruments et mesures proposés rend possible la diminution du taux de diagnostic erroné en clinique et permet ainsi de proposer des interventions aux enfants présentant un TDL. Il s'agira également de mettre sur pieds des études évaluant le pouvoir discriminant de ces outils et/ou d'une combinaison de ceux-ci auprès d'une population d'enfants bilingues en Suisse romande.

Schwob et Skoruppa, (soumis pour publication) ont notamment testé l'efficacité de plusieurs outils présentés dans cet article auprès d'une population bilingue français-portugais avec et sans TDL. Elles relatent l'intérêt de l'utilisation d'une répétition de non-mots et la faible qualité diagnostique d'autres outils testés. Davantage d'études doivent donc nécessairement être réalisées auprès de la population suisse au sujet de la validité de ces tâches. D'autres alternatives d'aide doivent également être développées afin de pouvoir soutenir les enfants présentant des difficultés transitoires dues à la situation de bilinguisme.

Bibliographie

- Altman, C., Goldstein, T., & Armon-Lotem, A. (2018). Vocabulary, Metalinguistic Awareness and Language Dominance Among Bilingual Preschool Children. *Frontiers in Psychology*, 9.
- Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English–Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, 37(1), 165-193. <https://doi.org/10.1017/S0142716415000466>
- Anaya, J., Peña, E., & Bedore, L. (2018). Conceptual Scoring and Classification Accuracy of Vocabulary Testing in Bilingual Children. *Lang Speech Hear Serv Sch*, 49(1), 85-97. https://doi.10.1044/2017_LSHSS-16-0081
- Anderson, R. T. (2012). First Language Loss in Spanish-Speaking Children. In B. A. Goldstein (éd.), *Bilingual Language Development and Disorders in Spanish-English Speakers*. Baltimore: Paul Brookes.
- Anderson, R. T. (2001). Learning an invented inflectional morpheme in Spanish by children with typical language skills and with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, 36(1), 1-19. <https://doi.org/10.1080/13682820118926>
- Armon-Lotem, S. (2012). Introduction : Bilingual children with SLI – the nature of the problem*. *Bilingualism: Language and Cognition*, 15(1), 1-4.
- Armon-Lotem, S., de Jong, J. & Meir, N. (2015). *Assessing Multilingual Children*. Bristol (Multilingual Matters). <https://doi.org/10.1017/S1366728911000599>
- ASHA (2004). *Bilingual Service Delivery*.

<https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589935225§ion=Overview>

- Barra, G., & McCabe, G. (2013). Oral Narrative Skills of Chilean Preschool Children. *Imagination, cognition and personality*, 32(4), 367-391. <https://dx.doi.org/10.2190/IC.32.4.d>
- Bassano, D., Labrell, D., Champaud C., Lemétayer, F., & Bonnet, P. (2005). Le DLPF : un nouvel outil pour l'évaluation du développement du langage de production en français. *Enfance*, 2(57).
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE : A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068-1080. <https://doi.org/10.1111/jcpp.12721>
- Boerma, T., & Blom, E. (2017). Assessment of bilingual children : What if testing both languages is not possible? *Journal of Communication Disorders*, 66, 65-76. <https://doi.org/10.1016/j.jcomdis.2017.04.001>
- Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment : Implications for clinical practice. *International Journal of Language & Communication Disorders*, 51(6), 626-638. <https://doi.org/10.1111/1460-6984.12234>
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A Quasi-Universal Nonword Repetition Task as a Diagnostic Tool for Bilingual Children Learning Dutch as a Second Language. *Journal of Speech, Language, and Hearing Research*, 58(6), 1747-1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058
- Breault, C., Béliveau, M.-J., Labelle, F., Valade, F., & Trudeau. (2019). Le trouble développemental du langage (TDL) : Mise à jour interdisciplinaire. *Neuropsychologie*

- Clinique et Appliquée, 3, 46-63. https://oaaqvousinforme.com/wp-content/uploads/2019/10/NCACN-Volume3_final-coup-de-chapeau.pdf
- Camilleri, B., & Law, J. (2007). Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology*, 9(4), 312-322. <https://doi.org/10.1080/14417040701624474>
- Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing Bias in Language Assessment: Processing-Dependent Measures. *JSLHR*, 40, 519-525.
- Centre National de la Recherche Scientifique (2019). *Fiches langues*. <https://lgidf.cnrs.fr/fiches-langues>
- Chiat, S. (2015). Non-Word Repetition. In S. Armon-Lotem, K. de Jong & N. Meir (ed.), *Assessing Multilingual Children*. (pp. 95-122). Bristol (Multilingual Matters).
- Coquet, F., Roustit, J., & Ferrand, P. (2009). *Evalo 2-6*. Isbergues:OrthoEdition
- Conti-Ramsden, G. (2003). Processing and Linguistic Markers in Young Children With Specific Language Impairment (SLI). *Journal of Speech, Language, and Hearing Research*, 46(5), 1029-1037. [https://doi.org/10.1044/1092-4388\(2003/082\)](https://doi.org/10.1044/1092-4388(2003/082))
- De Almeida, L., Ferré, S., Morin, E., Prévost, P., Dos Santos, C., Tuller, L., & Zebib, R. (2016). L'identification d'enfants bilingues avec Trouble Spécifique du Langage en France. *SHS Web of Conference*, 27, 1-16. <https://doi.org/10.1051/shsconf/20162710005>
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Bristol-Buffalo-Toronto (MM Textbooks).
- De Houwer, A., Bornstein, M. H., & Leach, D. B. (2005). Assessing early communicative ability: A cross-reporter cumulative score for the MacArthur CDI. *Journal of Child Language*, 32(4), 735-758. <https://doi.org/10.1017/S0305000905007026>
- De Weck, G. & Marro, P. (2010). *Les troubles du langage chez l'enfant. Description et évaluation*. Paris (Masson).

- Dos Santos, C., & Ferré, S. (2018). A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition*, 25(1), 58-71
<https://doi.org/10.1080/10489223.2016.1243692>
- Estienne, F., & Vander Linden, F. (2014). Les orthophonistes face au multilinguisme – résultats d'une enquête. In H. A. Bijleveld, F. Estienne & F. Vander Linden (éds.), *Multilinguisme et orthophonie réflexions à l'heure de l'Europe* (pp. 91-129). Issy-les-Moulineaux (Elsevier Masson).
- Fabiano-Smith, L., & Goldstein, B. (2010). Phonological Acquisition in Bilingual Spanish-English Speaking children. *Journal of Speech, Language, and Hearing Research*, 53, 160-178.
- Fenson, L., V. A. Marchman, D. J. Thal, P. S. Dale, J. S Reznick, and E. Bates. (2007). *The MacArthur-Bates Communicative Development Inventories* (2e éd). Baltimore, MD: Brookes
- Ferré, S., & Dos Santos, C. (2015). Comment évaluer la phonologie des enfants bilingues ? *Lidil. Revue de linguistique et de didactique des langues*, 51, 11-34.
<https://doi.org/10.4000/lidil.3678>
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., & Zebib, R. (2018). How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition*, 25(1), 85-101. <https://doi.org/10.1080/10489223.2016.1192635>
- Fitton, L., Hoge, R., Petscher, Y., & Wood, C. (2019). Psychometric Evaluation of the Bilingual English–Spanish Assessment Sentence Repetition Task for Clinical Decision Making. *Journal of Speech, Language, and Hearing Research*, 62(6), 1906-1922.
https://doi.org/10.1044/2019_JSLHR-L-18-0354
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balcuiniene, I., Bohnacker, U., & Walters, J. (2015). Assessment of Narrative Abilities in Bilingual Children. In S.

- Armon-Lotem, K. de Jong & N. Meir (ed.), *Assessing Multilingual Children*. (pp. 243-277). Bristol (Multilingual Matters).
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balciuniene, S., Bohnacker, U., & Walters, J. (2012). MAIN Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, 56.
- Govindarajan, K., & Paradis, J. (2019). Narrative abilities of bilingual children with and without Developmental Language Disorder (SLI) : Differentiation and the role of age and input factors. *Journal of Communication Disorders*, 77, 1-16.
<https://doi.org/10.1016/j.jcomdis.2018.10.001>
- Gray, S. (2003). Word-Learning by Preschoolers With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 46(1), 56-67.
[https://doi.org/10.1044/1092-4388\(2003/005\)](https://doi.org/10.1044/1092-4388(2003/005))
- Grégoire, J. (2006). Propriétés métriques des tests de langage et leurs implications pratiques. In F. Estienne & B. Piérart (éds.), *Les bilans de langage et de voix*. Paris (Masson).
- Grosjean, F. (1993). Le bilinguisme et le biculturalisme. Essai de définition. *TRANEL*, 19, 13-41.
- Grosjean, F. (2015). *Parler plusieurs langues. Le monde des bilingues*. Paris (Albin Michel).
- Guiberson, M., & Rodríguez, B. L. (2015). Nonword Repetition in Spanish-Speaking Toddlers with and without Early Language Delays. *Folia Phoniatica et Logopaedica*, 67(5), 253-258. <https://doi.org/10.1159/000442745>
- Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2010). Using Nonword Repetition Tasks for the Identification of Language Impairment in Spanish-English-Speaking Children: Does the Language of Assessment Matter? *Learning Disabilities Research & Practice*, 25(1), 48-58. <https://doi.org/10.1111/j.1540-5826.2009.00300.x>

- Gutiérrez-Clellen, V. F., & Pen A, E. (2001). Dynamic Assessment of Diverse Children: A Tutorial. *Language, Speech, and Hearing Services in Schools*, 32(4), 212-224. [https://doi.org/10.1044/0161-1461\(2001/019\)](https://doi.org/10.1044/0161-1461(2001/019))
- Gutiérrez-Clellen, V. F., & Quinn, B. (1993). Assessing Narratives of Children From Diverse Cultural/Linguistic Groups. *Language, Speech, and Hearing Services and Schools*, 24, 2-9.
- Hansen, P., Łuniewska, M., Simonsen, H. G., Haman, E., Mieszkowska, K., Kołak, J., & Wodniecka, Z. (2019). Picture-based vocabulary assessment versus parental questionnaires: A cross-linguistic study of bilingual assessment methods. *International Journal of Bilingualism*, 23(2), 437-456. <https://doi.org/10.1177/1367006917733067>
- Hasson, N., & Botting, N. (2010). Dynamic assessment of children with language impairments : A pilot study: *Child Language Teaching and Therapy*. <https://doi.org/10.1177/0265659009349982>
- 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>
- Haiden, M., & Knapp, A. (2012). OMER Outil linguistique pour l'évaluation du récit. *ZAS Papers in Linguistics*, 56.
- Helloin, M. C., & Thibault, M. P. (2006). *Exalang 3-6*. Grenade (Happyneuron).
- Kapantzoglou, M., Restrepo, M., & Thompson, M. (2011). Dynamic Assessment of Word Learning Skills : Identifying Language Impairment in Bilingual Children. *Language, speech, and hearing services in schools*, 43, 81-96. [https://doi.org/10.1044/0161-1461\(2011/10-0095\)](https://doi.org/10.1044/0161-1461(2011/10-0095))
- Kehoe, M. (2009). Plurilinguisme et logopédie : le défi de l'évaluation, *Langage et pratiques*, 44, 69-82.

- Keller, K., Troesch, L. et Grob, A. (2015). A large receptive-expressive gap in bilingual children. *Frontiers in Psychology*, 6(1284), 1-9.
<https://doi.org/10.3389/fpsyg.2015.01284>
- Kihm, A. (2016). Fiche langue le portugais européen. Site du CNRS.https://lgidf.cnrs.fr/sites/lgidf.cnrs.fr/files/images/PORTUGAIS_24_06_19_A4.pdf
- Kohnert, K. (2010). Bilingual Children with Primary Language Impairment : Issues, Evidence and Implications for Clinical Actions. *Journal of communication disorders*, 43(6), 456-473. <https://doi.org/10.1016/j.jcomdis.2010.02.002>
- Kohnert, K., Windsor, J., & Yim, D. (2006). Do Language-Based Processing Tasks Separate Children with Language Impairment from Typical Bilinguals? *Learning Disabilities Research & Practice*, 21(1), 19-29. <https://doi.org/10.1111/j.1540-5826.2006.00204.x>
- Law, J., McKean, C., Murphy, C.-A., & Thordardottir, E. (2019). *Managing Children with Developmental Language Disorder*. New-York (Routledge).
- Leclercq, A.-L., & Leroy, S. (2012). Introduction générale à la dysphasie : caractéristiques linguistiques et approches théoriques. In C. Maillart & M. A. Schelstraete (éds), *Les dysphasies : de l'évaluation à la rééducation* (pp. 5-25). Issy-les-Moulineaux : Elsevier Masson.
- Lefebvre, F. (2014). Multilinguisme : quels enjeux pour l'orthophonie. In H. A. Bijleveld, F. Estienne & F. Vander Linden (ed.), *Multilinguisme et orthophonie*. Issy les Moulineaux (Elsevier Masson).
- Leonard, L. (2014). *Children with Specific Language impairment* (2e éd). Cambridge (A Bradford book).

- Li'el, N., Williams, C., & Kane, R. (2019). Identifying developmental language disorder in bilingual children from diverse linguistic backgrounds. *International Journal of Speech-Language Pathology*, 21(6), 613-622. <https://doi.org/10.1080/17549507.2018.1513073>
- Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, K. de Jong & N. Meir (ed.), *Assessing Multilingual Children*. (pp. 95-125). Bristol (Multilingual Matters).
- McLeod, S., Verdon, S., Baker, E., Ball, M. J., Ballard, E., David Avivit, B., Bernhardt B. M., Bérubé, D., Blumenthal, M., Bowen, C., Brosseau-Lapré, F., Bunta, F., Crowe, K., Cruz-Ferreira Madalena, Davis Barbara, Fox-Boyer Annette, Gildersleeve-Neumann Christina, Grech, H., Goldstein, B., ... Zharkova, N. (2017). Tutorial : Speech Assessment for Multilingual Children Who Do Not Speak the Same Language(s) as the Speech-Language Pathologist. *American Journal of Speech-Language Pathology*, 26(3), 691-708. https://doi.org/10.1044/2017_AJSLP-15-0161
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder : Evidence from a population study. *Journal of Child Psychology and Psychiatry*, 57(11), 1247-1257. <https://doi.org/10.1111/jcpp.12573>
- Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick Incidental Learning (QUIL) of Words by School-Age Children With and Without SLI. *Journal of Speech, Language, and Hearing Research*, 38(2), 434-445. <https://doi.org/10.1044/jshr.3802.434>
- OFS- Office Fédéral de la statistique (2018). Langues. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/langues-religions/langues.html>
- O'Toole, C., Gatt, D., Hickey, T. M., Miękisz, A., Haman, E., Armon-Lotem, S., Rinker, T., Ohana, O., Santos, C. dos, & Kern, S. (2017). Parent report of early lexical production

in bilingual children: A cross-linguistic CDI comparison. *International Journal of Bilingual Education and Bilingualism*, 20(2), 124-145.

<https://doi.org/10.1080/13670050.2016.1179258>

Paradis, J. (2019). *English Second Language Acquisition from Early Childhood to Adulthood: The Role of Age, First Language, Cognitive, and Input Factors*.

Paradis, J., Genesee, F., & Crago, M.B. (2011). *Dual Language Development & Disorders: A handbook on bilingualism and second language learning* (2nd Edition). Baltimore (Brookes).

Paradis, J. (2007). Becoming bilingual. In P. Auer & L. Wei (ed.), *Handbook of Multilingualism and Multilingual Communication*, (pp. 15-44). Berlin (Mouton de Gruyter).

Paradis, J., Schneider, P., & Sorenson Ducan, T. (2013). Discriminating Children With Language Impairment Among English-Language Learners From Diverse First-Language Backgrounds. *Journal of Speech, Language, and Hearing Research*, 56, 971-981.
[https://doi.org/10.1044/1092-4388\(2012/12-0050\)](https://doi.org/10.1044/1092-4388(2012/12-0050))

Paradis, J., & Crago, M.B. (2001). The Morphosyntax of Specific Language Impairment in French: an extended Optional Default Account. *Language Acquisition*, 9(4), 269-300. https://doi.org/10.1207/S15327817LA0904_01

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2018). *Bilingual English- Spanish Assessment (BESA)*. Baltimore (Paul Brookes).

Peña, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders*, 51(2), 192-202.
<https://doi.org/10.1111/1460-6984.12199>

Peña, E. D., Gillam Ronald B., & Bedore Lisa M. (2014). Dynamic Assessment of Narrative Ability in English Accurately Identifies Language Impairment in English Language

- Learners. *Journal of Speech, Language, and Hearing Research*, 57(6), 2208-2220.
https://doi.org/10.1044/2014_JSLHR-L-13-0151
- Petersen, D. B., Chanthongthip, H., Ukrainetz, T. A., Spencer, T. D., & Steeve, R. W. (2017). Dynamic Assessment of Narratives : Efficient, Accurate Identification of Language Impairment in Bilingual Students. *Journal of Speech, Language, and Hearing Research*, 60(4), 983-998. https://doi.org/10.1044/2016_JSLHR-L-15-0426
- Rezzonico, S., Goldberg, A., Mak Katy, K.-Y., Yap, S., Milburn, T., Belletti, A., & Girolametto, L. (2016). Narratives in Two Languages : Storytelling of Bilingual Cantonese–English Preschoolers. *Journal of Speech, Language, and Hearing Research*, 59(3), 521-532. https://doi.org/10.1044/2015_JSLHR-L-15-0052
- Rezzoug, D., Bennabi-Bensekhar, M., & Moro, M. R. (2018). L'ELAL D'Avicenne. L'autre, cliniques, cultures et sociétés, 19(2).
- Rice, M. L. (2013). Language growth and genetics of specific language impairment. *International Journal of Speech-Language Pathology*, 15(3), 223-233.
<https://doi.org/10.3109/17549507.2013.783113>
- Roseberry, C. A., & Connell, P. J. (1991). The use of an invented language rule in the differentiation of normal and language-impaired Spanish-speaking children. *Journal of Speech & Hearing Research*, 34(3), 596-603. <https://doi.org/10.1044/jshr.3403.596>
- Ruiz-Felter, R., Cooperson, S., Bedore, L., & Peña, E. (2016). Influence of current input-output and age of first exposure on phonological acquisition in early bilingual Spanish-English-speaking kindergarteners. *Int J Lang Commun Disord*, 51(4), 368-383.
<https://doi.org/10.1111/1460-6984.12214>
- Sladkova, J., McCabe, A., & Cano, I. (2014). “ I don’t like speaking Spanish:” delayed narration in children of Latino immigrants. *Global Journal of Community Psychology Practice*, 5(2), 1-10.

- Scheidnes, M. (2020). Sentence repetition and non-word repetition in early total French immersion. *Applied Psycholinguistics*, 41(1), 107-131.
<https://doi.org/10.1017/S0142716419000420>
- Scheidnes, M., & Tuller, L. (2016). Assessing successive bilinguals in two languages : A longitudinal look at English-speaking children in France. *Journal of Communication Disorders*, 64, 45-61. <https://doi.org/10.1016/j.jcomdis.2016.10.001>
- Schwob, S., et Skoruppa, K. (Soumis pour publication). L'identification du trouble développemental du langage oral chez les enfants bilingues : les ressources à disposition des cliniciens. *Revue canadienne d'orthophonie et d'audiologie*.
- Simon, A., Ferradji, T., Serre, G., & Moro, M. R. (2010). On ne salue que les parents, les enfants sont toujours inclus. *Parcours des enfants tamouls ayant des troubles du langage*. *L'Autre*, Volume 11(1), 40-48.
- Thomas, S., Schulz, J., & Ryder, N. (2019). Assessment and diagnosis of Developmental Language Disorder: The experiences of speech and language therapists. *Autism & Developmental Language Impairments*, 4, 2396941519842812.
<https://doi.org/10.1177/2396941519842812>
- Thordardottir, E. (2015a). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology*, 17(2), 97-114.
<https://doi.org/10.3109/17549507.2014.923509>
- Thordardottir, E. (2015b). Proposed Diagnostic Procedures for Use in Bilingual and Cross-Linguistic Contexts. In S. Armon-Lotem, K. de Jong & N. Meir (ed.), *Assessing Multilingual Children*. (pp. 331-359). Bristol (Multilingual Matters)
- 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>

- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and Specificity of French Language and Processing Measures for the Identification of Primary Language Impairment at Age 5. *Journal of Speech, Language, and Hearing Research*, 54(2), 580-597. [https://doi.org/10.1044/1092-4388\(2010/09-0196\)](https://doi.org/10.1044/1092-4388(2010/09-0196))
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426-445. <https://doi.org/10.1177/1367006911403202>
- Thordardottir, E., & Namazi, M. (2007). Specific Language Impairment in French- Speaking Children: Beyond Grammatical Morphology. *Journal of Speech, Language, and Hearing Research*, 50, 698-715. [https://doi.org/10.1044/1092-4388\(2007/049\)](https://doi.org/10.1044/1092-4388(2007/049))
- Tsimpli, I. M., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, 37(1), 195-216. <https://doi.org/10.1017/S0142716415000478>
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., Santos, C. dos, Ibrahim, L. A., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, 53(4), 888-904. <https://doi.org/10.1111/1460-6984.12397>
- Tuller, L., (2015). Clinical Use of Parental Questionnaires in Multilingual Contexts. In S. Armon-Lotem, K. de Jong & N. Meir (ed.), *Assessing Multilingual Children*. (pp. 301-331). Bristol (Multilingual Matters).
- Tuller, L., Abboud, L., Ferré, S., Fleckstein, A., Prévost, P., Dos Santos, C., Scheidnes, L., & Zebib, R. (2015). Specific Language Impairment and Bilingualism: Assembling the Pieces. In C. Hamann & E. Ruigendijk (ed.), *Language Acquisition and Development*. (pp. 533-567). Cambridge (Cambridge Scholars Publishing).

- Vang Christensen, R. (2019). Sentence Repetition: A Clinical Marker for Developmental Language Disorder in Danish. *Journal of Speech, Language, and Hearing Research*, 62(12), 4450-4463. https://doi.org/10.1044/2019_JSLHR-L-18-0327
- Volpin, L., Rezzonico, S., & de Weck, G. (2017). Bilingual language assessment: a survey of speech-language therapy caseloads declared practices in French-speaking Switzerland. Paper presented at IASCL, Lyon, France, 18 July 2017.
- Vygotski, L. S. (1934/1985). *Pensée et Langage*. Paris (Editions sociales).
- Werker, J., Byers-Heinlein, K., & Fennell, C. (2009). Bilingual beginnings to learning words. *Philosophical Transactions of the Royal Society*, 364, 3649-3663. doi:10.1098/rstb.2009.0105
- Williams, C. J., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology*, 14(3), 292-305. <https://doi.org/10.3109/17549507.2011.636071>

Partie II : Problématique et méthode

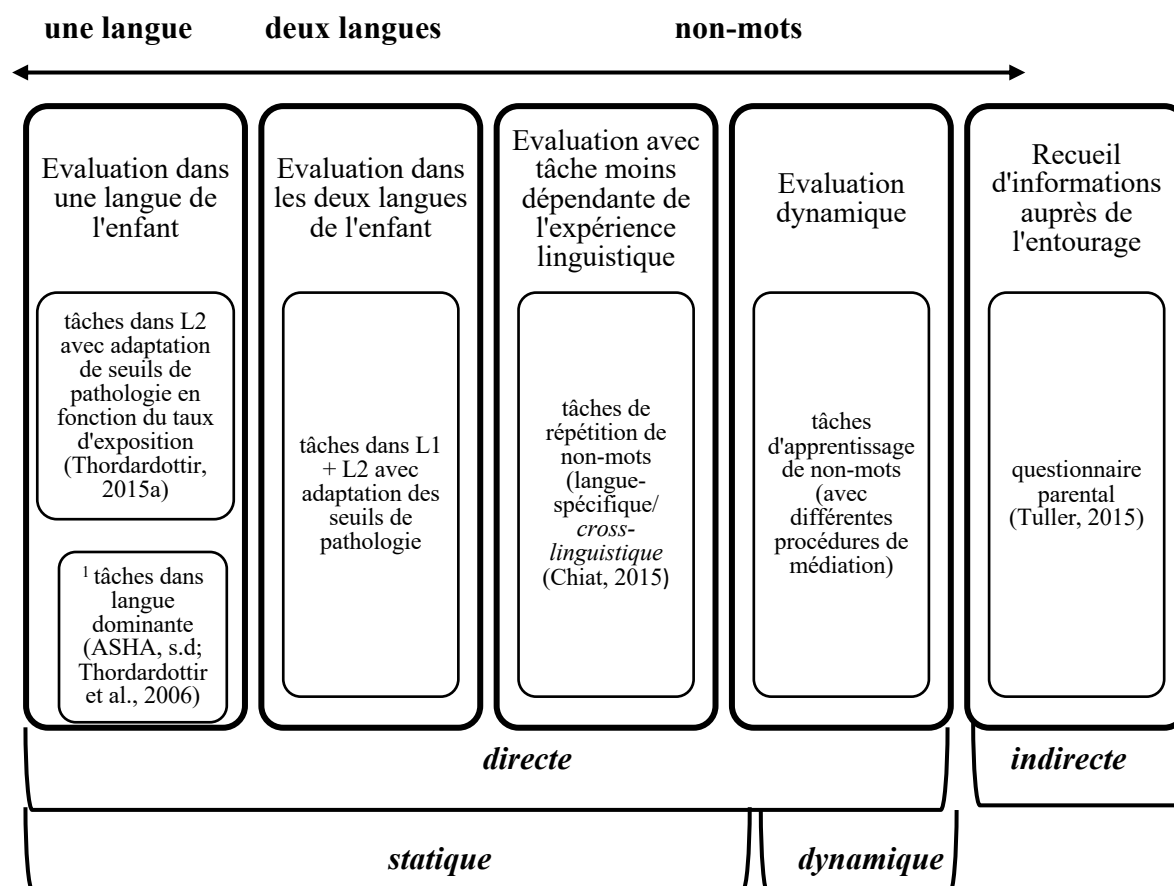
2. Problématique

Dans cette section, nous allons revenir sur certains éléments théoriques en mettant en lumière les lacunes présentes dans la littérature qui soutiennent notre intérêt de recherche et qui nous permettront d'exposer les aspects novateurs de celle-ci. Au point 2.1, nous rapporterons notre questionnement principal ainsi que plusieurs sous-questions de recherche. Finalement, nous exposerons nos hypothèses théoriques au moyen de plusieurs sous-sections.

En regard des données relevées par la littérature, nous constatons qu'il existe de nombreuses difficultés quant à l'évaluation du langage des enfants bilingues en clinique logopédique et que cette situation représente actuellement un des plus grands défis pour les logopédistes (Armon-Lotem, 2012). Nous avons évoqué dans notre premier article (Schwob, 2020), plusieurs pistes qui ont été investiguées dans la littérature pour résoudre ce problème. Nous en avons effectué une synthèse dans la Figure 3 présente ci-dessous. Nous allons commenter celle-ci dans les lignes qui la suivent.

Figure 3

Synthèse des outils et méthodes émanant de la littérature pour évaluer les enfants bilingues



Note. Abréviations : L1 : langue première de l'enfant ; L2 : langue seconde parlée par l'enfant bilingue ; ASHA : Amercian Speech-Language-Hearing Association.

¹L'évaluation dans la langue dominante n'a été que peu abordée dans la Partie I mais des informations supplémentaires seront fournies dans la Partie III (notamment dans l'article au point 4).

Cette Figure 3 rappelle les modes d'évaluation qui peuvent être utilisés par les logopédistes, en n'évoquant que ceux qui seront traités dans la suite de la thèse, et qui se situent sur un

continuum allant de l'évaluation dans une langue de l'enfant, passant par celle des deux langues jusqu'à l'autre extrême qui consiste à utiliser des non-mots présents dans aucune des langues parlées par l'enfant. À noter que des différences existent quant à l'extrême du continuum qui privilégie l'utilisation des non-mots puisque ceux-ci peuvent se baser sur des langues spécifiques ou avoir été créées selon une base *cross-linguistique*, nous reviendrons sur cette distinction tout au long de la thèse. Nous notons également que l'évaluation peut se faire de manière directe auprès de l'enfant ou de façon indirecte en utilisant un questionnaire parental.

Cette illustration met également en exergue les tâches préconisées et disponibles sur le terrain pour évaluer les enfants bilingues (ex. tâches dans la langue première ou seconde de l'enfant, répétitions de non-mots, tâches d'apprentissage ou encore questionnaire parental). La plupart des outils utilisés en clinique se centre actuellement sur une observation statique des capacités et difficultés langagières des enfants à un moment « x », sans prendre en compte leurs potentialités d'apprentissage (Camilleri et Law, 2007). L'évaluation dynamique est une option permettant de s'intéresser à la capacité d'apprentissage des enfants en leur faisant apprendre de nouvelles choses. Le contenu à apprendre (ex. nouveaux items) peut être des vrais mots inconnus de l'enfant (ex. Hasson et al., 2013) ou des non-mots (Kapantzoglou et al., 2012). Nous avons uniquement mis en lumière cette dernière proposition dans la Figure 3 car c'est l'option choisie dans cette thèse de doctorat (voir notamment chapitre 3 détaillant la méthodologie utilisée).

Dans cette thèse, nous avons choisi de vouer un intérêt particulier à l'évaluation dynamique puisque, à notre connaissance, seul le test pré-standardisé de Hasson et al. (Hasson, 2017 ; Hasson et al., 2013) est disponible comme test d'apprentissage ayant pour objectif le diagnostic de TDL. Plusieurs recherches attestent de l'utilité des tâches d'apprentissage pour des enfants monolingues et bilingues anglophones (ex. Camilleri et Law, 2007 ; Hasson et al., 2013) et

quelques recherches ont tenté une adaptation de celles-ci auprès d'enfants partageant d'autres combinaison de langues (ex. enfants bilingues ayant l'allemand comme deuxième ou troisième langue, Maragkaki et Hessels, 2017). De plus, il s'avère que seule l'étude de Camilleri et Law (2007) a tenté de comparer les performances à une évaluation dynamique des enfants monolingues et bilingues avec ou sans TDL. Nous avons donc décidé de développer deux situations d'évaluation dynamique innovantes et élaborées par nos soins permettant d'évaluer le potentiel d'apprentissage des enfants (Camilleri et Law, 2007). Les deux auront pour objectif le diagnostic (Hessels et Hessels-Schlatter, 2010) mais varieront quant aux situations et procédures d'aide utilisées. L'une d'entre-elle sera conçue comme une situation d'apprentissage de nouveaux mots en contexte informatisée avec la présence de l'expérimentateur mais qui n'interviendra que pour formuler des commentaires généraux, et quelques feedbacks correctifs seront donnés par l'ordinateur (tâche inspirée de Skoruppa, 2019). L'autre situation aura comme contexte la lecture de plusieurs histoires dans lesquelles l'enfant devra apprendre de nouveaux non-mots. L'expérimentateur proposera un étayage graduel (Campione et Brown, 1997) en fonction des difficultés rencontrées par l'enfant pour désigner ou dénommer les nouveaux items. Au sein des deux situations d'évaluation dynamique, les tâches d'apprentissage consisteront en l'acquisition de non-mots pour éviter les biais linguistiques (Kapantzoglou et al., 2012). Plus précisément, elles porteront sur les domaines lexicaux et (morpho)syntaxiques fréquemment relevés comme déficitaires chez les enfants avec un TDL (Bishop et al., 2017) en faisant apprendre des nouveaux mots (substantifs et verbes) et une nouvelle règle de flexion verbale aux enfants. Le domaine de la morphosyntaxe étant encore peu étudié dans le cadre des évaluations dynamiques (voir méta-analyse d'Orellana et al., 2019), il convient de s'y intéresser dans ce travail.

Pour confirmer notre population de recherche pour notre étude sur l'évaluation dynamique, nous avons décidé d'utiliser une série de sous-tests statiques également considérés dans la littérature comme pouvant évaluer adéquatement les enfants bilingues (ex. American Speech-Language-Hearing Association, s.d). Les tâches utilisées se situant différemment sur le continuum des modes d'évaluation (Figure 3). Bien que les études s'intéressant au contexte langagier bilingue se sont développées dès le XX^{ème} siècle et qu'elles se sont intéressées à ces divers outils, nous remarquons qu'encore peu d'entre-elles se sont penchées sur la comparaison des résultats aux tâches d'évaluation de divers sous-groupes (monolingue, bilingue, avec un DT et avec un TDL) et attelées à tester et comparer l'efficacité de ces tâches et modes d'évaluation auprès d'une population bilingue avec un TDL. Ces données étant pourtant nécessaires pour la question de l'identification des TDL au sein d'une population bilingue. De plus, les combinaisons de langues fréquemment étudiées concernent l'anglais et/ou l'espagnol (Peña et al., 2001) créant ainsi une impasse pour la pratique logopédique dans les pays où les combinaisons de langues à évaluer sont différentes. Nous souhaitons ainsi tester et comparer l'efficacité de ces tâches et modes d'évaluation, c'est-à-dire de nos sous-tests statiques et de nos situations d'évaluation dynamique, et déterminer leur précision diagnostique auprès d'une population bilingue français-portugais avec et sans TDL d'âge scolaire. Notre processus est donc constitué de plusieurs étapes (voir Figure 1 dans le préambule) puisque la plupart des tâches utilisées seront considérées comme des sous-tests statiques pour notre tâche d'intérêt qui est l'évaluation dynamique. Ainsi, nous avons comme objectif grâce aux analyses de précision diagnostique d'améliorer tout au long de cette recherche la classification initiale de notre population, en passant par une classification intermédiaire après avoir utilisé nos sous-tests langagiers et terminer par une classification complémentaire après la passation de nos situations d'évaluation dynamique. Puisque notre objectif poursuivi est complexe et novateur, nous avons décidé de procéder par étapes (Parties de la thèse) afin d'examiner les tâches et modes

d'évaluation les uns après les autres, en terminant par l'évaluation dynamique. Nous pourrions ainsi apporter des éléments de discussion et d'ouverture lors de notre dernière Partie (VI).

2.1 Questionnement principal

Nous allons, dans ce qui suit, énoncer notre principal questionnement ainsi que plusieurs sous-questions qui nous permettront de traiter celui-ci. Les sous-questions reflètent les étapes de la thèse ; la première sous-question s'intéresse aux sous-tests statiques et la deuxième à l'évaluation dynamique. Nous allons, dans ce chapitre, laisser de côté les questions plus spécifiques à chaque domaine qui seront exposées et étudiées dans les différents articles. Ainsi, notre proposition principale et les sous-questions se formulent ainsi :

Quelles sont les tâches langagières et/ou modes d'évaluation permettant une identification précise de la présence d'un TDL chez une population monolingue et bilingue français-portugais d'âge scolaire ?

Sous-question 1 : Parmi nos sous-tests statiques langagiers, quelles sont les tâches langagières (sous-tests standardisés- dénomination, compréhension syntaxique, répétitions de non-mots - ; questionnaire parental) et/ou modes d'évaluation (dans une ou deux langues, utilisant des non-mots langues-spécifiques ou *cross-linguistiques*) qui révèlent une différence de performance significative entre les enfants avec un DT et avec un TDL dans un contexte monolingue et bilingue ?

a Quelle tâche langagière et/ou mode d'évaluation fournit la meilleure précision diagnostique (pourcentage de sensibilité et spécificité) ? Permet-elle d'améliorer la classification initiale de nos sous-groupes ?

Sous-question 2 : Est-ce que nos deux situations d'évaluation dynamique permettent de révéler une différence de performance significative entre les enfants avec un DT et un TDL dans un contexte monolingue et bilingue ?

a Parmi nos deux situations d'évaluation dynamique, laquelle fournit la meilleure précision diagnostique ? Sera-t-elle en mesure d'améliorer la classification intermédiaire de nos sous-groupes ?

Tout d'abord, il s'agit de rappeler que notre intérêt premier concerne l'identification d'un TDL chez une population bilingue. Les résultats de notre population monolingue seront également nécessaires à prendre en considération, notamment pour les comparaisons entre enfants monolingues et bilingues. En effet, nous pourrons ainsi observer si les résultats des enfants varient en fonction de leur statut linguistique. Dans ce qui suit, nous allons d'abord présenter nos hypothèses théoriques portant sur nos sous-tests statiques langagiers en commençant par les modes d'évaluation (2.2.1) et en poursuivant par les tâches (2.2.2). Nous proposerons également plusieurs hypothèses théoriques au sujet de l'évaluation dynamique (2.3) mais renvoyons également le lecteur à la Partie V de cette thèse qui traite uniquement de cette thématique. Nous débuterons toujours par présenter nos hypothèses quant aux deux populations (monolingue et bilingue) avant de se centrer sur les enfants bilingues. À noter encore que le questionnaire parental, seule tâche émanant d'un mode d'évaluation indirecte, sera uniquement évoqué dans les hypothèses traitant des tâches (2.2.2).

2.2 Hypothèses concernant les sous-tests statiques langagiers

Les tâches utilisées comme des sous-tests statiques figurent dans l'illustration 3 présentée au début de la problématique. Il s'agit de sous-tests standardisés en français et en portugais, de répétitions de non-mots (langues-spécifiques et *cross-linguistique*) et d'un

questionnaire parental. Ces tâches varient en fonction de leur mode d'évaluation : dans une des langues de l'enfant, dans ses deux langues ou faisant usage de non-mots pouvant être créés sur la base de langues-spécifiques ou être *cross-linguistiques*. Nous détaillerons davantage ces tâches et leur choix dans la section de la méthode (chapitre 3).

2.2.1 Modes d'évaluation

Nous supposons que l'évaluation dans une seule langue de l'enfant (à savoir le français pour notre population), permettra de mettre en évidence des différences significatives entre les enfants avec un DT et un TDL mais également entre les enfants monolingues et bilingues. Ainsi, l'échec aux tâches en français (détails présents au point 2.2.2 pour les tâches) ne pourra pas être exclusivement expliqué par le statut clinique des participants. Plus précisément, comme énoncé dans la littérature, les tâches d'évaluation statiques sont sujettes au manque d'exposition à la langue testée (Hasson et Botting, 2010 ; Scheidnes et Tuller, 2016) et étant donné que notre population a des profils linguistiques hétérogènes (monolingue, bilingue ; simultané, consécutif⁹ ; dominant en français ou en portugais), nous supposons que les résultats inhérents à ces tâches seront difficilement interprétables, de façon similaire aux évaluations logopédiques actuelles sur le terrain. Spécifiquement pour les enfants bilingues, nous supposons que l'usage de seuils de pathologie adaptés en fonction du taux d'exposition à la langue testée (Thordardottir, 2015a) ou de l'évaluation dans la langue dominante de l'enfant (American Speech-Language-Hearing Association, s.d.), nous permettra de nuancer les différences obtenues entre les enfants monolingues et bilingues.

En référence aux différentes organisations (ex. American Speech-Language-Hearing Association, s.d.), nous supposons que l'évaluation dans les deux langues des enfants bilingues

⁹ La distinction entre ces deux types de bilinguisme est évoquée dans l'article de la Partie I, Schwob (2020).

(à savoir en français et en portugais) nous renseignera davantage sur le profil de l'enfant (Peña et al., 2016) et nous permettra d'associer davantage l'échec aux tâches aux profils cliniques des enfants. Nous supposons ainsi que les enfants bilingues avec un TDL échouent dans des sous-tâches de leurs deux langues. Toutefois, il est important de signaler que les tests qui seront utilisés proviendront de batteries françaises et portugaises qui ont été créées pour des enfants monolingues et qui ont ainsi des normes monolingues ce qui constitue réellement un problème pour l'évaluation des enfants bilingues (Boerma et Blom, 2017 ; Scheidnes et Tuller, 2016), et pouvant nous conduire à des erreurs dans l'identification d'un TDL bien que les deux langues soient évaluées. De plus, peu de tests standardisés sont à disposition du portugais et parmi ceux qui seront utilisés, l'étalonnage de référence ne couvrira malheureusement pas l'ensemble des âges de notre recherche (de Almeida et al., 2016). Nous supposons donc que bien que des différences significatives seront observables entre les performances des enfants avec un DT et un TDL dans ce mode d'évaluation, les valeurs de précision diagnostique des tâches ne seront pas convenables pour identifier un TDL chez notre population d'enfants bilingues français-portugais (inférieures à 80 % en prenant comme référence Plante et Vance, 1994). Nous reviendrons et exposerons davantage nos hypothèses portant sur les tâches dans la section suivante (2.2.2).

Nous présumons que les modes d'évaluation utilisant des non-mots permettront de réduire les biais potentiels liés aux langues parlées par les enfants (Kapantzoglou et al., 2012). En effet, nous supposons que ce mode d'évaluation permettra de mettre en exergue des différences significatives quant aux performances des enfants avec un DT et un TDL aux tâches de répétition de non-mots sans qu'elles puissent être attribuées à leur statut linguistique. Nous supposons que l'échec auxdites tâches pourra presque exclusivement être attribué au statut clinique des enfants. Dans la littérature, ce mode d'évaluation est vivement recommandé pour évaluer les enfants bilingues et fournit des valeurs de précision diagnostic très élevées (ex.

méta-analyse d'Estes et al., 2007 pour la répétition de non-mots). Nous supposons donc que le mode d'évaluation faisant usage de non-mots sera le plus efficace pour différencier les enfants avec un DT et un TDL monolingues et bilingues de notre recherche. Toutefois, nous sommes conscientes que la façon dont sont créés les non-mots exercent une influence sur les performances des enfants bilingues aux tâches d'évaluation. En effet, les non-mots peuvent être conçus sur la base de langues spécifiques et ainsi respecter la phonotactique, les phonèmes et les structures syllabiques de celles-ci. Ils peuvent aussi parfois être assez proches d'items des langues cibles. À contrario, les items peuvent être conçus pour être *cross-linguistiques*. Ainsi, les structures syllabiques et les phonèmes communs à la plupart des langues sont utilisées dans la création (ex. dans la tâche de Chiat, 2015). Nous supposons donc que ces éléments puissent exercer une influence sur les performances des enfants bien que nous utilisions un mode d'évaluation incluant des non-mots. Nous poursuivrons les réflexions quant aux conceptions des tâches dans la section qui suit et renvoyons également le lecteur à la partie IV traitant spécifiquement des RNM.

2.2.2 Tâches d'évaluation

Le chapitre 3 portant sur la méthode fournira des indications détaillées quant aux tâches choisies, toutefois, nous précisons ici brièvement que pour nos sous-tests statiques, nous avons choisi de proposer des tâches évaluant différents domaines et modalités langagières (Bishop et al., 2017) : des tâches de compréhension orale, de dénomination lexicale et des répétitions de non-mots. Parmi ces différentes tâches, nous pensons que les tâches de RNM créées sur la base de langues spécifiques (en français et en portugais), largement étudiées et présentes dans de nombreux tests cliniques, nous permettront d'obtenir des différences entre les enfants avec un DT et un TDL qu'ils soient monolingues ou bilingues (Coady et Evans, 2008 ; Dubreuil-Piché et al., 2017 ; Girbau et Schwartz, 2007). Plus précisément, malgré l'hétérogénéité linguistique

de notre population (monolingue, bilingue, bilingue simultané ou consécutif), le français et le portugais étant des langues proches phonologiquement (Kihm, 2016), nous supposons que l'échec à ce type de tâche pourra être associé à leur statut clinique plutôt qu'à leur statut linguistique. En outre, la littérature scientifique relate des difficultés de traitement phonologique et de mémoire à court-terme verbale des enfants monolingues et bilingues avec un TDL parlant diverses langues (ex. Dos Santos et Ferré, 2018). Cependant, comme évoqué succinctement au point 2.2.1, nous pensons que des différences inhérentes à la conception des tâches (ex. longueur et complexité des stimuli) seront susceptibles de faire varier les performances des enfants. Nous supposons ainsi que l'utilisation d'une tâche *cross-linguistique* telle que proposée par Chiat et collaborateurs (Chiat, 2015 ; Chiat et Polišenská, 2016), variant en termes de longueur syllabique et non en complexité phonologique, permettra d'être moins sujette au biais d'exposition et nous donnera la possibilité d'obtenir peu, voire aucune différence entre les enfants monolingues et bilingues de notre recherche. Toutefois, puisque cette tâche ne varie pas en fonction de la complexité phonologique (uniquement en longueur), nous supposons qu'elle pourra peut-être être aisée pour les deux groupes d'enfants avec un DT et un TDL. Enfin, nous formulons l'hypothèse que ces tâches de RNM conçues sur la base de langues spécifiques ou *cross-linguistique*, nous fourniront des valeurs de précision diagnostique jugées « adéquates » (environ 80 % en prenant comme référence Plante et Vance, 1994) en regard des larges tailles d'effets relevées dans la littérature (Estes et al., 2007). Néanmoins, dans le cas où certains enfants avec un TDL ne présentent pas de difficultés phonologiques, il semble nécessaire d'associer la RNM à d'autres tâches dans le bilan langagier (Bishop et al., 2017 ; Kapalková et al., 2013).

Au sujet des autres tâches issues de tests standardisés, nous pensons que les tâches de dénomination évaluant les lexiques productifs de l'enfant (en français et en portugais) ne seront que peu convaincantes pour l'identification d'un TDL. En effet, le vocabulaire étant largement

influencé par la quantité d'exposition aux langues (Thordardottir, 2011), nous supposons que même les enfants bilingues avec un DT pourront obtenir des performances relativement faibles à ce type de tâches et nous ne pourrons donc pas attribuer l'échec à celles-ci au statut clinique des enfants. La morphosyntaxe étant également impactée par le taux d'exposition aux langues (Thordardottir, 2015b), mais à moindre mesure que pour le lexique, nous supposons également que les résultats aux tâches de compréhension orale ne soient pas suffisamment précis quant à l'identification d'un TDL chez nos enfants bilingues (en dessous de 80 % de sensibilité et spécificité, Plante et Vance, 1994).

Nous supposons également que l'usage d'une tâche émanant d'un mode d'évaluation indirect telle que le questionnaire parental, nous permettra d'obtenir des informations complémentaires importantes en vue de l'identification d'un TDL, notamment la recension de plusieurs facteurs de risque (Tuller, 2015). Nous pensons toutefois qu'il ne pourra pas être utilisé seul en raison de son identification reconnue des parleurs tardifs plutôt que des TDL (Pearson, 2013). Pour rappel, trois quarts des enfants parleurs tardifs se développent ensuite normalement sans présenter de difficultés persistantes (Rescorla, 2011). De plus, il ne semble pas toujours facile pour les parents de répondre de façon fidèle aux questions posées (O'Toole et al., 2017) puisqu'il s'agit de questions rétrospectives pour lesquelles il semble plus aisé de répondre sur ce qu'ils connaissent maintenant plutôt que sur ce qu'ils savaient dans le passé.

2.3 Hypothèses à propos de l'évaluation dynamique

L'évaluation dynamique, telle que nous l'utiliserons dans ce travail, se situe dans un mode d'évaluation faisant recours à des non-mots afin d'éviter les biais linguistiques potentiels des langues parlées par les enfants (Kapantzoglou et al., 2012). C'est également la seule proposition de cette thèse qui revêtira un caractère dynamique puisque nous proposerons de nouveaux apprentissages langagiers aux enfants et nous ne nous concentrons pas sur leurs

habilités dans l'une ou l'autre de leurs langues. Nous pensons que l'évaluation dynamique qui prendra en considération les capacités d'apprentissage des enfants dans les domaines lexicaux et (morpho)syntaxiques, nous permettra d'obtenir des différences de performances entre les enfants avec un DT et un TDL (Gray, 2003 ; Hasson et al., 2013). Cette dernière aura l'avantage de nous donner une vue globale des capacités et fragilités langagières des enfants en fonction des différents domaines et modalités linguistiques testées. Nous supposons qu'au travers de nos deux situations d'évaluation dynamique, des différences seront éventuellement perceptibles entre les enfants monolingues et bilingues, en mettant en exergue un avantage pour les enfants bilingues en raison de leur supériorité dans des tâches impliquant les fonctions exécutives, notamment la mémoire de travail (ex. Adescope et al., 2010 ; Blom et al., 2014), bien qu'il n'y ait pas de consensus à ce sujet dans la littérature (ex. de Cat, 2018). Cependant, nous supposons que l'échec à ces tâches sera davantage influencé par le statut clinique des enfants (Gray, 2003). Nous pensons que nos tâches d'évaluation dynamique permettront également de mettre en lumière des pourcentages de sensibilité et de spécificité jugés « adéquats » (Plante et Vance, 1994), comme relevé par Petersen et al. (2020) ou encore par la méta-analyse d'Orellana et al. (2019). Nous faisons l'hypothèse que des différences quant aux situations d'apprentissage, aux procédures d'aide et aux tâches évaluatives utilisés seront susceptibles d'apparaître. Nous évoquons ici quelques-unes de nos idées et renvoyons les lecteurs à la Partie V de ce travail (chapitre 7) pour les hypothèses plus détaillées. À titre d'exemple, nous supposons que pour tous les enfants, les substantifs seront plus faciles à apprendre que les verbes (Oetting et al., 1995) et que les tâches de réception seront plus aisées à effectuer que les tâches de production (Kan et Windsor, 2010). Nous pensons également voir des différences de performances en fonction de la situation proposée (jeu autonome sur ordinateur vs. lecture d'histoires en interaction) et par la procédure de médiation apportée par l'adulte (avec uniques feedbacks correctifs dans la tâche de jeu sur ordinateur et indices graduels et interaction dans la situation

interactive de lecture d'histoires). Il est actuellement difficile de dire qu'elle sera la tâche la plus adaptée pour identifier les enfants avec un TDL puisque ces deux situations sont courantes dans la vie des enfants (Van Horn et Kan, 2015). Plus que la situation, nous supposons que c'est la procédure utilisée qui mettra en avant des différences. Nous supposons que la situation de jeu sur ordinateur sera moins sujette au statut linguistique des enfants puisque des non-mots leur sont présentés dans un contexte le plus neutre possible (faisant intervenir au minimum le français et l'expérimentateur). La deuxième situation, qui inclura des non-mots dans un contexte d'interaction en français (lecture d'histoires), sera susceptible d'être davantage influencée par le statut linguistique des enfants. Toutefois, elle aura également l'avantage d'évaluer le potentiel d'apprentissage de l'enfant lorsque celui-ci recevra un étayage graduel (Campione et Brown, 1987). Nous pensons, dans une vision interactionniste du développement du langage (Vygotski, 1934/1985), que les enfants pourront apprendre davantage et pourront progresser dans la situation de lecture d'histoires grâce à l'aide apportée par l'expérimentateur. Enfin, nous supposons que les deux groupes d'enfants pourront améliorer leurs performances mais que l'aide fournie sera plus importante pour les enfants avec un TDL (Maragkaki et Hessels, 2017 ; Petersen et al., 2020).

Nos situations d'évaluation dynamique auront l'avantage de se centrer sur plusieurs domaines langagiers (lexique, morphosyntaxe) fréquemment impactés dans le TDL, toutefois, en raison de l'hétérogénéité des TDL (Bishop et al., 2017 ; Thomas et al., 2019), nous supposons que c'est l'association de plusieurs tâches qui augmentera leur pouvoir discriminant et qui permettra ainsi une identification plus aisée et précise de la présence d'un TDL (Boerma et Blom, 2017). Une combinaison de plusieurs tâches utilisées dans cette thèse aurait l'avantage de nous faire obtenir un profil langagier plus complet des enfants (Peña et al., 2016). Toutefois, puisque seules les situations d'évaluation dynamique n'auront pas servies de sous-tests dans cette thèse de doctorat, nous annonçons d'ores et déjà renoncer à combiner nos tâches

d'évaluation dynamique avec nos tâches considérées comme des sous-tests langagiers pour tester leur précision diagnostique.

3. Méthode

Notre recherche est une étude transversale menée en Suisse romande auprès d'enfants ayant des origines différentes. Nous allons dans ce qui suit présenter notre population de recherche avec les critères d'inclusion et d'exclusion généraux à l'étude (3.1). Certaines spécificités méthodologiques seront détaillées dans les articles présentés dans les différents chapitres de ce travail (points 4, 6 et 7) mais la population de recherche a été la même durant toute notre étude. Nous préciserons cela dans une section traitant des classifications utilisées pour distinguer les enfants de notre recherche. Nous décrirons ensuite la procédure de recrutement. Nous poursuivrons en évoquant le choix et la conception de nos tâches d'évaluation (3.2). À nouveau, des détails quant à celles-ci seront abordés de façon précise dans les articles. Puis, nous continuerons en exposant notre procédure de passation (3.3). Enfin, nous terminerons cette section en parlant de la procédure de cotation (3.4) et de nos analyses statistiques (3.5).

En plus de nos données empiriques, nous avons décidé de réaliser une revue systématique de littérature ainsi qu'une méta-analyse sur la répétition de non-mots (voir point 5) dans le but de tester son efficacité diagnostique dans un contexte monolingue et bilingue. Les éléments méthodologiques s'y référant sont présentés de façon détaillée dans l'article en question dans la Partie IV de cette thèse de doctorat.

3.1 Participants

Au total, 72 enfants âgés entre 5 ;0 et 7 ;11 ans bilingues français-portugais ou monolingue français vivant en Suisse romande au moment du recueil de données ont pris part à la recherche principale de cette thèse de doctorat. Notre cohorte se compose de 36 enfants monolingues et 36 enfants bilingues. Parmi ceux-ci, 38 enfants ont été catégorisés comme ayant

un DT et 35 comme ayant un TDL. Dans ce qui suit, nous allons expliquer nos choix quant à la population choisie pour cette recherche. Nous allons ensuite évoquer les critères d'inclusion pour la constitution de nos sous-groupes.

Nous avons décidé de nous intéresser aux enfants entre 5 ;0 et 7 ;11 ans car il s'agissait d'enfants scolarisés au moment de la recherche. Nous avons fait ce choix pour plusieurs raisons. En premier lieu, certains enfants bilingues successifs ont ainsi eu la possibilité d'être exposés au français, qui est la langue de scolarisation en Suisse romande. Par ailleurs, les enfants avec un TDL sont fréquemment identifiés à ces âges (Bishop et al., 2017). En effet, bien que les difficultés apparaissent déjà en bas âge, pour parler de TDL, elles doivent être persistantes et avoir un impact sur la vie quotidienne des enfants (Bishop et al., 2017 ; Breault et al., 2019). Enfin, nous supposons également que les enfants scolarisés seraient plus habitués à un contexte d'évaluation que les enfants préscolaires, ce qui a motivé notre sélection.

Notre choix pour la langue portugaise s'explique car c'est la langue la plus représentée en Suisse romande après les langues nationales (Office Fédéral de la Statistique, 2020) et c'est la langue la plus fréquente dans le contexte logopédique de Suisse romande (Volpin et al., 2020). Bien qu'il n'existe que peu d'outils standardisés et normés en portugais (de Almeida et al., 2016), il a quand même été possible pour nous de trouver des tâches plutôt similaires pour réaliser nos évaluations (voir point 3.2). De plus, comme le français, le portugais est une langue romane et les deux langues comportent beaucoup de similarités (Kihm, 2016) notamment en ce qui concerne leur système phonologique, ce qui a facilité la conception de nos situations d'évaluation dynamique nécessitant l'usage de non-mots respectant les règles phonotactiques des deux langues et étant composés des phonèmes communs à celles-ci (voir 2.2.3). Nous avons décidé de ne pas dissocier le portugais d'Europe et le portugais du Brésil bien qu'il y ait des différences entre les deux, notamment orthographiques et au niveau de l'accent des locuteurs

(Pérez, 2000), deux domaines qui ne nous intéressent pas dans cette recherche. Enfin, notre choix pour le portugais s'explique également dans l'intérêt d'explorer une combinaison de langue (français-portugais) rarement étudiée dans les recherches scientifiques.

3.1.1 Critères pour la constitution des sous-groupes

De façon générale, aucun des enfants ne devait avoir un diagnostic ou avoir une suspicion de trouble du spectre autistique, de trouble cognitif, de trouble socio affectif, de trouble auditif. L'audition a notamment été vérifiée à l'aide d'un audiomètre durant le troisième rendez-vous (voir point 3.3) et les enfants devaient obtenir un seuil d'audibilité fixé à 20 décibels maximum dans les fréquences de 500, 1000, 2000 et 4000 Hertz. Un screening non-verbal a également été effectué afin d'exclure des difficultés de raisonnement non-verbal. Les enfants devaient obtenir des résultats dans la norme à cette épreuve issue du test Wide-Range de Glutting et al. (2000). L'absence de trouble du spectre autistique, de trouble socio-affectif et trouble cognitif a été vérifiée via des questions posées lors du premier entretien téléphonique afin de s'assurer de l'éligibilité de nos participants (voir point 3.3).

Statut linguistique

Les enfants de notre recherche vivaient tous en Suisse romande au moment de la recherche. Diverses indications linguistiques ont été récoltées via un questionnaire parental fourni à l'écrit aux parents (version raccourcie basée sur le PabiQ de Tuller, 2015¹⁰, voir Annexe E). Les enfants monolingues devaient au moment de la recherche entendre et parler plus de 85 % de français à la maison et les enfants bilingues devaient avoir un minimum de 25 % d'exposition au portugais et ne devaient pas être exposés à plus de 15 % à une troisième

¹⁰ Nous avons choisi de proposer une version raccourcie du questionnaire, car par expérience au sein de notre groupe de recherche, des versions longues étaient fréquemment que partiellement remplies par les parents. Voir 3.3.2 pour des détails quant au questionnaire.

langue. Les principales informations démographiques de notre population sont résumées dans le Tableau 1 ci-dessous.

Bien que divers facteurs auraient pu être intéressant à exploiter dans notre recherche, nous n'avons, par exemple, pas dissocié les enfants bilingues simultanés et consécutifs, notamment pour des raisons de taille d'échantillon. Le taux d'exposition aux langues a été considéré en prenant en compte l'exposition actuelle des enfants. Ainsi, nous avons pu déterminer la dominance actuelle des enfants à une de leurs langues. Cette information a été utilisée uniquement pour la Partie III de cette thèse portant sur l'évaluation dans les deux langues de l'enfant (voir article au point 4) et nous n'avons pas formulé d'hypothèses précises concernant l'exposition des enfants aux langues. L'intérêt premier de notre recherche étant l'évaluation dynamique au moyen d'apprentissages langagiers indépendants aux langues parlées par les enfants, nous n'avons pas choisi de consacrer une partie plus importante de la thèse à ces éléments.

Tableau 1

Principales données démographiques des participants

Statut clinique	Statut linguistique	Nombre	Genre		Age moyen en mois	Niveau moyen formation parent	Niveau moyen QI non- verbal	Type de bilinguisme		Dominance actuelle	
			F	G				Simultané	Consécutif	FR	PT
DT	Monolingue	20	12	8	77	4.3	52.5				
TDL	Monolingue	16	8	8	75	3.1	50.7				
DT	Bilingue	17	8	9	82	3.2	62.8	9	8	12	5
TDL	Bilingue	19	11	8	80	2.6	46.4	12	7	12	7

Note. Abréviations : DT = enfant avec un développement langagier typique, TDL = enfant avec un trouble développemental du langage oral, FR= français, PT = portugais, F = fille, G = garçon.

Statut clinique

Les enfants avec un DT ne devaient pas avoir été identifiés comme ayant un TDL ou avoir été suivis en logopédie. À contrario, les enfants avec un TDL devaient avoir été diagnostiqués par un logopédiste clinicien et être suivis en logopédie depuis au moins trois mois avant la participation à la recherche afin d’attester la persistance de leurs difficultés langagières. Bishop et al. (2017) mentionnent que les TDL sont hétérogènes et peuvent toucher divers domaines et modalités langagières (voir 1.1 pour chapitre théorique sur le TDL). Ainsi, les enfants de notre recherche présentaient des TDL divers et variés (ex. certains possédaient des difficultés touchant le versant réceptif et expressif alors que d’autres éprouvaient des difficultés uniquement pour la production) et touchant un ou plusieurs domaines langagiers (ex. phonologie, lexique, syntaxe, discours). L’homogénéité n’était pas un critère d’inclusion à notre étude car il aurait été trop complexe de recruter des profils relativement identiques d’enfants avec un TDL. Dans la suite du travail, nous avons ainsi décidé de faire usage de la formulation au singulier « enfants avec *un TDL* », bien que tous les enfants ne présentent pas les mêmes difficultés. Afin de préciser le statut clinique initial de nos participants, nous avons considéré les performances des enfants à différents sous-tests statiques langagiers (voir article aux points 4 et 6) comme référence pour être inclus ou non dans l’étude portant sur l’évaluation dynamique (7). Nous allons décrire brièvement notre procédure de classification ci-dessous.

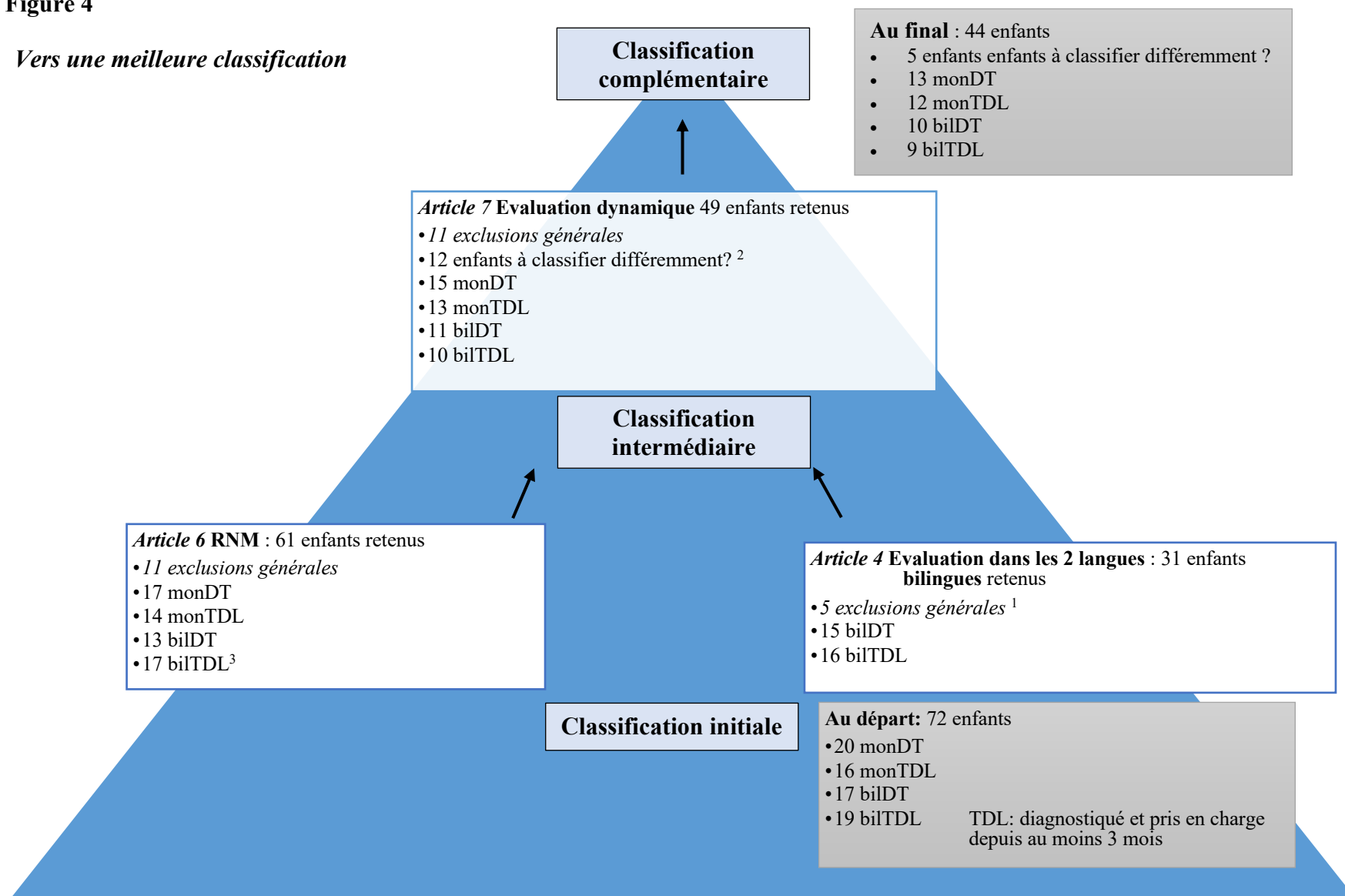
Classifications utilisées. Comme évoqué ci-dessus et dans la Figure 1 du préambule, notre travail se base initialement sur une classification des enfants avec un TDL effectuée par les logopédistes en charge des enfants. Les enfants avec un DT ont été recrutés via des écoles et centres parascolaires et leur développement langagier est présumé typique. Comme relevé dans la littérature (article au point 1.2), cette classification peut être difficile à établir et ainsi être imparfaite. Ainsi, il est possible que des erreurs de diagnostics soient présents dans ce

regroupement initial : soit, que des enfants avec un TDL ne présentent pas des difficultés relevant d'une telle pathologie et/ou que des enfants avec un DT éprouvent des difficultés langagières passées inaperçues jusqu'alors. Pour les enfants avec un TDL, nous aurions pu choisir de faire confiance aux logopédistes mais avons décidé de confirmer ou d'infirmer le diagnostic initial au moyen d'autres tâches et modes d'évaluation recommandés par la littérature scientifique (voir Figure 3). Notre but dans ce travail de doctorat a ainsi été d'améliorer la classification existante au moyen d'autres tâches et modes d'évaluation. C'est la raison pour laquelle plusieurs enfants de notre population ont été exclus de tout groupe durant la thèse. La Figure 4 ci-dessous résume la constitution des groupes en fonction des articles de recherche (présents aux chapitres 4, 6 et 7) et illustre le cheminement de notre travail. Dans nos articles traitant de nos sous-tests statiques (aux points 4 et 6), nous pouvons percevoir le début de notre réflexion. Nous avons mis en lumière la différence existante entre le jugement clinique (diagnostic actuel des logopédistes) et l'utilisation d'outils pour évaluer les deux langues de l'enfant bilingue. Ainsi, nous notons que 12 enfants semblent avoir été mal classifiés initialement par la première classification. Huit enfants classifiés comme ayant un TDL, n'ont pas échoué de tâches dans une ou leurs deux langues alors que les difficultés langagières de l'enfant bilingue avec un TDL doivent être perceptibles dans les deux langues (Kohnert, 2010). Plus précisément, un enfant monolingue avec un TDL n'a échoué aucun sous-tests en français et sept enfants bilingues avec un TDL n'ont échoué aucun sous-tests dans l'une ou les deux langues. Quatre enfants (deux monolingues et deux bilingues) avec un DT ont présenté des difficultés dans les tâches proposées. Ainsi, pour l'article présenté au chapitre 7 traitant de l'évaluation dynamique, nous avons, grâce à notre classification intermédiaire, fait le choix d'exclure les enfants pour qui le diagnostic n'était pas sûr. Après nos analyses de précision diagnostique d'une de nos situations d'évaluation dynamique considérée comme très précise (Plante et Vance, 1994), nous avons pu déterminer que cinq enfants pouvaient encore être

classifiés différemment de notre regroupement intermédiaire. En effet, trois enfants assignés au groupe TDL semblent mieux performer que l'ensemble du groupe (deux enfants bilingues et un enfant monolingue) et deux enfants monolingues avec un DT éprouvent des difficultés.

Figure 4

Vers une meilleure classification



Note. Abréviations : mon = monolingue, bil = bilingue, DT = enfant avec un développement langagier typique, TDL = enfant avec un trouble développemental du langage oral.

¹Taux d'exposition insuffisant, échec aux tâches de raisonnement non-verbal ou à l'audiomètre, problème technique.

²Exclusion pour non-conformité aux sous-tests statiques langagiers (articles 4 et 6). Ex. échec de la part d'un enfant avec un DT ou réussite pour un enfant avec un TDL.

³un enfant bilingue avec un TDL a été ajouté dans cette recherche car il avait échoué initialement à l'audiomètre puis après consultation chez un médecin ORL aucune difficulté auditive n'a pu être attestée et il a donc été ajouté dans notre population pour les recherches suivantes.

3.1.2 Procédure de recrutement

Les enfants avec un DT ont été recrutés par le biais de la diffusion d’annonces auprès des écoles et des centres parascolaires. Concernant les enfants avec TDL, les logopédistes de la Suisse romande ont été informés de la recherche via l’association professionnelle des logopédistes diplômés (ARLD) et ils ont ainsi pu identifier les enfants susceptibles de participer à l’étude. Des assistantes de recherche ainsi que des étudiants en première année de Master en Logopédie à l’Université de Neuchâtel nous ont également apporté leur aide dans la phase de recrutement.

Les familles intéressées à participer à la recherche recevaient un document d’information, un questionnaire parental et une déclaration de consentement libre et éclairé, validés par la commission d’éthique locale (commission d’éthique du canton de Vaud ; n° de projet : 2017-01881, autorisation accordée le 07.02.2018) en français ou en portugais selon leur choix. Un entretien téléphonique a ensuite été réalisé avec les parents afin de répondre à leurs éventuelles questions et de s’assurer de prime abord de l’éligibilité des participants. Cet appel permettait également de convenir du lieu et de l’horaire des rendez-vous (ceux-ci sont détaillés au point 3.3).

3.2 Matériels

Pour faire les différents choix qui seront énoncés ci-dessous, nous avons réalisé un pilotage de nos différentes tâches d’évaluation (sous-tests statiques et situations d’évaluation dynamique). Huit enfants ont participé à cette phase pilote. Quatre enfants présentaient un DT et quatre avaient été diagnostiqués comme ayant un TDL. Trois enfants étaient bilingues et cinq étaient monolingues. Ils ont participé à cinq séances d’évaluation car nous avions initialement pensé proposer les évaluations dynamiques à deux reprises. À la suite de ce pilotage, nous avons décidé de réduire la durée des évaluations en nous limitant à trois séances en proposant à une

seule reprise chacune des situations d'évaluation dynamique. Des remaniements portant sur nos tâches ont également été réalisés (ex. consignes, items). Nous avons exclu une de nos tâches standardisées, la répétition de phrases, car celle en portugais (issue du test PALPA-P de Castro et al., 2007) était très longue et que nous avons obtenu un effet plancher pour les différents groupes d'enfants (avec un DT et un TDL).

Toutes les tâches qui ont été utilisées pour l'étude principale sont plus amplement détaillées dans les articles se focalisant sur l'une ou l'autre d'entre-elles. Nous allons présenter brièvement les tâches provenant de tests standardisés ou reprises de précédentes études ainsi que le questionnaire parental utilisé. Nous terminerons par présenter nos situations d'évaluation dynamique.

3.2.1 Sous-tests langagiers

Différentes tâches issues de tests standardisés disponibles sur le marché et permettant d'observer les difficultés langagières fréquemment relevées chez les enfants avec un TDL (Bishop et al., 2017) ont été proposées aux participants. Elles ont été utilisées comme des sous-tests langagiers auprès de notre population afin s'assurer de la constitution de nos groupes dans notre étude portant sur l'évaluation dynamique (article présenté au point 7). Les tâches choisies sont : la dénomination pour le lexique productif, la compréhension orale de phrases pour la morphosyntaxe réceptive ainsi que des tâches de répétition de non-mots pour le traitement phonologique et la mémoire à court-terme verbale. Les domaines lexicaux et syntaxiques sont fréquemment évalués pour identifier un TDL (Bishop et al., 2017 ; Denman et al., 2017) et la répétition de non-mots semble fréquemment discriminante dans le domaine de la phonologie (Boerma et Blom, 2017 ; Estes et al., 2007). En français, les tâches de dénomination lexicale et de compréhension immédiate proviennent du test « Évaluation du Langage Oral » (ELO) (Khomsi, 2001) et la répétition de non-mots du test Exalang, en utilisant la version Exalang 3-

6 (Helloin et Thibault, 2006) ou Exalang 5-8 (Thibault et Helloin, 2010) en fonction de l'âge de l'enfant. Nous avons tenté de sélectionner des tâches évaluant les mêmes domaines en portugais. Les tâches de dénomination lexicale ainsi que de compréhension syntaxique sont reprises du test « Avaliação da Llinguagem Oral » (ALO) (Sim-Sim, 2006) et la répétition de non-mots de la batterie « Provas de Avaliação da Linguagem e da Afasia em Português » (PALPA-P) (Castro et al., 2007). À noter que pour les tests lusophones, peu de tranches d'âges sont couvertes par la normalisation (ex. quatre et six ans ne sont pas couverts dans le PALPA-P). En ce qui concerne les répétitions de non-mots issus des tests standardisés en français et en portugais, celles-ci sont considérées comme étant des tâches langues-spécifiques (français et portugais), puisque les items ont été créés en respectant la phonotactique, les phonèmes et les structures syllabiques présentes dans les langues. Les tâches varient en termes de longueur syllabique et de complexité phonologique (c'est-à-dire de structures syllabiques employées) et permettent ainsi d'évaluer la qualité de la précision phonologique des enfants ainsi que leurs capacités en mémoire à court-terme verbale. Des détails quant à ces tâches sont fournis dans l'article présenté au chapitre 6 de ce travail. En plus des répétitions de non-mots issues des tests standardisés en français et en portugais, nous avons également choisi d'utiliser la tâche *cross-linguistique* provenant de Chiat et collaborateurs (Chiat, 2015 ; Chiat et Polisenská, 2016). Les auteurs ont créé des items respectant les propriétés communes à une majorité de langue (phonèmes, structures syllabiques) et se veulent donc être « quasi-universels ». Les items de la tâche ne varient pas en termes de complexité phonologique mais en longueur syllabique et visent à tester spécifiquement la mémoire à court-terme verbale. Nous avons donc repris les items proposés par l'auteure et ses collègues et avons vérifié qu'aucun non-mots ne corresponde à des mots en français et en portugais.

Adaptation de seuils de pathologie

Dans notre Partie III, nous avons tenté d'utiliser une adaptation des seuils de pathologie en fonction du taux d'exposition langagier de l'enfant comme le suggère Thordardottir (2015a). Selon cette auteure, si l'enfant bilingue est évalué dans ses deux langues séparément et qu'il est comparé à des normes monolingues, les seuils de déviation à la norme doivent être abaissés par rapport à ceux définis pour l'enfant monolingue. En se basant sur des recherches récentes (ex. Thordardottir, 2011) qui comparent des enfants monolingues et bilingues simultanés apprenant le français ou l'anglais, Thordardottir (2015a) formule quelques propositions pour les enfants bilingues simultanés : quand l'enfant bilingue peut être évalué dans sa langue dominante (calculé si possible sur l'entièreté de sa vie), il doit présenter des performances en dessous de -1.5 ou -1.75 écart-type (ET) à deux tâches d'évaluation pour qu'un TDL soit suspecté ; un enfant qui possède un taux d'exposition équilibré dans ses deux langues doit performer en dessous de -1.75 ou -2 ET dans deux sous-tests pour présenter potentiellement un TDL ; enfin, dans le cas où l'enfant peut être évalué uniquement dans sa langue minoritaire, ses scores doivent se situer entre -2.25 et -2.5 ET.

3.2.2 Questionnaire parental

Les parents des participants bilingues ont rempli une version revisitée et raccourcie du questionnaire PabiQ en français ou en portugais (Tuller, 2015 ; voir annexe E pour le français). Nous avons décidé de fournir ce document à l'écrit aux parents afin de ne pas consacrer une séance supplémentaire pour compléter le questionnaire et, dans les cas où les séances se sont déroulées dans le contexte scolaire, nous n'avons pas rencontré les parents. Nous avons cherché à connaître l'âge du début d'exposition ainsi que le taux d'exposition actuelle aux langues par l'enfant. Nous avons ainsi pu déterminer sa langue dominante actuelle (nous étant utile pour la Partie III de cette thèse). Le questionnaire relevait également l'âge des premiers mots et la

combinaison de deux mots, l'inquiétude parentale par rapport au langage de l'enfant et la présence de difficultés langagières dans l'entourage familial, dans le but de pouvoir déterminer des facteurs de risque de TDL (Tuller, 2015) (voir chapitre 4 pour les différents résultats). Enfin, nous avons également cherché à connaître le niveau de formation des parents. Nous avons laissé de côté certaines informations intéressantes telles que la qualité d'exposition ou encore l'exposition aux langues dans la prime enfance puisque ces informations n'auraient pas été traitées dans cette thèse de doctorat, s'attendant déjà à de nombreuses investigations. Toutefois, nous évoquerons, dans la partie consacrée aux limites de cette recherche, les impacts de nos choix.

3.2.3 Situations d'évaluation dynamique

Les enfants ont également pris part à deux situations d'évaluation dynamique qui se sont focalisées sur les domaines lexicaux et (morpho)syntaxiques. Les enfants ont ainsi dû apprendre de nouveaux substantifs, verbes et structures syntaxiques (flexion verbale) créées par nos soins. Les non-mots à apprendre ont été conçus afin de ne correspondre à aucun mot de la langue française et portugaise pour éviter tous les biais linguistiques possibles (Kapantzoglou et al., 2012). Nous avons également vérifié qu'ils soient composés des phonèmes communs aux deux langues, qu'ils soient bien distincts les uns des autres et qu'ils respectent les règles phonotactiques du français et du portugais (longueur et complexité notamment). Les détails quant à la création des items et la liste de ceux-ci sont présents dans l'Annexe A (et dans l'article présenté au chapitre 7). Les situations d'évaluation dynamique ont été créées sur Microsoft PowerPoint et les personnages et objets à l'aide de plusieurs sites internet (*animatron* pour le fond des histoires, ; *nounproject* pour les personnages et objets, et le logiciel Microsoft Paint pour modifier ceux-ci). Les situations d'apprentissage poursuivaient un objectif de diagnostic (Hessels et Hessels-Schlatter, 2010) et des procédures ont été mises en place en vue d'une

potentielle standardisation. Nos deux situations d'évaluation dynamique variaient quant aux procédures utilisées. Une des situations d'évaluation dynamique a été conçue comme un jeu sur ordinateur dans lequel des nouveaux mots étaient appris à l'enfant grâce à une voix préenregistrée et identique selon les enfants. Le script ainsi que des captures d'écrans du jeu sont présents dans l'Annexe B. Pour concevoir ce jeu, nous nous sommes largement inspirées de la procédure utilisée et détaillée par Skoruppa (2019). L'expérimentateur était présent durant la tâche et fournissait des feedbacks généraux notamment pour motiver l'enfant à poursuivre la tâche. Des commentaires correctifs étaient donnés par l'ordinateur et la voix préenregistrée. Les non-mots à apprendre ont été enregistrés par une locutrice suédoise afin que l'accent tonique soit sur la dernière syllabe comme cela peut être le cas en français et en portugais, et qu'aucun des enfants ne soient avantagés quant à l'intonation et l'accent utilisés pour produire les items. Dans l'autre situation, l'expérimentateur jouait un rôle actif en racontant plusieurs histoires à l'enfant dans laquelle il intégrait des nouveaux non-mots. Les histoires émanaient de notre imagination et sont présentées et accompagnées par des illustrations dans l'Annexe C. En plus de sa présence et de son interaction durant toute la durée de la situation, l'expérimentateur fournissait des indices graduels (Campione et Brown, 1987) à l'enfant durant les tâches d'évaluation de désignation et de dénomination en respectant un canevas préétabli et identique pour tous les participants (voir Annexe H et article au chapitre 7). En effet, pour évaluer la capacité d'apprentissage des enfants, plusieurs tâches d'évaluation ont été créées et suivaient les deux situations d'apprentissage : une désignation pour le lexique réceptif, une tâche d'évocation et de dénomination pour le lexique productif, une tâche de jugement d'acceptabilité pour évaluer la précision des représentations phonologiques et sémantiques des enfants ainsi qu'une tâche de rappel différé pour évaluer la mise en mémoire à long terme des items. Les détails de création de ces tâches sont présents dans l'Annexe D (et dans l'article au chapitre 7).

Tous les tests ont été administrés par une locutrice native de la langue testée familière avec l'évaluation logopédique ; il s'agissait de l'auteure de la thèse ainsi que d'une étudiante lusophone en Master de Logopédie. Des étudiantes en première année de Master en Logopédie participant au séminaire de « travaux pratiques de recherche » ont également apporté leur aide dans la passation de certains sous-tests statiques. Seule la tâche d'évaluation dynamique comprenant la lecture d'histoires a toujours été effectuée par l'auteure de la thèse, étant également logopédiste sur le terrain, afin de ne pas faire varier l'étayage et le style d'aide fourni aux enfants.

3.3 Procédure de passation

Les participants ont pris part à trois séances d'évaluation d'environ une heure (résumées dans le Tableau 2 ci-dessous) majoritairement à leur domicile familial. Quelques participants ont été vu dans les cabinets de logopédie, à l'école ou dans les structures parascolaires mais il s'agit d'exception. Chaque séance a été espacée entre quatre (minimum) et 15 jours (maximum) d'intervalle en fonction des disponibilités communes.

Tableau 2

Descriptif des séances d'évaluation

Séances		Tâches		
Séance 1	RNM en français + <i>cross-linguistique</i>	Tâche de raisonnement non-verbal	Évaluation dynamique 1	
Séance 2	Rappel différé séance 1	Dénomination en français	Compréhension orale en français	Évaluation dynamique 2
Séance 3	Rappel différé séance 2	Audiomètre	Tests en portugais	

Les séances étaient filmées au moyen d'une caméra vidéo et du logiciel de capture d'écran *Filmora Screen* afin de pouvoir procéder aux transcriptions et analyses après les évaluations. Le premier rendez-vous consistait à faire connaissance avec l'enfant et ses parents et à répondre aux éventuelles questions. Nous profitons également de faire signer le consentement parental si ce n'était pas encore fait à ce moment. L'enfant et l'expérimentateur s'installaient ensuite dans un endroit calme afin de commencer les évaluations. Celles-ci débutaient par les répétitions de non-mots en français et *cross-linguistique*. S'ensuivait la tâche de raisonnement non-verbal. La séance se terminait par une des situations d'évaluation dynamique (celles-ci ont été contrebalancées). Lors du second rendez-vous, un rappel différé des items présentés dans la tâche d'évaluation dynamique a été effectué. Nous proposons également deux sous-tâches du test standardisé en français, à savoir la dénomination et la compréhension orale de phrases. La séance se terminait par la deuxième situation d'évaluation dynamique. Le dernier rendez-vous comprenait le rappel différé des items présentés dans la deuxième tâche d'évaluation dynamique ainsi que l'audiomètre afin de s'assurer des capacités

auditives des enfants. Nous avons décidé de réaliser cet examen lors de la dernière séance car l'appareil utilisé est méconnu des enfants et est parfois surprenant (notamment à cause de l'utilisation de grands casques). Nous souhaitions ainsi avoir établi un rapport de confiance avec l'enfant avant de pouvoir effectuer ce test. Pour les enfants bilingues, la séance se terminait par les évaluations langagières en portugais en présence de notre assistante lusophone. Nous avons décidé de réaliser ces épreuves lors de la dernière séance afin de faire venir notre assistante qu'à une seule reprise notamment. À la fin de cette dernière séance, les enfants recevaient un livre en guise de compensation et récompense.

3.4 Procédure de cotation et analyse

À la suite des évaluations, des transcriptions et des analyses ont été effectuées à l'aide des enregistrements vidéo et des captures d'écran. En ce qui concerne les épreuves issues de tests langagiers et cognitifs, nous avons calculé les scores des enfants et nous les avons comparés à l'étalonnage de référence fourni par les tests. Pour les tâches de répétition de non-mots (en français, en portugais et *cross-linguistique*), nous avons procédé à la transcription phonétique selon l'alphabet phonétique international des items ainsi qu'à l'analyse des items selon une catégorisation des erreurs que nous avons préalablement établie (voir Annexe F et point 6). En ce qui concerne l'évaluation dynamique, une transcription des items produits par les enfants a été réalisée en utilisant également l'alphabet phonétique international. Nous avons également mis sur pieds plusieurs catégories d'analyses pour les erreurs réalisées par les enfants (voir Annexe G). Nous avons également transcrit et codé toutes les interventions de l'expérimentateur selon plusieurs catégories d'analyses (voir également l'Annexe G). Nous avons calculé le nombre d'items correctement produit/désigné par l'enfant par situation (jeu sur ordinateur, lecture d'histoires), par tâches évaluatives (ex. dénomination, évocation, désignation) et par items (substantifs, verbes, phrases). Nous avons également cherché à

compter le nombre d'aide fournit dans la situation de lecture d'histoires en comptabilisant le nombre d'indices données par l'expérimentateur dans les tâches de dénomination et de désignation. Nous avons également analysé la complexité des indices fournis aux enfants (l'étayage étant graduel). Enfin, dans la situation de lecture d'histoires, nous avons pu comparer les performances des enfants avant et après étayage. Les différentes analyses ont été conduites en comparant nos différents groupes d'enfants (avec et sans TDL, monolingues et bilingues). En prenant en considération toutes les performances individuelles, nous avons également calculé les pourcentages de précision diagnostique de nos deux situations d'évaluation dynamique en nous basant sur Plante et Vance (1994) qui considère un test comme précis à partir de 80 % de sensibilité et spécificité. Nous renvoyons le lecteur aux différents articles présentés dans cette thèse (points 4, 6 et 7) afin d'avoir davantage de détails sur les analyses effectuées.

3.4.1 Accord inter-juge

Plusieurs accords inter-juge ont été réalisés sur les données de RNM en français et *cross-linguistique* ainsi que sur les données d'évaluation dynamique. Les deuxièmes juges, respectivement des étudiants de Master en Logopédie et de Bachelor en Lettres et Sciences humaines, avaient été formés préalablement à l'analyse de nos données avant de procéder au codage à l'aveugle. Nous avons utilisé le calcul de Kappa de Cohen's pour les données binaires et le calcul du coefficient de corrélation interclasse pour les données quantitatives à intervalles sur le logiciel SPSS (Version 27.0, IBM Corp., 2020). Les niveaux d'accords sont présentés dans les articles respectifs traitant des différentes thématiques : au point 6 pour les RNM et au point 7 pour les situations d'évaluation dynamique.

3.5 Analyses statistiques

Pour traiter nos données empiriques, nous avons réalisé des analyses statistiques. Les analyses statistiques descriptives et inférentielles ont été réalisées en utilisant les logiciels SPSS (Version 27.0, IBM Corp., 2020) ainsi que R (Version 4.0.3, R Development Core Team, 2019). Les tests utilisés dépendent de nos données (ex. distribution normale ou non) et du nombre de participants (Field, 2018). Les différences quant aux tests utilisés sont détaillées dans les articles (voir points 4, 6 et 7) présentés au fil de cette thèse de doctorat.

Partie III : L'évaluation dans les deux langues de l'enfant

4. Article de recherche

L'identification du trouble développemental du langage oral chez les enfants bilingues français-portugais : les ressources à disposition des orthophonistes

Schwob. S. et Skoruppa, K.

État : accepté pour publication le 9 février 2021 dans « *Revue canadienne d'orthophonie et d'audiologie* »

Résumé

L'évaluation du langage oral des enfants bilingues constitue actuellement un défi pour les orthophonistes. La présente étude a pour objectif de dresser un état des lieux des ressources disponibles sur le terrain et de tester leur efficacité auprès d'une population constituée d'enfants bilingues français-portugais en Suisse romande. Trente et un enfants bilingues âgés entre 5;0 et 7;11 ans (15 avec un développement typique et 16 présentant un trouble développemental du langage) ont réalisé plusieurs épreuves langagières issues de tests standardisés en français et en portugais. Nous avons testé différents modes d'évaluation se situant sur un continuum allant de l'évaluation dans une langue à l'évaluation dans les deux langues de l'enfant et passant par des méthodes alternatives, 1) tests en français avec normes monolingues avec seuils de pathologie habituellement utilisés en clinique, 2) tests en français avec adaptation des seuils de pathologie en fonction de l'exposition langagière selon les recommandations de Thordardottir (2015b), 3) tests dans la langue dominante de l'enfant sans puis avec adaptation des seuils de pathologie (Thordardottir, 2015b), 4) évaluation dans les deux langues de l'enfant. Les parents ont également rempli un questionnaire (basé sur Tuller, 2015) s'intéressant à leurs pratiques langagières et permettant de dégager des facteurs de risque de difficultés langagières. Les résultats mettent en exergue l'atout discriminant de la tâche de répétition de non-mots en français avec adaptation des seuils de pathologie en fonction du taux d'exposition obtenant une sensibilité de 81 % et une spécificité de 80 %. Les autres sous-tests quelque que soit le mode d'évaluation se montrent moins discriminants. Certaines données issues du questionnaire parental ont également montré des différences entre les deux groupes d'enfants mais n'ont pas permis d'améliorer le taux de discrimination. Nous discuterons des impacts de ces résultats pour la clinique orthophonique.

Mots clés

Évaluation orthophonique bilingue, trouble développemental du langage oral, bilinguisme

Introduction

Les troubles développementaux du langage oral et leur évaluation orthophonique

La majorité des enfants, qu'ils soient exposés à une ou plusieurs langues, développent un langage oral sans difficultés apparentes (Kohnert, 2010). Toutefois, environ 10 % des enfants âgés entre 4 et 5 ans sont touchés par des troubles du langage (Norbury et al., 2016). Chez les trois quarts environ, il s'agit de « troubles développementaux du langage oral » (TDL, voir Bishop et al., 2017) qui apparaissent sans cause apparente, c'est-à-dire en l'absence de conditions biomédicales sous-jacentes et de déficience intellectuelle attestée (Norbury et al., 2016). Malgré des décennies de recherche concernant ces troubles, leur étiologie reste floue et leurs causes sont difficilement définissables mais semblent multifactorielles, incluant des facteurs génétiques (Leonard, 2014).

Pour identifier un TDL, les tests de langage standardisés sont importants en clinique car ils permettent des comparaisons des enfants à une norme développementale (Leonard, 2014). Les écarts considérés comme pathologiques dépendent des tests et sous-tests utilisés et varient entre -1 et -2 écart-types¹¹ par rapport à la moyenne de la tranche d'âge en question (Thordardottir et al., 2011). Les tests standardisés doivent être normalisés, validés, fidèles et précis (Lafay et Cattini, 2018). Leur validité discriminante est liée à leur sensibilité (pourcentage d'enfants avec troubles identifié correctement) et leur spécificité (pourcentage d'enfants sans trouble rejeté correctement : ex. Thordardottir et al., 2011). L'indice de Youden (1950) permet de combiner la spécificité et la sensibilité afin de déterminer le meilleur compromis entre les deux valeurs ainsi que le meilleur score seuil. Selon Plante et Vance (1994), un test peut être considéré comme discriminant s'il permet de classer adéquatement entre 80 et 100 % de la population. En pratique, la valeur discriminante attestée par la plupart

¹¹ Certains tests fournissent des scores mesurés en écart-types (*ET*) et d'autres en percentiles (*P*).

des tests concerne généralement une population monolingue, et n'a été examinée que rarement chez des enfants bilingues.

L'évaluation orthophonique chez les enfants bilingues

Le bilinguisme

Dans cet article, nous adopterons la définition du bilinguisme par l'usage, qui est actuellement la plus utilisée en clinique et en recherche francophone. Ainsi, Grosjean (2015) définit le bilinguisme comme « la maîtrise d'au moins une compétence linguistique dans une autre langue » (2015, p.16), utilisée pour communiquer dans la vie de tous les jours. Selon cette définition, plus de 50 % de la population dans le monde serait bilingue (Grosjean, 2015), et plus de deux tiers de la population en Suisse, où notre étude a été menée (Office fédéral de la statistique, 2020).

Bien souvent, en fonction des opportunités d'apprentissage et d'utilisation, les locuteurs bilingues peuvent avoir une langue dominante (ex. Kohnert, 2010). Cette dominance peut varier en fonction de l'âge, des contextes et des opportunités d'apprentissage. Même si les premières grandes étapes (premiers mots et combinaisons de mots) sont franchies au même moment chez les enfants bilingues et monolingues, leur développement linguistique présente aussi des différences quantitatives et qualitatives (Zurer Pearson, 2013). En plus des différences interindividuelles entre les enfants, le développement des enfants bilingues est encore plus hétérogène et dépend de nombreux facteurs additionnels. Nous pouvons évoquer la quantité et la qualité de l'exposition aux langues qui sont primordiales pour que les enfants développent leurs langues de manière harmonieuse (Paradis, 2019). Obtenir une exposition suffisante (Simard et al., 2015) peut être problématique dans certains contextes bilingues, surtout pour une langue jugée comme minoritaire par la société (Dubreuil-Piché et al., 2017). La typologie et la proximité des langues de l'enfant peuvent aussi faire différer le développement linguistique

et la quantité d'exposition nécessaire à son apprentissage (Paradis, 2019). De plus, les éventuels écarts entre les enfants bilingues peuvent aussi varier en fonction du domaine linguistique. Par exemple, Thordardottir (2015a) relève que le développement morphosyntaxique bilingue est influencé par la quantité d'exposition, mais à moindre mesure que pour le vocabulaire (Thordardottir, 2011). La modalité linguistique joue également un rôle ; l'écart entre vocabulaire réceptif et expressif pouvant être plus grand chez les enfants bilingues que chez les enfants monolingues (Keller et al., 2015). Finalement, le statut des langues, le parcours migratoire (Di Meo et al., 2014), la motivation, le style d'apprentissage et la confiance en soi de l'enfant (Paradis et al., 2011) peuvent avoir une incidence sur le développement langagier de l'enfant bilingue. Vu cette multitude de facteurs influençant le niveau de compétences des enfants bilingues dans leurs différentes langues, il n'est pas étonnant qu'il soit difficile d'établir des normes développementales bilingues (Zurer Pearson, 2013) et que l'identification d'un TDL chez les enfants bilingues soit souvent caractérisé comme le plus grand défi dans la clinique orthophonique actuelle (Armon-Lotem, 2012).

Les défis actuels de l'évaluation orthophonique des enfants bilingues

L'apprentissage de plusieurs langues ne semble pas être un avantage ou un désavantage en ce qui concerne l'apparition d'un TDL (Kohnert, 2010). Cependant, il est souvent difficile pour les cliniciens de faire la différence entre des difficultés relevant d'un TDL et des difficultés passagères dues à une exposition insuffisante à une langue (Camilleri et Law, 2007). En effet, les enfants bénéficiant de moins d'exposition à la langue d'évaluation sont fréquemment mal catégorisés ou sur-identifiés car ils échouent les épreuves proposées sans que cela puisse être imputé à de réelles difficultés langagières (Camilleri et Law, 2007). De plus, certains patrons d'erreurs sont similaires chez des bilingues apprenants et dans le contexte de pathologie (Armon-Lotem, 2012). En français, par exemple, les omissions des objets clitics sont

considérées comme un marqueur de pathologie mais également comme des erreurs fréquemment réalisées par les enfants bilingues sans TDL peu exposés au français (Tuller et al., 2018). Chez ces enfants, les erreurs sont, par conséquent, temporaires et diminueront au gré des expériences langagières de l'enfant. À contrario, dans un contexte de pathologie, les erreurs sont persistantes et indépendantes du degré d'exposition langagière. Inversement, la sous-identification d'un TDL est aussi un risque, consistant à attribuer les difficultés langagières des enfants bilingues à un manque d'exposition ou à l'acquisition de plusieurs langues, même si les difficultés sont massives et persistantes et relèvent, en réalité, d'une pathologie (Grimm et Schulz, 2014).

Pour éviter ces erreurs d'identification, plusieurs organisations (ex. American Speech-Language-Hearing Association., s. d.) déconseillent fortement d'évaluer les enfants bilingues dans une seule langue et recommandent de tester les deux langues puisque les difficultés langagières sont présentes dans chacune de leurs langues (Kohnert, 2010) et également pour obtenir un profil langagier complet (Peña et al., 2016). Cependant, relativement peu de recherches ont comparé différents modes d'évaluation chez les enfants bilingues, et une évaluation strictement bilingue s'avère actuellement difficile en clinique orthophonique (Estienne et Vander Linden, 2014), pour plusieurs raisons.

Premièrement, il n'existe peu ou pas de tests standardisés dans de nombreuses langues, y compris certaines langues européennes comme le portugais par exemple (de Almeida et al., 2016). Un deuxième problème concerne les aspects culturels. En effet, pour le français, il existe un large échantillon de tests créés et normés auprès d'une population française. Toutefois, même l'utilisation de ces tests dans des pays voisins (ex. Suisse) peut déjà conduire à une évaluation biaisée à cause d'importantes différences culturelles et régionales (Grégoire, 2006) ; et ce biais ne risque que d'augmenter dans le cas d'enfants bilingues dont les familles viennent de cultures plus éloignées (White et Jin, 2011). Il faut également relever que les orthophonistes

ne parlent souvent pas la deuxième langue de l'enfant et qu'il n'est pas aisé d'avoir à disposition et/ou de pouvoir financer des interprètes compétents dans le domaine orthophonique (Boerma et Blom, 2017). Un dernier défi, et non des moindres, est que la grande majorité des tests n'est pas normée pour les enfants bilingues. Quelques exceptions concernent le bilinguisme anglais-espagnol pour lequel il existe quelques outils d'évaluation normés sur une population bilingue (ex. BESA de Peña et al., 2018). La conception et la validation de tels instruments n'est pas aisée puisqu'il faut prendre en considération, comme évoqué ci-dessus, que l'expertise langagière dépend de nombreux facteurs (Paradis, 2019). Il est donc difficile de collectionner des échantillons de références bilingues suffisamment larges et homogènes pour déterminer ce qu'un enfant bilingue devrait, dans un contexte d'exposition spécifique, connaître dans telle ou telle langue.

Ainsi, de nombreuses solutions hybrides ont été proposées pour la clinique (ex. Thordardottir, 2015b), que l'on peut envisager sur un continuum entre une évaluation strictement « monolingue » versus « bilingue » (ex. Kehoe, 2009) et qui sont expliquées ci-dessous.

Quelques solutions pour la clinique

Évaluation d'une seule langue avec alternatives.

Adaptation du seuil de pathologie. Un taux d'exposition suffisant est primordial pour développer des compétences dans une langue. Selon Thordardottir (2015b), l'orthophoniste peut adapter les critères de déviation à la norme des tests monolingues en fonction du taux d'exposition des langues de l'enfant bilingue, en allant de -1.5 *ET* à - 2 *ET* pour une évaluation dans la langue dominante de l'enfant (au lieu de -1.25 *ET* pour un enfant monolingue) à -2.25 à -2.5 *ET* pour une évaluation dans une langue non-dominante.

Évaluation de la langue dominante. Lorsque la dominance linguistique peut être clairement identifiée chez l'enfant (ex. via un questionnaire parental), une alternative est d'évaluer uniquement cette langue (American Speech-Language-Hearing Association., s. d. ; Thordardottir et al., 2006) en adaptant les seuils de pathologie comme décrit ci-dessus (Thordardottir, 2015b). Bien entendu, des biais ne sont non négligeables comme la non-obtention d'un profil complet des capacités et difficultés langagières de l'enfant.

Recueil d'informations auprès de l'entourage. Le questionnaire parental semble être un outil complémentaire dans le contexte d'évaluation des enfants bilingues. Il fournit des informations sur la qualité et la quantité d'input reçu dans les langues, sur le développement du langage et permet de relever la présence de facteurs de risque d'un TDL (Boerma et Blom, 2017), tels que l'apparition tardive des premiers mots et/ou de la combinatoire, des difficultés langagières dans l'entourage familial et/ou des capacités langagières considérées comme faibles par les parents (Tuller, 2015). Ces informations sont précieuses pour identifier les enfants parleurs tardifs (Zurer Pearson, 2013), dont la prévalence s'élève à environ 13 % (Rescorla, 2011). Toutefois, trois quarts de ces enfants parleurs tardifs se développent ensuite normalement sans présenter de TDL (Rescorla, 2011). C'est pour cette raison que le questionnaire parental n'est pas utilisé seul mais de façon complémentaire avec d'autres outils lors de l'évaluation.

Utilisation de tâches indépendantes d'une langue spécifique. De nombreux auteurs proposent d'utiliser des tâches qui ne soient pas ou peu affectées par les connaissances préalables ou l'expérience de l'enfant dans une langue précise (Campbell et al., 1997). La tâche la plus souvent évoquée dans ce contexte et qui est incluse dans de nombreux tests cliniques disponibles dans plusieurs langues est la répétition de non-mots¹², qui sera détaillée ci-dessous.

¹² Nous allons utiliser de manière synonyme les termes de « pseudo-mots », « non-mots » et de « logatomes » tous deux étant considérés comme du matériel verbal dépourvu de lexicalité mais étant prononçable.

Notons que pour le présent article qui s'intéresse aux outils déjà disponibles sur le terrain, nous avons dû laisser de côté certaines possibilités récentes et prometteuses, telles que l'évaluation dynamique (voir Hasson et Botting, 2010), la réponse à l'intervention (voir Lidz et Peña, 2009) et les tâches narratives (voir Boerma et Blom, 2017).

La tâche de répétition de non-mots. L'échec à une tâche de répétition de non-mots (RNM) est actuellement reconnu comme étant un marqueur de TDL chez une population monolingue et bilingue de différentes langues (Chiat, 2015 ; Dubreuil- Piché et al., 2017). L'utilisation d'unité dépourvu de sens (ex. /sibumakila/, exemple tiré de Chiat, 2015) permet de s'assurer qu'aucun des enfants testés n'ait déjà rencontré ces mots et ainsi d'éviter les biais linguistiques (Chiat, 2015). Toutefois, bien que les résultats à la tâche de RNM soient moins affectés par les connaissances langagières préalables de l'enfant, les scores semblent, tout de même, affectés par les spécificités de ses langues (Messer et al., 2010). En effet, plusieurs études (ex. Boerma et al., 2015) montrent que les scores obtenus par les enfants sont dépendants de leur expérience langagière dans la langue de passation. Malgré cela, certains auteurs (ex. de Almeida et al., 2016) obtiennent des résultats de spécificité et sensibilité élevés pouvant justifier la tâche de RNM comme étant discriminante et peu dépendante du taux d'exposition langagier. Les différences quant aux résultats obtenus peuvent, en partie, s'expliquer par des différences méthodologiques (Boerma et Blom, 2017).

Objectifs visés par l'étude

La présente étude souhaite mettre en pratique les différentes alternatives d'évaluation présentées dans l'introduction auprès d'une population d'enfants parlant le français et le portugais âgée entre 5;0 et 7;11 ans afin de déterminer quelle tâche et quel mode d'évaluation sont les plus prometteurs en clinique pour cette population.

La combinaison de ces deux langues n'est pas seulement la plus fréquente chez les bilingues en Suisse romande (Office fédéral de la statistique, 2020) mais aussi chez les enfants pris en charge en orthophonie (Volpin et al., 2020). Nous allons comparer les performances des enfants bilingues avec un TDL avéré et avec un développement typique (DT) à différentes épreuves dans les deux langues. Étant donné que les enfants de cette étude participeront également à d'autres épreuves dans le cadre d'une recherche doctorale plus large, nous nous limiterons à trois tâches souvent utilisées en clinique, évaluant différents domaines et modalités langagières nous permettant d'observer les difficultés langagières souvent relevées chez les enfants TDL (Bishop et al., 2017). Les tâches choisies sont : la dénomination pour le lexique productif, la compréhension orale de phrases pour la morphosyntaxe réceptive ainsi qu'une tâche de répétition de non-mots pour le traitement phonologique. Les domaines lexicaux et syntaxiques sont fréquemment évalués pour identifier un TDL (Denman et al., 2017) et la répétition de non-mots semble fréquemment discriminante dans le domaine de la phonologie (Boerma et Blom, 2017). Nous allons aussi analyser les réponses données par les parents à un questionnaire parental (basé sur Tuller, 2015), vu son utilisation dans plusieurs études (ex. de Almeida et al., 2016).

Pour les évaluations standardisées, les modes d'évaluation qui seront utilisées se situent sur un continuum allant d'une évaluation dans une langue (français) à une évaluation dans les deux langues (français et portugais). Plus précisément, nous débuterons par une évaluation en français avec l'utilisation de seuils de pathologie habituellement utilisés en clinique pour les enfants monolingues. Ensuite, deux méthodes alternatives à une évaluation dans une langue et dans deux langues seront réalisées : l'évaluation en français avec les seuils de pathologie ajustés en fonction du taux d'exposition langagier (selon Thordardottir, 2015b) ainsi que l'évaluation dans la langue dominante des enfants sans puis avec adaptation des seuils de pathologie

(Thordardottir, 2015b). Enfin, nous effectuerons également une évaluation dans les deux langues de l'enfant (français et portugais).

En plus de constater s'il existe une différence de performance significative entre les enfants bilingues avec un DT et avec un TDL pour chaque outil et mode d'évaluation par des tests inférentiels, nous allons également déterminer leur valeur discriminante, en calculant leur sensibilité et leur spécificité pour différents seuils, dont le meilleur sera choisi par le calcul de l'indice de Youden (1950).

Méthode

Participants et procédure de recrutement

Au total 36 enfants âgés entre 5;0 et 7;11 ans bilingues français-portugais vivant en Suisse romande ont pris part à la recherche. Ils devaient avoir un minimum de 25 % d'exposition au portugais et ne devaient pas être exposés à plus de 15 % à une troisième langue (données récoltées via un questionnaire parental). Les enfants ne devaient pas avoir été identifiés, avoir une suspicion de trouble du spectre autistique, de trouble cognitif, socio-affectif et auditif. Ainsi, les résultats de cinq enfants parmi les 36 participants ont dû être écartés, soit parce que leurs performances au screening audiométrique (seuil d'audibilité fixé à 20dB maximum dans les fréquences de 500, 1000, 2000 et 4000 Hertz) ($n = 2$) ou au screening non-verbal effectué via le test Wide-Range de Glutting et al. (2000) ($n = 1$) étaient en dessous de la norme, soit parce qu'ils n'avaient pas un taux d'exposition suffisant au portugais ($n = 2$). Parmi les 31 enfants retenus, 16 présentaient un TDL avéré : ils étaient diagnostiqués et suivis par un orthophoniste depuis au moins trois mois. Les 15 autres enfants avaient un DT et ne devaient jamais avoir été identifiés comme ayant un TDL et/ou avoir été suivis en orthophonie.

Les familles intéressées à participer à la recherche recevaient un document d'information, un questionnaire parental et une déclaration de consentement libre et éclairé,

validés par la commission d'éthique locale (CER-Vaud ; n° de projet : 2017-01881, autorisation accordée le 07.02.2018). Les séances d'évaluation se sont majoritairement déroulées au domicile familial. Cette étude s'intégrant dans une recherche plus large de doctorat, les participants ont participé à 3 séances d'environ une heure incluant également d'autres tâches langagières.

Matériel

Sous-tests langagiers

Les enfants bilingues ont été évalués dans chacune de leurs langues à l'aide de sous-tests utilisant les trois tâches expliquées ci-dessus. Ils sont issus de standardisés utilisés largement en clinique ayant des normes monolingues pour les âges investigués dans la présente étude. En français, deux sous-tests du test « Évaluation du Langage Oral » (ELO) (Khomsî, 2001), les tâches de dénomination lexicale et de compréhension immédiate ont été utilisées. La répétition de logatomes du test Exalang a été sélectionnée, en utilisant la version Exalang 3-6 (Helloin et Thibault, 2006) ou Exalang 5-8 (Thibault et Helloin, 2010) en fonction de l'âge de l'enfant. Nous avons tenté de sélectionner des tâches évaluant les mêmes modalités en portugais. Nous avons utilisé deux sous-tests issus du test « Avaliação da Llinguagem Oral » (ALO) (Sim-Sim, 2006) : les tâches de dénomination lexicale ainsi que de compréhension syntaxique. Pour la répétition de non-mots, le sous-test issu de la batterie « Provas de Avaliação da Linguagem e da Afasia em Português » (PALPA-P) (Castro et al., 2007) a été choisi bien que la population sur laquelle la standardisation a été réalisée est moindre (de Almeida et al., 2016). A noter que pour les tests lusophones, peu de tranches d'âges sont couvertes par la normalisation (ex. 4 et 6 ans ne sont pas couverts dans le PALPA-P). Tous les tests ont été administrés par une locutrice native de la langue testée familière avec l'évaluation

orthophonique ; il s'agissait de la première auteure ainsi que d'une étudiante lusophone en Master de Logopédie.

Questionnaire parental

Les parents des participants ont rempli une version revisitée du questionnaire PabiQ (Tuller, 2015). Afin de pouvoir calculer le taux d'exposition et déterminer la langue dominante de l'enfant, les parents indiquaient la fréquence d'utilisation du français et du portugais à la maison et à l'école sur une échelle à cinq gradients : « presque toujours (90-10 %) », « souvent (65-90 %) », « la moitié du temps (35-65 %) », « rarement (10-35 %) », « jamais (0-10 %) ». Le questionnaire relevait également l'âge du début d'exposition aux langues, des premiers mots et la combinaison de deux mots ainsi que l'inquiétude parentale par rapport au langage de l'enfant (oui/non), la présence de difficultés langagières dans l'entourage familial (oui/non) et le niveau de formation des parents (échelles de 1 à 5 : 1 = école primaire, 2 = école secondaire, 3 = apprentissage, 4 = maturité gymnasiale, 5 = Hautes-Écoles).

Procédure de cotation et d'analyse

Sous-tests langagiers

En fonction des sous-tests et de l'âge de l'enfant, des scores ont été calculés et comparés à l'étalonnage de référence afin d'obtenir des scores de déviation à la moyenne. Pour les tests ELO (Khomsî, 2001) et Exalang 5-8 (Thibault et Helloin, 2010), l'âge de l'enfant est exprimé en termes de classes scolaires françaises. Puisque les systèmes scolaires français et suisse sont différents et que plusieurs enfants de notre échantillon étaient scolarisés dans des écoles spécialisées, nous avons converti les classes scolaires en équivalent d'âge chronologique (ex. grande section maternelle = 5 ans). Nous avons ainsi toujours comparé les enfants à l'étalonnage correspond à leur âge chronologique. Au niveau lexical, certains synonymes

fréquemment utilisés en Suisse ont été acceptés comme réponse correcte (ex. poussette pour landau). Les critères d'arrêt proposés en fonction de l'âge de l'enfant ont été respectés. Concernant les tests lusophones, c'est l'âge chronologique qui a été utilisé. Pour le test PALPA-P, il faut noter que la standardisation a eu lieu sur une population assez restreinte (de Almeida et al., 2016). Puisque la normalisation des tests lusophones ne couvre pas l'ensemble des âges de notre recherche, nous avons donc réalisé des interpolations et extrapolations linéaires (formule utilisée : $\text{moyenne âge 1} + (\text{âge 1} / \text{âge 2} - \text{âge 1}) * (\text{moyenne âge 2} - \text{moyenne âge 1})$). Plus précisément, pour le sous-test de Palpa-P, les scores de 3 enfants ont été définis par interpolation linéaires et pour les sous-tests de ALO, les scores de 20 enfants ont été obtenus par interpolation et/ou extrapolation linéaire.

Pour cette étude, le critère de déviation à la norme retenu pour l'évaluation monolingue a été fixé au percentile 10 soit à $-1,28 ET$ (en référence à Thordardottir et al., 2011). Par contre, dans les autres modes de cotation, les données du questionnaire nous permettent de réaliser une adaptation des seuils de déviation à la norme monolingue en prenant en considération les recommandations de Thordardottir (2015b). En fonction de l'exposition langagière des enfants à ses deux langues, nous avons adopté les seuils suivants : $-1.5 ET$ (percentile 7) lorsque l'enfant bilingue est évalué dans sa langue dominante et $-2.25 ET$ (percentile 1) lorsque l'enfant bilingue est évalué dans sa langue non-dominante.

Questionnaire parental

L'exploitation du questionnaire parental a permis de dégager l'âge d'exposition, le type de bilinguisme et les taux d'exposition aux langues des enfants. Le type de bilinguisme a été déterminé grâce à l'âge du premier contact aux langues, en utilisant le seuil des 3 ans (Paradis, 2007) comme critère de distinction entre bilinguisme simultané et successif. En fonction de la

réponse du parent à la fréquence d'utilisation des langues à la maison et à l'école, l'enfant a été classé comme dominant en français ou en portugais.

Mesures statistiques

Afin de déterminer s'il existe des différences significatives entre les différents groupes d'enfants, des mesures statistiques non-paramétriques ont été effectuées (test de Mann-Whitney et de X^2) puisque nos données ne suivent pas une distribution normale (au test Shapiro-Wilk, les p valeurs de nos variables dépendantes sont en dessous de $p = .05$, nous avons dû ainsi rejeter l'hypothèse de distribution normale). Nous avons également généré des courbes Roc (Dunn, 2014) comprenant les différents sous-tests utilisés et permettant d'apprécier visuellement l'aire qui se situe sous les courbes et qui représente la précision des sous-tests. Nous avons ainsi pu obtenir les pourcentages de sensibilité et spécificité (selon Plante et Vance, 1994.) Enfin, grâce aux données de sensibilité et spécificité, les indices de Youden (1950) des sous-tests utilisés ont aussi été calculés afin de déterminer les valeurs seuils optimales qui permettent de discriminer le mieux les participants ayant un TDL. Afin de déterminer s'il y avait un effet de l'âge sur les scores en percentile, des tests non-paramétriques de corrélation Tau-B de Kendall ont été effectués.

Résultats et discussion

Participants- données générales

Le Tableau 1 (ci-dessous) résume les données démographiques générales de nos participants. Nous relevons qu'il n'y a pas de différences significatives quant à l'âge médian ($U = 100, p = .446$), au sexe ($X^2 = 0.45, p = .83$), au niveau de formation des parents ($U = 95, p = .338$) et à la langue dominante ($X^2 = 0.20, p = .88$) des participants des deux groupes.

Tableau 1

Description des 31 participants, données démographiques et concernant le bilinguisme

		Groupe	
		DT	TDL
Sexe	filles	9	9
	garçon	6	7
Age médian		86 (62 à 90)	73 (64 à 95)
Niveau d'éducation des parents		3	2.5
Langue dominante	français	9	10
	portugais	6	6

Note. $n = 31$ ($n = 15$ avec un DT et $n = 16$ avec un TDL).

Les données ont été recensées via le questionnaire parental (basé sur Tuller, 2015). L'âge médian a été indiqué en mois. L'étendue des âges par groupe a également été fournie entre parenthèse.

Évaluation en français, données comparées aux normes pour les enfants monolingues francophones

Seuils de pathologie habituellement utilisés en clinique

Tout d'abord, nous nous attelons aux résultats obtenus par les enfants bilingues lorsque le critère de déviation est fixé au percentile 10, comme nous le ferions pour des enfants monolingues (Thordardottir et al., 2011). Les scores standards médians exprimés en percentile par groupe ainsi que les valeurs discriminantes (sensibilité et spécificité) des épreuves réalisées sont présentées dans le Tableau 2 (première colonne) et dans la figure 1, ci-dessous, qui permet d'en apprécier la variance. Seul le sous-test de RNM permet de différencier significativement les enfants bilingues avec et sans TDL ($U = 51, p = .006$), avec une sensibilité de 81 % et une spécificité 73 %. Nous notons également que les deux autres sous-tests s'avèrent non-discriminants (tâche de dénomination : $U = 87, p = .202$; tâche de compréhension immédiate : $U = 78, p = .101$).

Tableau 2

Scores standards médians exprimés en percentile obtenus par les enfants avec un DT et avec un TDL aux sous-tests aux sous-tests langagiers dans leurs deux langues, résultats statistiques au test Mann-Whitney et valeurs discriminantes des sous-tests avec/sans adaptation des seuils de pathologie

Sous-tests	Mdn DT Mdn TDL		DT/TDL	Critère de déviation		Critère de déviation		Critère de déviation	
				sans adaptation des		avec adaptation des		avec adaptation des	
				seuils, $p = 10$		seuils, $p = 7$ ou $p = 1^a$		seuils, $p = 4^b$	
				Sensibilité	Spécificité	Sensibilité	Spécificité	Sensibilité	Spécificité
Dénomination en français	37.50	22.50	$U = 87, p = .202$	40 %	63 %	19 %	100 %	25 %	100 %
Dénomination en portugais	25	15	$U = 88, p = .216$	37 %	73 %			31 %	86 %
Compréhension en français	58.33	31.25	$U = 78, p = .101$	31 %	100 %	25 %	100 %	25 %	100 %
Compréhension en portugais	70	15	$U = 72, p = .060$	50 %	73 %			37 %	86 %

Sous-tests	Mdn DT	Mdn TDL	DT/TDL	Critère de déviation		Critère de déviation		Critère de déviation	
				sans adaptation des		avec adaptation des		avec adaptation des	
				seuils, $p = 10$		seuils, $p = 7$ ou $p = 1^a$		seuils, $p = 4^b$	
				Sensibilité	Spécificité	Sensibilité	Spécificité	Sensibilité	Spécificité
RNM en français	30	1.04	$U = 51, p = .006^{**}$	81 %	73 %	81 %	80 %	56 %	73 %
RNM en portugais	94	63.25	$U = 54, p = .008^{**}$	6 %	93 %			6 %	93 %

Note. Abréviations : Mdn = médiane, RNM= répétition de non-mots. $N = 31$ ($n = 15$ avec un DT et $n = 16$ avec un TDL).

Les épreuves sont issues des tests ELO (Khomsî, 2011), Exalang (Helloin et Thibault, 2006 ; Thibault et Helloin, 2010), ALO (Sim-Sim, 2006) et Palpa-p (Castro et al., 2007).

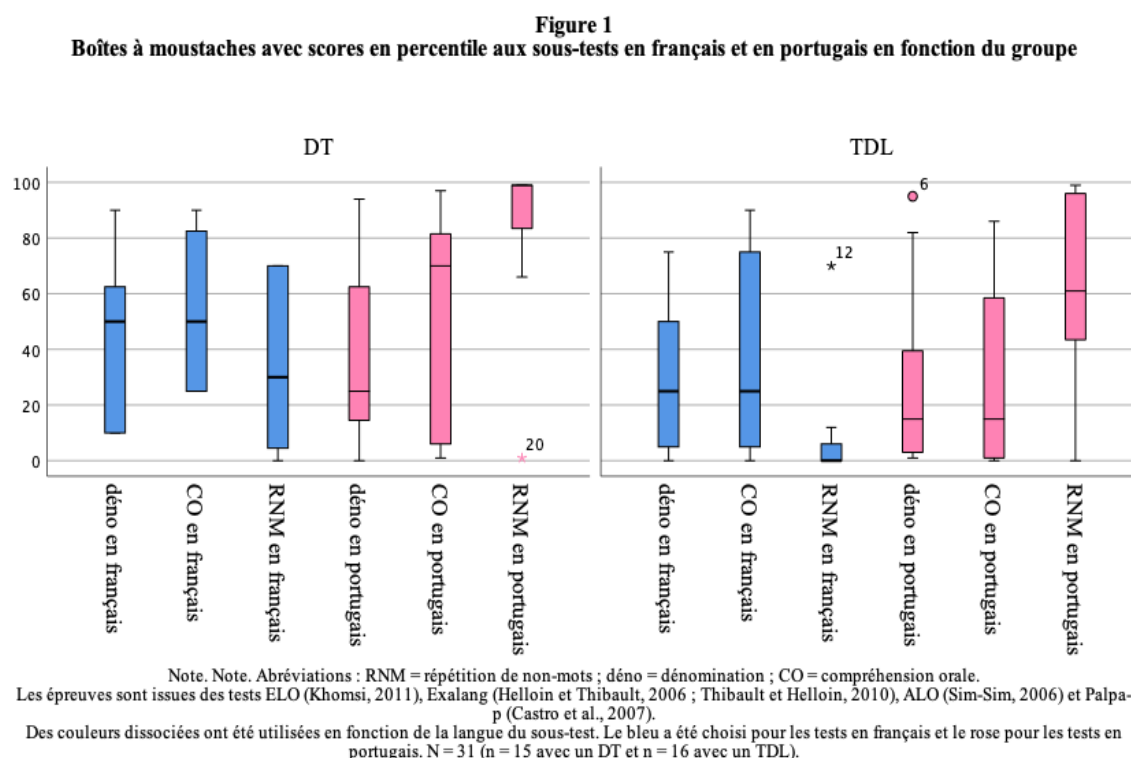
^aAdaptation des critères de déviation selon Thordardottir (2015b) : percentile 7 lorsque l'enfant est testé dans sa langue dominante et percentile 1 quand l'enfant est testé dans sa langue non-dominante.

^bLe critère de déviation a été fixé au percentile 4, comme seuil alternatif considérant ainsi le bilinguisme comme équilibré.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 1

Boîtes à moustaches avec scores en percentile aux sous-tests en français et en portugais en fonction du groupe



À noter, qu’aucune corrélation significative entre l’âge des participants et les scores standards obtenus n’a été trouvée pour les sous-tests en français (tâche de dénomination : $\tau_b = -0.14$, $p = .91$; tâche de compréhension orale : $\tau_b = 0.096$, $p = .48$; tâche de répétition de logatomes : $\tau_b = 0.037$, $p = .79$).

Seuils de pathologie ajustés en fonction de l’exposition langagière

Dans le Tableau 2, exposé ci-dessus (colonne du milieu), nous notons l’obtention de résultats plutôt similaires à ceux réalisés sans adaptation des seuils. Nous relevons que la tâche la plus sensible (81 %) et spécifique (80 %) est toujours la RNM en français. Sa valeur

discriminante se voit légèrement augmentée avec l'adaptation des seuils en fonction de la dominance langagière.

Évaluation de la langue dominante de l'enfant, données comparées aux normes pour les enfants monolingues dans cette langue

Lorsque nous prenons en compte les résultats des sous-tests réalisés dans la langue dominante de chaque enfant, nous relevons des différences significatives entre les enfants avec un DT et avec un TDL pour les sous-tests de compréhension orale ($U = 67.5$, $p = .037$) avec une sensibilité entre 43 % (sans adaptation) et 37 % (avec adaptation) et une spécificité de 80 % avec et sans adaptation des seuils de pathologie (voir tableau 3).

Tableau 3

Scores standards médians exprimés en percentile obtenus par les enfants avec un DT et avec un TDL aux sous-tests langagiers dans leur langue dominante, résultats statistiques au test Mann-Whitney, valeurs discriminantes des sous-tests avec/sans adaptation des seuils de pathologie

Sous-tests	Mdn DT	Mdn TDL	DT/TDL	Critère de déviation sans adaptation des seuils, $p = 10$		Critère de déviation avec adaptation des seuils, $p = 7^a$	
				Sensibilité	Spécificité	Sensibilité	Spécificité
Dénomination langue dominante	52.50	34.33	$U = 80, p = .119$	31 %	93 %	25 %	93 %
Compréhension orale langue dominante	75	25	$U = 67.5, p = .037 *$	43 %	80 %	37 %	80 %
RNM langue dominante	70	9.66	$U = 73, p = .066$	50 %	80 %	50 %	80 %

Note. Abréviation : Mdn = médiane, RNM= répétition de non-mots. $N = 31$ ($n = 15$ avec un DT et $n = 16$ avec un TDL).

Les épreuves sont issues des tests ELO (Khomsî, 2011), Exalang (Helloin et Thibault, 2006 ; Thibault et Helloin, 2010), ALO (Sim-Sim, 2006) et Palpa-p (Castro et al., 2007).

^a Le critère de déviation a été fixé au percentile 7 puisque l'enfant est évalué dans sa langue dominante.

* $p < .05$. ** $p < .01$. *** $p < .001$

À travers l'exposition de ces résultats, nous relevons que seul un sous-test dans la langue dominante s'avère discriminer les deux groupes d'enfants (compréhension orale) comme lors d'une évaluation monolingue (tâche de RNM). Toutefois, nous notons que les pourcentages de sensibilité et de spécificité de la meilleure tâche (RNM en français) sont plus élevés lorsqu'une évaluation monolingue est effectuée (sensibilité : 81 % ; spécificité : 80 %).

Évaluation dans les deux langues

Nous nous attelons maintenant à l'analyse des résultats issus d'une évaluation dans les deux langues de l'enfant. Nous observons dans le tableau 2 (dernière colonne) que les tâches de RNM permettent de discriminer les enfants avec et sans TDL puisque nous notons des différences significatives entre les deux groupes d'enfants ($U = 51, p = .006$ pour la RNM en français ; $U = 54, p = .008$ pour la RNM lusophone). En ce qui concerne la tâche en portugais, la sensibilité s'élève à 6 % et la spécificité se situe à 93 % (avec et sans adaptation des seuils). Nous relevons donc que cette tâche permet de différencier les groupes d'enfants avec un DT et avec un TDL mais qu'elle s'avère que peu discriminante. Nous observons une corrélation significative entre les scores standards en percentile à la tâche de RNM en portugais et l'âge des enfants ($\tau_b = 0.333, p = .014$). À l'inverse, il n'existe aucune corrélation significative entre l'âge des enfants et leurs résultats à la tâche de dénomination lusophone : $\tau_b = 0.051, p = .69$ et l'âge des enfants et les scores standards obtenus à la tâche de compréhension orale lusophone : $\tau_b = 0.225, p = .084$. Au niveau graphique (Figure 1), nous pouvons relever l'important effet plafond des enfants avec un DT à la tâche de RNM. Les tâches de dénomination et de compréhension orale sont non discriminantes ayant des différences non significatives entre les deux groupes.

En réalisant une évaluation dans les deux langues, nous relevons que deux sous-tests s'avèrent discriminants (les deux RNM) contre un seul lors d'une évaluation dans la langue

dominante (compréhension langue dominante) et lors d’une évaluation en français (RNM en français). Toutefois, nous notons que seule la tâche de RNM en français fournit des scores de sensibilité et spécificité acceptable (avec adaptation des seuils : 80 % de sensibilité et 81 % de spécificité).

Synthèse : courbes Roc et indices de Youden pour les tests standardisés

Afin d’identifier mathématiquement les seuils en percentile idéaux pour chaque sous-test, nous avons réalisé des courbes Roc (Dunn, 2014), présentées dans la Figure 2 ci-dessous, et nous avons calculé les indices de Youden (1950) (formule utilisée : « sensibilité + spécificité -1 »), listés dans le Tableau 4. Les résultats montrent que parmi les sous-tests utilisés c’est toujours la tâche de RNM en français qui est la plus sensible et spécifique quel que soit le seuil adopté.

Figure 2

Courbes Roc pour les différents sous-tests en français et en portugais

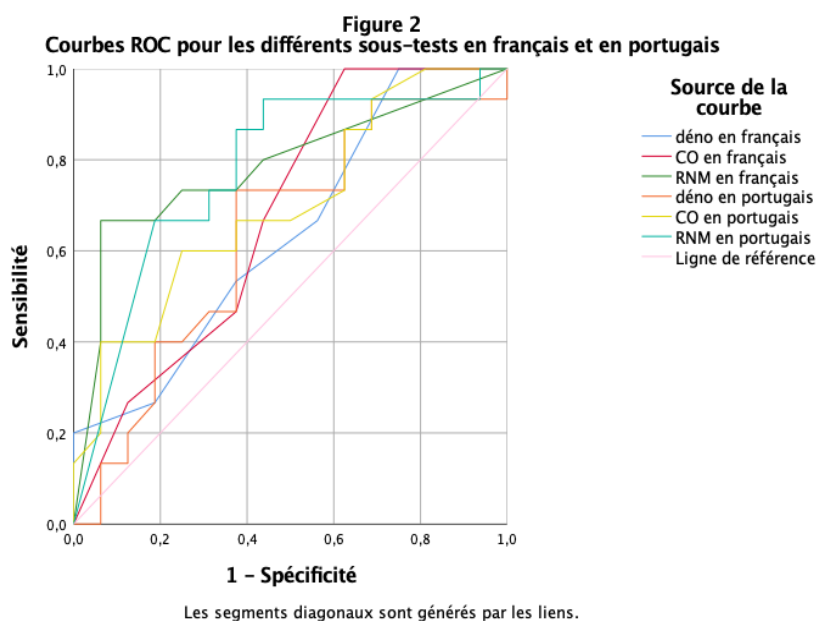


Tableau 4*Scores à l'indice de Youden et seuils optimaux*

Sous-tests	Sensibilité optimale	Spécificité optimale	Indice Youden	Seuil optimal
Dénomination en français	100 %	25 %	0.25	Percentile 5
Compréhension en français	100 %	37 %	0.37	Percentile 17.5
RNM en français	73 %	93 %	0.67	Percentile 14
Dénomination en portugais	73 %	62 %	0.35	Percentile 20.5
Compréhension en portugais	60 %	75 %	0.35	Percentile 58.5
RNM en portugais	93 %	56 %	0.49	Percentile 63.5

Note. Abréviation : RNM= répétition de non-mots.

Les épreuves sont issues des tests ELO (Khomsî, 2011), Exalang (Helloin et Thibault, 2006 ; Thibault et Helloin, 2010), ALO (Sim-Sim, 2006) et Palpa-p (Castro et al., 2007)

Nous notons que pour obtenir ces meilleurs seuils, les scores de sensibilité et de spécificité peuvent être disparates. Par exemple, pour le sous-test de dénomination en français, nous notons qu'une sensibilité à 100 % et une spécificité de 25 % est optimale avec un seuil se situant au percentile 5. Le compromis entre sensibilité et spécificité dépend toujours de l'objectif du test. En effet, nous relevons que la sensibilité et la spécificité de la tâche de RNM en français était de 80 % et de 81 % lorsqu'une évaluation monolingue était choisie avec une adaptation de seuils. En privilégiant le seuil issu de l'indice de Youden (= percentile 14), nous obtenons une sensibilité moins élevée (73 %) mais une spécificité plus grande (93 %).

Utilisation du questionnaire parental

Nous n'observons aucune différence significative quant à l'âge des premiers mots chez les enfants avec un DT et avec un TDL ($U = 142, p = .401$). Par contre, l'âge de la combinatoire est significativement plus élevé chez les enfants avec un TDL que chez les enfants avec un DT ($U = 184, p = .011$). En prenant comme seuil la médiane des deux groupes (= ou > 22 mois), nous relevons que la mesure de la combinatoire est sensible à 75 % et spécifique à 73 %. Le tableau 5 ci-dessous résume les informations recueillies auprès des parents.

Tableau 5*Données issues du questionnaire parental et valeurs discriminantes*

Questionnaire	Mdn	DT	TDL	DT/TDL	Sensibilité	Spécificité
Âge des 1 ^{ers} mots	13.33	11.66 (8 à 24)	15.42 (7 à 24)	$U = 142, p = .401$	56 %	66 %
Âge de la combinatoire	22.00	18.00 (9 à 36)	25.80 (14 à 36)	$U = 184, p = .011 *$	75 %	73 %
Difficultés dans l'entourage		1	4	$X^2 = 0.675, p = .411$	25 %	75 %
IP		4	12	$X^2 = 14.846, p < .001 * * *$	75 %	93 %
RNM en français + IP ^a					56 %	100 %
RNM en français et/ou IP ^b					93 %	60 %

Note. Abréviations : Mdn = médiane, IP = inquiétude parentale. Les données sont issues du questionnaire parental basé sur Tuller (2015).

Les âges médians des premiers mots et de la combinatoire sont exprimés en mois. Pour calculer la sensibilité et la spécificité, les médianes des deux groupes ont été déterminé comme seuil. Pour les difficultés langagières dans l'entourage familial et l'inquiétude parentale ce sont le nombre d'occurrence par groupe qui a été relevé.

$N = 31$ ($n = 15$ avec un DT et $n = 16$ avec un TDL).

^aLe sous-test de RNM et les réponses quant à l'inquiétude parentale sont mis en commun.

^bLes scores du sous-test de RNM et/ou les réponses données quant à l'inquiétude parentale sont mis en commun.

* $p < .05$. ** $p < .01$. *** $p < .001$

Les parents se sont également exprimés sur leur inquiétude quant au langage de leur enfant. Nous relevons que significativement plus de parents d'enfants avec un TDL (12/16) ont été/sont inquiets que les parents des enfants avec un DT (1/15 ; $X^2 = 14.846$, $p < .001$). Ce critère d'inquiétude parentale est sensible à 75 % et spécifique à 93 %. Nous nous sommes également intéressées à la présence ou non de difficultés dans l'entourage des enfants. Parmi les familles, 4 parents d'enfants avec un TDL relèvent des difficultés langagières dans l'entourage, alors que seul un parent d'enfant avec un DT en observe, mais cette différence n'est pas significative ($X^2 = 0.675$, $p = .411$).

Combinaison des mesures : questionnaire parental et RNM en français

Nous notons que la RNM en français avec adaptation des seuils de pathologie en fonction de la dominance langagière et l'inquiétude parentale recensée via le questionnaire parental fournissent les meilleurs scores de sensibilité et spécificité. Nous exploitons donc dans le Tableau 5 (ci-dessus), la combinaison de ces mesures. La première combinaison consiste à observer les scores de sensibilité et spécificité lorsque le sous-test de RNM et les réponses quant à l'inquiétude parentale sont mis en commun. Cette combinaison fournit un score 56 % de sensibilité et de 100 % de spécificité. Nous avons également essayé de combiner les scores à ce sous-test et/ou les réponses données quant à l'inquiétude parentale. Dans ce cas de figure, la sensibilité s'élève à 93 % pour une spécificité à 60 %. Pris séparément, les scores à la RNM et les réponses données quant à l'inquiétude fournissent des meilleures données quant à la sensibilité et à la spécificité, cette combinaison ne permet donc pas l'amélioration de l'identification des enfants avec et sans TDL.

Discussion générale et conclusion

Les outils et alternatives à disposition des cliniciens

Dans cet article, nous avons répertorié et testé plusieurs tâches et méthodes à disposition des cliniciens pour distinguer les enfants bilingues avec un DT et avec un TDL. Nous avons débuté par la réalisation d'évaluations en français avec des seuils de pathologie habituellement utilisés en clinique pour les enfants monolingues. Nous avons trouvé des différences significatives entre les deux groupes d'enfants seulement en ce qui concerne la tâche de RNM. Les données vont dans le sens des études de nombreux auteurs (ex. Chiat, 2015) qui ont mis en exergue l'intérêt de la tâche de RNM dans l'évaluation langagière des enfants bilingues. Nous avons aussi observé que cette tâche était, sans surprises, la plus sensible et la plus spécifique. Nous avons ensuite suivi les recommandations de Thordardottir (2015b) en ajustant les seuils de pathologie en fonction de l'exposition langagière des enfants. Fait intéressant, nous notons que la spécificité du sous-test de RNM s'est vu améliorer grâce à l'adaptation des seuils passant de 73 % à 80 % de spécificité. La capacité du sous-test à rejeter les sujets avec un DT s'est ainsi renforcée. Nous pouvons donc mentionner que ce sous-test démontrant une sensibilité et une spécificité équivalente ou supérieure à 80 % est considéré comme discriminant (Plante et Vance, 1994). Nous nous sommes ensuite penchées sur la réalisation des sous-tests dans la langue dominante de l'enfant et nous observons ici que l'utilisation du sous-test de compréhension orale dans la langue dominante s'est avérée efficace pour distinguer les enfants avec un DT et avec un TDL mais que son pourcentage de sensibilité n'est pas acceptable (Plante et Vance, 1994). Nous avons ensuite opté pour une évaluation dans les deux langues des enfants qui nous a montré l'intérêt des tâches de RNM en français et en portugais pour distinguer les enfants avec un DT et avec un TDL. Toutefois, parmi ces tâches, seule la RNM en français fournit des scores de sensibilité (80 %) et de spécificité (81 %) jugés comme étant discriminants (Plante et Vance, 1994). La complexité phonologique (structure syllabique) est identique entre

les deux tâches et la longueur des stimuli varient également dans les deux (jusqu'à 3 syllabes). Toutefois, nous pouvons observer que la majorité des items en portugais possède une seule syllabe ce qui peut expliquer la plus grande facilité des enfants pour cette tâche. Grâce aux courbes Roc et aux indices de Youden, nous avons pu déterminer les seuils optimaux en percentile pour chaque sous-test. Nous relevons qu'à l'exception de la tâche de RNM en français, l'indice de Youden est faible et nous notons de grandes disparités entre les scores de sensibilité et de spécificité. Il ne semble, en fait, n'y avoir aucun bon indice pour les autres tâches.

Les réponses parentales en ce qui concerne leur inquiétude et l'âge de la combinatoire montrent également des différences significatives entre les enfants avec un DT et avec un TDL. Au sujet de l'inquiétude parentale, cette observation doit être nuancée puisque les parents d'enfants avec un TDL se disent plus inquiets que les parents d'enfants avec un DT mais il faut garder à l'esprit que les enfants avec un TDL bénéficient d'une prise en charge orthophonique. Il y a donc un biais non négligeable concernant les réponses données par les parents au sujet de leur inquiétude. Il est intéressant de noter que les réponses des parents quant à leur inquiétude et à l'âge de la combinatoire possèdent des scores de sensibilité et de spécificité meilleurs que certains sous-tests langagiers utilisés dans notre étude. Nous avons ensuite testé une combinaison de deux mesures soit la RNM en français ainsi que la réponse des parents quant à leur inquiétude parentale. Nous ne trouvons, dans cet article, pas d'intérêt à combiner ses deux mesures puisque les pourcentages de sensibilité et de spécificité s'amointrissent.

Les limites

A la fin de cet article, nous relevons le faible pouvoir discriminant de la plupart des sous-tests utilisés en français et en portugais auprès de notre population bilingue. Nous pouvons ainsi nous questionner sur le choix des tests utilisés. En effet, en français, nous aurions pu

choisir des tests plus récents bien qu'ils soient moins utilisés en clinique ou qu'ils ne permettent pas une évaluation rapide. En portugais, nous n'avions malheureusement que peu de choix et nous sommes conscientes des fragilités des sous-tests utilisés notamment en ce qui concerne le nombre d'enfants utilisé pour la standardisation et la normalisation des tests (de Almeida et al., 2016) et le peu de tranches d'âge couvert par celles-ci.

Au sujet des modes d'évaluation, nous arrivons également au constat que malgré qu'une évaluation dans les deux langues de l'enfant ait été réalisée (extrême de notre continuum), il n'est pas évident d'en tirer des conclusions. En effet, nous relevons que parmi les 16 enfants identifiés comme ayant un TDL, sept enfants n'ont échoué aucune tâche en portugais alors que les scores des sous-tests en français étaient clairement en dessous de la norme même lorsqu'une adaptation de seuils de pathologie était opérée. Parmi ces sept enfants, nous relevons que quatre d'entre eux possèdent le portugais comme langue dominante, ce qui peut expliquer qu'ils soient plus performants dans cette langue. L'évaluation des enfants bilingues étant actuellement problématique en clinique, nous laisse supposer que le portugais n'a pas ou peu été évalué chez ces sept enfants. Toutefois, nous notons, en prenant connaissance des résultats de la tâche de RNM en français jugée comme la plus sensible et spécifique de notre recherche, que ces sept enfants ont clairement échoué celle-ci avec des scores se situant entre le percentile 0 et 7. Nous prenons conscience de la difficulté des cliniciens lorsqu'ils doivent identifier la présence d'un TDL chez ces enfants. Nous relevons également que trois enfants avec un DT échouent à un ou plusieurs sous-tests en français et en portugais. Ces trois enfants ont le portugais comme langue dominante. Il est possible qu'ils aient, en effet un TDL jamais détecté, car ils auraient développé de bonnes stratégies de compensation et que leurs difficultés apparaissent dans des tâches spécifiques pour évaluer leurs capacités langagières, comme c'est le cas dans notre étude. Un bilan orthophonique complet a été suggéré aux parents de ces enfants.

Ces différents modes d'évaluation ont ainsi été testés et nous ne nous ont pas permis de tirer des conclusions claires puisqu'aucun d'entre eux n'a pas permis de discriminer clairement et unanimement les enfants avec un TD et TDL bilingues. Il convient de rappeler que les comparaisons ont toutes été effectuées à partir de normes monolingues avec seuils de pathologie ajustés ou non en fonction du taux d'exposition de l'enfant à la langue testée et que celles-ci semblent insatisfaisantes pour l'évaluation des enfants bilingues. Toutefois, même lors d'une évaluation langagière d'enfants monolingues, les tests avec normes monolingues présentent des failles. Plus précisément, faire usage de tests standardisés constituent une évaluation statique à un moment précis et ne permet pas d'apprécier le profil complet des difficultés et capacités de l'enfant qu'il soit monolingue ou bilingue (Hasson et Botting, 2010). De plus, il convient de rappeler l'usage problématique des normes monolingues dans d'autres régions ou pays que ceux de l'étalonnage de référence (Grégoire, 2006).

Conclusion et perspectives futures

Nous pouvons évoquer plusieurs implications cliniques de notre étude. Les données fournissent des renseignements sur les tâches et les modes d'évaluation à préconiser pour évaluer les enfants bilingues français-portugais. Nous pouvons mentionner qu'une tâche de RNM en français est une tâche à effectuer lors des bilans de ces enfants en regard de sa précision diagnostique. Une tâche de RNM dans la langue de la société constitue donc une solution intéressante lorsque le matériel dans l'autre langue de l'enfant présente des lacunes et/ou ne couvre pas l'ensemble des âges, d'autant plus si les deux langues sont phonologiquement proches (ex. français et portugais). Bien que la RNM en français soit créée sur la base de la phonologie du français, elle semble tout de même peu affectée par l'expérience linguistique du sujet (de Almeida et al., 2016). Pour les langues qui n'ont pas beaucoup de matériel normé (ex. le portugais), une évaluation dans les deux langues de l'enfant permet tout de même d'obtenir

des données qualitatives intéressantes pour le bilan (ex. l'enfant ne possède que peu de stock lexical dans les deux langues) et pour des pistes d'intervention (ex. renforcer le stock lexical dans les deux langues) mais ne nous donnent pas plus de certitudes quant à l'identification d'un TDL.

Nous relevons indispensable que d'autres études analysent le pouvoir discriminant d'autres sous-tests en français et en portugais sur le marché. Nous constatons également que pour identifier de façon fiable la présence d'un TDL, d'autres mesures et outils doivent être utilisés conjointement à la RNM et à l'exploitation des données issues d'un questionnaire parental. Hormis la conception de tests avec normes bilingues qu'il est difficile à mettre en œuvre (cf. introduction théorique), d'autres recherches se sont notamment penchées sur l'utilisation d'outils d'évaluation se centrant sur le potentiel d'apprentissage des enfants, c'est le cas de l'évaluation dynamique. Plusieurs recherches (ex. Camilleri et Law, 2007) ont mis en avant des différences dans la capacité d'apprentissage des enfants avec un DT et avec un TDL monolingues et bilingues. Ce type d'évaluation semble être un outil prometteur et il est encore nécessaire que des études déterminent sa qualité diagnostique. Il serait également intéressant d'analyser si une combinaison alliant RNM, questionnaire parental et évaluation dynamique serait suffisamment discriminante pour différencier les enfants avec un DT et avec un TDL. Nous pourrions également nous intéresser au pouvoir discriminant de la narration chez des enfants bilingues avec et sans TDL. L'étude de Boerma et Blom (2017) atteste déjà l'intérêt d'une combinaison de tâches alliant RNM, questionnaire parental et tâche narrative avec une précision diagnostique de 97 % pour les enfants bilingues hollandais-turque avec un DT et avec un TDL. Il conviendrait alors de s'intéresser à une telle combinaison pour discriminer les enfants bilingues avec un DT et avec un TDL partageant d'autres combinaisons de langues.

Les résultats obtenus confirment ainsi la nécessité de la poursuite de recherches scientifiques mais ils ont également permis de dégager quelques pistes pour les cliniciens. Nous

avons également mis en lumière le travail d'évaluation rendu complexe par la situation de bilinguisme pour les cliniciens, notamment en ce qui concerne l'identification d'un TDL.

Remerciements

Nous remercions tous les participants et leurs familles d'avoir pris part à cette recherche, ainsi que les collègues, étudiants et membres d'institutions nous ayant permis d'entrer en contact avec les familles. Nous remercions également nos assistantes-étudiantes lusophones pour leur contribution dans cette étude. Enfin, merci aux relecteurs anonymes de la revue pour leurs commentaires éclairés.

Références

- Altman, C., Goldstein, T. et Armon-Lotem, A. (2018). Vocabulary, metalinguistic awareness and language dominance among bilingual preschool children. *Frontiers in Psychology*, 9(1953). <https://doi.org/10.3389/fpsyg.2018.01953>
- Armon-Lotem, S. (2012). Introduction: Bilingual children with SLI – the nature of the problem*. *Bilingualism: Language and Cognition*, 15(1), 1-4. <https://doi.org/10.1017/S1366728911000599>
- American Speech-Language-Hearing Association. (s. d.). *Bilingual Service Delivery*. Repéré le 18 mars 2020 à, <https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589935225§ion=Overview>
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A. et Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068-1080. <https://doi.org/10.1111/jcpp.12721>
- Boerma, T. et Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, 66(1), 65-76. <https://doi.org/10.1016/j.jcomdis.2017.04.001>
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F. et Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research*, 58(6), 1747-1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058
- Camilleri, B. et Law, J. (2007). Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology*, 9(4), 312-322. <https://doi.org/10.1080/14417040701624474>

- Campbell, T., Dollaghan, C., Needleman, H. et Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech and Language Hearing Research*, 40(3), 519-525. <https://doi.org/10.1044/jslhr.4003.519>
- Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R. et Coltheart, M. (2007). *Provas de avaliação da linguagem e da afasia em português*. CEGOC.
- Chiat, S. (2015). Non-Word Repetition. Dans S. Armon-Lotem, K. de Jong et N. Meir (dir.), *Assessing multilingual children* (p. 95-122). Multilingual Matters.
- De Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L. et Zebib, R. (2016). L'identification d'enfants bilingues avec trouble spécifique du langage en France. *SHS Web of Conferences*, 27(10005), 1-16. <https://doi:10.1051/shsconf/20162710005>.
- Denman, D., Speyer, R., Munro, N., Pearce, W., Chen, Y. W., Cordier, R. (2017). Psychometric properties of language assessments for children aged 4-12 years: a systematic review. *Frontiers in Psychology*, 1515(8), 1-28. <https://doi:10.3389/fpsyg.2017.01515>
- Di Meo, S., Sanson, C., Simon, A., Bossuroy, M., Rakotomala, L., Rezzoug, D., Serre, G., Baudet, T. et Moro, M. R. (2014). Le bilinguisme des enfants migrants. Dans H. A. Bijleveld, F. Estienne et F. Vander Linden (dir.), *Multilinguisme et orthophonie. Réflexions et pratiques à l'heure de l'Europe* (p. 149-171). Elsevier Masson.
- Dubreuil-Piché, S., Mayer-Crittenden, C. et Lachance, J. (2017). Nonword repetition and sentence imitation as clinical markers for primary language impairment in bilingual French-English- and English-French speaking children in Northern Ontario: A narrative review. *Diversity of research in health journal* 1(1), 61-69. <https://pubs.biblio.laurentian.ca/index.php/drhj-rdrs/article/view/19>
- Dunn, G. (2014). Statistics in psychiatry. Dans M. Lovric (dir.), *International encyclopedia of statistical science* (p. 1136-1138). Arnold.

- Estienne, F. et Vander Linden, F. (2014). Les orthophonistes face au multilinguisme - résultats d'une enquête. Dans H. A. Bijleveld, F. Estienne et F. Vander Linden (dir.), *Multilinguisme et orthophonie. Réflexions et pratiques à l'heure de l'Europe* (p. 183-198). Elsevier Masson.
- Grégoire, J. (2006). Propriétés métriques des tests de langage et leurs implications pratiques. Dans F. Estienne et B. Piérart (dir.), *Les bilans de langage et de voix* (p. 14-26). Masson.
- Grimm, A. et Shulz, P. (2014). Specific language impairment and early second language acquisition: the risk of over- and underdiagnosis. *Child Indicators Research*, 7(4), 821-841. <https://doi.org/10.1007/s12187-013-9230-6>
- Grosjean, F. (2015). *Parler plusieurs langues. Le monde des bilingues*. Albin Michel.
- Glutting, J., Adams, W. et Sheslow, D. (2000). *Wide range intelligence test*. Wide Range.
- Hasson, N. et Botting, N. (2010). Dynamic assessment of children with language impairments: A pilot study. *Child Language Teaching and Therapy*, 26(3), 249-272. <https://doi.org/10.1177/0265659009349982>
- Helloin, M. C. et Thibault, M. P. (2006). *Exalang 3-6*. Happyneuron.
- Kay, J., Lesser, R. et Coltheart (2007). *Provas de avaliação da linguagem e da afasia em português*. Hogrefe.
- Keller, K., Troesch, L. et Grob, A. (2015). A large receptive-expressive gap in bilingual children. *Frontiers in Psychology*, 6 (1284), 1-9. <https://doi.org/10.3389/fpsyg.2015.01284>
- Kehoe, M. (2009). Plurilinguisme et logopédie: le défi de l'évaluation, *Langage et pratiques*, 44, 69-82. https://arld.ch/fileadmin/user_upload/Documents/ARLD/WWW/Editeurs/Logopedistes/Langages_pratiques/44-textes.pdf

- Khomsî. A. (2001). *Évaluation du langage oral*. ECPA
- Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of communication disorders*, 43(6), 456-473. <https://doi.org/10.1016/j.jcomdis.2010.02.002>
- Lafay, A. et Cattini, J. (2018). Analyse psychométrique des outils d'évaluation mathématique utilisés auprès des enfants francophones. *Canadian Journal of Speech-Language Pathology and Audiology*, 42(2), 127-144.
https://cjslpa.ca/files/2018_CJSLPA_Vol_42/No_02/CJSLPA_Vol_42_No_2_2018_MS_1129.pdf
- Lidz, C. et Peña, E. (2009). Response to intervention and dynamic assessment: Do we just appear to be speaking the same language. *Seminars in Speech and language*, 30(2), 121-133. <https://doi.10.1055/s-0029-1215719.ISSN0734-0478>
- Leonard, L. (2014). *Children with specific language impairment* (2^e éd.). A Bradford book.
- Messer, M., Leseman, P., Boom J. et Mayo, A. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, 105(4), 306-323.
<https://doi.org/10.1016/j.jecp.2009.12.006>
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G. et Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, 57(11), 1247-1257. <https://doi.org/10.1111/jcpp.12573>
- Office Fédéral de la Statistique. (2020, 27 février). *Langues principales les plus fréquentes*.
<https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/langues-religions/langues.assetdetail.11947533.html>
- Paradis, J. (2007). Becoming bilingual. Dans P. Auer et L. Wei (dir.), *Handbook of*

- multilingualism and multilingual communication*, (p. 15-44). Mouton de Gruyter.
- Paradis, J. (2019). English second language acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors. Dans M. Brown et B. Dailey (dir.), *Proceedings of the 43rd annual Boston University Conference on Language Development* (p. 11-26). Cascadia Press.
- <http://www.lingref.com/buclid/43/BUCLD43-02.pdf>
- Paradis, J., Genesee, F. et Crago, M. B. (2011). *Dual language development & disorders: A handbook on bilingualism and second language learning* (2^e éd.). Brookes.
- Peña, E. D., Bedore, L. M. et Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders*, 51(2), 192-202.
- <https://doi.org/10.1111/1460-6984.12199>
- Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A. et Bedore, L. M. (2018). *Bilingual English-Spanish assessment*. Paul Brookes.
- Plante, E. et Vance, R. (1994). Selection of Preschool Language Tests. *Language, Speech, and Hearing Services in Schools*, 25(1), 15-24. <https://doi.org/10.1044/0161-1461.2501.15>
- Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews*, 17(2), 141-150. <https://doi.org/10.1002/ddrr.1108>
- Simard, N., Mayer-Crittenden, C., Robillard, M. et Boissonneault, J. (2015). Sous l'influence : l'effet du contact des langues à la petite école. Dans J. Alem et G. Kpazai (dir.), *Actes de la 22^e journée : sciences et savoirs- mission social des universités* (p. 59-61). Erudit.
- <https://www.erudit.org/fr/livres/actes-des-colloques-de-lacfas-sudbury/actes-22e-journee-sciences-savoirs-mission-sociale-universites/004254co.pdf>
- Sim-Sim, I. (2006). *Avaliação da linguagem oral - um contributo para o conhecimento do desenvolvimento linguístico das crianças portuguesas* (4^e éd.), Fundação Calouste

Gulbenkian.

- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426-445.
<https://doi.org/10.1177/1367006911403202>
- Thorardorttir, E. (2015a). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology*, 17(2), 97-114.
<https://doi.org/10.3109/17549507.2014.923509>
- Thordardottir, E. (2015b). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. Dans S. Armon-Lotem, K. de Jong et N. Meir (dir.), *Assessing multilingual children* (p. 331-359). Multilingual Matters.
- Thordardottir, E., Rothenberg, A., Rivard, M. E., et Naves. R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, 4(1), 1-21.
<https://doi.org/10.1080/14769670500215647>
- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N. et Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, 54(2), 580-597.
[https://doi.org/10.1044/1092-4388\(2010/09-0196\)](https://doi.org/10.1044/1092-4388(2010/09-0196))
- Thibault, M. P. et Helloin, M. C. (2010). *Exalang* 5-8. Happyneuron.
- Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. Dans S. Armon-Lotem, K. de Jong et N. Meir (dir.), *Assessing multilingual children* (p. 301-331). Multilingual Matters.
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., Dos Santos, C., Ibrahim, L. A. et Zebib, R. (2018). Identifying language impairment in bilingual children in

- France and in Germany. *International Journal of Language & Communication Disorders*, 53(4), 888-904. <https://doi.org/10.1111/1460-6984.12397>
- Volpin, L., Rezzonico, S. et de Weck, G. (2020). Enfants bilingues et prise en charge logopédique : panorama des pratiques déclarées. *Glossa*, 129(1) 16-31.
- White, C. et 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>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32-35. [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)
- Zurer Pearson, B. (2013). Distinguishing the bilingual as a late talker from the late talker who is bilingual. Dans L. Rescorla et P. Dale (dir.), *Late talkers* (p. 67-91). Paul Brookes.

Partie IV : La répétition de non-mots dans l'évaluation logopédique des enfants monolingues et bilingues

5. Article de recherche

Using Non-Word Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-analysis

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P. et Skoruppa, K.

État : publié

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P. et Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, ePub Ahead of Print
https://doi.org/10.1044/2021_JSLHR-20-00552

Review Article

Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis

Salomé Schwob,^a  Laurane Eddé,^a Laure Jacquin,^a Mégane Leboulanger,^a Margot Picard,^a Patricia Ramos Oliveira,^a and Katrin Skoruppa^a

Purpose: A wealth of studies has assessed the diagnostic value of the nonword repetition task (NVRT) for the detection of developmental language disorder (DLD) in the clinical context of speech and language therapy, first in monolingual children and, more recently, in bilingual children. This review article reviews this literature systematically and conducts a meta-analysis on the discriminative power of this type of task in both populations.

Method: Three databases were used to select articles based on keyword combinations, which were then reviewed for relevance and methodological rigor based on internationally recognized checklists. From an initial pool of 488 studies, 46 studies were selected for inclusion in the systematic review, and 35 of these studies could be included in a meta-analysis.

Results: Most of the articles report significant discrimination between children with and without DLD in both monolingual and bilingual contexts, and the meta-analysis shows a large mean effect size. Three factors (age of the child, linguistic status, and language specificity of the task)

yielded enough quantitative data for further exploration. Subgroups analysis shows variance in effect sizes, but none of the three factors, neither their interactions, were significant in a metaregression. We discuss how other, less explored factors (e.g., nature of the stimuli, scoring methods) could also contribute to differences in results. Sensitivity and specificity analyses reported in 33 studies confirmed that, despite possible effect size differences, the diagnostic accuracy of the NVRT is generally near thresholds considered to be discriminatory. It generally increases when it is combined with other tasks (e.g., parental questionnaire).

Conclusions: This review indicates that the NVRT is a promising diagnostic tool to identify children with DLD in monolingual and bilingual contexts with a large mean effect size. However, it seems necessary to choose the precise NVRT materials based on the children's language background and to complement the assessment sessions with other tools in order to ensure diagnosis and to obtain complete language profile of the child.

The ability to adequately repeat nonwords relies on many processes such as auditory perception, phonological encoding and assembly, storage of phonological representations, motor planning, and the ability to articulate (Coady & Evans, 2008). Thus, it is unsurprising that many studies have shown that nonword repetition task (NVRT) is a good clinical indicator of developmental language disorders, which may affect these

processes (see, for instance, the recent literature review by Coady & Evans, 2008). The accuracy with which phonologically plausible nonwords are repeated is a relatively reliable index of language level (Szwedczyk et al., 2018) and a clinical marker of developmental language disorder (DLD; Boerma et al., 2015; Conti-Ramsden et al., 2001). Beyond general performance differences between children with and without DLD (Archibald, 2008), the presence and nature of errors in NVRT seem to reflect difficulties in short-term verbal memory and phonological accuracy (Ferré & Dos Santos, 2015; Li'el et al., 2019). Indeed, verbal short-term memory is often reported to be affected by language difficulties (Gray, 2004).

Compared to other assessments used in speech and language therapy (SLT), NVRT has the advantage of not

^aInstitut des sciences logopédiques, Pierre-à-Mazel 7, 2000 Neuchâtel, Université de Neuchâtel, Suisse

Correspondence to Salomé Schwob: salome.schwob@unine.ch

Editor-in-Chief: Stephen M. Camarata

Editor: Filip Smolik

Received September 18, 2020

Revision received February 11, 2021

Accepted May 9, 2021

https://doi.org/10.1044/2021_JSLHR-20-00552

Disclosure: The authors have declared that no competing interests existed at the time of publication.

directly involving the test subject's linguistic knowledge (e.g., lexicon and sentence structure). It is more independent from linguistic experience, since it assesses the ability to process new information (Archibald, 2008). Thus, this task can already be administered to very young children (Guiberson & Rodríguez, 2015) and has also been praised as a promising tool for children from multilingual backgrounds with very limited exposure to the test language (e.g., Boerma et al., 2015; Chiat, 2015).

In a global context where multiple languages coexist, and where the number of individuals speaking several languages continues to increase (Grosjean, 2015), the question of the correct language assessment for these individuals is a central concern. In addition to the many interindividual differences between children, disparities in language development in a context of bilingualism depend on further factors. For instance, both quantity and quality of exposure crucially impact the development of the bilingual child's language skills (Paradis, 2019), which can also be influenced by the migration path, social and motivational factors (Di Meo et al., 2014; Paradis et al., 2011). Given this multitude of aspects influencing the level of linguistic knowledge in bilingual children, establishing bilingual developmental norms is not straightforward (Pearson, 2013) and the diagnosis of DLD in bilingual children is often characterized as the greatest challenge for the present-day SLT clinic (Armon-Lotem, 2012). Indeed, it is often difficult for clinicians to differentiate between persistent language difficulties caused by DLD and transient difficulties due to insufficient exposure to one language (Camilleri & Law, 2007). The language difficulties of children falling into these two scenarios are strikingly similar: Typically developing (TD) children beginning to learn a new language often make morphosyntactic errors traditionally associated with DLD in monolingual children (Armon-Lotem, 2012), for instance, omissions of clitic objects in French (Tuller et al., 2018). Therefore, the identification of DLD in bilingual children requires the construction of a new set of tools, of which NWRT seems to be an important part (Boerma & Blom, 2017). However, it is important to note some controversies on this issue. Although NWRT scores are less affected by the child's prior language knowledge, some studies show that they are not completely free from language specificities. Indeed, a child is more likely to correctly repeat nonwords created on the basis of the phonological characteristics of their own language. Several studies (e.g., Boerma et al., 2015; Kohnert et al., 2006) have found that children's scores depend largely on their experience in the language in which the NWRT was created, leading some authors to advocate the construction of a quasi-universal NWRT (e.g., Chiat, 2015). Other authors (e.g., De Almeida et al., 2016; Thordardottir & Brandeker, 2013) found no influence of language experience on the task scores used.

Aims of the Article

In view of the numerous studies and the various controversies on the subject (see also Coady & Evans, 2008), the

primary objective of this article is to provide a general overview of the current state of evidence regarding NWRT and DLD diagnosis, both in monolingual and bilingual children. Coady and Evans's (2008) literature review and Estes et al.'s (2007) meta-analysis already provided first overviews of the use of the task with monolingual children, but to our knowledge, this article is the first systematic review and meta-analysis that specifically focuses on bilingual children with DLD.

In particular, we will attempt to answer the following research questions:

1. Does the NWRT reveal significant differences in performance between children with TD versus DLD, in a monolingual versus bilingual context?
 - a. What is the mean effect size of the studies?
 - b. What are the sensitivity and specificity values of the NWRT found by the studies in the review?
2. What are the internal and external characteristics of children that may influence their NWRT scores and the discriminating power of the task?
 - a. Does language status (monolingual vs. bilingual) influence the discriminating power of the task and the effect size?
 - b. Does the language specificity of the task (language specific vs. quasi-universal) influence the effect size?
 - c. Is the age of the children a moderator factor of effect size?
 - d. Which other factors do studies report to influence performance on this task?
3. Which other tasks correlate with NWRT scores?
4. Which other tasks, in combination with NWRT, improve the discriminative power and thus facilitate the distinction between children with TD versus DLD?

In order to zoom in on the most relevant age range for DLD diagnosis, we decided to focus on studies using data from children up to the age of 8;11 (years;months). However, despite the fact the diagnosis of DLD is typically not made until around the fourth year of life (Bishop et al., 2017; Norbury et al., 2016), we decided not to use a lower age limit and to include studies that have focused on children at risk for DLD and late talkers, generally defined as producing less than 50 words and having no word combinations at 24 months of age (Desmarais et al., 2008; Pearson, 2013) in our qualitative part.

Method

Article Selection

A literature search was carried out using three databases: PsycINFO, Scopus, and ASHA (American Speech-Language-Hearing Association). The PsycINFO and Scopus databases were selected because of the large number of publications available, and ASHA, because of its specificity

to SLT. In order to conduct an exhaustive search of the existing literature, the following two combinations of English keywords were entered into the three databases: (a) nonword repetition, language impairment or language disorders, and assessment; and (b) nonword repetition, language impairment or language disorders, diagnosis. The same search was also conducted with the terms in French and in German, but no results were found in all three databases. For each database, and for all combinations of keywords, the bibliographic search was conducted by two researchers in order to ensure the reliability of the results. In case of disagreement, the results were reviewed jointly. The search was conducted between November 2019 and June 2020. In order to exclude articles that did not correspond to the interests of this review, criteria were used to mask the following publications: publications prior to 2000; publications whose language of publication was neither English, French, nor German; publications whose population consisted exclusively of children over the age of 8;11 (or, by extension, adolescents/adults).

The flow chart (see Figure 1) details the exact number of articles selected at each stage of the literature search. At this first stage of the selection process, 488 references were identified from the three databases and 16 from external sources. After removal of duplicates, a total of 226 studies were retained. Two of the authors of this study then reviewed the titles and abstracts of the articles to ensure their relevance. In case of disagreement, the studies were reviewed jointly. The following studies were then excluded: (a) systematic reviews and meta-analysis, (b) studies whose research design ultimately did not use NWRT, and (c) studies whose population consisted of children with disorders associated with DLD (e.g., deafness, attention deficit disorder) or of children who did not fit clinical diagnostic criteria for DLD or late talkers (e.g., “at-risk” children with scores below the 50th percentile, which was deemed as too broad a criterion). A total of 74 articles were then included. A critical evaluation process was then conducted using the INESS recommendations (Martin et al., 2013) to identify quality studies, as well as the PRISMA (Moher et al., 2009) checklist. Then, the methodological part and the section concerning the results of the studies were scrutinized. A robust and sufficiently described methodological framework was a prerequisite for our systematic review. We therefore verified that the studies (a) had a sample size of more than 10 children; that (b) the groups of participants were described, defined, and matched; that (c) the NWRT used was described (e.g., stimuli); that (d) the data analysis methodology was presented; and finally that (e) inferential statistical analyses had been performed and described. Each article was evaluated on the basis of these criteria by two authors, so as to ensure reliability among judges in the final selection of articles actually included in this review.

A total of 46 articles were thus selected and reviewed in their entirety, of which 35 were used to conduct the meta-analysis. The remaining articles were excluded from the meta-analysis because they either did not contain the information needed to calculate effect sizes, or the clinical

population was not clearly identified as DLD, but at-risk children or late talkers.

Analysis

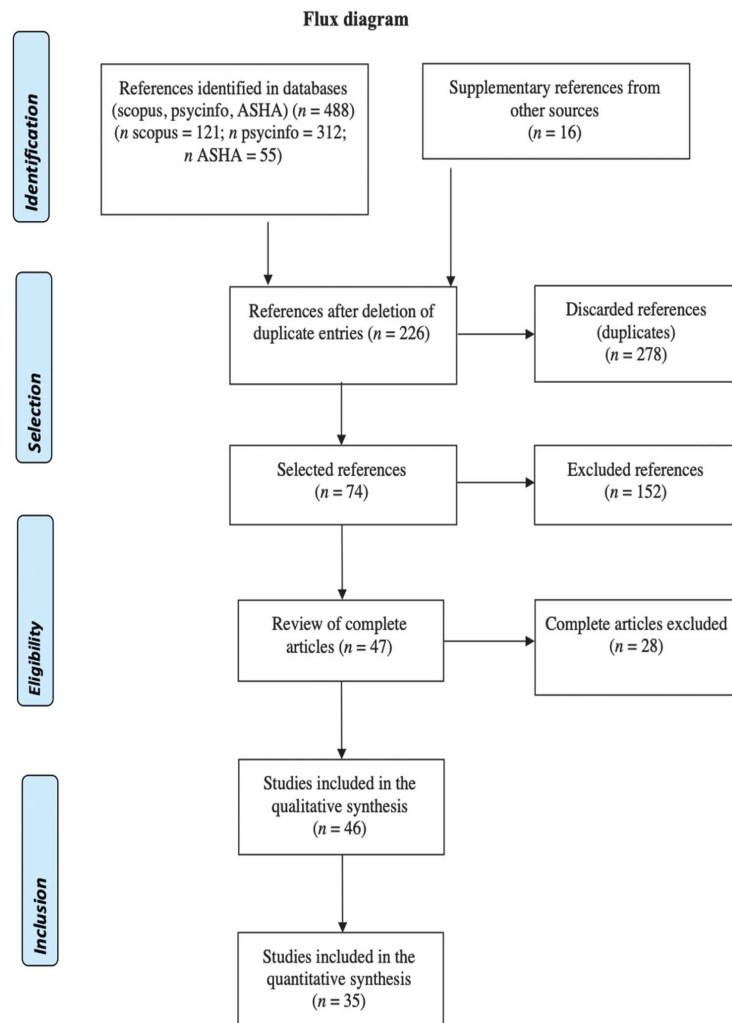
To conduct the meta-analysis, we opted for the Comprehensive Meta-Analysis software (Bornstein et al., 2013). We calculated an effect size (Cohen’s d) for each sample by subtracting each DLD group’s mean nonword repetition score from its TD group’s mean score and dividing the difference by the mean within-group (pooled) standard deviation. With reference to Cohen (1988), the effect size is considered small when $d = .20$, medium when $d = .50$, and large when $d = .80$. All effect sizes have been corrected for bias in the estimation of the sample size. We therefore applied the formula of Hedges and Becker (1986) for the correction factor. A homogeneity test of effects was used to determine whether all effects were from the same population. Since our data are heterogeneous (the I^2 index showed a large heterogeneity of 79.4%), the overall effect size was calculated using random effects models. In addition, subgroups analysis (language specificity of the task and linguistic status) and a metaregression with moderators (age of children, language specificity of the task, and linguistic status) was applied to determine whether study characteristics explain the variation in effect size between groups. To test for publication bias, we created a funnel plot and we conducted an Egger’s regression test.

Results and Discussion

We will first report some general characteristics of the studies included in our systematic review and meta-analysis. We will then present our quantitative meta-analysis on group differences between children with TD and DLD and perform a mean effect size calculation and check for publication bias. We will then conduct subgroups analysis and metaregression to understand what contributes to the variance in effect sizes. We will continue by presenting other possible influencing factors qualitatively. Finally, we will examine the discriminative values (sensitivity and specificity) of this task in monolingual and bilingual assessment contexts, as well as correlations and combinations with other language tasks.

General Characteristics of the Included Studies

Some basic methodological features (test languages and stimuli, age ranges, and clinical status of the participants) of each study are summarized in Table 1 below. It is worth noting that the majority (32 of the 46 studies) involved an exclusively monolingual population, five involved an exclusively bilingual population, and nine involved a comparison between a monolingual and a bilingual population. Table 1 also shows the different language profiles of the children participating in the studies identified. We note that the most studied language is English and that other languages are examined in several papers (e.g., French and Dutch in particular). Other languages and profiles (e.g., bilingualism)

Figure 1. PRISMA flux diagram illustrating the literature research steps and quantitative results.

are also studied (e.g., Vietnamese monolinguals in the study by Pham & Ebert, 2020).

Quantitative Meta-Analysis

First, we conducted an overall meta-analysis of effect sizes. A forest plot summarizing these results is depicted in Figure 2, which shows the estimated effect sizes and the 95% confidence interval (CI) for each study. All individual effect sizes can be found in the Supplemental Material S1. The mean effect size across the studies is large $g = 1.57$ (95% CI [1.37, 1.72], $p < .001$), indicating that the children with DLD consistently performed much poorer than children with TD.

We then conducted subgroups analysis to observe whether effect sizes vary with the linguistic status of the children (monolingual or bilingual or with the language specificity of the task (language-specific or quasi-universal). Table 2 shows slight variations in mean effect size across subgroups. Specifically, the effect size is larger for the monolingual subgroups (1.61) than for the bilingual subgroups (1.36). Similarly, the use of the quasi-universal task provides a higher mean effect size (1.73) than the language-specific task (1.52).

To examine the statistical significance of these variations, we performed a metaregression. Including three moderators often reported in the literature as influencing factor (language status, age, and language specificity of the task) for which sufficient quantitative data were available (> 10

Table 1. Clinical and methodological characteristics and general aims of the studies on nonword repetition task (NVRT) included in this review ($n = 46$).

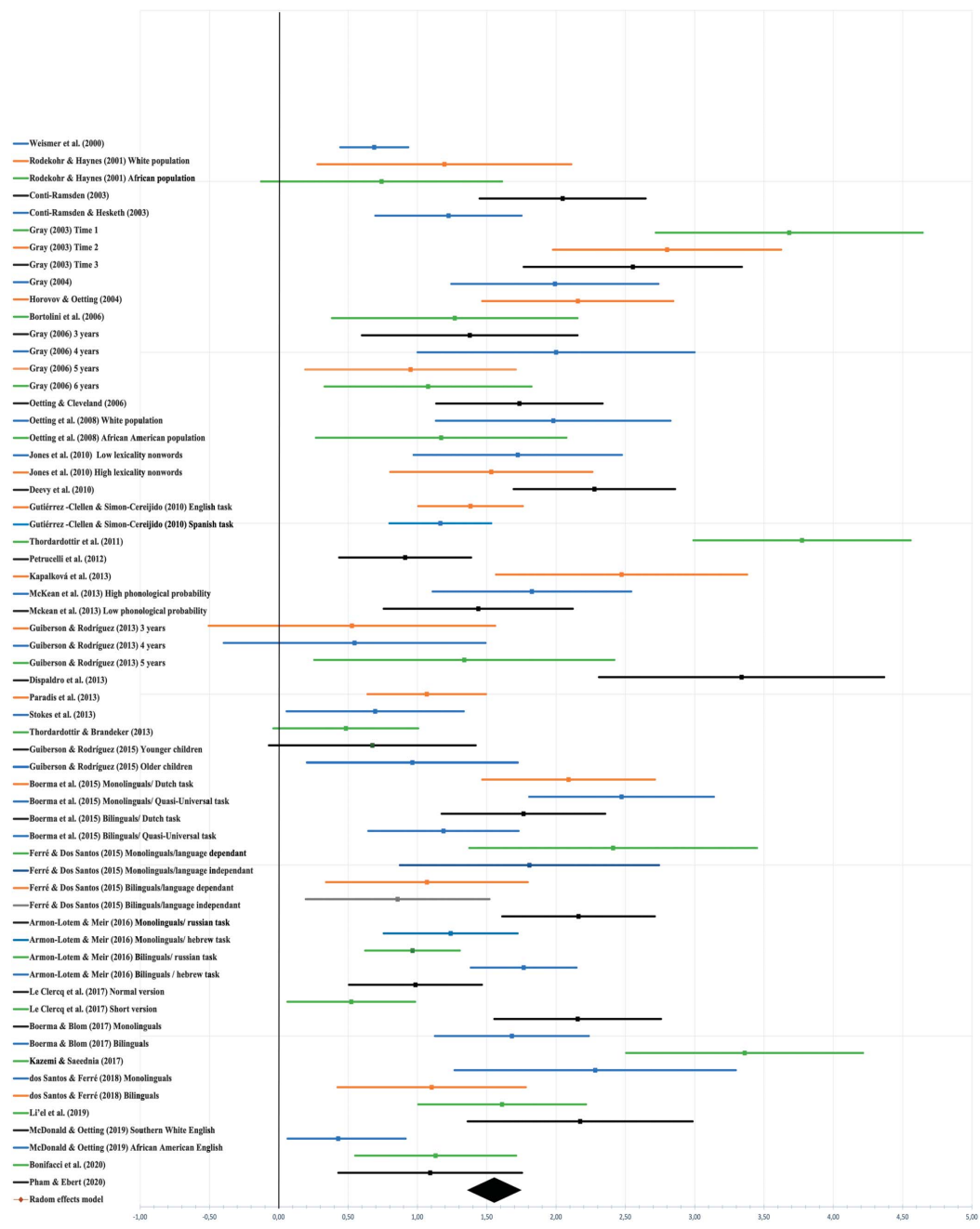
Study	Sample size	Age	Linguistic profile	NVRT used
Weismer et al. (2000)	$n = 581$	7;1–8;11 (years;months)	English monolinguals	Dollaghan & Campbell (1998)
Rodekhorst & Haynes (2001)	$n = 40$	7;0–7;3	English monolinguals	Dollaghan & Campbell (1998)
Conti-Ramsden (2003)	$n = 64$	4;4–5;10	English monolinguals	Gathercole & Baddeley (1996)
Conti-Ramsden & Hesketh (2003)	$n = 64$	DLD: 4;4–5;10/LT: 2;4–3;7	English monolinguals	Gathercole & Baddeley (1996)
Gray (2003)	$n = 44$	4;0–5;11	English monolinguals	Gathercole et al. (1994)
Gray (2004)	$n = 40$	4;0–5;11	English monolinguals	Dollaghan & Campbell (1998)
Horohov & Oetting (2004)	$n = 54$	5–7 years	English monolinguals	Montgomery (1995)
Washington & Craig (2004)	$n = 81$	3;10–4;11	English monolinguals	Dollaghan & Campbell (1998)
Thal et al. (2005)	$n = 64$	4;0–4;6	English monolinguals	Dollaghan & Campbell (1998)
Bortolini et al. (2006)	$n = 22$	3;7–5;6	Italian monolinguals	own task
Gray (2006)	$n = 106$	3–6 years	English monolinguals	Dollaghan & Campbell (1998)
Oetting & Cleveland (2006)	$n = 83$	4;0–6;0	English monolinguals	Dollaghan & Campbell (1998)
Chiat & Roy (2007)	$n = 483$	2;0–4;0	English monolinguals	“PRT” (Roy & Chiat, 2004)
Oetting et al. (2008)	$n = 95$	4;0–6;0	English monolinguals	Dollaghan & Campbell (1998)
Archibald & Joannisse (2009)	$n = 400$	5;3–9;4	English monolinguals	Dollaghan & Campbell (1998)
Jones et al. (2010)	$n = 36$	5;7–6;7	English monolinguals	Gathercole & Baddeley (1996) and own task
Deevy et al. (2010)	$n = 76$	4;1–5;11	English monolinguals	Dollaghan & Campbell (1998)
Gutiérrez-Ciellen & Simon-Cerejido (2010)	$n = 144$	3;0–7;0	English–Spanish bilinguals	Dollaghan & Campbell (1998) and own task
Thordardottir et al. (2011)	$n = 92$	4;1–5;9	French monolinguals	Courcy (2000)
Petrucelli et al. (2012)	$n = 95$	5;0–5;8	English monolinguals	Gathercole & Baddeley (1996)
Kapalkova et al. (2013)	$n = 32$	3–5 years	Slovak monolinguals	own task
McKean et al. (2013)	$n = 50$	3;0–5;11	English monolinguals	Ebert et al. (2008)
Guiberson & Rodriguez (2013)	$n = 44$	3;0–5;10	Spanish monolinguals	Dispalidro et al. (2009, 2011)
Dispalidro et al. (2013)	$n = 34$	3;11–5;8	Italian monolinguals	“CTOPP” (Wagner et al., 1999)
Paradis et al. (2013)	$n = 178$	4;10–8;7	bilinguals English–other languages	own task
Stokes et al. (2006)	$n = 44$	DLD: 4;2–7;7/same-age TD: 4;1–6;9/younger	Cantonese monolinguals	
Thordardottir & Brandeker (2013)	$n = 140$	TD: 2;11–3;6 4;4–5;9	French monolinguals, bilinguals French–other languages	Courcy (2000), Thordardottir et al. (2011), and Gathercole et al., (1994)
Tuller et al. (2013)	$n = 29$	5–7 years	French–Arabic bilinguals	French LITMUS
Topbas et al. (2014)	$n = 60$	3;6–8;3	Turkish monolinguals	own task
Guiberson & Rodriguez (2015)	$n = 65$	2;0–2;11	Spanish monolinguals	own task
Boerma et al. (2015)	$n = 120$	4;6–7;3	Dutch monolinguals, bilinguals Dutch–other languages	Rispens & Baker (2012), Dutch LITMUS
Ferre & Dos Santos (2015)	$n = 67$	5;2–8;5	French monolinguals, French–Arabic/English bilinguals	French LITMUS

(table continues)

Table 1. (Continued).

Study	Sample size	Age	Linguistic profile	NWRT used
De Almeida et al. (2016)	<i>n</i> = 136	5;4–8;11	French monolinguals, French Portuguese/Arabic/ Turkish bilinguals	French LITMUS and “Palpa-P” (Castro et al., 2007)
Armon-Lotem & Meir (2016)	<i>n</i> = 231	5;5–6;8	Hebrew and Russian monolinguals, Hebrew–Russian bilinguals	Shortened version of Armon-Lotem & Chiat (2012)
De Almeida et al. (2017)	<i>n</i> = 137	5;4–8;11	French monolinguals, French Portuguese/Arabic/ Turkish bilinguals	French LITMUS and “Palpa-P” (Castro et al., 2007)
Le Clercq et al. (2017)	<i>n</i> = 88	5;4–8;10	Dutch monolinguals	Normal version and shortened version of Rispen & Baker (2012)
Boerma & Blom (2017)	<i>n</i> = 132	5;0–6;0	Dutch monolinguals, Dutch–Berber/Moroccan/ Turkish bilinguals	Quasi-universal LITMUS (Chiat, 2015)
Hodges et al. (2017)	<i>n</i> = 52	2;1–2;11	English monolinguals	“TENR” (Stokes & Klee, 2009) and “MITT” (Hodges et al., 2017)
Kazemi & Saeednia (2017)	<i>n</i> = 51	3;7–5;3	Persian monolinguals	Afshar et al. (2013)
Marini et al. (2017)	<i>n</i> = 293	2;2–2;11	Italian monolinguals	“BVL” (Marini et al., 2015)
Tuller et al. (2018)	<i>n</i> = 227	5;6–8;11	French and German monolinguals, French/German– Arabic/Portuguese/Turkish bilinguals	French and German LITMUS
Dos Santos & Ferré (2018)	<i>n</i> = 67	5;2–8;5	French monolinguals, French/Arabic/English bilinguals	French LITMUS
Li’el et al. (2019)	<i>n</i> = 61	5;0–6;0	bilinguals English–other languages	“CTOPP” (Wagner et al., 1999)
McDonald & Oetting (2019)	<i>n</i> = 106	4;11–6;2	English monolinguals	Dollaghan & Campbell (1998)
Pham & Ebert (2020)	<i>n</i> = 104	5;2–6;2	Vietnamese monolinguals	Pham et al. (2018)
Bonifacci et al. (2020)	<i>n</i> = 55	6;4–7;4	bilinguals Italian–other languages	“PVN 5–11” (Bisiacchi et al., 2005)

Note. *N* = 46, listed by order of publication date. DLD = developmental language disorders; LT = late talkers; TD = typical development.

Figure 2. Forest plot. $N = 35$.

studies). However, as can be concluded from the results in Table 3, the three moderators do not significantly influence effect size as assumed, neither do their interactions. Thus, although there may be small trends with respect to language status and the language specificity of the stimuli in the subgroup

analysis, these should be interpreted with caution as they are not statistically confirmed, which may of course be due to the small and unequal sample size of the studies.

It may seem surprising at first that age does not have a significant impact, but it is worth noting that, here, we do

Table 2. Mean effect size and subgroups analysis.

Task/group	Hedges's <i>g</i>	Standard error	Variance	Lower limit	Upper limit	Z value	<i>p</i> value
Language-specific	1.52	0.09	0.01	1.33	1.71	15.51	<i>p</i> < .001***
Quasi-universal	1.73	0.18	0.03	1.36	2.10	9.15	<i>p</i> < .001***
Bilinguals	1.36	0.10	0.01	1.16	1.57	12.95	<i>p</i> < .001***
Monolinguals	1.61	0.11	0.01	1.37	1.84	13.64	<i>p</i> < .001***
Overall	1.57	0.08	0.00	1.40	1.73	18.19	<i>p</i> < .001***

Note. Some studies were counted multiple times because they detailed their results according to several tasks or groups of children (e.g., age, language status). See details in Supplemental Material S1. For language specificity of the task, *N* = 59 (language-specific task, *n* = 53; quasi-universal task, *n* = 6). One study (looking at two groups of children, monolingual and bilingual) could not be clearly identified as having used a language-specific or quasi-universal task, that is, Dos Santos and Ferré (2018), those who used tasks that included language-dependent and language-independent items. For linguistic status of the children, *N* = 61 (bilingual children, *n* = 13, monolingual children, *n* = 48).

****p* < .001.

not examine the influence of age on the children's scores per se, but on performance differences between children with DLD and TD of the same age. Furthermore, the NWRT stimuli, often taken from test batteries, are generally chosen to fit with the age range of the children studied, and many tasks contain (sometimes age-specific) abortion criteria in order to make them maximally discriminative for all ages. Thus, the discriminative power of NWRT does not seem to vary across the age range examined (2;00–8;11), and we can therefore confirm that NWRT is a reliable indicator of DLD across development in young children.

Finally, to test for publication bias, we created a funnel plot (in the Supplemental Material S2). We noted missing data mostly on the left side of the figure, indicating that studies with small effect sizes or with nonsignificant data are missing. The Egger's regression was statistically significant ($t = 4.13$, $p < .001$), indicating a publication bias. We can think of several explanations for this fact: First, NWRT is a well-established and robust clinical task used in many test batteries that are designed to maximize the effect of DLD; thus, studies not finding such an effect at least in monolingual children would be very hard to publish. Second, despite our efforts to contact authors in the field and to participate in relevant conferences, we were not able to find and include unpublished gray literature on the topic.

However, it should be noted that the main point of our review was not to establish whether NWRT can be used to distinguish TD and DLD children, which has been done before, but to compare how it fares for monolingual and bilingual children as well as analyze the influence of other factors.

Qualitative Analysis

In the following, we will qualitatively assess other factors that may influence NWR scores, but that could not be included in the meta-analysis because of lack of data. These factors are mentioned by few studies, and often the details (means and standard deviations) relating to them do not appear in the articles. Again, we focus specifically on the influence of these factors on the discriminative power of the task, rather than scores per se.

Stimuli Characteristics

Referring to Table 1 (presented earlier), we note that different tasks were used. Several studies used language-specific tasks like Dolloghan and Campbell's (1998) task or the NWRT developed by Gathercole et al. (1994) and Gathercole and Baddeley (1996). Some studies created their own NWRT (e.g., Bortolini et al., 2006) or drew on other

Table 3. Moderators analysis.

Covariate	Coefficient	Standard error	95% Lower	95% Upper	Z value	<i>p</i> value
Intercept	1.94	0.50	0.95	2.93	3.85	<i>p</i> < .001***
Linguistic status	0.14	0.23	-0.31	0.61	0.62	<i>p</i> = .53
Task type	0.08	0.42	-0.75	0.92	0.21	<i>p</i> = .83
Linguistic Status × Task Type	0.44	0.59	-0.71	1.60	0.75	<i>p</i> = .45
Age	-0.00	0.00	-0.02	0.00	-1.24	<i>p</i> = .21

Note. Some studies were counted multiple times because they detailed their results according to several tasks or groups of children (e.g., age, language status). See details in Supplemental Material S1. For language specificity of the task, *N* = 59 (language-specific task, *n* = 53; quasi-universal task, *n* = 6). One study (looking at two groups of children, monolingual and bilingual) could not be clearly identified as having used a language-specific or quasi-universal task, that is, Dos Santos and Ferré (2018), those who used tasks that included language-dependent and language-independent items. For linguistic status of the children, *N* = 61 (bilingual children, *n* = 13, monolingual children, *n* = 48). For age of children, *N* = 61. Test of the model: $Q = 3.95$, $df = 4$, $p = .41$.

****p* < .001.

tasks previously designed or derived from test batteries (for details, refer to Table 1). A few studies have also used Chiat's (2015) quasi-universal task and combinations of language-dependent and language-independent items (e.g., Dos Santos & Ferré, 2018).

Different characteristics of the NWRT stimuli have been shown to influence children's scores. Turns out, 46 studies have focused on syllable structure (Dos Santos & Ferré, 2018; Ferré & Dos Santos, 2015; Hodges et al., 2017; Jones et al., 2010; Marini et al., 2017). This varies widely among the different studies. Some studies used a single syllabic structure (CV or CVC), while others combined different syllabic structures within nonwords (e.g., CV + CVC, CV + CCV). Dos Santos and Ferré (2018) found that, in the presence of complex consonant groups, the performance of DLD children decreases in both monolinguals and bilinguals. Children with DLD have more difficulties than children with TD when the structures are complex (Jones et al., 2010). Dos Santos and Ferré also specified that the complexity of items can be influenced by whether or not they resemble the child's language. Therefore, on the basis of the LITMUS task, these authors create stimuli that are dependent on and independent of the language spoken by the child. The results show that with the exception of monolingual children with DLD, none of the groups had a significant difference between the two types of stimuli of the test. Other authors such as Hodges et al. (2017) have focused on the acquisition of phonemes in relation to syllabic structure. They found that the effect of complexity appears only when the initial consonants are those acquired early and not late. This is valid for children with DLD and with TD, even if for the latter the effect is less pronounced.

Nonword length is another factor that is often considered important. In the studies reviewed here, it varies between one and five syllables. Some authors have chosen not to go beyond four syllables, given the age of the children in their sample (e.g., Jones et al., 2010). Of the 46 studies in our literature review, 18 found an effect of the length of nonword (Boerma et al., 2015; Bortolini et al., 2006; Chiat & Roy, 2007; Dispaldro et al., 2013; Guiberson & Rodríguez, 2013; Guiberson & Rodríguez, 2015; Horohov & Oetting, 2004; Jones et al., 2010; Marini et al., 2017; McDonald & Oetting, 2019; McKean et al., 2013; Oetting & Cleveland, 2006; Petruccelli et al., 2012; Rodekohr & Hayne, 2001; Thal et al., 2005; Thordardottir & Brandeker, 2013; Topbaş et al., 2014; Weismer et al., 2000). In general, both children with TD and children with DLD have more difficulty with NWRT as stimulus length increases (Bortolini et al., 2006; Chiat & Roy, 2007; Dispaldro et al., 2013; Guiberson & Rodríguez, 2013; Horohov & Oetting, 2004; Jones et al., 2010; Marini et al., 2017; Oetting & Cleveland, 2006; Petruccelli et al., 2012; Rodekohr & Haynes, 2001; Thal et al., 2005; Topbaş et al., 2014). Boerma et al. (2015) noted that monolingual and bilingual children with DLD perform less well than those with TD on all syllable lengths. Some authors (Dispaldro et al., 2013; McKean et al., 2013; Petruccelli et al., 2012) claimed that there is a significant difference between children with DLD and TD only when

the length of nonwords increases, reflecting the limited short-term verbal memory capacity of children with DLD. One study found that children with DLD were significantly less accurate than children with TD on nonwords with two and three syllables of simple structures (Jones et al., 2010). The study by Petruccelli et al. (2012) identified that late-talking children's scores of fall between the scores of children with TD and children with DLD. Children with DLD and late-talking children have more difficulties than children with TD, especially on three- and four-syllable nonwords (Weismer et al., 2000). Finally, some authors claim that, for preschool children, it is not relevant to use four- and five-syllable nonwords in the NWRT because of floor effects (Guiberson & Rodríguez, 2015).

Furthermore, five of the 46 studies examined possible effects of the wordlikeness of the nonwords used (Chiat & Roy, 2007; Gray, 2003; Hodges et al., 2017; Jones et al., 2010; McKean et al., 2013). McKean et al. (2013) found a wordlikeness effect for children with TD and with DLD. Jones et al. (2010) confirmed these results, although the effect is much stronger for children with DLD than for children with TD. Hodges et al. (2017) observed this, but with late-talking children instead of children with DLD. Gray (2003) reflected that the wordlikeness effect may signal a greater influence of prior language knowledge. Finally, lexical status seems to influence performance as well, as shown by Chiat and Roy (2007) in their study comparing repetition of words and nonwords. Chiat and Roy (2007) also looked at prosody. They noted that accented syllables are almost never omitted. Children with DLD omitted accented syllables only in long nonwords (bisyllables and trisyllables). However, children with DLD omit nonaccented syllables that follow accented syllables more easily, but this is not the case for children with TD.

Scoring Method

Different scoring methods were used, depending on the tasks and on authors' preferences (e.g., Gray, 2003, 2006). Two main scoring methods were identified in the review studies. First, a relatively straightforward count of percentage of items correct (PIC) was used by 21 studies (Archibald & Joanisse, 2009; Armon-Lotem & Meir, 2016; Boerma & Blom, 2017; Bonifacci et al., 2020; Chiat & Roy, 2007; Conti-Ramsden, 2003; Conti-Ramsden & Hesketh, 2003; De Almeida et al., 2016; De Almeida et al., 2017; Gray, 2003; Gray, 2006; Guiberson & Rodríguez, 2015; Horohov & Oetting, 2004; Jones et al., 2010; Li'el et al., 2019; Marini et al., 2017; Paradis et al., 2013; Petruccelli et al., 2012; Thal et al., 2005; Tuller et al., 2013, 2018).

Second, 12 studies in the review count the percentage of phonemes correct (PPC; Deevy et al., 2010; Gray, 2004; Gutiérrez-Clellen & Simon-Cerejido, 2010; Hodges et al., 2017; McDonald & Oetting, 2019; Oetting & Cleveland, 2006; Oetting et al., 2008; Rodekohr & Haynes, 2001; Thordardottir & Brandeker, 2013; Thordardottir et al., 2011; Washington & Craig, 2004; Weismer et al., 2000). The precise way in which the PPC is calculated can influence the method's discriminating performance. For example, Deevy et al. (2010) compared two PPC methods (e.g., by

treating phonemes out of inventory either as errors or as unscorable) and observed differences in their percentages of sensitivity and specificity.

Eight out of 46 studies combined different scoring methods: For example, Topbaş et al. (2014) combined PIC and PPC and added the percentage of vowels correct. McKean et al. (2013) calculated the number of incorrect phonemes in each nonword, and Dos Santos and Ferré (2018) also looked at the number of errors present in each item. These three studies then linked this score to the PIC or PPC. A final study conducted in Cantonese (Stokes et al., 2006) attempted to assign a point for each correct nucleus and consonant. This shows that, depending on the language, in this case Cantonese, the scoring method can differ for linguistic reasons. Bortolini et al. (2006) combined the PIC and the number of errors per segment. Finally, Kapalková et al. (2013) and Pham and Ebert (2020) combined several measures (e.g., between three and five depending on the study). They observed that all the measures used made it possible to dissociate children with DLD from those with TD. A few studies (Boerma et al., 2015; Deevy et al., 2010; Dispaldro et al., 2013; Guiberson & Rodríguez, 2013; Le Clercq et al., 2017; Pham & Ebert, 2020) used several scoring methods (including PIC and PPC) to compare them. The authors found that both methods discriminated between children with TD and DLD. However, the PIC had better sensitivity (e.g., 97% in Boerma et al., 2015; 71% in Guiberson & Rodríguez, 2013; and 100% in Dispaldro et al., 2013) in contrast to the PPC (63%, 48%, and 94%, respectively). Furthermore, Boerma et al. (2015) found that PIC was more accurate for monolingual children (in the quasi-universal NWRT), whereas no difference was found between PPC and PIC for bilingual children. Le Clercq et al. (2017) observed similar results for the effectiveness of PIC and PPC. Finally, PIC scoring is faster and therefore seems interesting for the clinic (Dispaldro et al., 2013; Pham & Ebert, 2020).

Error Analysis

As noted above, beyond a generally lower performance on NWRT, children with DLD may also make different errors than children with TD. Kapalková et al. (2013) found more phonological errors in Slovak children with DLD, including some atypical errors such as nasal and alveolar assimilations. In a study in which children were considered at risk for DLD (children were referred to speech language therapy because they were suspected of having a DLD), older at-risk children scored similarly to younger children with TD, omitting syllables more often than children with TD of their age (Chiat & Roy, 2007). However, only few studies make use of a complete error analysis, which is very time consuming, and none of them analyzed its potential to increase overall diagnostic accuracy.

Sociocultural Characteristics

Two studies (Chiat & Roy, 2007; Washington & Craig, 2004) found no effect of gender and socioeconomic status on children's performance in the NWRT. Study by

Boerma et al. (2015) also found no effect of socioeconomic status on NWRT scores. In addition, Chiat and Roy (2007) found that there was no effect of day care type and languages spoken at home on scores.

Discriminative Value of the NWRT

Going beyond group differences, we will now turn to a clinically highly relevant parameter, the discriminative power of the NWRT to detect DLD in individuals among monolingual and bilingual children. We note that 33 out of 46 studies, summed up in the Supplemental Material S3, report the sensitivity and specificity of the NWRT. We note large methodological differences in the pathology thresholds adopted (sometimes expressed as scores, percentile, standard deviation) and, in the tasks, scoring methods and inclusion criteria of DLD used. The discriminative value of the NWRT also varies largely, with sensitivity ranging from 40% to 100% and specificity from 48% to 100%. We note that the majority of studies consider NWRT to be accurate since they have sensitivity and specificity percentages between 80% and 100% (e.g., Plante & Vance, 1994). However, 14 studies obtained at least one result considered to be diagnostically inaccurate. Seven of these studies involved a monolingual population. We note that some authors (e.g., Conti-Ramsden, 2003) have attempted to vary the threshold score (based on receiver operating characteristic curves; Dunn, 2014) to be reached but the percentages did not improve. It should be noted that the study by Thal et al. (2005) carried out on a monolingual population obtained a low sensitivity percentage (40%) for the NWRT alone, but this result improves significantly when the NWRT is combined with other tools. This point will be discussed more fully in later sections. Among the seven studies reporting poor discriminant results in a bilingual population, the majority used language-specific tasks (e.g., Dos Santos & Ferré, 2018; Gutiérrez-Clellen & Simon-Cerejido, 2010), sometimes influencing the subject's ability to repeat according to their experience in the language of the task (Boerma et al., 2015; Kohnert et al., 2006). The quasi-universal LITMUS task (Chiat, 2015) seems promising because it provides results that are still considered accurate in bilinguals. With this bilingual population, the deviation thresholds sometimes have to be modified in order to find the optimal discriminating values for children with and without DLD (e.g., Thordardottir, 2015). Indeed, Armon-Lotem and Meir (2016) also noted that, depending on the language of the tasks used, sensitivity and specificity scores can sometimes improve (e.g., the case for Hebrew but not for Russian). Finally, three studies (Deevy et al., 2010; Guiberson & Rodríguez, 2013; Pham & Ebert, 2020) noted differences in sensitivity and specificity depending on the scoring method. It is therefore not possible to conclude whether it is better to use an NWRT in the language of the society and/or family. Indeed, the skills of the bilingual child differ according to the time and tasks proposed (Gutiérrez-Clellen & Simon-Cerejido, 2010). Moreover, the latter can sometimes be designed with

various methodological properties (Dos Santos & Ferré, 2018). In order to obtain the best diagnostic accuracy, it therefore seems appropriate to carry out NWRTs in both of the child's languages or to use a quasi-universal task (Tuller et al., 2018).

NWRT and Other Assessment Tools

Correlations Between NWRT and Other Measures

Some authors reported correlations between NWRT scores and other verbal and nonverbal measures (summary presented in the Supplemental Material S4). For example, some studies observed a significant correlation between NWRT scores and vocabulary measures in English, Spanish, Italian, and French (e.g., Thordardottir & Brandeker, 2013). However, two further studies, Stokes et al. (2006) and Thordardottir et al. (2011), found no correlation between vocabulary measures and NWRT scores, neither for children with TD nor with DLD who speak Cantonese and French. Thordardottir and Brandeker (2013) and De Almeida et al. (2017) also found significant correlations between NWRT scores and other measures from standardized French and English language tests tapping into morphosyntax (sentence repetition) and phonology (word repetition). In addition to the aforementioned correlation with vocabulary, Horohov and Oetting (2004) reported significant correlations between NWRT scores and standardized measures of English syntax (syntax comprehension). Finally, Paradis et al. (2013) observed correlations between NWRT and tasks testing inflectional morphology in English for the group of children with TD. These authors, however, observed no correlation between these standardized test measures and NWRT scores for the group of children with DLD.

A number of studies put NWRT in relation with other tasks requiring repetition. Dispaldro et al. (2013) noted a correlation between the ability to repeat nonwords and the ability to repeat words in Italian. De Almeida et al. (2016, 2017) and Thordardottir and Brandeker (2013) observed correlations between children's scores on NWRT and a sentence repetition task in bilinguals with TD and DLD in French, Portuguese, and English. Conversely, the study by Stokes et al. (2006), focusing on Cantonese, found no correlation between NWRT and sentence repetition. Petrucci et al. (2012), on the other hand, found no significant correlation between NWRT scores and scores on number and word recall in English. Gray (2003) noted correlations between the NWRT task and a number repetition task (assessed at different points in time) and other standardized language assessments in English. Other research has investigated links between NWRT and nonverbal tasks, with conflicting results. Oetting et al. (2006, 2008) noted a correlation between NWRT scores and a nonverbal reasoning test, whereas Boerma et al. (2015) found no correlation between nonverbal IQ and NWRT.

Several authors are also interested in possible correlations with parental observations. Tuller et al. (2018) found

that high NWRT scores were most strongly correlated with positive early language development, as established through analysis of a parental questionnaire. Similarly, De Almeida et al. (2016, 2017) noted a correlation between NWRT and a language impairment risk index calculated through a parental questionnaire.

To sum up, NWRT seems to be generally correlated to other assessment measures, in particular those concerning oral language, such as vocabulary, morphosyntax, and early developmental indexes. However, the data are heterogeneous and, to date, no firm generalizations can be made regarding the link between NWRT and other verbal and nonverbal measures.

Combining NWRT With Other Measures

Nine out of 46 studies have addressed the value of combining NWRT with other empirical tools to increase diagnostic accuracy for DLD (see in the Supplemental Material S5). For example, NWRT can be combined with a sentence repetition task (Armon-Lotem & Meir, 2016), a comprehension test (Oetting & Cleveland, 2006), or a receptive vocabulary test (Thordardottir et al., 2011¹). Boerma and Blom (2017) also obtained an increase in discriminant value by associating the NWRT with a narrative task, but only in the group of bilingual children. Bortolini et al. (2006), in their study of Italian, found that sensitivity and specificity improve when NWRT is combined with morphosyntactic tasks, such as third-person plural inflection and a clitic object task. Four studies (Boerma & Blom, 2017; Bonifacci et al., 2020; Li'el et al., 2019; Paradis et al., 2013) showed that the combination of NWRT with a parental questionnaire reporting early language milestones and parental concern can improve the accuracy of diagnosis in different groups of children (monolinguals–bilinguals/children with TD or DLD). Thus, combining NWRT with other tasks can be an effective means to increase diagnostic accuracy, especially in situations where the discriminative value NWRT is not sufficient on its own.

General Discussion and Conclusions

The aim of this work was to report and synthesize recent results on the use of NWRT in monolingual and bilingual SLT assessments. The various studies show that, in general, children with DLD (or at risk of developing DLD) score significantly lower than children with TD, which can be related, among other things, to impairment of their short-term verbal memory. Beyond these memory deficits, children with DLD often have poorer phonological abilities and have more difficulty developing fine and accurate phonological representations (McKean et al., 2013). We also note that children with DLD are less accurate in repetition and make more errors (Jones et al., 2010). The phonological errors of

¹The study by Thordardottir et al. (2011) is included in Supplemental Material S5, but the new sensitivity and specificity scores with combined tasks are not described in the reference article.

these children are sometimes atypical, such as nasal and alveolar assimilations (Kapalková et al., 2013). However, given the number of NWR tasks used in the studies and the associated methodological differences, it is difficult to determine what exactly each type of NWR measures (verbal short-term memory vs. phonological accuracy). We note that children with clearly diagnosed DLD show much more difficulties with all NWR types than their peers without DLD. The large effect size found in our meta-analysis confirms that the NWR appears to be a robust and recommendable tool for detecting DLD in monolingual and bilingual children.

In order to explore why our meta-analysis data were heterogeneous, we examined the influence of three factors known to influence NWR performance, linguistic status, language specificity of the stimuli, and age. Indeed, subgroups analysis showed that effect sizes appear to vary according to the linguistic status of the children (with larger effects for monolinguals) and the language specificity of the task (with larger effects for quasi-universal tasks). However, these factors were not significant moderators in a metaregression, which may be due to the small and unequal number of studies included. In addition, with respect to age, discriminative power does not seem to vary by age in children up to age 8;11, who represent the main target group for DLD assessment. However, it is possible that an inclusion of older children (for instance, in the context of reading difficulties) would yield an influence of the age factor.

Other child-internal factors were analyzed qualitatively such as sociocultural, socioeconomic, gender, and day care attendance, and they do not appear to have an impact on children's scores (e.g., Chiat & Roy, 2007). It appears then that the NWR is relatively transparent to the child's characteristics and could be used with a relatively heterogeneous population.

We also had to address several well-known factors relating to stimuli and task characteristics qualitatively, due to the limited number of studies that report this type of details. We noted that an effect of syllabic length of the nonword items has been demonstrated many times, in particular on the performance of children with DLD (e.g., Boerma et al., 2015; Thordardottir & Brandeker, 2013). Although analyzed systematically in fewer studies, an effect of syllabic complexity has also been shown (e.g., Dos Santos & Ferré, 2018). Wordlikeness and prosody have been studied infrequently and can also be important (e.g., Chiat & Roy, 2007). We were also able to highlight differences in the scoring method. The two most commonly used methods are PIC (counting correct items) and PPC (counting correct phonemes). Conveniently, it appears that PIC is faster, easier to use, and more discriminating (Le Clercq et al., 2017). We can therefore think that these factors, which are more complex to analyze quantitatively, could constitute other interesting moderators to deal with. In particular, Estes et al., (2007) had shown that the length of stimuli was an important moderator of effect size.

Despite significant methodological differences (e.g., type of task and stimuli, age of children, analysis),

we noted that they unanimously mention the NWR's high discriminative power for detection of DLD in monolingual and bilingual children. Most studies reported acceptable sensitivity and specificity (over 80% following Plante & Vance, 1994). The studies that obtained poor discriminative accuracy mainly concern language-specific tasks, where the scores obtained by children may depend on their language experience in the language of the task (Boerma et al., 2015; Kohnert et al., 2006). One solution for this issue may be the use of a quasi-universal task, like the one devised by Chiat (2015), which still allows for good discrimination. However, we noted that some of the poor discriminative accuracy scores also come from studies in a monolingual population (e.g., Conti-Ramsden, 2003). We can therefore hypothesize that methodological choices regarding the design of certain tasks can also contribute to low accuracy percentages.

Finally, it also appears that performance in NWR is frequently correlated with that on other tasks and that it could be combined with other tools to increase diagnostic accuracy. Several studies associated NWR with other tasks such as sentence repetition, expressive and receptive lexical tasks, narrative tasks, or a parental questionnaire, which all increased diagnostic accuracy to a certain extent. In absence of a clearly superior combination, clinician may be able to choose associated tasks depending on the child's profile of difficulty, the family's request, and what is available in a specific language. It should also be kept in mind that the NWR highlights phonological and short-term verbal memory difficulties in children with DLD (Ferré & Dos Santos, 2015; Gray, 2004; Li'el et al., 2019). However, some children with DLD may not have phonological difficulties, making the complementary use of other tools in assessment sessions essential (Kapalková et al., 2013). This recommendation fits well with the long-standing clinical tradition to combine different diagnostic sources in order to establish with relative certainty the actual presence of DLD in a child (Bishop & McDonald, 2009), whether monolingual or bilingual. Moreover, combining tasks focused on the evaluation of procedural skills, as the NWR, with tests assessing the child's declarative knowledge (e.g., lexical, grammatical, narrative skills), is also important for guiding intervention.

To conclude, we note that the NWR is a tool that has been found to be effective in the studies reviewed and thus deserves a fixed place in language assessments of monolingual and bilingual children. Of course, as already mentioned, the type of task and stimuli need to be carefully chosen to fit the age and linguistic profile of the child, and NWR should be complemented with other tools in order to obtain the most complete language profile possible and to corroborate any diagnostic indications from NWR by different sources.

Acknowledgments

We would like to thank our colleague, Nuzhat, for her revision.

References

- Afshar, M. R., Gorbani, A., Jalilevand, N., & Kamali, M. (2013). Providing the non-word repetition test and determining its validity and reliability and comparing phonological working memory in 4 to 6 Farsi-speaking normal and SSD children in Tehran City. *Journal of Research in Rehabilitation Sciences*, 9(5), 899–911.
- Archibald, L. (2008). The promise of nonword repetition as a clinical tool. *Canadian Journal of Speech-Language Pathology and Audiology*, 32(1), 21–28. https://cjslpa.ca/files/2008_CJSLPA_Vol_32/No_01_1-68/Archibald_CJSLPA_2008.pdf
- Archibald, L., & Joanisse, M. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, 52(4), 899–914. [https://doi.org/10.1044/1092-4388\(2009/08-0099\)](https://doi.org/10.1044/1092-4388(2009/08-0099))
- Armon-Lotem, S. (2012). Introduction: Bilingual children with SLI—The nature of the problem. *Bilingualism: Language and Cognition*, 15(1), 1–4. <https://doi.org/10.1017/S1366728911000599>
- Armon-Lotem, S., & Chiat, S. (2012). How do sequential bilingual children perform on non-word repetition tasks? In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th annual Boston University Conference on Language Development* (pp. 53–62). Cascadia Press.
- Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, 51(6), 715–731. <https://doi.org/10.1111/1460-6984.12242>
- Bishop, D. V. M., & McDonald, D. (2009). Identifying language impairment in children: Combining language test scores with parental report. *International Journal of Language & Communication Disorders*, 44(5), 600–615. <https://doi.org/10.1080/13682820802259662>
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry*, 58(10), 1068–1080. <https://doi.org/10.1111/jcpp.12721>
- Bisiacchi, P. S., Cendron, M., Gugliotta, M., Tressoldi, P. E., & Vio, C. (2005). *BVN 5–11: Batteria di valutazione neuropsicologica per l'età evolutiva* [BVN 5–11: Neuropsychological assessment battery for the developmental age]. Centro Studi Erickson.
- Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible. *Journal of Communication Disorders*, 66(1), 65–76. <https://doi.org/10.1016/j.jcomdis.2017.04.001>
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research*, 58(6), 1747–1760. https://doi.org/10.1044/2015_JSLHR-L15-0058
- Bonifacci, P., Atti, E., Casamenti, M., Piani, B., Porrelli, M., & Mari, R. (2020). Which measures better discriminate language minority bilingual children with and without developmental language disorder? A study testing a combined protocol of first and second language assessment. *Journal of Speech, Language, and Hearing Research*, 63(6), 1898–1915. https://doi.org/10.1044/2020_JSLHR-19-00100
- Bornstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). *Comprehensive meta-analysis Version 3*. Biostat.
- Bortolini, U., Arfè, B., Caselli, C., Degasperis, L., Deevy, P., & Leonard, L. (2006). Clinical markers for specific language impairment in Italian: The contribution of clitics and non-word repetition. *International Journal of Language & Communication Disorders*, 41(6), 695–712. <https://doi.org/10.1080/13682820600570831>
- Camilleri, B., & Law, J. (2007). Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology*, 9(4), 312–322. <https://doi.org/10.1080/14417040701624474>
- Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R., & Coltheart, M. (2007). *Provas de avaliação da linguagem e da afasia em português*. CEGOC.
- Chiat, S. (2015). Non-word repetition. In S. Armon-Lotem, K. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95–122). Multilingual Matters. <https://doi.org/10.21832/9781783093137-008>
- Chiat, S., & Roy, P. (2007). The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*, 50(2), 429–443. [https://doi.org/10.1044/1092-4388\(2007/030\)](https://doi.org/10.1044/1092-4388(2007/030))
- Coady, J., & Evans, J. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, 43(1), 1–40. <https://doi.org/10.1080/13682820601116485>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, 46(5), 1029–1037. [https://doi.org/10.1044/1092-4388\(2003/082\)](https://doi.org/10.1044/1092-4388(2003/082))
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *The Journal of Child Psychology and Psychiatry*, 42(6), 741–748. <https://doi.org/10.1111/1469-7610.00770>
- Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for SLI: A study of young language-learning children. *International Journal of Language & Communication Disorders*, 38(3), 251–263. <https://doi.org/10.1080/1368282031000092339>
- Courcy, A. (2000). *Conscience phonologique et apprentissage de la lecture* [Phonological awareness and reading acquisition] (Unpublished doctoral dissertation, Université de Montréal).
- De Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., & Zebib, R. (2016). L'identification d'enfants bilingues avec trouble spécifique du langage en France. *SHS Web of Conferences*, 27, 10005. <https://doi.org/10.1051/shsconf/20162710005>
- De Almeida, L., Ferré, S., Morin, E., Prévost, P., Dos Santos, C., Tuller, L., Zebib, R., & Barthez, M.-A. (2017). Identification of bilingual children with specific language impairment in France. *Linguistic Approaches to Bilingualism*, 7(3–4), 331–358. <https://doi.org/10.1075/lab.15019.alm>
- Deevy, P., Weil, L. W., Leonard, L. B., & Goffman, L. (2010). Extending use of the NRT to preschool-age children with and without specific language impairment. *Language, Speech, and Hearing Services in Schools*, 41(3), 277–288. [https://doi.org/10.1044/0161-1461\(2009/08-0096\)](https://doi.org/10.1044/0161-1461(2009/08-0096))
- Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language &*

- Communication Disorders*, 43(4), 361–389. <https://doi.org/10.1080/13682820701546854>
- Di Meo, S., Sanson, C., Simon, A., Bossuoy, M., Rakotomala, L., Rezzoug, D., Serre, G., Baudet, T., & Moro, M. R. (2014). Le bilinguisme des enfants migrants. In H. Bijleveld, F. Estienne, & F. Vander Linden (Eds.), *Multilinguisme et orthophonie. Réflexions et pratiques à l'heure de l'Europe* (pp. 149–171). Elsevier Masson.
- Dispaldro, M., Benelli, B., Marcolini, S., & Stella, G. (2009). Real-word repetition as a predictor of grammatical competence in Italian children with typical language development. *International Journal of Language & Communication Disorders*, 44(6), 941–961. <https://doi.org/10.3109/13682820802491794>
- Dispaldro, M., Deevy, P., Altoé, G., Benelli, B., & Leonard, L. B. (2011). A cross-linguistic study of real-word and non-word repetition as predictors of grammatical competence in children with typical language development. *International Journal of Language & Communication Disorders*, 46(5), 564–578. <https://doi.org/10.1111/j.1460-6984.2011.00008.x>
- Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-word and nonword repetition in Italian-speaking children with specific language impairment: A study of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, 56(1), 323–336. [https://doi.org/10.1044/1092-4388\(2012\)11-0304](https://doi.org/10.1044/1092-4388(2012)11-0304)
- Dollaghan, C., & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136–1146. <https://doi.org/10.1044/jslhr.4105.1136>
- Dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition*, 25(1), 58–71. <https://doi.org/10.1080/10489223.2016.1243692>
- Dunn, G. (2014). Statistics in psychiatry. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1136–1138). Springer Science & Business Media.
- Ebert, K. D., Kalanek, J., Cordero, K. N., & Kohnert, K. (2008). Spanish nonword repetition: Stimuli development and preliminary results. *Communication Disorders Quarterly*, 29(2), 67–74. <https://doi.org/10.1177/1525740108314861>
- Estes, K. G., Evans, J., & Else-Quest, N. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177–195. [https://doi.org/10.1044/1092-4388\(2007\)015](https://doi.org/10.1044/1092-4388(2007)015)
- Ferré, S., & Dos Santos, C. (2015). Comment évaluer la phonologie des enfants bilingues. *Revue de linguistique et de didactique des langues*, 51(1), 11–34. <https://doi.org/10.4000/lidil.3678>
- Gathercole, S. E., & Baddeley, A. D. (1996). *The children's test of nonword repetition*. The Psychological Corp.
- Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, 2(2), 103–127. <https://doi.org/10.1080/09658219408258940>
- Gray, S. (2003). Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, 36(2), 129–151. [https://doi.org/10.1016/S0021-9924\(03\)00003-0](https://doi.org/10.1016/S0021-9924(03)00003-0)
- Gray, S. (2004). Word learning by preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 47(5), 1117–1132. [https://doi.org/10.1044/1092-4388\(2004\)083](https://doi.org/10.1044/1092-4388(2004)083)
- Gray, S. (2006). The relationship between phonological memory, receptive vocabulary, and fast mapping in young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 45(5), 955–969. [https://doi.org/10.1044/1092-4388\(2006\)069](https://doi.org/10.1044/1092-4388(2006)069)
- Grosjean, F. (2015). *Parler plusieurs langues: Le monde des bilingues*. Albin Michel.
- Guiberson, M., & Rodríguez, B. L. (2013). Classification accuracy of nonword repetition when used with preschool-age Spanish-speaking children. *Language, Speech, and Hearing Services in Schools*, 44(2), 121–132. [https://doi.org/10.1044/0161-1461\(2012\)12-0009](https://doi.org/10.1044/0161-1461(2012)12-0009)
- Guiberson, M., & Rodríguez, B. L. (2015). Nonword repetition in Spanish-speaking toddlers with and without early language delays. *Folia Phoniatrica et Logopaedica*, 67(5), 253–258. <https://doi.org/10.1159/000442745>
- Gutiérrez-Ciellen, V., & Simon-Cerejido, G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish-English speaking children: Does the language of assessment matter. *Learning Disabilities Research & Practice*, 25(10), 48–58. <https://doi.org/10.1111/j.1540-5826.2009.00300.x>
- Hodges, R., Munro, N., Baker, E., McGregor, K., & Heard, R. (2017). The monosyllable imitation test for toddlers: Influence of stimulus characteristics on imitation, compliance and diagnostic accuracy. *International Journal of Language & Communication Disorders*, 52(1), 30–45. <https://doi.org/10.1111/1460-6984.12249>
- Horohov, J. E., & Oetting, J. B. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics*, 25(1), 43–65. <https://doi.org/10.1017/S0142716404001031>
- Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in specific language impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language, and Hearing Research*, 53(6), 1642–1655. [https://doi.org/10.1044/1092-4388\(2010\)09-0222](https://doi.org/10.1044/1092-4388(2010)09-0222)
- Kapalková, S., Polišenská, K., & Vicensová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: Novel scoring methods. *International Journal of Language & Communication Disorders*, 48(1), 78–89. <https://doi.org/10.1111/j.1460-6984.2012.00189.x>
- Kazemi, Y., & Saeednia, S. (2017). The clinical examination of non-word repetition tasks in identifying Persian-speaking children with primary language impairment. *International Journal of Pediatric Otorhinolaryngology*, 93(1), 7–12. <https://doi.org/10.1016/j.ijporl.2016.11.028>
- Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals. *Learning Disabilities Research & Practice*, 21(1), 19–29. <https://doi.org/10.1111/j.1540-5826.2006.00204.x>
- Le Clercq, C. M. P., Van Der Schroeff, M. P., Rispens, J. E., Ruytjens, L., Goedegebure, A., Van Ingen, G., & Franken, M.-C. (2017). Shortened nonword repetition task (NWR-S): A simple, quick, and less expensive outcome to identify children with combined specific language and reading impairment. *Journal of Speech, Language, and Hearing Research*, 60(8), 2241–2248. https://doi.org/10.1044/2017_JSLHR-L-16-0060
- Li'el, N., Williams, C., & Kane, R. (2019). Identifying developmental language disorder in bilingual children from diverse linguistic backgrounds. *International Journal of Speech-Language Pathology*, 21(6), 613–622. <https://doi.org/10.1080/17549507.2018.1513073>
- Marini, A., Marotta, L., Bulgheroni, S., & Fabbro, F. (2015). *Batteria per la valutazione del linguaggio in bambini dai 4 ai 12 anni*. Giunti O.S.

- Marini, A., Ruffino, M., Sali, M. E., & Molteni, M. (2017). The role of phonological working memory and environmental factors in lexical development in Italian-speaking late talkers: A one-year follow-up study. *Journal of Speech, Language, and Hearing Research*, 60(12), 3462–3473. https://doi.org/10.1044/2017_JSLHR-L15-0415
- Martin, V., Renaud, J., & Dagenais, P. (2013). *Les normes de production des revues systématiques: Guide méthodologique*. National Institute of Excellence in Health and Social Services.
- McDonald, J. L., & Oetting, J. B. (2019). Nonword repetition across two dialects of English: Effects of specific language impairment and nonmainstream form density. *Journal of Speech, Language, and Hearing Research*, 62(5), 1381–1391. https://doi.org/10.1044/2018_JSLHR-L18-0253
- McKean, C., Letts, C., & Howard, D. (2013). Developmental change is key to understanding primary language impairment: The case of phonotactic probability and nonword repetition. *Journal of Speech, Language, and Hearing Research*, 56(5), 1579–1594. [https://doi.org/10.1044/1092-4388\(2013\)12-0066](https://doi.org/10.1044/1092-4388(2013)12-0066)
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, 6(7), Article e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Montgomery, J. (1995). Examination of phonological working memory in specifically language impaired children. *Applied Psycholinguistics*, 16(4), 355–378. <https://doi.org/10.1017/S0142716400065991>
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *The Journal of Child Psychology and Psychiatry*, 57(11), 1247–1257. <https://doi.org/10.1111/jcpp.12573>
- Oetting, J. B., & Cleveland, L. H. (2006). The clinical utility of nonword repetition for children living in the rural south of the US. *Clinical Linguistics & Phonetics*, 20(7–8), 553–561. <https://doi.org/10.1080/02699200500266455>
- Oetting, J. B., Cleveland, L. H., & Cope, R. F. (2008). Empirically derived combinations of tools and clinical cutoffs: An illustrative case with a sample of culturally/linguistically diverse children. *Language, Speech, and Hearing Services in Schools*, 39(1), 44–53. [https://doi.org/10.1044/0161-1461\(2008\)005](https://doi.org/10.1044/0161-1461(2008)005)
- Paradis, J. (2019). English second language acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors. In M. Brown & B. Dailey (Eds.), *Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 11–26). Cascadia Press. <http://www.lingref.com/bucld/43/BUCLD43-02.pdf>
- Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development & disorders: A handbook on bilingualism and second language learning* (2nd ed.). Brookes.
- Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, 56(3), 971–981. [https://doi.org/10.1044/1092-4388\(2012\)12-0050](https://doi.org/10.1044/1092-4388(2012)12-0050)
- Pearson, B. Z. (2013). Distinguishing the bilingual as a late talker from the late talker who is bilingual. In L. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 67–91). Brookes.
- Petrucelli, N., Bavin, E. L., & Bretherton, L. (2012). Children with specific language impairment and resolved late talkers: Working memory profiles at 5 years. *Journal of Speech, Language, and Hearing Research*, 55(6), 1690–1703. [https://doi.org/10.1044/1092-4388\(2012\)11-0288](https://doi.org/10.1044/1092-4388(2012)11-0288)
- Pham, G., & Ebert, K. D. (2020). Diagnostic accuracy of sentence repetition and nonword repetition for developmental language disorder in Vietnamese. *Journal of Speech, Language, and Hearing Research*, 63(5), 1521–1536. https://doi.org/10.1044/2020_JSLHR-19-00366
- Pham, G., Ebert, K. D., Dinh, K. T., & Dam, Q. (2018). Nonword repetition stimuli for Vietnamese-speaking children. *Behavior Research Methods*, 50(1), 1311–1326. <https://doi.org/10.3758/s13428-018-1049-0>
- Plante, E., & Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools*, 25(1), 15–24. <https://doi.org/10.1044/0161-1461.2501.15>
- Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research*, 55(3), 683–694. [https://doi.org/10.1044/1092-4388\(2011\)10-0263](https://doi.org/10.1044/1092-4388(2011)10-0263)
- Rodekohr, R. K., & Haynes, W. O. (2001). Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders*, 34(3), 255–272. [https://doi.org/10.1016/S0021-9924\(01\)00050-8](https://doi.org/10.1016/S0021-9924(01)00050-8)
- Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, 47(1), 223–234. [https://doi.org/10.1044/1092-4388\(2004\)019](https://doi.org/10.1044/1092-4388(2004)019)
- Stokes, S. F., & Klee, T. (2009). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech, Language, and Hearing Research*, 52(4), 872–882. [https://doi.org/10.1044/1092-4388\(2009\)08-0030](https://doi.org/10.1044/1092-4388(2009)08-0030)
- Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research*, 49(2), 219–236. [https://doi.org/10.1044/1092-4388\(2006\)019](https://doi.org/10.1044/1092-4388(2006)019)
- Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, 179, 23–36. <https://doi.org/10.1016/j.cognition.2018.06.002>
- Thal, D. J., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language, and Hearing Research*, 48(6), 1481–1495. [https://doi.org/10.1044/1092-4388\(2005\)103](https://doi.org/10.1044/1092-4388(2005)103)
- Thordardottir, E. (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, K. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 331–359). Multilingual Matters. <https://doi.org/10.21832/9781783093137-014>
- 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>
- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at

- age 5. *Journal of Speech, Language, and Hearing Research*, 54(2), 580–597. [https://doi.org/10.1044/1092-4388\(2010/09-0196\)](https://doi.org/10.1044/1092-4388(2010/09-0196))
- Topbaş, S., Kaçar-Kütükcü, D., & Kopkalli-Yavuz, H.** (2014). Performance of children on the Turkish nonword repetition test: Effect of word similarity, word length, and scoring. *Clinical Linguistics & Phonetics*, 28(7–8), 602–616. <https://doi.org/10.3109/02699206.2014.927003>
- Tuller, L., Abboud, L., Ferré, S., Fleckstein, A., Prérévost, P., dos Santos, C., Scheidnes, M., & Zebib, R.** (2013). Specific language impairment and bilingualism: Assembling the pieces. In C. Hamann & E. Ruigendijk (Eds.), *Language acquisition and development: Proceedings of GALA*. Cambridge Scholars Publishing. https://www.unige.ch/fapse/logopedie/files/8314/2245/9669/12_article_2_P_Prevost.pdf
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., dos Santos, C., Abed Ibrahim, L., & Zebib, R.** (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, 53(4), 888–904. <https://doi.org/10.1111/1460-6984.12397>
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A.** (1999). *Comprehensive test of phonological processing*. Pro-Ed.
- Washington, J. A., & Craig, H. K.** (2004). A language screening protocol for use with young African American children in urban settings. *American Journal of Speech-Language Pathology*, 13(4), 329–340. [https://doi.org/10.1044/1058-0360\(2004/033\)](https://doi.org/10.1044/1058-0360(2004/033))
- Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M.** (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 43(4), 865–878. <https://doi.org/10.1044/jslhr.4304.865>

Supplementary Material S1

Individual effect sizes

Study name	Statistics for each study							
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	P-Value	Weight
Weismer et al. (2000)	0.68	0.12	0.01	0.44	0.93	5.47	$p < .001$	2.12
Rodekohr & Haynes (2001)	1.19	0.46	0.21	0.27	2.11	2.55	$p = .01$	1.38
<i>White population</i>								
Rodekohr & Haynes (2001)	0.74	0.44	0.19	-0.12	1.61	1.66	$p = .09$	1.43
<i>African population</i>								
Conti-Ramsden (2003)	2.04	0.30	0.09	1.44	2.64	6.68	$p < .001$	1.76
Conti-Ramsden & Hesketh (2003)	1.22	0.27	0.07	0.69	1.75	4.53	$p < .001$	1.83
Gray (2003)	3.68	0.49	0.24	2.71	4.64	7.48	$p < .001$	1.33
<i>Time 1</i>								
Gray (2003)	2.80	0.42	0.17	1.97	3.62	6.66	$p < .001$	1.49
<i>Time 2</i>								
Gray (2003)	2.55	0.40	0.16	1.76	3.34	6.34	$p < .001$	1.53

Study name	Statistics for each study							
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	P-Value	Weight
<i>Time 3</i>								
Gray (2004)	1.99	0.38	0.14	1.24	2.73	5.21	$p < .001$	1.58
Horovov & Oetting (2004)	2.15	0.35	0.12	1.46	2.84	6.12	$p < .001$	1.65
Bortolini et al. (2006)	1.26	0.45	0.20	0.38	2.15	2.80	$p < .001$	1.41
Gray (2006)	1.37	0.39	0.15	0.59	2.15	3.46	$p < .001$	1.54
<i>3 years</i>								
Gray (2006)	0.89	0.47	0.22	-0.02	1.82	1.90	$p = .05$	1.37
<i>4 years</i>								
Gray (2006)	0.95	0.38	0.15	0.18	1.71	2.44	$p = .01$	1.56
<i>5 years</i>								
Gray (2006)	1.07	0.38	0.14	0.32	1.82	2.82	$p < .001$	1.58
<i>6 years</i>								
Oetting & Cleveland (2006)	1.73	0.30	0.09	1.13	2.33	5.65	$p < .001$	1.76
Oetting et al. (2008)	1.97	0.43	0.18	1.13	2.82	4.58	$p < .001$	1.46

Study name	Statistics for each study							
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	P-Value	Weight
<i>White population</i>								
Oetting et al. (2008)	1.17	0.46	0.21	0.26	2.07	2.53	$p = .01$	1.40
<i>African American population</i>								
Jones et al. (2010)	1.72	0.38	0.14	0.97	2.47	4.48	$p < .001$	1.57
<i>Low lexicality nonwords</i>								
Jones & al. (2010)	1.53	0.37	0.13	0.80	2.26	4.11	$p < .001$	1.60
<i>High lexicality nonwords</i>								
Deevy et al. (2010)	2.27	0.29	0.08	1.69	2.86	7.64	$p < .001$	1.78
Gutiérrez -Clellen & Simon-Cereijido (2010)	1.38	0.19	0.03	1.00	1.76	7.16	$p < .001$	2.01
<i>English task</i>								
Gutiérrez -Clellen & Simon-Cereijido (2010)	1.16	0.18	0.03	0.79	1.53	6.19	$p < .001$	2.02
<i>Spanish task</i>								
Thordardottir et al. (2011)	3.77	0.40	0.16	2.98	4.55	9.42	$p < .001$	1.54
Petrucelli et al. (2012)	0.91	0.24	0.05	0.43	1.38	3.74	$p < .001$	1.90

Study name	Statistics for each study							
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	P-Value	Weight
Kapalková et al. (2013)	2.47	0.46	0.21	1.56	3.37	5.34	$p < .001$	1.39
McKean et al. (2013)	1.82	0.36	0.13	1.10	2.54	4.98	$p < .001$	1.62
<i>High phonological probability</i>								
McKean et al. (2013)	1.43	0.34	0.12	0.75	2.12	4.12	$p < .001$	1.66
<i>Low phonological probability</i>								
Guiberson & Rodríguez (2013)	0.52	0.52	0.27	-0.50	1.56	0.99	$p = .31$	1.25
<i>3 years</i>								
Guiberson & Rodríguez (2013)	0.546	0.48	0.23	-0.39	1.49	1.13	$p = .25$	1.35
<i>4 years</i>								
Guiberson & Rodríguez (2013)	1.33	0.55	0.30	0.25	2.42	2.41	$p = .01$	1.20
<i>5 years</i>								
Dispaldro et al. (2013)	3.33	0.52	0.27	2.30	4.36	6.35	$p < .001$	1.26
Paradis et al. (2013)	1.06	0.21	0.04	0.63	1.49	4.87	$p < .001$	1.95
Stokes et al. (2013)	0.69	0.32	0.10	0.06	1.33	2.12	$p = .03$	1.71

Study name	Statistics for each study							
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	P-Value	Weight
Thordardottir & Brandeker (2013)	0.48	0.26	0.07	- 0.04	1.00	1.80	$p = .07$	1.85
Guiberson & Rodríguez (2015) <i>Younger children</i>	0.67	0.38	0.14	- 0.07	1.42	1.76	$p = .07$	1.58
Guiberson & Rodríguez (2015) <i>Older children</i>	0.96	0.38	0.15	0.20	1.72	2.48	$p = .01$	1.56
Boerma et al. (2015) <i>Monolinguals/ Dutch task</i>	2.09	0.31	0.10	1.46	2.71	6.56	$p < .001$	1.73
Boerma et al. (2015) <i>Monolinguals / Quasi-Universal task</i>	2.47	0.34	0.11	1.80	3.13	7.26	$p < .001$	1.68
Boerma et al. (2015) <i>Bilinguals/ Dutch task</i>	1.76	0.30	0.09	1.17	2.35	5.85	$p < .001$	1.77
Boerma et al. (2015) <i>Bilinguals / Quasi-Universal task</i>	1.18	0.27	0.07	0.64	1.73	4.28	$p < .001$	1.83
Ferré & Dos Santos (2015)	2.41	0.53	0.28	1.37	3.45	4.54	$p < .001$	1.25

Study name	Statistics for each study							
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	P-Value	Weight
<i>Monolinguals/language dependant</i>								
Ferré & Dos Santos (2015)	1.80	0.47	0.22	0.87	2.74	3.78	$p < .001$	1.36
<i>Monolinguals/language independant</i>								
Ferré & Dos Santos (2015)	1.06	0.37	0.13	0.34	1.79	2.87	$p < .001$	1.60
<i>Bilinguals/language dependant</i>								
Ferré & Dos Santos (2015)	0.85	0.33	0.11	0.19	1.52	2.52	$p = .01$	1.68
<i>Bilinguals/ language independant</i>								
Armon-Lotem & Meir (2016)	2.16	0.28	0.07	1.60	2.71	7.68	$p < .001$	1.82
<i>Monolinguals/ Russian task</i>								
Armon-Lotem & Meir (2016)	1.23	0.24	0.06	0.75	1.72	5.02	$p < .001$	1.89
<i>Monolinguals/ Hebrew task</i>								
Armon-Lotem & Meir (2016)	0.96	0.17	0.03	0.62	1.30	5.52	$p < .001$	2.04
<i>Bilinguals / Russian task</i>								
Armon-Lotem & Meir (2016)	1.76	0.19	0.03	1.38	2.14	9.05	$p < .001$	2.00

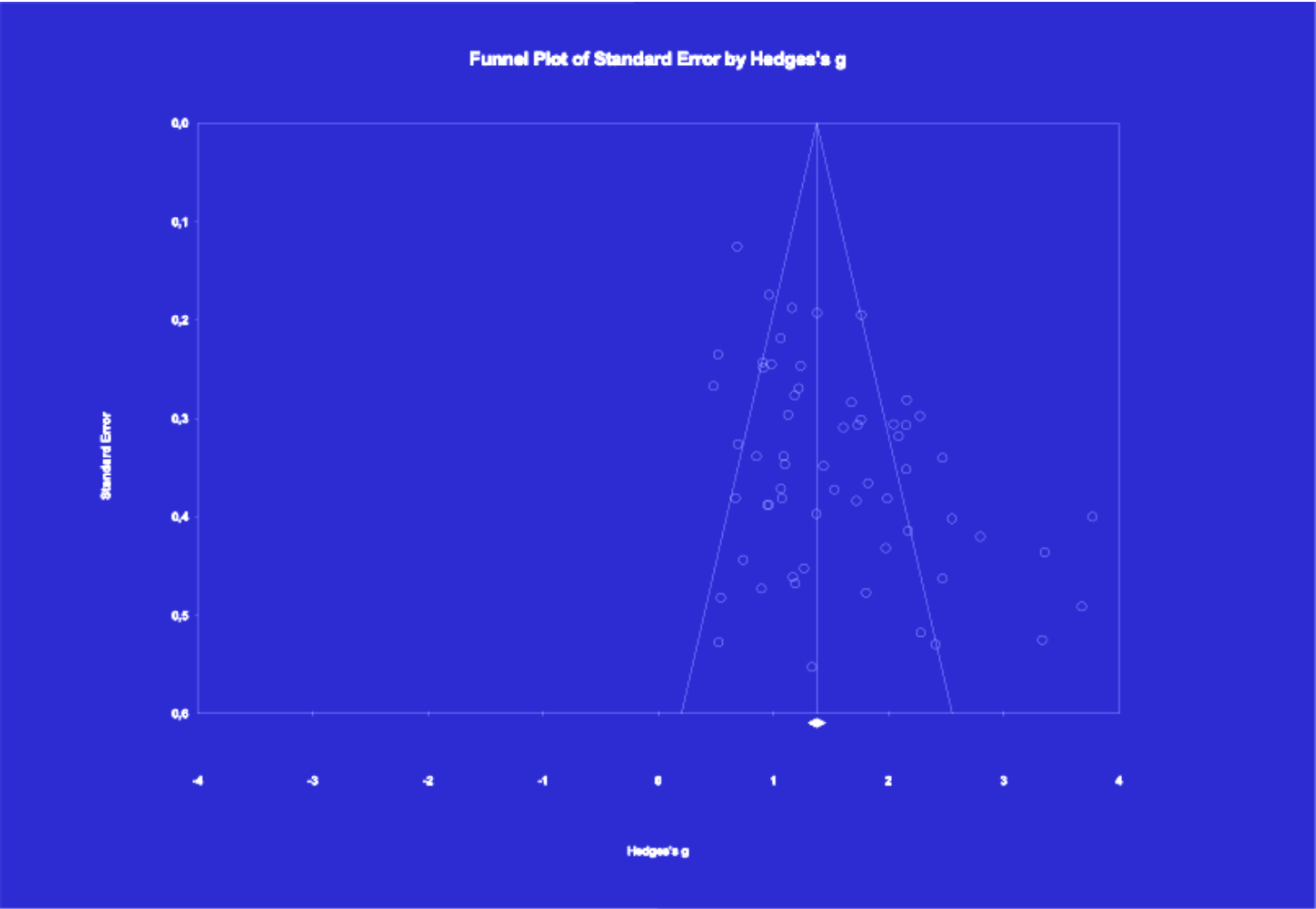
Study name	Statistics for each study							
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	P-Value	Weight
<i>Bilinguals/ Hebrew task</i>								
Le Clercq et al. (2017)	0.98	0.24	0.06	0.50	1.46	4.02	$p < .001$	1.90
<i>Normal version</i>								
Le Clercq et al. (2017)	0.52	0.23	0.05	0.06	0.98	2.21	$p = .02$	1.92
<i>Short version</i>								
Boerma & Blom (2017)	2.15	0.30	0.09	1.55	2.75	7.01	$p < .001$	1.76
<i>Monolinguals</i>								
Boerma & Blom (2017)	1.68	0.28	0.08	1.12	2.23	5.92	$p < .001$	1.81
<i>Bilinguals</i>								
Kazemi & Saeednia (2017)	3.35	0.43	0.19	2.50	4.21	7.69	$p < .001$	1.45
dos Santos & Ferré (2018)	2.28	0.51	0.26	1.26	3.29	4.40	$p < .001$	1.27
<i>Monolinguals</i>								
dos Santos & Ferré (2018)	1.10	0.34	0.12	0.42	1.78	3.17	$p < .001$	1.66
<i>Bilinguals</i>								

Study name	Statistics for each study							
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	P-Value	Weight
Li'el & al. (2019)	1.61	0.30	0.09	1.00	2.21	5.20	$p < .001$	1.75
McDonald & Oetting (2019) <i>Southern White English</i>	2.17	0.41	0.17	1.36	2.98	5.24	$p < .001$	1.50
McDonald & Oetting (2019) <i>African American English</i>	0.91	0.24	0.06	0.42	1.40	3.68	$p < .001$	1.89
Bonifacci et al. (2020)	1.13	0.29	0.08	0.54	1.71	3.81	$p < .001$	1.78
Pham & Ebert (2020)	1.09	0.33	0.11	0.42	1.75	3.22	$p < .001$	1.68
Radom effects model	1.54	0.09	0.00	1.37	1.72	17.22	$p < .001$	

Note.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Supplementary Material 2



Supplementary Material S3

Discriminative values (sensitivity and specificity)

Study	DLD criteria	Threshold	Sensitivity	Specificity
Conti-Ramsden (2003)	DLD recruited via SLTs or specialist	<i>p</i> .25	66 %	100 %
	teachers in language. Evaluation	<i>p</i> .16	59 %	100 %
	through a standardized language test.			
Conti-Ramsden &	identical to Conti-Ramsden (2003)	<i>p</i> .25	66 %	85 %
Hesketh (2003)		<i>p</i> .16	59 %	85 %
Gray (2003)	DLD recruited via SLTs, scoring more		Time 1 : 95 %	Time 1 : 100 %
	than -1.5 <i>SD</i> on two standardized	na	Time 2 : 91 %	Time 2 : 95 %
	language tests.		Time 3 : 95 %	Time 3 : 86 %
Thal et al. (2005)	Late talkers identified at 16 months as		40 %	86 %
	having limited vocabulary (below the	na		
	10 percentile) based on parental			
	reports.			

Study	DLD criteria	Threshold	Sensitivity	Specificity
Bortolini et al., (2006)	DLD diagnosed and followed by a SLT, scoring at or below -1.5 <i>SD</i> on more than one subtest, with and MLU between 2.1 and 4.8.	na	82 %	82 %
Oetting & Cleveland (2006)	Children referred to SLT services or identified by the teacher as having	na	56 %	92 %
Oetting et al. (2008)	low language skills, scoring below -1 <i>SD</i> in two standardized tests.	70 %	56 %	92 %
Archibald & Joanisse (2009)	Children scoring below the 15 percentile on the NWR and SR tasks, as well as below -1 <i>SD</i> on a standardized test.	<i>p.15</i>	40-46 %	60-63 %
Deevy et al. (2010)	DLD recruited via SLTs, scoring	> 66%	PPC method 1 : 86 %	PPC method 1 : 91 %
	below -3.25 <i>SD</i> on a standardized test.	> 68%	PPC method 2 : 79 %	PPC method 2 : 89 %

Study	DLD criteria	Threshold	Sensitivity	Specificity
Gutiérrez-Clellen & Simon-Cereijido (2010)	Assessment protocol included evidence of clinical concern, 20 % of ungrammatical utterances and low performance on monolingual and bilingual standardized tests (< 60 % correct responses).	70 %	English NWRT : 55 % Spanish NWRT : 61 %	English NWRT : 82 % Spanish NWRT : 82 %
Thorardorttir et al. (2011)	DLD recruited by SLTs who provided a diagnosis and data from a parental report.	-1 <i>SD</i>	85 %	86 %
		-1.28 <i>SD</i>	85 %	88 %
		-2 <i>SD</i>	77 %	93 %
		82 % /- 0.88 <i>SD</i>	92 %	86 %
Dispaldro et al. (2013)	DLD diagnosed by SLT, score below the norm on standardized tests, MLU analysis, score below -2 <i>SD</i> on third person object clitic pronoun use task	65 %	PIC : 100 % PPC : 94 %	PIC : 100 % PIC : 94 %
		93%		

Study	DLD criteria	Threshold	Sensitivity	Specificity
Guiberson & Rodríguez (2013)	Identification of the DLD by a bilingual SLT, parental concerns about language development, language expression scores more than - 1.5 <i>SD</i> on a standardized test.	na	PIC : 71% PPC : 48 %	PIC : 74 % PPC : 78 %
Kapalková et al. (2013)	Children with a DLD had a documented history of language impairment since the age of 2–3 years, with a diagnosis based both on clinical judgement and a qualitative assessment of children's language (no standardized tests available in Slovak).	8 items correct	PPC : 94 %	PPC : 100 %
Stokes et al., (2013)	Diagnosis of DLD by SLT and scores -1.2 <i>SD</i> on a standardized test.	na	77 %	97 %

Study	DLD criteria	Threshold	Sensitivity	Specificity
Thordardottir & Brandeker (2013)	DLD children diagnosed by SLT, scores below the mean on a French assessment, clinical history.	82 %	monolinguals : 92 % bilinguals : 85 %	monolinguals : 100 % bilinguals : 79 %
Tuller et al. (2013)	Children with a DLD followed by an SLT, assessment in both languages and use of deviation criteria by Thordardottir (2015).	60 %	90 %	89,5 %
Topbaş et al. (2014)	DLD diagnosed by SLTs and scores on a standardized test more than -1 <i>SD</i> .	-2 <i>SD</i>	89 %	87 %
Boerma et al. (2015)	DLD diagnosed by SLT, global score of – 2 <i>SD</i> or lower on a Dutch assessment battery with score of -1.5	77.7	monolinguals : quasi-universal LITMUS: 83 %	monolinguals : quasi-universal LITMUS: 90 %
	<i>SD</i> or lower on at least two of the four subtests.	72.7	Dutch LITMUS: 93 % bilinguals:	Dutch LITMUS: 93 % bilinguals:
		78.1		

Study	DLD criteria	Threshold	Sensitivity	Specificity
			quasi-universal LITMUS:	93 % (for both LITMUS
		63.8	83 %,	tasks)
			Dutch LITMUS: 63 %	
		78.1	RNM	monolinguals+bilinguals:
			monolinguals+bilinguals:	quasi-universal
		72.7	quasi-universal LITMUS:	LITMUS: 92 %
			83 %	Dutch LITMUS: 77 %
			Dutch LITMUS: 87 %	
			monolinguals: 90 %	monolinguals: 92 %
			bilinguals:	bilinguals:
Ferré & Dos Santos (2015)	DLD children recruited and followed	-2 SD		
	by SLTs, parental report to determine			
	risk factors, standardized assessment	-1 SD	73 %	82 %
	in the child's mother tongue and in	-1,28 SD	68 %	89 %
	French. with deviation criteria by	-1,65 SD	59 %	91 %
	Thordardottir (2015).	-2 SD	59 %	95 %

Study	DLD criteria	Threshold	Sensitivity	Specificity
Guiberson & Rodríguez (2015)	SLT diagnosed language delay, parental concern about language development, score of -1.5 <i>SD</i> or more on a task from a Spanish standardized test.	na	80 %	62 %
Armon-Lotem & Meir (2016)	Monolingual children with a DLD: diagnosis by SLT, parental report (e.g., parental concern), scores – 1.25 <i>SD</i> or more based on monolingual norms on Russian or Hebrew tests.	.71	monolinguals: Russian NWRT: 86 % Hebrew NWRT: 93 %	monolinguals: Russian NWRT: 90 % Hebrew NWRT: 66 %
	Bilingual children with a DLD: diagnosed by SLT or history of disorder reported by parents/teachers, scores – 1.25 <i>SD</i> or more based on	.71 .86	bilingual children with monolingual norms: Russian NWRT: 44 % Hebrew NWRT: 96 %	bilingual children with monolingual norms: Russian NWRT: 86 % Hebrew NWRT: 48 %
			bilingual children with monolingual norms: Russian NWRT: 70 %	bilingual children with monolingual norms: Russian NWRT: 76 %

Study	DLD criteria	Threshold	Sensitivity	Specificity
	bilingual norms on Russian and Hebrew tests.	.79 .71	Hebrew NWRT: 81 % Both combined: 65 %	Hebrew NWRT: 89 % Both combined: 94 %
De Almeida et al. (2016)	DLD children were followed by SLTs, standardized tests in each language, with deviation criteria by Thordardottir (2015).	80 %	monolinguals: 88 % bilinguals: 81%	monolinguals: 83 % bilinguals: 79 %
De Almeida et al. (2017)	identical to De Almeida & al. (2016)	80 %	monolinguals: 88 % bilinguals: 81%	monolinguals: 83 % bilinguals: 79 %
Hodges et al. (2017)	Below 15th percentile on a parental vocabulary report, within normal limits for receptive language (> 16th percentile on a standardized test).	60 % 51%	MITT : 92 % TENR : 90 %	MITT : 77 % TENR : 87 %

Study	DLD criteria	Threshold	Sensitivity	Specificity
Kazemi & Saeednia (2017)	DLD children recruited via SLTs, clinical judgment by two SLTs (no standardized test in Iran).	34 : PSC	95 %	98 %
		12,5 : PIC	90 %	96 %
Le Clercq et al. (2017)	DLD children recruited from special education schools, scoring below -1.5 <i>SD</i> in at least two language domains, or at least -2 <i>SD</i> below the general mean measured by Dutch language standardized tests.	11.4 % 78.6 %	short task: 83 %	short task: 95 %
			complete task: 87 %	complete task: 88 %
Dos Santos & Ferré (2018)	DLD children recruited and followed by SLTs, parental report to determine risk factors, standardized assessment in the child's mother tongue and in French with deviation criteria by Thordardottir (2015).	-2 <i>SD</i>	monolinguals: 90%	monolinguals: 92 %
			bilinguals:	bilinguals:
		-1 <i>SD</i>	73 %	82 %
		-1,28 <i>SD</i>	68 %	89 %
		-1,65 <i>SD</i>	59 %	91 %
		-2 <i>SD</i>	59 %	95 %

Study	DLD criteria	Threshold	Sensitivity	Specificity
Li'el et al. (2018)	DLD children diagnosed and followed by SLTs or on the waiting list. Parental questionnaire (ALDeQ, Paradis & al., 2010) used to assess risk factors.	≤ 7 correct items	89 %	81 %
Tuller et al. (2018)	DLD are followed by SLTs. Assessments in both languages: scores below the mean in two language domains of the dominant language, using the deviation criteria by Thordardottir (2015).	77.5 59.9	French LITMUS: 88 % German LITMUS: 92 %	French LITMUS: 92 % German LITMUS: 90 %
McDonald & Oetting (2019)	DLD children scored below -1 <i>SD</i> on an English syntax and lexical subtest.	.65 .76	53 % 77 %	98 % 74 %
Pham & Ebert (2020)	DLD children scored below – 1 <i>SD</i> on at least one of the parent or teacher concern measures and on at least three	0.20 0.56 0.85	PIC : 40 % PSC : 80 % PPC : 90 %	95 % 79 % 79 %

Study	DLD criteria	Threshold	Sensitivity	Specificity
	of the six direct language measures, and in at least two of the three language domains that were measured (vocabulary, grammar, and narratives).			
Bonifacci et al. (2020)	DLD children were diagnosed by SLTs.	na	50 %	91 %

Note. $n = 33$, listed by order of publication date. Abbreviations: PIC = percentage items correct, PPC = percentage phonemes correct, PSC = percentage syllables correct; SLT = speech and language therapist; DLD: children with a development language disorder; SD = standard deviation; MLU = mean length utterance; SR = sentence repetition.

For the inclusion criteria for DLD children, performance is indicated with an arrhythmia sign (-) indicating below-average scores.

Supplementary Materials S4

Correlations between NWRT and Other Measures

Study	Other measures	Correlation
Guiberson & Rodríguez (2015)	Receptive and productive vocabulary	$r = 0.52, p < .001$
Horovov et Oetting (2004)	Receptive and productive vocabulary	$p < .001$
	Sentence/ syntax comprehension	$p < .001$
Marini et al. (2017)	Articulation	$r = .470, p < .001$
	Productive vocabulary	$r = .328, p < .001$
	Semantic fluency	$r = .338, p < .001$
	Receptive vocabulary	$r = .268, p < .001$
	Language developpment survey	$r = .425, p < .001$
	Family history of difficulties	$r = -.199, p < .001$
	Literacy questionnaire	$r = .222, p < .001$
Chiat et Roy (2007)	Receptive vocabulary	$p < .001$
Stokes et al. (2013)	Sentence repetition	$rs = .32, .37, .25, p > .05$
	Receptive and productive vocabulary	ns
	Nonverbal IQ	ns
	French sentence repetition	$r = 0.624, p < .001$

Study	Other measures	Correlation
Thorardottir & Brandeker (2013)	English sentence repetition	$r = 0.555, , p < .001$
	English receptive vocabulary	English NWRT: $r = 0.575, p < .001$
		French NWRT: $r = 0.200, p = .177$
	French receptive vocabulary	French NWRT: $r = 0.483, p = .001$
De Almeida et al. (2017)		English NWRT: $r = 0.032, p = .833$
	Sentence repetition	$p < .0001$
	Word repetition	DT: $r_s = .249, p = .053$ DLD: $r_s = .574, p = .008$
	No-risk index	$r_s = .466, p < .0001$
Paradis et al. (2013)	Syntax	TD: $r = .242, p < .01$ DLD: $r = .412, p > .05$
	Narration	TD: $r = .239, p < .01$ DLD: $r = .133, p > .05$
	Receptive vocabulary	TD: $r = .269, p < .01$ DLD: $r = .291, p > .05$

Study	Other measures	Correlation
Dispaldro et al. (2013)	Word repetition	TD: $r = 0.29, p = .262$ DLD: $r = 0.83, p < .001$
Petrucelli et al. (2012)	Number recall/repetition	ns
Gray (2003) ^a	Number recall/repetition	TD: $r = 0.66, p < .001$ DLD: $r = 0.54, p < .01$
	Nonverbal task of cognitive abilities	TD: $r = 0.26, p > .05$ DLD: $r = 0.09, p > .05$
	Morphosyntax et syntax	TD: $r = -0.57, p < .001$ DLD: $r = 0.49, p < .01$
	Phonology	TD: $r = 0.63, p < .001$ DLD: $r = -0.05, p > .05$
	Receptive vocabulary	TD: $r = 0.38, p < .01$ DLD: $r = 0.45, p > .05$
Oetting et Cleveland (2006)	Nonverbal reasoning	$r = .56, p < .01$
	Articulation task	$r = .60, p < .01$
Oetting et al. (2008)	Nonverbal reasoning	$r = .58, p < .05$
Thordardottir et al. (2011)	Receptive vocabulary	ns
	Receptive test (vocabulary, syntax, morphosyntax)	correlated

Study	Other measures	Correlation
	Sentence repetition	correlated
Boerma et al. (2015)	Nonverbal IQ	ns
De Almeida et al. (2017)	Sentence repetition	$p < .0001$
	Word repetition	TD: $r_s = .249, p = .053$ TDL: $r_s = .574, p = .008$
	No-risk index	$r_s = .466, p < .0001$
Tuller et al. (2018)	Positive early development	German NWRT: $r = .493, p < .001$ French NWRT: $r = .512, p < .001$
	Early exposure to L2	German NWRT: $r = -.398, p = .004$ French NWRT: $r = .086, p = .424$
	Current L2 richness	German NWRT: $r = .189, p = .189$ French NWRT: $r = .030, p = .783$

Note. $n = 17$.

^a These tasks were performed more than once, this table only shows the data for the best correlation results. For more details, please refer to the original article.

^b The statistical details were not further specified.

* $p < .05$ ** $p < .01$ *** $p < .001$

Supplementary Material S5

Combination of NWRT and other measures

Study	Threshold	Sensitivity (NWRT alone)	Specificity (NWRT alone)	Other tasks	Threshold	Sensitivity (combined)	Specificity (combined)
Thal et al. (2005)	na	40 %	86 %	Productive vocabulary	na	65 %	86 %
Bortolini et al. (2006)	na	82 %	82 %	(a) Plural 3rd person	na	82 %	100 %
				(b) Object clitics		91 %	100 %
				Both a+b		91 %	91 %
Oetting & Cleveland (2006)	na	56 %	92 %	Nonverbal reasoning	na	81 %	94 %
Oetting et al. (2008)	70 %	56 %	92 %	Nonverbal reasoning	70 % 91 %	81 %	94 %
Thordardottir et al. (2011)	82 %	92 %	86 %	Receptive vocabulary	. 100	98 %	/
Paradis et al. (2013)	/	/	/	Model 2 ^a	na	91 %	92 %

Study	Threshold	Sensitivity (NWRT alone)	Specificity (NWRT alone)	Other tasks	Threshold	Sensitivity (combined)	Specificity (combined)
Armon-Lotem & Meir (2016)	. 71	Russian monolingual	Russian monolingual			Russian monolingual	Russian monolingual
		86 %	90 %	Number recall	. 4	57 %	90 %
				Sentence and number recall	. 4 . 88	50 %	100 %
	. 71	Hebrew monolingual	Hebrew monolingual			Hebrew monolingual	Hebrew monolingual
		93 %	66 %	Number recall	. 4	92 %	86 %
				Sentence and number recall	. 4 . 88	92 %	97 %
	. 71	bilinguals ^c	bilinguals				

Study	Threshold	Sensitivity (NWRT alone)	Specificity (NWRT alone)	Other tasks	Threshold	Sensitivity (combined)	Specificity (combined)
		70 % Russian	76 % Russian	Number recall	. 4	61 %	92 %
		NWRT	NWRT	Sentence and number recall	. 4	59 %	95 %
					. 88		
		81 % Hebrew	89 % Hebrew	Number recall			
		NWRT	NWRT	Sentence and number recall	. 4	67 %	85 %
					. 4	67 %	95 %
					. 88		
Boerma & Blom (2017)	/	/	/		na	monolingual	monolingual
				Development index		94 %	93 %
				Development index and narration		94 %	93 %
						bilingual	bilingual
				Development index		90 %	87 %

Study	Threshold	Sensitivity (NWRT alone)	Specificity (NWRT alone)	Other tasks	Threshold	Sensitivity (combined)	Specificity (combined)
				Development index and narration		97 %	97 %
Bonifacci et al. / (2020)	/	/	/	Model 3A ^b		91.4 %	100 %

Note. $n = 9$. Abbreviation: NWRT = non-word repetition task

^aThe study includes different model variants with different percentages of sensitivity and specificity. We have mentioned only one of these models (among the best). Model 2 includes a parental questionnaire, a NWRT, a morphosyntactic production task, a narrative task.

^bThe study includes different model variants with different percentages of sensitivity and specificity. We have mentioned only one of these models (among the best). Model 3A includes: a receptive lexicon task, a parental questionnaire, a morphosyntactic comprehension task, a NWRT and a morphosyntactic production task.

^cas categorized with bilingual norms.

6. Note de recherche

**Detecting Developmental Language Disorder in Monolingual and Bilingual Children:
Comparison of Language-Specific and Quasi-Universal Non-Word Repetition Tasks in
French and Portuguese**

Schwob, S. et Skoruppa, K

État: accepté pour publication le 26.10.2021 par la revue “*Journal of Speech, Language,
and Hearing Research*”

Abstract

Purpose: Over the last decades, many studies have documented the clinical potential of non-word repetition tasks for detecting developmental language disorder in mono- and bilingual children, by unveiling their difficulties in short-term memory and phonological accuracy. However, the precise nature of the non-words to be used and the best scoring methods remain under debate. Some authors (e.g., Gutiérrez-Clellen & Simon-Cereijido 2010) support the use of "language-specific" non-words designed for a given test language in standardized tests. Other authors (e.g., Chiat, 2015) advocate the use of “crosslinguistic” stimuli, thus allowing assessment independently of the languages spoken by the child.

Method: This article compares two language-specific tasks (French vs. Portuguese) and a crosslinguistic non-word repetition task (Chiat, 2015) in a population of 5- to 7-year-old monolingual and bilingual children. Group comparisons (children with vs. without developmental language disorder; monolingual vs. bilingual children), an error analysis, sensitivity and specificity calculations (assessed according to the recommendations of Plante & Vance 1994) and Youden's indices (1950) are reported.

Results: All three tasks significantly differentiate children with and without developmental language disorder, but did not show an effect for bilingualism, which is encouraging for the bilingual assessment (Antoničević-Elliott et al., 2019). As expected, an influence of children's age and length and complexity of the stimuli was also found. The language-specific French task was found to be the most sensitive (max. 88 %) and specific (max. 92 %), the crosslinguistic task also reached good accuracy percentages for the bilingual group (max. 82 % sensitivity and 84 % specificity).

Conclusions: This article adds to the evidence that non-word repetition tasks are promising tools for the identification of monolingual and bilingual children with developmental language disorder.

Introduction

Language assessment of bilingual children is a major challenge in speech and language therapy for a number of reasons: a) there are few or no standardized tests in many languages (Williams & McLeod, 2012); b) speech and language therapists often do not speak one of the children's languages and c) linguistically trained interpreters are hard to find (Boerma & Blom, 2017); and, last but not least, d) most tests are not normed for bilingual children (de Almeida et al., 2017).

One solution is to use tasks that are not or only minimally affected by the child's prior knowledge or experience of a specific language (Coady & Evans, 2008). The task most often mentioned in the literature and included in many clinical tests is non-word repetition (NWR): Children have to repeat word-like syllable sequences that are pronounceable but do not mean anything in their language, e.g., /klipafu/. Failure on an NWR task is recognized as a marker of developmental language disorder (DLD) in monolingual and bilingual populations (Coady & Evans, 2008; Girbau & Schwartz, 2007). The presence and nature of errors in the NWR task are thought to reflect difficulties in short-term verbal memory and phonological accuracy (Dos Santos & Ferré, 2018). Non-words with high phonological complexity are particularly difficult for DLD children whether they are monolingual or bilingual (Dos Santos & Ferré, 2018; Estes et al., 2007). Non-word length also appears to be a criterion for difficulty but impacting both children with DLD and with a typical language development (TD) (e.g., Marini et al., 2017).

However, although NWR task scores are arguably little affected by a child's prior lexical and syntactic knowledge, they are still influenced by the phonological specificities of the test language. A child is more likely to repeat non-words correctly if they share phonological properties of his or her own language (Boerma et al., 2015). Using a NWR created according to the phonological structure of a specific language thus produces controversial results, which can, at least partly, be explained by methodological differences (Boerma & Blom 2017). Indeed,

several studies (for a meta-analysis, see Schwob et al., to appear) report that scores obtained by bilingual children in language specific NWR tasks depend on their exposure of the test language, thus diminishing their discriminative power for DLD. However, since some authors (e.g., de Almeida et al., 2017) still obtain high specificity and sensitivity values, the language-specific NWR task may be regarded as discriminative even for bilingual children.

In order to avoid biases related to language exposure and language properties, a “crosslinguistic” NWR (CL) task has been devised by Chiat and colleagues (Chiat, 2015; Chiat & Polisenská, 2016). This task includes non-words based on frequent syllable structures (consonant-vowel, CV, units such as /ba/) and frequent phonemes of the world’s languages (see Chiat, 2015), in order to avoid disadvantaging speakers of specific languages. Boerma et al. (2015), among others, evaluated the clinical value of this task in monolingual Dutch and bilingual children with Dutch as a second language aged 5-6 years and the results seem promising, since bilingualism did not affect performance on this task and provides percentages of 83 % sensitivity and 93 % of specificity for bilingual children. However, due to its quasi-universal design, this task taps mostly into phonological memory and cannot reveal difficulties with specific complex phonological structures (e.g., consonant clusters) that are typical for children with DLD.

In addition to the type of stimuli, other factors are known to influence performance on NWR tasks. The scoring method is often reported to be influential. Interestingly, the number of correct whole words can be more specific and sensitive than a more detailed count of the number of correct phonemes (Boerma et al., 2015). Age is often mentioned as an influencing factor: younger children have more difficulty repeating non-words (de Almeida et al., 2017). Here, we set out to compare the clinical potential of three different NWR tasks (two language-specific ones and one crosslinguistic) in French- and Portuguese-speaking children (a frequent language combination in Western Switzerland; Federal Statistical Office, 2020) aged 5;0 to

7;11 years. We will also compare different scoring methods (words vs. phonemes), perform an error analysis and analyze potential effects of further factors relating to the child (age) and the stimuli (syllable length, phonological complexity).

Method

Participants

Sixty-one children aged 5; 0 to 7; 11 living in the French-speaking part of Switzerland and with no history or suspicion of developmental disorders, socio-affective or hearing impairment participated. According to language exposure data collected via a parental report (based on Tuller, 2015), 31 children were exposed to French at least 85 % at time of the testing and thus classified as monolingual. The remaining 30 bilingual children had a minimum of 25 % exposure to Portuguese at time of the study (and no more than 15 % exposure to a third language). Among these bilingual children, 19 have French as their dominant language and 11 have Portuguese; 12 are successive bilingual and 18 are simultaneous bilingual. Thirty-one children had a confirmed DLD (17 bilingual and 14 monolingual): they had been diagnosed and followed by a speech and language therapist for at least three months to attest to the persistence of difficulties. The remaining thirty children (17 monolingual and 13 bilingual) were reported to have a TD, were never identified as having a DLD and had never received speech and language therapy.

All children successfully passed an audiometric screening (audibility thresholds ≤ 20 dB at 0.5, 1, 2 and 4 kHz) and a non-verbal intelligence screening using the matrices subtest of the Wide-Range Intelligence Test by Glutting et al. (2000). The parental questionnaire also asked the level of education of the parents (scales from 1 to 5: 1 = elementary school, 2 = high school, 3 = apprenticeship, 4 = high school graduation, 5 = high school). An additional eleven children participated but were excluded because they failed the audiometry ($n = 3$) or the non-verbal

intelligence ($n = 2$) screening or did not fit the exposure criteria ($n = 2$), or due to a technical problem at the time of testing ($n = 4$).

Children with and without DLD in monolingual (mon) and bilingual (bil) groups were matched individually on age ($H = 5.19, p = .158$; median age mon TD : 72 months; mon DLD : 71.5; bil TD : 87; bil DLD : 76), non-verbal IQ ($H = 3.71, p = .294$; median non-verbal IQ mon TD : 50; mon DLD : 52.5; bil TD : 70; bil DLD : 45) and gender ($X^2(3, n = 61) = 1.05, p = .789$; 37 girls and 24 boys), and these factors did not differ significantly between the four groups. However, we found a significant difference in parental education between the bilingual DLD (median: 2.5) and the monolingual TD children ($H = 13.53, p = .004$; median: 4.75), the other groups did not differ significantly in this respect (median mon DLD : 3; bil TD : 3).

Recruitment procedure

Participants were recruited through their schools and speech and language therapists. Families interested in participation received an information sheet, a parental report and an informed consent form, validated by the local ethics commission (Swiss ethics Vaud; project no.: 2017-01881, authorization granted on 07.02.2018). Most of the evaluation sessions took place in the family home or in school. As part of a larger PhD research, participants participated in three sessions of about one hour each, which also included other language tasks. The NWR tasks were always done at the beginning of the session and were administered by a native speaker of the test language who was familiar with language assessment; the first author for French and a bilingual Portuguese French student in a master's program in speech and language therapy.

Tasks

Two language-specific NWR tasks were chosen from standardized tests: Exalang 5-8 (Thibault & Helloin, 2010) for French and PALPA-P Portuguese (Castro et al., 2007) for Portuguese (only completed by the bilingual children). The French task consists of 16 two- or three-syllable items and the Portuguese task is composed of 24 one- to three-syllable non-words. The two tasks have a phonological complexity that can be described on a 3-gradient scale based on Dos Santos & Ferré (2018) and Mateus et al., (2005) (see the Supplementary Material 9 for details of the different tasks) : complexity 0 = simple structure like consonant-vowel, CV: /ladõ/, /fêṭ/); complexity 1 = intersyllabic cluster, beginning of non-word by vowel or consonant-semi vowel: /bidwano/, /ədɔdɐ/); and complexity 2 = intrasyllabic cluster: /drigõbir/, /trɛf/). The 16 crosslinguistic non-words were chosen from the lists provided by Chiat (2015), including items of two to five simple syllables (CV). The items for this latter task were pre-recorded by a foreign (Swedish) speaker with final stress, so as to avoid advantaging the monolingual language group by using French as a test language.

Scoring procedure

Two scoring methods were applied by counting different units, that is, a) the total number of incorrect whole words and b) the number of incorrect phonemes per word. For the latter, we used the procedure of Dollaghan and Campbell (1998). Thus, each phoneme (consonant or vowel) was scored as incorrect with respect to its target phoneme, taking into consideration its position in the word. For the error analysis, we based our analysis on Kapalková et al. (2013) and we categorized our errors according to different levels: phoneme level errors (e.g., consonant substitution : /zilu/ for /zibu/; vowel and consonant additions: /zibubu/ for /zibu/), syllable level errors (e.g., vowel omission : /zib/ for /zibu/; cluster reduction : /lover/ for /flovir/) and word level errors (e.g., regressive or progressive assimilation: /bibu/

for /zibu/; lexicalization: in French /caramel/ for /caranel/; creation, when more than 4 wrong phonemes: /nikalasa/ for /nugitala/; metathesis: /zubi/ for /zibu/).

Inter-rater reliability

A second judge categorized 10 % of our sample on the CL and French NWRT. Inter-judge agreement indices were then calculated using the statistical software SPSS (Version 27.0, IBM Corp., 2020). Near perfect agreement (according to Landis & Koch, 1977) in Cohen's kappa tests was found for whole word scoring (.98 for French NWRT and 1 for CL NWRT). Excellent inter-rate reliability (according to Koo & Li, 2016) in terms of intra-class correlation coefficient (ICC) was found for scoring in terms of number of incorrect phonemes (.99 for CL NWRT and 1 for French NWRT). For categorization in terms of error types (for both tasks), ICCs indices also indicate strong to near-perfect agreement (between .66 to 1) (Koo & Li, 2016).

Statistical analyses

In order to determine whether there were significant differences between the different groups of children, univariate and multivariate linear models and repeated measures ANOVAs (with the Greenhouse-Geisser correction) were conducted using SPSS version 27 (IBM Corp., 2020), after transformation into ranks according to the method proposed by Conover and Iman (1981) because our data did not follow a normal distribution. For small number of missing data on parental education of monolinguals children ($n = 10$), we performed multiple imputations (Donders et al., 2006) to generate these observations and control for them in our statistical model. Differences in terms of error categories were analyzed using Fisher exact tests. We also generated Roc curves (Dunn, 2014) for the different tasks and different scoring method, from which we derived sensitivity and specificity (which were interpreted according to Plante &

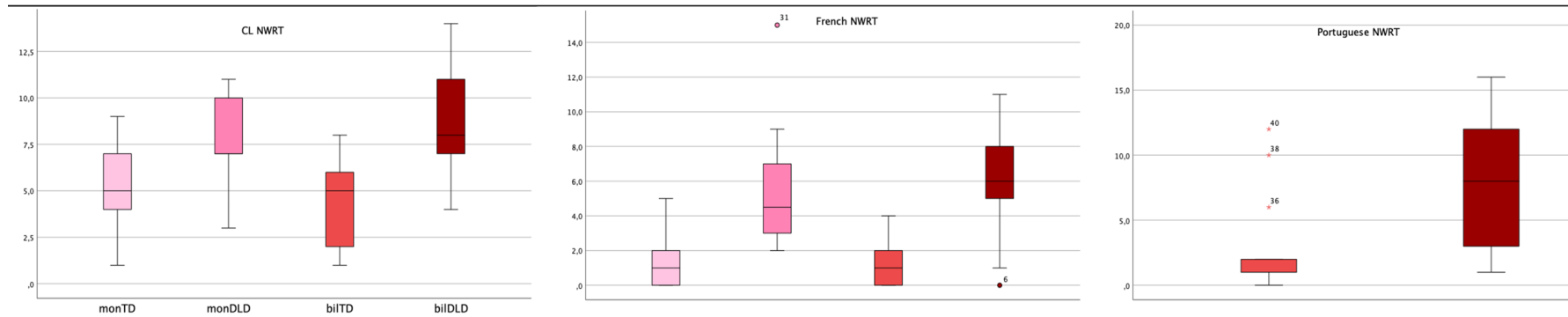
Vance, 1994.) Finally, Youden's indices (1950) were also calculated in order to determine the optimal threshold to distinguish between participants with and without DLD.

Results

Effect of disorder

Considering the number of incorrect words produced in the different groups (Figure 1), we note that children with DLD perform significantly worse than children without DLD in all three tasks when counting the number of incorrect words: French NWR: $F(1, 54) = 65.85, p < .001$, Portuguese NWR: $F(1, 25) = 15.5, p = .010$ and CL NWR: $F(1, 54) = 18.61, p < .001$. We find similar result taking into account the number of incorrect phonemes (see S1 in Supplementary Materials). We note that performance significantly increases with age on the three NWR tasks (CL NWR: $F(1, 54) = 5.78, p = .020$; French NWR: $F(1, 54) = 15.49, p < .001$; Portuguese NWR: $F(1, 25) = 18.61, p < .001$). Older children (TD or DLD) produce fewer errors (words or phonemes). Scatterplots displaying individual data can be found in the Supplementary Materials (see S2, S3 and S4).

Figure 1
Box plots showing the number of incorrect words in the three tasks



Note. Abbreviations: CL = Crosslinguistic task; TD = children with a typical language development; DLD = children with a developmental language disorder; mon = monolingual children; bil = bilingual children.

The repetition tasks are taken from the Exalang (Thibault & Helloin, 2010) and Palpa-p (Castro et al., 2007) test batteries.

Effect of bilingualism

Looking at the number of incorrect words produced by monolingual and bilingual children, we find no significant difference between the two groups in the French and CL tasks (French NWR: $F(1, 54) = 3.34, p = 0.073$; CL NWR: $F(1, 54) = 0.18, p = .673$). Similar results are observable by taking into consideration the number of incorrect phonemes (see S1 in Supplementary Material). We also found no interaction between language status and the presence of a disorder in our tasks (French NWR: $F(1, 54) = .00, p = 0.986$; CL NWR ($F(1, 54) = 1.14, p = .290$)).

Effect of stimuli characteristics

We found significant effects of syllabic length for the three tasks (CL NWR: $F(2.5, 152.9) = 26.23, p < .001$; French NWR: $F(1, 59) = 22.79, p < .001$; Portuguese NRW: $F(1.7, 49.1) = 4.59, p = .018$) and phonological complexity for the language-specific tasks (French NWR: $F(1.7, 100.4) = 25.97, p < .001$; Portuguese NWR: $F(1.7, 47.8) = 22.34, p < .001$) in our results (see S5 in Supplementary Materials). We note that long and complex words are more difficult for all children. However, we also find significant interactions between disorder and length in the CL and French tasks, when we consider the number of incorrect phonemes (CL NWR: $F(1.28, 75.74) = 9.41, p = .001$; French NWR: $F(1, 59) = 8.98, p = .004$). Children with a DLD make more errors than TD children on long words than on words with fewer syllables. No interaction between length and disorder was found for the Portuguese NWRT (e.g., for count in incorrect phonemes: $F(1.8, 52.1) = 1.84, p = .171$). Finally, we see significant interactions between disorder and phonological complexity in the language-specific tasks (French NWR: $F(1.7, 100.4) = 15.75, p < .001$; Portuguese NWR: $F(1.7, 47.8) = 3.82, p = .035$). In the French task, DLD children make more mistakes than TD children on complex words. While in the

Portuguese task, the difference between DLD and TD children is larger for simple words than for complex words (the simple ones are also the most frequent in this task).

Error analysis

As noted before, children with DLD produce more errors than children with TD. We now turn to analyzing whether the distribution of errors differs between the two groups (see S6 in the Supplementary Material). For the NRW task in French, we find significant differences in the proportion of errors between TD and DLD children at the phoneme ($p < .001$), syllable ($p = 0.040$) and word ($p < .001$) level. For the phoneme level, we find significant differences in the proportion of errors for consonant ($p < .001$) and vowel substitutions ($p = .007$) and for consonant ($p = .031$) and vowel additions ($p = .007$). For the syllable level, we find significant differences in the proportion of errors of vowel omissions ($p = .038$) and cluster reduction ($p = .023$). For the word level, we note differences only for assimilation errors ($p = .009$). We note statistical differences in Portuguese NWR for only two levels (phoneme: $p = .001$; word: $p = .047$). We note differences in consonant ($p = .026$) et vowel substitutions ($p = .010$), for vowel additions ($p = .002$) and for lexicalizations ($p = .025$). In the CL task, although there are many errors, we observe no significant differences by level (phoneme: $p = .224$; syllable: $p = .461$; word: $p = .116$).

Discriminative value

Finally, we examined the discriminative value of the tasks. After generating ROC curves (Dunn, 2014) for the different tasks and different scoring method (see S7 in the Supplementary Material), we calculated Youden's indices for all possible thresholds. We note that only the French task (with scoring in number of incorrect words) provides sensitivity and specificity

scores (see Table 1) that are to have good discriminant accuracy (Plante & Vance, 1994) with 88 % of sensitivity and 92 % of specificity.

Table 1

Discriminating values of tasks according to monolingual and bilingual children with a typical language development and children with a developmental language disorder.

Tasks	Scoring method	Population	Sensitivity	Specificity	Threshold	Youden Index
French NWR	words	all	88 %	92 %	2.5	0.80
French NWR	phonemes	all	77 %	100 %	5.5	0.77
CL NWR	words	all	82 %	76 %	6.5	0.59
CL NWR	phonemes	all	77 %	73 %	11.5	0.50
French NWR	words	MON	92 %	76 %	2.5	0.69
French NWR	words	BIL	88 %	92 %	2.5	0.80
French NWR	phonemes	MON	100 %	70 %	2.5	0.70
French NWR	phonemes	BIL	88 %	100 %	5	0.88
CL NWR	words	MON	85 %	64 %	6.5	0.50
CL NWR	words	BIL	82 %	76 %	6.5	0.59
CL NWR	phonemes	MON	85 %	58 %	8.5	0.44
CL NWR	phonemes	BIL	82 %	84 %	11.5	0.67

Tasks	Scoring method	Population	Sensitivity	Specificity	Threshold	Youden Index
Portuguese NWR	words	BIL	82 %	76 %	2.5	0.59
Portuguese NWR	phonemes	BIL	82 %	76 %	44.5	0.59

Note. Abbreviations: NRW = non-word repetition task; CL = crosslinguistic task; MON = monolingual children; BIL = bilingual children.

$N = 61$ ($n = 31$ children with a DLD and 30 children with a TD).

The repetition tasks are taken from the Exalang (Thibault & Helloin, 2010) and Palpa-p (Castro et al., 2007) test batteries.

We then separated the monolingual and bilingual groups in the two tasks that were tested in both and calculated the discriminative values according to linguistic status (see Table 1 and S7 and S8 in Supplemental Materials). Considering Youden's indices (1950), we note that the two tasks tend to be more discriminating for the bilingual population than for the monolingual population in our research. We also observe that the CL task obtains sensitivity (82 %) and specificity (76 %) percentages approaching the Plante and Vance (1994) standards, and the diagnostic accuracy of this task improves for the group of bilingual children with a count of incorrect phonemes (82 % of sensitivity and 84 % of specificity).

General discussion and conclusion

This article assessed the diagnostic accuracy of NWR tasks in identifying developmental disorders in bilingual and monolingual school-age children speaking French and Portuguese. In line with other recent studies (e.g., Kapalková et al., 2013), the first analyses show that three tasks distinguish the group of children with DLD from children with TD on a group level. No differences were, however, found between the monolingual and bilingual groups, which is encouraging for the use of these tasks in bilingual language assessment (Antoniejevic-Elliott et al., 2019). These results can partly be explained by the phonologies of French and Portuguese, which are very similar (including potential sources of difficulty for children, such as the presence of nasal vowels and consonant clusters, Kihm, 2016). However, in the French NWRT, we observe a tendency: monolingual children seem to perform slightly better than bilingual children, which indicates that we should still be cautious in using language-specific NWRT even if the two languages shared by bilingual children are close.

We also looked at other factors that may influence NWR task scores. We observe that age, length and phonological complexity are influencing factors for all three tasks, a result consistent with previous studies (e.g., de Almeida et al., 2017). The use of complex materials,

i.e., stimuli that vary in length and complexity, seems to be a good way to dissociate children with and without DLD. While the crosslinguistic task does not test language-specific complex elements of phonological structures, we observe that it is in this task where TD and DLD children make the most errors, probably because it contains the longest items (up to five syllables) and involves phonological memory. It is important to note that the use of a pre-recorded voice of a Swedish speaker for the CL NWRT certainly made the task complex for the different groups of children (monolingual and bilingual TDs and DLDs) which may also explain the presence of many errors. However, in this task, the number of errors produced is large, but the proportions are not significantly different between TD and DLD children. In the two language-specific tasks, whose items vary in length and complexity, we note significantly different proportions depending on the clinical status of the children. In particular, we observed more frequent errors (especially vowels additions and substitutions) in children with DLD in these two tasks.

Finally, we looked at the discriminative value of the tasks. We tested the use of two types of scoring methods: the number of incorrect words and the number of incorrect phonemes since the methods of analysis seem to have an impact on diagnostic quality (e.g., Boerma et al., 2015). The best option seems to be to use the number of incorrect words since it yields better Youden indices (1950), but only the language-specific French task provides a good discriminatory accuracy (Plante & Vance, 1994). In fact, this task comes from a test whose psychometric qualities were controlled (Thibault & Helloin, 2010). The Portuguese task, on the other hand, seems less sensitive and specific. Interestingly, the best threshold found in this study does not correspond to the norm values, which yield even worse Youden's indices (see Schwob & Skoruppa, to appear), highlighting the weak discriminative power of this task which has been normed with few individuals (de Almeida et al., 2017). Finally, the CL task also seems promising, particularly for bilinguals, especially since it is usable for most languages

(Antoniejevic-Elliott et al., 2019). Moreover, this task is particularly interesting for languages that do not have standardized tools available on the market. Since the items vary considerably in length, this task appears to assess verbal short-term memory more than phonological representation. We note, however, that at the age of our participants, the specific phonological complexities of the Portuguese and French language have an influence on the children's scores and should thus be taken into consideration.

In addition, when we are interested in length effects, a count by the number of incorrect phonemes was more sensitive, because it allowed us to identify different strategies according to the groups of children: children with a DLD struggle to repeat long non-words and produce the beginning or do not repeat at all; conversely, children with TD tend to try to produce the non-word (even if it is long) with some phonemes errors. These differences are not noticeable when we count the number of incorrect words, as both solutions yield a “0” score.

In this brief article, we have focused only on NWR tasks, although it is widely acknowledged that the variety of assessment tasks also contributed to improving diagnostic power (Boerma & Blom, 2017). Thus, it would be interesting to replicate this study with larger groups of participants and we should also be interested in combining NWR tasks with other tools that have been identified as promising (e.g., dynamic assessment: Hasson & Botting, 2010; narrative tasks: Boerma & Blom, 2017). However, as a first step, we were able to show in this article that the use of NWR tasks to assess the monolingual or bilingual French-Portuguese child is appropriated and thus merits its place in the oral language assessment.

Acknowledgments

We would like to thank all participants and their families for taking part in this research, as well as colleagues, students and members of institutions who allowed us to contact the families. We also thank our Portuguese-speaking student assistants for their contribution to this study. Parts of a prior version of this text were translated from French by the free version of the software Deep-L.

References

- Antonijevic-Elliott, S., Lyons, R., O' Malley, M. P., Meir, N., Haman, E., Banasik, N., Carroll, C., McMenamin, R., Rodden, M., & Fitzmaurice, Y. (2019). Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics & Phonetics*, 34(4), 293-311.
<https://doi:10.1080/02699206.2019.1637458>
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi- universal nonword repetition task as a diagnostic tool for bilingual children learning dutch as a second language. *Journal of Speech, Language, and Hearing Research*, 58(6), 1747-1760. https://doi:10.1044/2015_JSLHR-L-15-0058
- Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, 66(1), 65-76.
<https://doi:10.1016/j.jcomdis.2017.04.001>
- Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R. et Coltheart, M. (2007). *Provas de avaliação da linguagem e da afasia em português*. CEGOC.
- Chiat, S. (2015). Non-Word Repetition. In S. Armon-Lotem, K. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children* (pp. 95-122). Multilingual Matters.
- Chiat, S., & Polisenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualisms status on children's performance. *Journal of Speech, Language, and Hearing Research*, 59(5), 1179-1189.
https://doi:10.1044/2016_JSLHR-L-15-0293
- Coady, J., & Evans, J. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *Journal of Language & Communication Disorders*, 43(1), 1-40. <https://doi:10.1080/13682820601116485>

- Conover, W., & Iman, R. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3), 124-129.
<https://doi:10.1080/00031305.1981.10479327>
- De Almeida, L., Ferré, S., Morin, E., Prévost, P., Dos Santos, C., Tuller, L., Zebib, R., & Barthez, M.-A. (2017). Identification of bilingual children with specific language impairment in France. *Linguistic Approaches to Bilingualism*, 7, 331-358.
<https://doi:10.1075/lab.15019.alm>
- Dollaghan, C., & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136-1146.
<https://doi.org/10.1044/jslhr.4105.1136>
- Dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition*, 25(1), 58-71.
<https://doi.org/10.1080/10489223.2016.1243692>
- Dunn, G. (2014). Statistics in psychiatry. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1136-1138). Arnold.
- Estes, K. G., Evans, J., & Else-Quest, N. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177-195.
<https://doi:1092-4388/07/5001-0177>
- Federal Statistical Office (2020, February 27). Most frequent main languages.
<https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/langues-religions/langues.assetdetail.11947533.html>
- Girbau, D., & Schwartz, R. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, 42(1), 59-75. <https://doi:10.1080/13682820600783210>

- Glutting, J., Adams, W., & Sheslow, D. (2000). *Wide range intelligence test*. Wide Range.
- Gutiérrez-Clellen, V., & Simon-Cereijido, G. (2010). Using nonword repetition tasks for the identification of language impairment in spanish-english speaking children: Does the language of assessment matter. *Learning Disabilities Research & Practice*, 25(10), 48-58. <https://doi.org/10.1111/j.1540-5826.2009.00300.x>
- Hasson, N. et Botting, N. (2010). Dynamic assessment of children with language impairments: A pilot study. *Child Language Teaching and Therapy*, 26(3), 249-272. <https://doi.org/10.1177/0265659009349982>
- IBM Corp. (2020). *IBM SPSS Statistics for Macintosh* (Version 27.0) [Computer software]. IBM Corp.
- Kihm, A. (2016). *Fiche langue le portugais européen*. CNRS website. https://lgidf.cnrs.fr/sites/lgidf.cnrs.fr/files/images/PORTUGAIS_24_06_19_A4.pdf
- Kapalková, S., Polisenská, K., & Vicenová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: novel scoring methods. *International Journal of Language and Communication Disorders*, 48(1), 78-89. <https://doi.org/10.1111/j.1460-6984.2012.00189.x>
- Marini, A., Ruffino, M., Sali, M. E., & Molteni, M. (2017). The role of phonological working memory and environmental factors in lexical development in Italian-speaking late talkers: A one-year follow-up study. *Journal of Speech, Language, and Hearing Research*, 60(12), 3462-3473. https://doi.org/10.1044/2017_JSLHR-L-15-0415
- Mateus, M. M., Falé, I., & Freitas, M. (2005). *Fonética e Fonologia do Português*. Lisboa: Universidades Aberta.
- Plante, E., & Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools*, 25(1), 15-24. <https://doi.org/10.1044/0161-1461.2501.15>

- Schwob, S., & Skoruppa, K. (to appear). L'identification d'un trouble développemental du langage oral chez les enfants bilingues français-portugais : Les ressources à disposition des orthophonistes. *Revue canadienne d'orthophonie et d'audiologie*.
- Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P. et Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, ePub Ahead of Print https://doi.org/10.1044/2021_JSLHR-20-00552
- Thibault, M. P., & Helloin, M. C. (2010). *Exalang 5-8*. Happyneuron.
- Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, K. de Jong & N. Meir (Eds.), *Assessing multilingual children* (pp. 301-331). Multilingual Matters.
- Williams, C., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech Language Pathology*, 14(3), 292-305. <https://doi:10.3109/17549507.2011.636071>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32-35. [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)

Supplementary Material 1

Factors influencing performance on repetition tasks

Tasks	Scoring method	Disorder	Bilingualism	Disorder * Bilingualism	Age	Age * Disorder	Age * Bilingualism
French NWR	words	$F(1, 54) = 65.85,$ $p < .001$ ***	$F(1, 54) =$ $3.34, p = .073$	$F(1, 54) =$ $0.00, p = .986$	$F(1, 54) =$ $15.49, p < .001$ ***	French NWR	words
French NWR	phonemes	$F(1, 54) = 75.53,$ $p < .001$ ***	$F(1, 54) =$ $2.26, p < .138$	$F(1, 54) =$ $0.00, p = .942$	$F(1, 54) =$ $14.57, p < .001$ ***	French NWR	phonemes
Portuguese NWR	words	$F(1, 25) = 15.5,$ $p = .010$ **	/	/	$F(1, 25) = 18.61$ $p < .001$ ***	Portuguese NWR	words
Portuguese NWR	phonemes	$F(1, 25) = 20.54,$ $p < .001$ ***	/	/	$F(1, 25) =$ $32.14, p <$ $.001$ ***	Portuguese NWR	phonemes

Tasks	Scoring method	Disorder	Bilingualism	Disorder *	Age	Age * Disorder	Age * Bilingualism
CL NWR	words	$F(1, 54) = 18.61,$ $p < .001$ ***	$F(1, 54) =$ 0.18, $p = .673$	$F(1, 54) =$ 1.14, $p = .290$	$F(1, 54) = 5.78,$ $p = .020$ *	CL NWR	words
CL NWR	phonemes	$F(1, 54) = 14.32,$ $p < .001$ ***	$F(1, 54) =$ 0.10, $p = .749$	$F(1, 54) =$ 2.19, $p = .144$	$F(1, 54) = 4.07,$ $p = .048$ *	CL NWR	phonemes

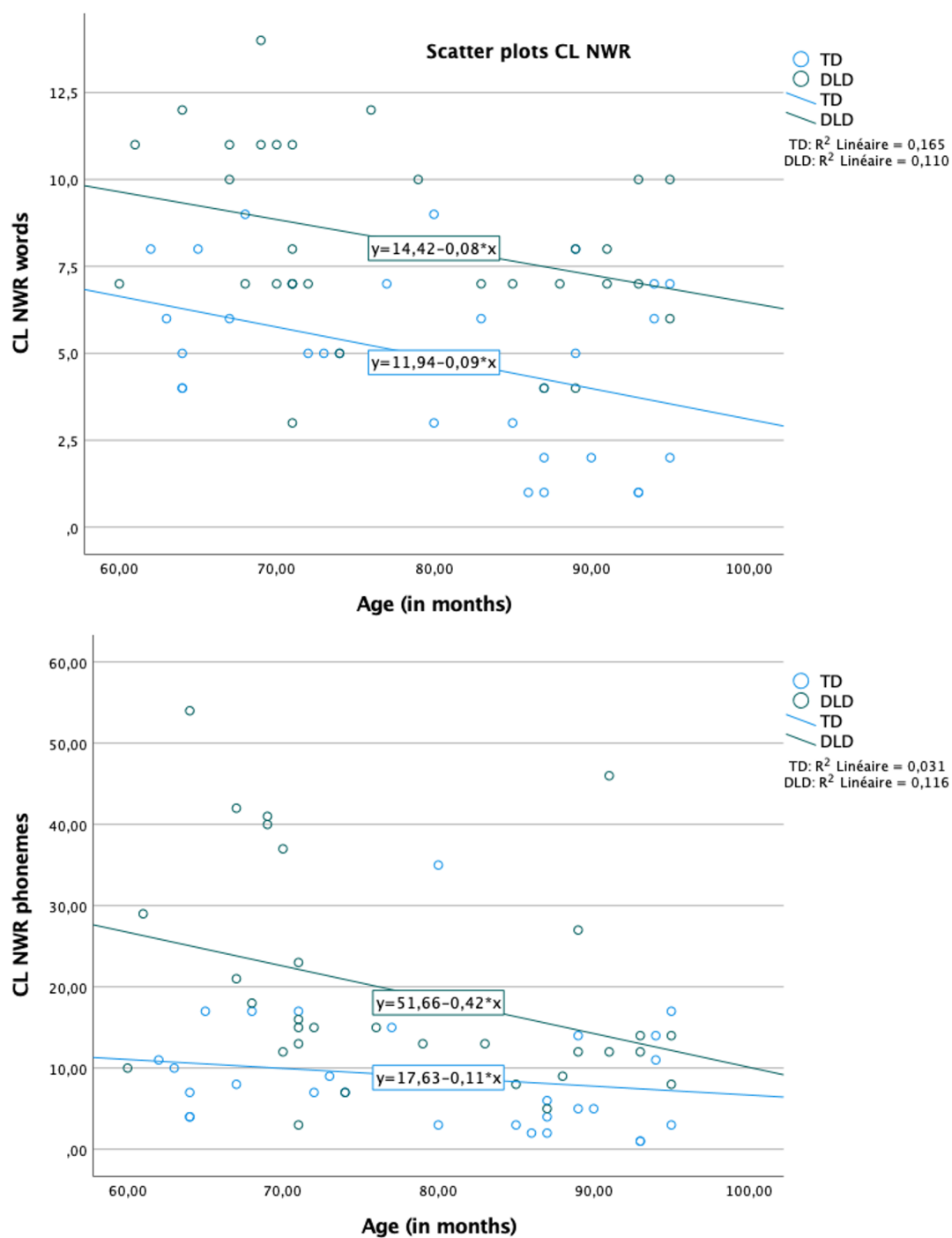
Note. Abbreviations: NRW = non-word repetition task; CL= crosslinguistic

$N = 61$ ($n = 31$ children with a DLD and 30 children with a TD / $n = 30$ bilingual children and 31 monolingual children).

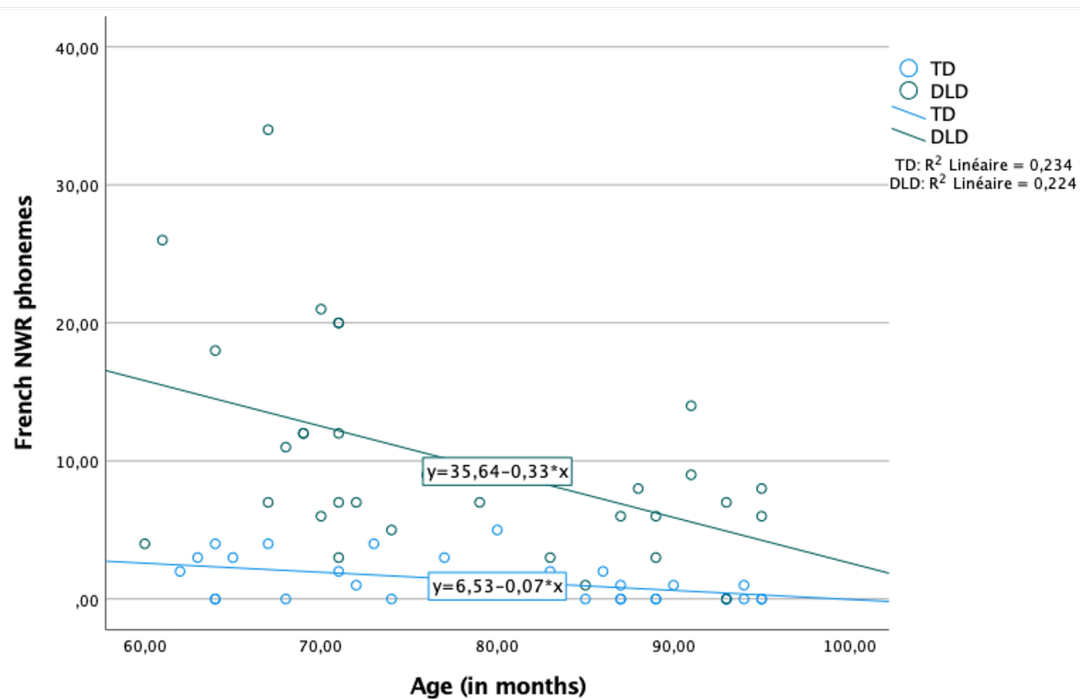
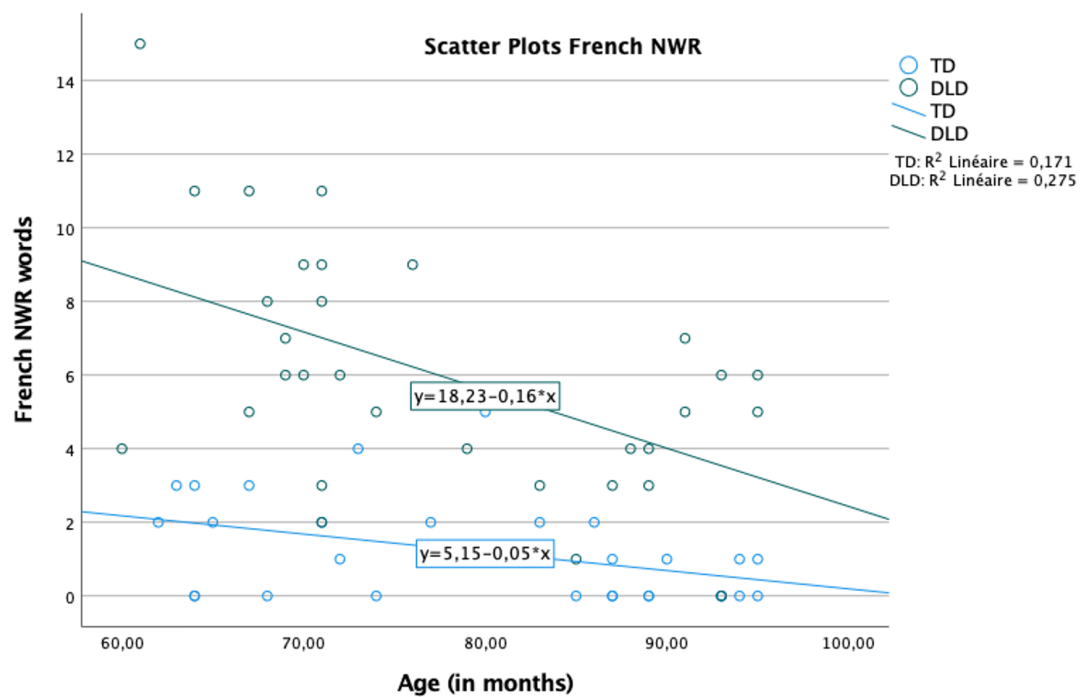
The repetition tasks are taken from the Exalang (Thibault & Helloin, 2010) and Palpa-p (Castro et al., 2007) test batteries.

* $p < .05$. ** $p < .01$. *** $p < .001$.

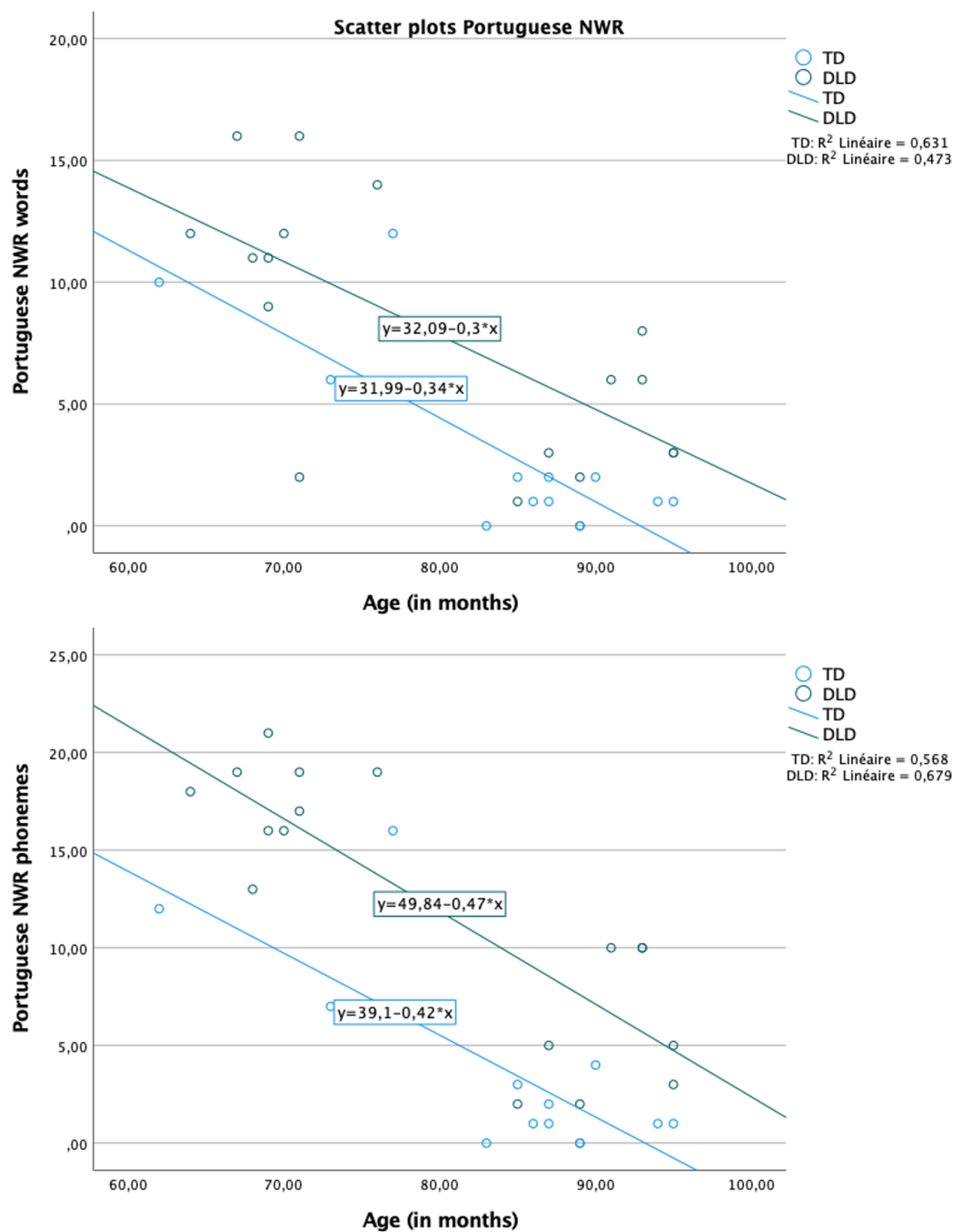
Supplementary Material 2



Supplementary Material 3



Supplementary Material 4



Supplementary Material 5

Effect of stimuli characteristics

Tasks	Scoring method	Length	Length * Disorder	Phonological complexity	Phonological complexity * Disorder
French NWR	word	$F(1, 59) = 22.79, p < .001^{***}$	$F(1, 59) = 2.31, p = .133$	$F(1.7, 100.4) = 25.97, p < .001^{***}$	$F(1.7, 100.4) = 15.75, p < .001^{***}$
French NWR	phoneme	$F(1, 59) = 21.44, p < .001^{***}$	$F(1, 59) = 8.98, p = .004^{**}$	$F(1.64, 97.21) = 9.75, p < .001^{***}$	$F(1.64, 97.21) = 6.53, p = .002^{**}$
Portuguese NWR	word	$F(1.7, 49.1) = 4.59, p = .018^{*}$	$F(1.7, 49.1) = 1.58, p = .217$	$F(1.7, 47.8) = 22.34, p < .001^{***}$	$F(1.7, 47.8) = 3.82, p = .035^{*}$
Portuguese NWR	phoneme	$F(1.8, 52.1) = 5.97, p = .005^{**}$	$F(1.8, 52.1) = 1.84, p = .171$	$F(1.5, 44.1) = 33.85, p < .001^{***}$	$F(1.5, 44.1) = 7.26, p = .004^{**}$
CL NWR	word	$F(2.5, 152.9) = 26.23, p < .001^{***}$	$F(2.5, 152.9) = 5.20, p = .003^{**}$	/	/
CL NWR	phoneme	$F(1.28, 75.74) = 39.36, p < .001^{***}$	$F(1.28, 75.74) = 9.41, p = .001^{***}$	/	/

Note. Abbreviations: NWR = non-word repetition task; CL = crosslinguistic

$N = 61$ ($n = 31$ children with a DLD and 30 children with a TD / $n = 30$ bilingual children and 31 monolingual children).

The repetition tasks are taken from the Exalang (Thibault & Helloin, 2010) and Palpa-p (Castro et al., 2007) test batteries.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Supplementary Material 6

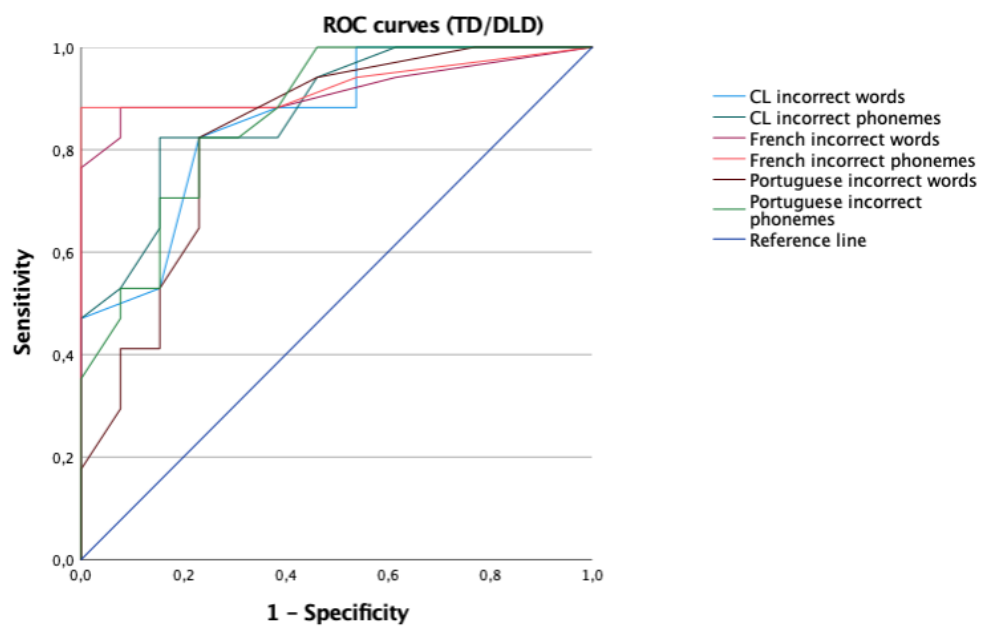
Errors in NWR tasks

Errors types	French NWR			Portuguese NWR			CL NWR		
	Sum TD	Sum DLD	¹ TD/DLD	Sum TD	Sum	TD/DLD	Sum TD	Sum DLD	TD/DLD
	DLD								
<i>Phoneme level</i>	20 (54 %)	162 (61 %)	$p < .001$	17 (40 %)	91 (49 %)	$p = .001$	124 (57 %)	222 (47 %)	$p = .224$
			***			***			
Consonant substitutions	10 (27 %)	95 (36 %)	$p < .001$	12 (28 %)	46 (25 %)	$p = .026 *$	59 (27 %)	122 (26 %)	/

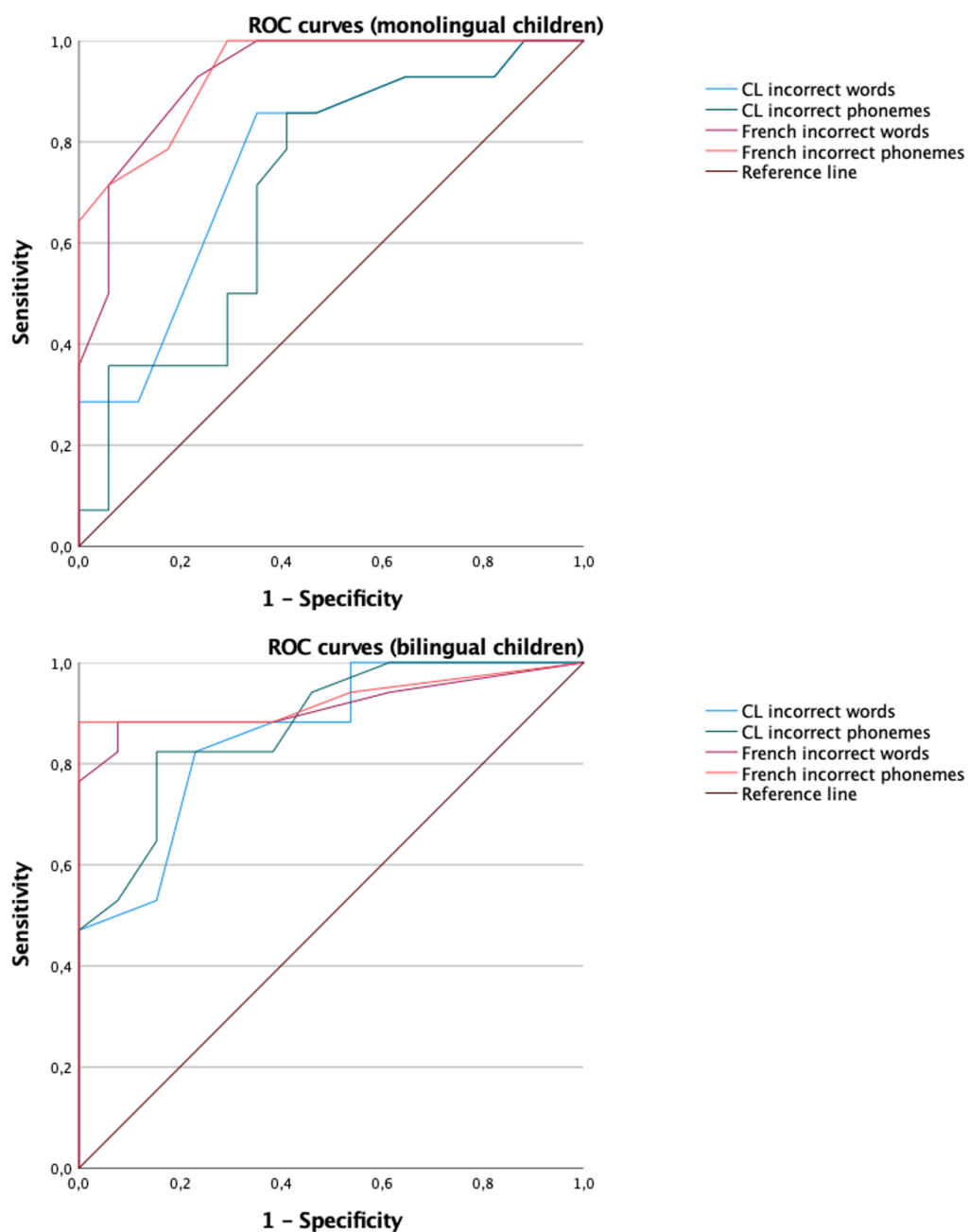
Vowel substitutions	4 (11 %)	29 (11 %)	$p = .007$	4 (9 %)	22 (12 %)	$p = .010$	42 (19 %)	70 (15 %)	/
			**			**			
Consonant additions	6 (15 %)	25 (9 %)	$p = .031 *$	1 (2 %)	9 (5 %)	$p = .092$	20 (9 %)	24 (5 %)	/
Vowel additions	0 (0 %)	13 (5 %)	$p = .007$	0 (0 %)	14 (8 %)	$p = .002 **$	3 (1.5 %)	6 (0 %)	/
			**						
<i>Syllable level</i>	9 (24 %)	70 (26 %)	$p = .040 *$	17 (40 %)	55 (30 %)	$p = .237$	54 (25 %)	180 (37 %)	$p = .461$

Errors types	French NWR			Portuguese NWR			CL NWR		
	Sum TD	Sum DLD	¹ TD/DLD	Sum TD	Sum	TD/DLD	Sum TD	Sum DLD	TD/DLD
					DLD				
Consonant omissions	5 (14 %)	36 (14 %)	$p = .060$	3 (8 %)	4 (2 %)		19 (9 %)	94 (20 %)	/
Vowel omissions	1 (3 %)	9 (3 %)	$p = .038 *$	9 (21 %)	16 (9 %)		35 (16 %)	86 (18 %)	/
Cluster reduction	3 (8 %)	25 (9 %)	$p = .023 *$	5 (12 %)	35 (19 %)		/	/	/
Word level	8 (22 %)	34 (13 %)	$p < .001$ ***	9 (20 %)	40 (21 %)	$p = .047 *$	40 (18 %)	75 (16 %)	$p = .116$
Assimilations	3 (8 %)	15 (6 %)	$p = .009$ **	4 (9 %)	7 (3 %)	$p = .828$	17 (8 %)	36 (8 %)	/
lexicalizations	4 (11 %)	9 (3 %)	$p = .418$	4 (9 %)	25 (13 %)	$p = .025 *$	1 (0.5 %)	1 (0 %)	/
Creations	0 (0 %)	6 (2 %)	$p = .081$	0 (0 %)	0 (0 %)	/	7 (3 %)	22 (5 %)	/
Metathesis	1 (3 %)	4 (2 %)	$p = .354$	1 (2 %)	8 (4 %)	$p = .283$	15 (7 %)	16 (3 %)	/

Supplementary Material 7



Supplementary Material 8



Supplementary Material 9

Phonological complexity criteria for non-words complexity of NWR tasks

Tasks	Non-words	Syllabic structure	Level of complexity
CL NWR	/zibu/	CV-CV	0
CL NWR	/lita/	CV-CV	0
CL NWR	/magi/	CV-CV	0
CL NWR	/nuli/	CV-CV	0
CL NWR	/sipula/	CV-CV-CV	0
CL NWR	/bamudi/	CV-CV-CV	0
CL NWR	/malitu/	CV-CV	0
CL NWR	/lumika/	CV-CV-CV	0
CL NWR	/zipalida/	CV-CV-CV-CV	0
CL NWR	/nugitala/	CV-CV-CV-CV	0
CL NWR	/kasuluni/	CV-CV-CV-CV	0
CL NWR	/litisagu/	CV-CV-CV-CV	0
CL NWR	/sibumakila/	CV-CV-CV-CV-CV	0
CL NWR	/dulikasumu/	CV-CV-CV-CV-CV	0
CL NWR	/maluziguba/	CV-CV-CV-CV-CV	0
CL NWR	/litapimudi/	CV-CV-CV-CV-CV	0
French NWR	/abo/	V-CV	1
French NWR	/tizjɔ̃/	CV-CSV	1
French NWR	/zobwɛ̃/	CV-CSV	1
French NWR	/ʃurval/	CVC-CVC	1
French NWR	/fløvir/	CCV-CVC	2
French NWR	/rɛ̃poli/	CV-CV-CV	0
French NWR	/ladɔ̃/	CV-CV	0
French NWR	/mystafo/	CVC-CV-CV	1
French NWR	/pufo/	CV-CV	0
French NWR	/saljɛ̃/	CV-CSV	1
French NWR	/fajame/	CV-SV-CV	0
French NWR	/drigɔ̃bir/	CCV-CV-CVC	2
French NWR	/vjolasɛl/	CSV-CV-CVC	1
French NWR	/gramɔ̃tar/	CCV-CV-CVC	2

Tasks	Non-words	Syllabic structure	Level of complexity
French NWR	/caranel/	CV-CV-CVC	1
French NWR	/bidwano/	CV-CSV-CV	1
Portuguese NWR	/mɐviɐ/	CV-CV-V	0
Portuguese NWR	/sɛlgɐ/	CVC-CV	0
Portuguese NWR	/nɛit/	CVVC	0
Portuguese NWR	/pɑn/	CVCC	0
Portuguese NWR	/ɔdevɐ/	V-CV-CV	1
Portuguese NWR	/kluzɐ/	CCV-CV	2
Portuguese NWR	/izudɐ/	V-CV-CV	1
Portuguese NWR	/fɛnoɐ/	CV-CV-V	0
Portuguese NWR	/tilm/	CVCC	0
Portuguese NWR	/ɛrdɔl/	VC-CVC	1
Portuguese NWR	/sɛij/	CVVC	0
Portuguese NWR	/gretu/	CCV-CV	2
Portuguese NWR	/rɛnɐɔ/	CV-CV-V	0
Portuguese NWR	/fɛt/	CVC	0
Portuguese NWR	/pɛrsu/	CVC-CV	0
Portuguese NWR	/ɛdɔdɐ/	V-CV-CV	1
Portuguese NWR	/gɔt/	CVC	0
Portuguese NWR	/dail/	CVVC	0
Portuguese NWR	/blizɐ/	CCV-CV	2
Portuguese NWR	/brɛm/	CCVC	2
Portuguese NWR	/ɛʒizu/	V-CV-CV	1
Portuguese NWR	/luvɛj/	CV-CVV	0
Portuguese NWR	/vuidu/	CV-V-CV	0
Portuguese NWR	/trɛf/	CCVC	2
Portuguese NWR	/mɛrij/	CV-CVC	0

Note. Abbreviations: NRW = non-word repetition task; CL = Crosslinguistic task; C = consonant; V = vowel; SV = semi-vowel.

The repetition tasks come from the Exalang (Thibault & Helloin, 2010) and Palpa-p (Castro et al., 2007) tests.

The two language-specific tasks have a phonological complexity that can be described on a 3-gradient scale based on Dos Santos & Ferré (2018) and Mateus et al., (2005): complexity 0 = simple structure like consonant-vowel, CV: /ladõ/, /fêr/); complexity 1 = intersyllabic cluster, beginning of non-word by vowel or consonant-semi vowel: /bidwano/, /ədɔde/); and complexity 2 = intrasyllabic cluster: /drigõbir/, /trɛf/).

6.1 Informations additionnelles

Les données quant à la tâche de répétition de non-mots inter-linguistique sont également partagées internationalement afin de contribuer à un article multi-auteurs. Les données ont été transmises à Kamila Poliřenská et Shula Chiat en mars 2021. Nous avons veillé à transmettre uniquement les données des participants ayant accepté de participer à des recherches dans le futur (cf. formulaire de consentement validé par la commission d'éthique).

Partie V : L'usage de l'évaluation dynamique au sein du bilan logopédique

7. Article de recherche

Evaluating language learning abilities to detect developmental language disorder in monolingual and bilingual children: A comparison of dynamic assessment situations with and without adult mediation.

Schwob, S. et Skoruppa, K.

État: en révision

Abstract

Purpose: Bilingual children's language skills are strongly influenced by exposure to each of their languages, among other linguistic, environmental and cognitive factors (e.g., Paradis, 2019). In the speech and language therapy clinic, it is difficult to disentangle developmental language disorders from insufficient exposure (Hasson et al., 2013). Dynamic assessment, which directly tests the learning potential of children (Camilleri & Law, 2007), offers a promising solution for this dilemma.

Method: This study compares the clinical potential of two dynamic assessment situations, varying the learning setting (more and less adult mediation; autonomous computer game and interactive story reading with graduated cues), and the linguistic domains and materials involved (nouns, verbs and inflections). Forty-nine French monolingual and French-Portuguese bilingual children with and without developmental language disorder, aged 5;0 to 7;11 years, participated.

Results: Performance in most assessment tasks of the two situations differed according to clinical status (developmental language disorder vs. typical development) with medium and large effect sizes, but not according to language background (except for some productive tasks), highlighting the clinical value of our situations and the near absence of linguistic biases. Overall, a combination of tasks of the interactive dynamic assessment situation yielded the highest clinical classification power, with 91 % sensitivity and 85 % of specificity.

Conclusions: This study adds to evidence that dynamic assessment is a promising task for identifying bilingual and monolingual children with developmental language disorder, particularly when the situation involves interaction, and the experimenter provides graduated cues to the child.

Introduction

Language development and assessment in bilingual children

At present, the majority of children in the world grow up with more than one language (Special Eurobarometer 386, European Commission, 2012). In Switzerland, where our research was conducted, bilinguals represent around two thirds of the population and language combinations frequently include several of the four national languages or may be combined with other languages such as Portuguese, which is the most common additional language in French-speaking part of Switzerland (Federal Statistical Office, 2020).

The primary developmental milestones (first words and word combinations) are achieved at the same time in bilingual and monolingual children (Nicoladis & Genesee, 1997). However, bilingual children's development is even more heterogeneous than that of monolinguals due to a variety of additional influences: the quantity (e.g., number of waking hours of exposure) and quality of language exposure (especially exposure in interactive activities according to a socio-cognitive theory of language development, Vygotski, 1978), as well as the typology and proximity of the child's languages (Paradis, 2019). Furthermore, there are variations by language domain – for instance, bilingual lexical development (Thordardottir, 2011) seems to be more influenced by exposure than morphosyntactic development Thordardottir (2015) – and language modality – the receptive-expressive vocabulary gap may be greater for bilingual children than for monolingual children (Keller et al., 2015). Finally, language status, migration background, motivation, learning style, and self-confidence (Paradis, 2019; Paradis et al., 2011) may affect the language development of the bilingual child.

In principle, bilingual children are no more likely to be affected by a developmental language disorder (DLD, Bishop et al., 2017) than their monolingual peers, with an estimated prevalence of 7.8 % in preschool children (Norbury et al., 2016). However, in the case of bilingual development, it is currently difficult for speech and language therapists (SLTs) to

differentiate between a true DLD and temporary language difficulties due to insufficient exposure to the language being tested (Camilleri & Law, 2007) or to the various influencing factors of bilingual development (see above). Qualitatively, the nature of language difficulties is also often similar during the initial stages of learning a new language and in pathological contexts (Armon-Lotem, 2012). Bilingual children are therefore sometimes over-identified, when any failure on assessment tasks is interpreted systematically as a language disorder, or under-identified when all difficulties are considered to be the inevitable result of lack of language exposure (Camilleri & Law, 2007). It is often only the persistence of the difficulties, that indicates the presence of a DLD and demonstrates clear difficulties with language learning (Gray, 2003). In order to reduce these identification errors, several avenues of assessment have been proposed in the scientific literature. Because it appears to be least biased by children's level of exposure, the non-word repetition task (NWRT) has proven to be of discriminative value in assessing DLD in both monolingual and bilingual contexts (Estes et al., 2007; Schwob et al., 2021). However, such a single task is not sufficient to obtain a full profile of the child's difficulties (Peña et al., 2016) and children with DLD without phonological processing difficulties can pass it. Dynamic assessment (DA), which directly probes the learning potential of children (Camilleri & Law, 2007) rather than knowledge of one particular language per se, has also been proposed as a less biased solution to evaluate children speaking different languages and coming from different cultures (Peña et al., 2014).

Methods of dynamic assessment

As opposed to the static character of classical tests, a dynamic procedure can be proposed to assess particularly children with learning difficulties, speaking different languages, or coming from different cultures (Budoff et al., 1987). Based on the postulates of Vygotski's socio-cognitive theory (1978; notably the concept of “zone of proximal development”),

assessment of learning potential was set up in the mid-20th century (See Lidz, 1987 for historical perspective) in the field of psychology, mainly to compensate for the cultural and language bias of IQ tests. However, even before that, ideas similar to what is now called DA were evoked by authors around the world (e.g., Rey's assessment of educability, 1934; Woodrow's learning ability, 1946). There is considerable confusion about what constitutes DA, and the characterization of different subtypes is often unclear (Hessels-Schlatter & Hessels, 2009). Furthermore, Hessels-Schlatter & Hessels (2009) explain that DA can take distinct forms depending on the objectives pursued: diagnosis or intervention. For diagnosis, a learning task with standardized training is usually proposed, with training or aids that can be more or less numerous and take several forms (Hessels-Schlatter & Hessels, 2009; Hessels & Hessels-Schlatter, 2010). Given that the present study aims at detecting language disorders in bilingual children, we will focus on such diagnostic applications in the SLT domain, leaving aside more intervention-focused dynamic clinical work which, following Feuerstein's (1979) framework, focusses more on the patient's learning processes and modifiability (see, for example, studies of Peña and colleagues on narratives, e.g., Peña et al., 2007, for an SLT application).

In the following, we will first explain different DA formats (and combinations thereof) following the classification of Hessels-Schlatter & Hessels (2009); Hessels & Hessels-Schlatter (2010), and then zoom in on the linguistic domains and materials that can be used. We will focus on SLT studies, several of which are included in Orellana et al.'s (2019) meta-analysis of dynamic assessment of bilingual children in SLT. A first format consists of three phases- static pre-test, training, and static post-test (e.g., Hasson et al., 2013; Petersen et al., 2020), which allows for comparing pre- and post-tests, as well as measuring the help given during training, using, for example, modifiability measurement scales (e.g., Petersen et al., 2020). Another procedure is to train the child on new, invented language materials without pre-test and to evaluate their learning gain only via a post-test. For example, Roseberry & Connell (1991)

taught a novel morphological rule to children for 15-20 minutes. They then offered a post-test to see if the rule had been understood and could be applied by the children. A third frequently used method is to give cues to the child based on the errors produced (e.g., Camilleri & Law, 2007; Hasson et al., 2013). For example, in the DAPPLE test by Hasson et al. (Hasson, 2017; Hasson et al., 2013), the authors use a pre-test, training, and post-test procedure and during training, they offer graduated cues based on the responses given by the child, such as semantic (Alt et al., 2006), prosodic and gestural cues (Weismer & Hesketh, 1993) to help the child learn new words. In this way, one can also calculate the number or type of aids needed by the child to be able to reach the expected response. Finally, a last procedure involves minimal corrective feedback during the training phase. The use of verbal language can be reduced in this type of task (Hessels, 1997) in order to neutralize potential linguistic biases. For example, Skoruppa (2019) used a computerized procedure so that the experimenter intervenes only for general feedback during the task. Corrective feedback was provided by the computer tool based on the child's responses.

To sum up, learning test situations can take many forms and depend in particular on the procedure used and the purpose of the test. Nowadays, children are exposed to new language materials both during human interaction and through digital media, and both seem to lead to learning beyond infancy: Van Horn & Kan (2015) compared unfamiliar words learning in two frequent situations in preschool children's lives, a TV cartoon and live story reading. The authors did not find any differences in the situations since the training procedure was identical between the two situations (but see Lytle et al. 2018, for a discussion of the superiority of interactive contexts in very young children). Thus, it seems that it is rather the procedure used and the objective pursued that has an influence rather than the situation itself, as long as it is motivating for the child (Budoff, 1987).

The purpose of the present study is to compare the clinical potential of two DA procedures (more vs. less adult mediation) in two learning situations (autonomous computer game vs. interactive story reading) in order to assess which procedure and situation is best suited to highlight differences between monolingual and bilingual children with TD and DLD.

Word learning

Turning now to different linguistic domains, word learning skills are a good and well-studied candidate for DA: Many studies involving “fast mapping” or “quick incidental learning” have shown that after only a few presentations of new word-meaning mappings, TD children already have phonological and semantic representations of them (Brackenbury & Fey, 2003). Indeed, the lexical domain appears to provide good results in distinguishing children with TD and DLD as well as high diagnostic accuracy (Petersen et al., 2020), since children with DLD have more difficulty learning new words (or non-words, Kapantzoglou et al., 2012). In their meta-analysis on word learning in TD vs. DLD children, Kan & Windsor (2010) obtained a medium mean effect size. The latter require more presentations and more cues for learning new words, which has been related to their typical deficits in short-term verbal memory (Gray, 2003) and difficulties with semantic analysis (Nash & Donaldson, 2005). Kan & Windsor (2010) noted that in addition to clinical status, children's age, task design (receptive vs. productive tasks, item types, number of item presentations, cues provided) and children's cognitive profile also appear to be important moderators in explaining the differences between groups. In general, this meta-analysis found that children appear to perform better if they are older (e.g., school-age vs. preschool age group), if they have a high non-verbal IQ and if they are assessed through receptive tasks involving nouns (vs. verbs). Receptive assessment tasks appear to be easier because they are less demanding (Gray, 2003) and in word learning, understanding is thought to precede production. One possible reason for the proficiency of

learning nouns vs. verbs is that in the acquisition of Indo-European languages, nouns are learned earlier than verbs, the latter often being considered more difficult (Waxman et al., 2013). Other explanation is that children are more interested in the salient semantic characteristics of objects such as form or animacy vs. the salient characteristics of actions such as movement (Alt et al., 2004). In particular, Alt et al. (2004) found that the DLD children in their study had more difficulty learning verbs vs. nouns than the TD children, which may be due to the difficulty of the syntactic role of the verbs. To learn verbs, however, both groups seem to be sensitive to the presence of morphological markers, such as the addition of a morphophonological rule to facilitate learning (Alt et al., 2004; Gray, 2003).

Differences between monolingual and bilingual population in learning new words have also been investigated. Research on children has shown an advantage for bilingual children in certain executive functions (e.g., working memory) and metalinguistic skills (e.g., ability to manipulate syllables and phonemes, Adesope et al., 2010) that can facilitate their learning, especially in more challenging contexts. Because of their exposure to two languages, bilingual children may have developed linguistic processing systems that are simply more efficient than their monolingual peers, potentially resulting in a bilingual advantage in word learning (Alt et al., 2019). However, it is important to mention that there is no consensus about a real advantage of bilingual children in terms of their executive functioning (de Cat et al., 2018). With respect to language learning studies in children with TD, studies have reported both bilingual advantages and disadvantages, highlighting the need for clinical DA tasks to have both a monolingual and a bilingual control group: Advantages were found in an adjective learning task (Yoshida et al., 2011) and in a perceptual verb learning task (Skoruppa, 2019), and could be linked to better inhibition of salient noun interpretations in bilinguals (Yoshida et al., 2011). Disadvantages were found for picture naming of English-like novel words, possibly explained by the greater familiarity of monolingual children with the English sound system (Alt et al.,

2013) and in a mispronunciation task involving English phonology (Alt et al., 2019), possibly explained by the greater leniency of bilingual children due to their greater exposure to accented speech. However, many studies find no differences between monolingual and bilinguals' learning of new words (e.g., receptive tasks in Alt et al., 2013, noun learning in Skoruppa, 2019), highlighting the clinical interest of this task to detect true language difficulties beyond insufficient exposure in children with DLD.

Morphosyntactic learning

Fewer learning studies have focused on the morphosyntactic domain, which is widely recognized as one of the biggest challenges for children with DLD (Bishop et al., 2017). In their meta-analysis, Orellana et al. (2019) retain only one article in this domain, a study by Roseberry & Connell (1991), which consists of learning an invented morphological rule (an invented morpheme meaning “part of”). Significant performance differences were found between bilingual TD and DLD children in understanding and producing the new morphosyntactic rule. Hasson et al. (2013) proposed an expressive syntax task, with a mediation and gradual scaffolding phase designed to increase the number of syntactic elements children use in sentence in comparison to a pre-test. The authors noted that less mediation is required for TD vs. DLD children, and that TD children improved more at posttest. However, there were already differences between TD and DLD children at the static pre-test, which complicates the comparison between the two groups.

Looking at language status, Kaushanskaya et al. (2017) found no differences in the ability to learn a new derivational morpheme marking part-whole distinction in English nouns between monolingual and bilingual TD children. This highlights the clinical potential of such a task since it does not seem to be influenced by language status. Kohnert & Danahy (2007) assumed that the performance of bilingual children depends on the morphological systems of

the languages they speak. Performance on such tasks would therefore not be influenced by language status itself, but by the morphological system of their languages. It is therefore important to pay attention to these elements when creating new tasks.

To sum up, there is a growing interest for different types of DA in the SLT literature, although few studies have directly compared different DA procedures and learning situation and linguistic domains. However, so far, no conclusive clinical applications have been constructed. Apart from Hasson et al.'s (Hasson, 2017; Hasson et al., 2013) pre-standardized DAPPLE test, few tasks from SLT studies can be considered as learning tests with a diagnostic purpose. Moreover, the lexical domain has been particularly researched, but more work is needed in the morphosyntactic domain, which is frequently affected by DLD. Moreover, only the study by Camilleri & Law (2007) compared monolingual and bilingual populations with and without DLD.

Aims of the study

The present study aims to fill, at least partially, these gaps identified in the literature. Here, we will compare two DA situations with a population of French- and Portuguese-speaking children aged between 5;0 and 7;11 years with and without previously diagnosed DLD. We chose this language combination because it is one of the most frequent ones in French-speaking Switzerland (Federal Statistical Office, 2020), especially in an SLT context (Volpin et al., 2020). We selected this age range because children with DLD are frequently identified at these ages (Bishop et al., 2017).

We are interested in word learning, both of nouns and verbs, but also in the area of morphosyntax acquisition (we teach a novel inflection rule), which is less studied in DA research. In addition to productive tasks, we want to offer receptive tasks because they seem to be less subject to participants' language status (Alt et al., 2013). We also want to compare two

ecological learning situations that are frequent in children's lives (interactive story reading and autonomous computer game), and that vary in the procedure used and adult assistance provided. Specifically, we want to determine if there are significant differences between these different situations according to clinical (TD vs. DLD) and linguistic (monolingual vs. bilingual) status. Our objective being also the diagnosis, we are going to generate Roc curves and calculate the Youden indices (1950) in order to be able to define the diagnostic accuracy of our situations and tasks. We will attempt to answer the following research questions:

1. In order to aim for a diagnosis of DLD, do any of our two DA situations show differences in the performance (as per number of correct items on a posttest) of participants according to their clinical status (TD/DLD)? Are there also differences according to their linguistic status (monolingual/bilingual)?
2. What are the differences according to the type of assessments tasks (e.g., productive, or receptive) and the linguistic material (e.g., nouns, verbs, inflections)?
3. How much cues, and what kind of cues can facilitate learning in monolingual and bilingual children with TD vs. DLD in the interactive story reading situation?
4. Beyond group differences, what are the values for sensitivity and specificity for the situations and tasks show significant differences between TD and DLD children in (1)?

Method

Participants

Forty-nine children aged 5;0 to 7;11 living in the French-speaking part of Switzerland participated. They had no history or suspicion of autistic spectrum disorder, cognitive, socio-affective or hearing impairment. According to language exposure data collected via a parental report (based on Tuller, 2015), 28 children were exposed to French at least 85 % at the time of research and thus classified as monolingual. The remaining 21 bilingual children had a

minimum of 25 % exposure to Portuguese at the time of research (and no more than 15 % exposure to a third language). Children are simultaneous or successive bilinguals, but we have not dissociated these children in our research. Sixteen bilingual children had French as their current dominant language and five had Portuguese. The Table 1 below summarizes the demographic information of our participants.

In order to minimize the chicken-and-egg problem of initial diagnosis common in the field of bilingual assessment, we decided to use a strict two-step identification procedure in order to classify children as having a confirmed DLD. To meet the criteria for inclusion in our study, DLD children had to have been clinically diagnosed by a SLT and received therapy for at least three months (in order to rule out spurious initial over-identification). They also had to score clearly below age-equivalent norms on standardized language tests (see details below) that were run at the same time as our experimental tasks, equal to or below $-1.28 SD$ (according to Thordardottir et al., 2011) or percentile 10 on different subtests. Bilingual children had to fail at least one subtest in each language to confirm difficulties in French and in Portuguese related to a DLD (Peña et al., 2016) and to exclude pure lack of exposure. To be equally sure about the status of children classified as having a TD, they should never have been identified as having DLD and never have received SLT and fail no standardized pre-test.

In addition, all children had to pass an audiometric screening (audibility thresholds $\leq 20\text{dB}$ at 0.5, 1, 2 and 4 kHz) and a non-verbal intelligence screening, using the matrices subtest of the Wide-Range Intelligence Test by Glutting et al. (2000). A parental questionnaire also asked for the level of education of the parents (scales from 1 to 5: 1 = elementary school, 2 = high school, 3 = apprenticeship, 4 = high school graduation, 5 = high school). An additional twenty-three children participated but were excluded because they failed the audiometry ($n = 3$) or the non-verbal intelligence ($n = 2$) screening, did not fit the language exposure criteria ($n = 2$) or due to a technical problem at the time of testing ($n = 4$). Twelve children (3

monolinguals, 9 bilinguals) had to be excluded because they did not fit the inclusion criteria based on the language pre-tests: four children presumed to have a TD had difficulties in at least one language subtest and eight children diagnosed with DLD did not fail any subtests in at least one of the two languages, highlighting once more the difficulty of DLD diagnosis in bilinguals.

Children with and without DLD in monolingual and bilingual groups were matched individually on age ($H = 5.80, p = 0.121$), non-verbal IQ ($H = 4.20, p = 0.240$) and gender ($X^2(3, n = 49) = 1.14, p = 0.767$), and these factors did not differ significantly between the four groups. However, we found a significant difference in parental education between the bilingual DLD (mean: 2.65) and the monolingual TD children (mean: 4.30; $H = 14.27, p = .003$), the other groups did not differ significantly in this respect.

Table 1

Participant demographic information

Clinical status	Linguistic status	Number	Gender		Age (months)	Parental education	Non- verbal IQ	Bilingualism		Actual dominance	
			G	B				simultaneous	successive	FR	PT
TD	Monolingual	15	11	4	75	4.30	53.20				
DLD	Monolingual	13	8	5	76	3.10	47.30				
TD	Bilingual	11	6	5	85	3.00	64.90	7	4	9	2
DLD	Bilingual	10	7	3	81	2.60	48.00	7	3	7	3

Note. Abbreviations: G = girls ; B = boys; FR = French; PT = Portuguese.

Recruitment procedure

Participants were recruited through their schools and SLTs. Families interested in participation received an information sheet, a parental report and an informed consent form, validated by the local ethics commission (Swiss ethics Vaud; project no.: 2017-01881, authorization granted on 07.02.2018). Most of the evaluation sessions took place in the family home or in school. As part of a larger PhD project, children participated in three sessions which also included another task (crosslinguistic NWRT, see Schwob & Skoruppa, submitted).

Measures used to identify children with DLD

We have chosen three tasks often used in the clinic, evaluating different domains and language modalities that may be difficult for DLD children (Bishop et al., 2017): picture naming for the productive lexicon, oral comprehension of sentences for receptive morphosyntax, and a NWRT for phonological processing. The lexical and syntactic domains are frequently assessed to identify a DLD (Denman et al., 2017) and NWRT seems to be discriminant in the field of phonology (Gray, 2003). The tasks were taken from two French tests: Evaluation of Oral Language (“ELO”, Khomsi, 2001) for naming and comprehension and Exalang for non-word repetition (Helloin & Thibault, 2006; Thibault & Helloin, 2010) and from two Portuguese tests: The Evaluation of Oral Language test (“ALO”, Sim-Sim, 2006) for naming and comprehension and from Language and Aphasia Assessment Tests in Portuguese (“PALPA-P”, Castro et al., 2007) for NWRT. We noted that in French, children with DLD failed the tasks in the following proportions: $n = 8$ in naming, $n = 10$ in sentence comprehension and $n = 23$ in NWRT. In Portuguese: $n = 10$ in naming task, $n = 12$ in sentence comprehension

and $n = 1$ in NWRT (details on our pre-tests and classification can be found in Schwob & Skoruppa, in press).

Dynamic assessment materials

We developed two DA situations that will be detailed in the following. Each situation was run in a separate session and consisted of having the children learn new invented nouns, verbs and inflections in different phases, and then evaluating their learning via different tasks. In the interactive story reading situation, we were also interested which cues given to children would facilitate their learning. The order of presentation stimuli (non-word list A or B) and learning situations (autonomous computer game or interactive story reading) was counterbalanced across children.

Stimuli

Auditory stimuli were eight non-words per list that do not exist in French or Portuguese, in order to avoid exposure bias (Kapantzoglou et al., 2011). We took care to ensure that they were well distinct from each other and varied in length (one and two syllables) and in complexity (with or without cluster). We compose our non-words with vowels common to both languages (/a,i,o,u/) as well as early-acquired consonants common to both languages (/l, t, f, v, b, p, s, k, n/; following Chiat, 2015). Our items respect the phonotactic patterns of both French and Portuguese. The stimuli and their illustrations are presented below in Table 2. Specifically, there were four nouns to be learned per situation. Two were designed as masculine and ended with the vowel /i/ (e.g., /lati/) and two were feminine and ended with the vowel /u/ (e.g., /plifu/). Furthermore, two nouns were also represented by aliens and two were inanimate objects. Four verbs were learned per situation. These were created to represent actions that cannot be named by a verb in both French and Portuguese (e.g., turning and descending at the same time). The

verbs ended with the consonant /t/ (e.g., /flinat/). We hypothesized that these phonological cues could help children acquire these new words (Alt et al., 2004; Gray, 2003). Two verbs were transitive, and two verbs were intransitive. Finally, the children also learned sentences by associating previously learned nouns and verbs. We generated a total of eight sentences per situation. In order to test morphosyntactic learning skills, we added an inflectional rule not present in either Portuguese or French: when the gender of the subject was masculine, the verb ended with the consonant /t/ (e.g., /lati flinat/), whereas if the subject was feminine, the verb ended with the consonant /m/ (e.g., /plifu flinam/). Note that in both languages, the gender of subject can be indicated by their determiner and by adjective and noun endings but is not marked on the verb.

Table 2

Stimuli

Stimuli	Semantics	Length	¹ Complexity	Picture/description	List	Items
/lati/	Alien masculine	2	0		A	Noun
/fli/	Object	1	1		A	Noun
/plifu/	Alien feminine	2	1		A	Noun
/stabi/	Alien masculine	2	1		B	Noun
/blu/	Object	1	1		B	Noun
/otu/	Alien feminine	2	0		B	Noun
/fakat/	Intransitive action	2	0	Go down and turn at the same time	A	Verb

Stimuli	Semantics	Length	¹ Complexity	Picture/description	List	Items
/vat/	Transitive action	1	0	Move a leaf across the floor	A	Verb
/aplut/	Transitive action	2	1	Make a house disappear	A	Verb
/upit/	Intransitive action	2	0	Make a roll to the left	B	Verb
/lut/	Transitive action	1	0	Approach a magic stick	B	Verb
/flinat/	Transitive action	2	1	Enlarge a balloon from a distance	B	Verb
/lati fakat/					A	Sentence
/plifu fakam/					A	Sentence
/lati vat/					A	Sentence
/plifu vam/					A	Sentence
/lati aplut/					A	Sentence
/plifu aplum/					A	Sentence
/stabi upit/					B	Sentence
/otu/ upim/					B	Sentence
/stabi lut/					B	Sentence
/otu lum/					B	Sentence
/stabi flinat/					B	Sentence
/otu flinam/					B	Sentence

Note. Abbreviations: length = syllable length; complexity = phonological complexity.

¹Complexity was determined by assigning 0 when the non-word had a simple consonant-vowel structure and 1 point when the structure included a cluster.

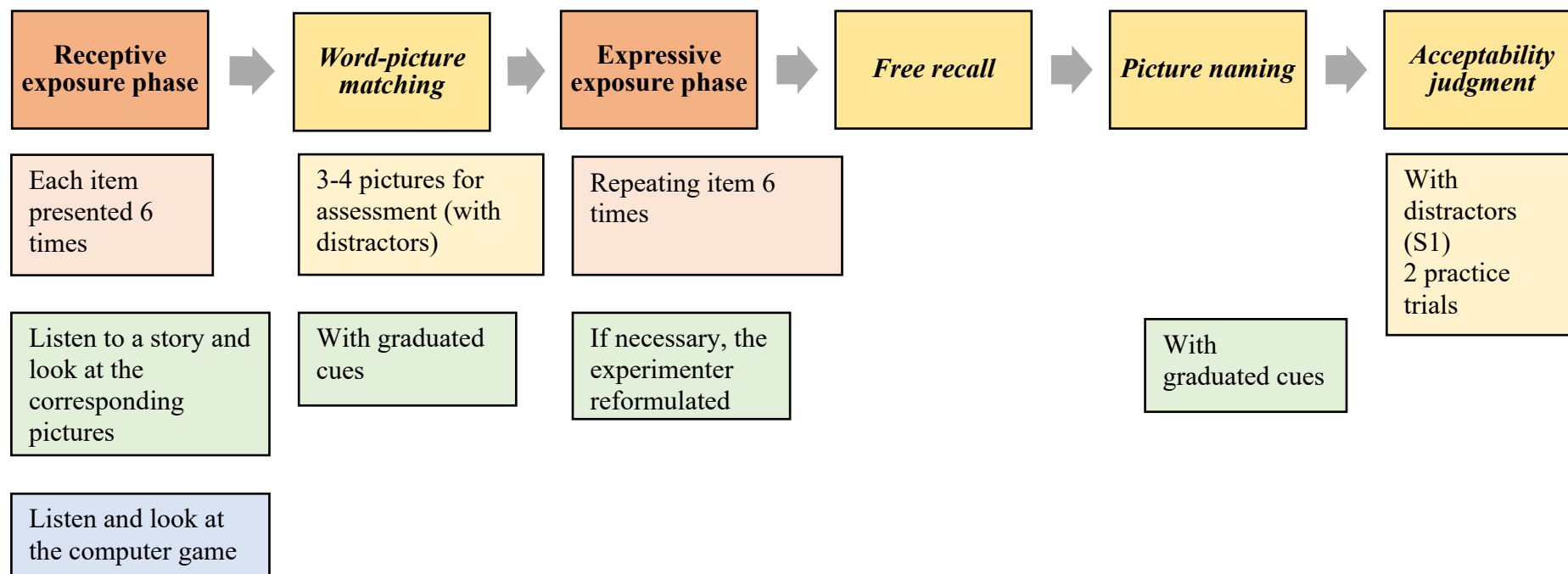
Dynamic assessment procedure

The children learned these new items in different phases common to both situations (see Figure 1 and details for situations in the following). All instructions were given in French (either pre-recorded in the case of the autonomous computer game or by the experimenter in the interactive story reading situation). During the first *receptive exposure phase*, children were presented with the new items passively. In the autonomous computer game situation, the child watched the images appear and move on the screen and heard corresponding sound files (non-words), and in the interactive reading of a story, the child listened to the story in French including the new words, while watching pictures and illustrative movements. The items were presented six times each in this first phase (following the recommended exposure in the meta-analysis of Kan & Windsor, 2010). Immediately after these first exposures, children were asked to complete a *word-picture matching task*, to evaluate if the new items were acquired receptively. This task included three pictures for the assessment of nouns and verbs (two distractors and the correct item) and four pictures for the assessment of sentences (three distractors and the correct item). The children then participated in a second *expressive exposure phase*, which consisted of repeating the new items six times each. In the interactive story reading situation, when repetition errors were detected, the experimenter reformulated them. After this, the children completed a *free recall task* and a *picture naming task*, to evaluate if they could produce the names of the items learned. Finally, an *acceptability judgment task* was proposed to evaluate the accuracy of the children's semantic and phonological representations. An audio item was presented to the child, and they had to judge whether it was indeed the item presented visually. Semantic (e.g., /lati/ as a distractor for /plifu/), phonological (e.g., /plifi/ for

/plifu/) and syntactic (for sentences; order distractor, e.g., /vam otu/ for /otu vam/; inflection distractor, e.g., /otu vat/ for /out vam/) distractors were inserted (see complete list in Supplementary Material 1). Before starting this task, there were two practice trials with unrelated distractors and corrective feedback. These different phases were repeated once for each stimulus type (one block for nouns, one for verbs and one for sentences). Each situation (with these different phases and item types) required approximately 45 minutes, including a short break. Finally, at the beginning of the following session (4- 15 days later), the recall task was repeated to test long term retention (*delay recall task*).

Figure 1

Exposure phases and assessment procedures



Note. Yellow was used to represent the assessment tasks and orange to identify the exposure phase. The specificities according to the situations are mentioned in blue for the autonomous computer game situation and in green for the interactive story reading situation.

Situation 1: autonomous computer game

Our computer game was created with Microsoft Power-point (2018) presented on a laptop in front of the child. Both the child and the experimenter were wearing headphones. The child was immersed in an alien world where they would learn a new “language”. The new items were recorded by a Swedish speaker, so that none of the children would have an advantage in terms of the intonation and accent used to produce the items. The speaker produced a tonic accent on the last syllable, as is possible in both languages. During the adventure, the children were accompanied by an alien teacher, Mrs. Flamap, who introduces them to the different items on the screen and rewards them with stars at the end of each learning phase. The experimenter was only present to record the child’s receptive responses via mouse-click and to encourage the child to proceed, but no specific assistance was provided. Some corrective feedbacks were given by the computer (e.g., in the acceptability judgment tasks). This situation has the advantage of involving only minimal French (for the simple instructions), thus avoiding linguistic bias as much as possible.

Situation 2: interactive story reading

The other situation was a more interactive situation between the child and the experimenter, even if the pictures were also presented on a computer screen. The experimenter had an active role (Hasson & Joffe, 2007) performing the live reading of three stories (one for nouns, verbs and sentences) in French. The three stories formed a logical sequence (aliens who visit other planets and learn magic gymnastics figures; see an example in Supplementary Material 2). The child and the adult interacted during the narrations and the adult provided graduated cues during the assessment tasks (Campione & Brown, 1987; see hierarchy of mediation below). Stories are told in French, into which the experimenter inserted the new items. Each time a new non-word was presented, the experimenter accentuated it verbally and

pointed to the corresponding item on the screen (Weismer & Hesketh, 1993). The stories also offer semantic cues (three cues per non-word, see Supplementary Material 3 for examples) to facilitate learning. For this situation, the experimenter was the same for all the children (the first author, who is a trained SLT).

Mediation hierarchy

During the word-picture matching and naming tasks of the interactive story reading situation, the experimenter provided graduated scaffolding (Campione & Brown, 1987) for the child to observe their learning potential when assistance was provided. The support provided is similar to that in SLT (de Weck & Marro, 2010) and could be more or less numerous and vary in terms of complexity of cues (see Supplementary Material 3 for all cues). The aids provided start with very general cues and then become more and more specific and detailed. In the following, we will describe the level of cues for the different tasks.

During ***word-picture matching***, the experimenter would begin by asking the child again to point (cue 1) and encouraging him on his performance (cue 2). Then, semantic cues (cue 3) (e.g., *c'est un extraterrestre, il est jaune, il a des cheveux...*, 'he is yellow, he is tall, he has hair...' for /stabi/) were provided so that the child could find the correct item. For the sentence-picture matching, the experimenter proceeded to give these semantic cues first on the noun and then on the verb.

In the ***picture naming task***, the experimenter also began by repeating the question (cue 1) and provide encouragement (cue 2). Next came semantic cues (cue 3) (e.g., *c'est un extraterrestre jaune avec des cheveux*, 'it's a yellow alien with hair'), and then a cloze cue with mounting intonation, consisting in the determiner (*le/la ... ?*, 'the?') for nouns and the pronoun (*il/elle ... ?*, 'he/she?') for the verb (cue 4). Finally, the onset of the item was given (e.g., *c'est un l...?*, 'this is a l....?', for /lati/) (cue 5). When naming sentences, in addition to

repeating the request, the experimenter also asked to name either the noun or the verb if possible (e.g., *tu te souviens de son nom? Qu'est-ce qu'il fait?* 'do you remember his name? or what he does?'). The scaffolding then went progressively through the sentence, first with the noun determiner (e.g., *c'est un ...*, 'it is a...') (cue 4), then the noun determiner with the first phoneme of the noun (e.g., *un l...* 'a l...') (cue 5), the noun plus the onset of the verb (e.g., *un lati fl...*, 'a lati fl...') (cue 6) and finally the noun and the verb, but without inflection (e.g., *un lati flina...*, 'a lati flina...') (cue 7). In any case, when the child could not produce the requested items despite the different levels of support, the experimenter would formulate the correct answer (e.g., *un lati flinat*, 'a lati flinat') for the child to repeat.

Scoring procedure

We calculated the number of correct items by phase and item types assigning one point when the item was correctly produced or designated, and zero when there were one or more errors. In the interactive reading situation, a score before and after scaffolding was calculated. The number and levels of cues were also counted.

Inter-rater reliability

Inter-judge agreement on 10 % of our sample was calculated on several categories: correct item (0/1) in the different situations and tasks before and after scaffolding, and the number of mediations. Inter-judge agreement indices were then calculated, using the statistical software SPSS (Version 27.0, IBM Corp., 2020) with Cohen's Kappa test for scoring in terms of correct items and with the intra-class correlation coefficient (ICC) for scoring in terms of number of scaffoldings. Cohen's Kappa indicates a reliable inter-rater agreement of .98, a near perfect agreement (Landis & Koch, 1977) and ICC is .95, indicating good to excellent inter-rater reliability (Koo & Li, 2016).

Statistical analyses

For our descriptive statistics we have made violin plots including box plots via the R software (R Development Core Team, 2019). We would have liked to perform multivariate statistical analyses but unfortunately, our data did not respect the required assumptions (e.g., normal distribution, large sample size; Field, 2018). Thus, we performed binomial generalized linear models using SPSS version 27 (IBM Corp., 2020), whose purpose is to quantify the relationship between several independent variables (e.g., clinical, and linguistic status, age, non-verbal IQ) and a dependent variable that follows a binomial distribution. We also made sure to control the education level of the parents and the order of the tasks that could influence our results. We reported effect sizes for clinical and linguistic status outcomes by using the Cohen's d using the unstandardized regression coefficient (B), our groups sample size (TD-DLD / mon-bil) and the standard deviation of our dependent variable. A Cohen's d effect size of .80 or above is considered large, whereas effect sizes of .20 are considered small, and those of .50 are considered medium sized. Based on our initial sample size of 72 children, we had calculated a statistical power of $>.80$, but after all exclusions, the post-hoc statistical power based on the size of our groups as well as the effect sizes, we obtain an average statistical power of .69. Therefore, our results must be interpreted with caution, and it would be appropriate to reproduce such a study with a larger sample size (Cohen, 1988), although it is equal or superior to previous studies in the field (see Orellana et al., 2019). Since we had missing data on parental education of monolingual children ($n = 10$), we performed multiple imputations (Donders et al., 2006) to generate these observations and control for them in our statistical model. We also generated Roc curves (Dunn, 2014) from which we derived sensitivity and specificity values, which we interpreted according to Plante & Vance (1994). More precisely, with values of 80 % and above, the test is considered accurate. Finally, Youden's indices (1950) were also calculated for these situations, via the formula: "sensitivity + specificity – 1", in order to

determine the optimal threshold to distinguish between participants with and without DLD (best values would be approaching 1).

Results and discussion

Factors explaining the number of correct items in the DA

We performed our analyses by situations (autonomous computer game, interactive story reading), tasks (e.g., naming, recall) and item type (nouns, verbs and sentences). All inferential statistics were in Supplementary Material 4 for the computer game and the interactive story reading situation.

Autonomous computer game situation

Using generalized linear models, we note an effect of clinical status with an advantage for the TD children over DLD children on the word-picture matching task for nouns ($W^2 = 4.59$, $p = .032$) and for sentences ($W^2 = 4.97$, $p = .026$) and in the acceptability judgment for sentences ($W^2 = 5.19$, $p = .023$) in the autonomous computer game situation. The effect size associated with these comparisons are medium and large. An effect of language status was found for the word-picture matching for nouns ($W^2 = 6.34$, $p = .012$) and the acceptability judgment for verbs ($W^2 = 4.04$, $p = .044$) with an advantage for bilingual children vs. monolingual children with medium effect sizes. We also observe an effect of language status in the recall task for verbs ($W^2 = 9.17$, $p = .002$) with, in this task, an advantage of monolingual children with a small effect size. The remaining tasks do not show significant differences between children with TD and DLD or bilingual and monolingual children (see graphics in the Supplementary Material 6). In the descriptive statistics (Table 3 below), we note the low number of responses produced by the children in the productive tasks, especially when the items are verbs or sentences. In these

production tasks, we find an effect of age with an advantage for older children in the noun and verb recall (noun: $W^2 = 6.21, p = 0.013$; verb: $W^2 = 27.63, p < .001$), and noun naming task ($W^2 = 7.11, p = .008$). The non-verbal IQ had only an influence on the recall for verbs ($W^2 = 11.13, p = .001$) with an advantage for children with high non-verbal IQ.

Interactive story reading situation

In this situation, using also generalized linear models, we find an effect of clinical status in a lot of tasks (see graphical representations are in Supplementary Material 6). We find an advantage for TD children in the word-picture matching of verbs ($W^2 = 5.01, p = .025$) and sentences ($W^2 = 12.47, p < .001$), in recall of any item (nouns: $W^2 = 9.15, p = .002$; verbs: $W^2 = 32.66, p < .001$; sentences: $W^2 = 7.98, p = .005$), in naming of verbs ($W^2 = 18.71, p < .001$) and sentences ($W^2 = 4.79, p = .029$) and in all acceptability judgment tasks (nouns: $W^2 = 13.10, p < .001$; verbs: $W^2 = 31.25, p < .001$; sentences: $W^2 = 9.07, p = .003$). Most effect sizes are medium or large (details in Supplementary Material 4). Few items were retained by the children in the delayed recall (see Table 3) and no significant differences were observed between the groups of children. Linguistic status has an influence in the verb word-picture matching ($W^2 = 5.72, p = .017$), with an advantage for monolingual children and in the noun acceptability judgment task, with an advantage for bilingual children ($W^2 = 6.90, p = .009$). It is also in this latter task that we see an interaction between the clinical and linguistic status ($W^2 = 5.24, p = .022$): monolingual children with DLD have the most difficulty here. An effect of linguistic status is also observed in the verb recall ($W^2 = 4.41, p = .036$) and noun and verb naming task (nouns: $W^2 = 8.30, p = .004$; verbs: $W^2 = 3.90, p = .048$). In these productive tasks, we observe an advantage of monolingual over bilingual children. Effect sizes for language status are small or medium except for the acceptability judgment task for nouns and the verb recall, which are large. As in the previous situation, we find an effect of age in the productive tasks. We see the

age effect in the recall tasks (nouns: $W^2 = 28.15, p < .001$; verbs: $W^2 = 25.59, p < .001$; sentences: $W^2 = 8.39, p = .004$), in two of the naming tasks (nouns: $W^2 = 17.73, p < .001$; verbs: $W^2 = 9.45, p = .002$) and in delayed recall of nouns ($W^2 = 4.31, p = .038$). Older children are able to produce more words or sentences on these assessments. An age effect was also seen in the acceptability judgment task involving nouns ($W^2 = 4.72, p = .030$) and in word-picture matching for verbs ($W^2 = 8.46, p = .004$) and for sentences ($W^2 = 8.16, p = .004$). The children's non-verbal IQ appeared to be influential only in the noun and verb word-picture matching tasks (nouns: $W^2 = 4.62, p = 0.031$; verbs: $W^2 = 8.07, p = .004$). Children with higher non-verbal IQ perform better on these tasks.

Table 3

Descriptive analyses

Situations	Tasks	Items	TD MON	TD BIL	DLD MON	DLD BIL
Autonomous DA	Word-picture matching	Nouns	2.27 (0.88)	2.82(0.60)	1.62(1.04)	2.40(0.96)
Autonomous DA	Word-picture matching	Verbs	1.67(1.17)	2.55(0.68)	1.38(1.04)	1.90(1.28)
Autonomous DA	Word-picture matching	Sentences	2.40(1.50)	3.27(1.95)	1.92(1.11)	2.10(0.73)
Autonomous DA	Recall	Nouns	1.00(1.00)	1.18(0.75)	0.54(0.96)	0.90(1.28)
Autonomous DA	Recall	Verbs	0.87(0.91)	0.64(0.80)	0.46(0.77)	0.40(0.69)
Autonomous DA	Recall	Sentences	0.27(0.59)	0.09(0.30)	0.00(0.00)	0.10(0.82)
Autonomous DA	Naming	Nouns	0.80(0.77)	1.36(1.02)	0.31(0.48)	0.70(0.82)
Autonomous DA	Naming	Verbs	0.53(0.64)	0.36(0.50)	0.15(0.37)	0.30(0.46)
Autonomous DA	Naming	Sentences	0.13(0.35)	0.00(0.00)	0.00(0.00)	0.00(0.00)
Autonomous DA	Acceptability judgment	Nouns	6.33(1.67)	8.27(1.00)	5.54(2.84)	6.50(1.35)
Autonomous DA	Acceptability judgment	Verbs	5.73(1.22)	6.18(1.07)	5.15(2.07)	5.90(1.66)
Autonomous DA	Acceptability judgment	Sentences	9.40(1.68)	9.64(1.69)	7.92(1.70)	8.30(2.00)
Autonomous DA	Delayed recall	Nouns	0.20(0.56)	0.36(0.67)	0.15(0.37)	0.30(0.48)

Situations	Tasks	Items	TD MON	TD BIL	DLD MON	DLD BIL
Autonomous DA	Delayed recall	Verbs	0.00(0.00)	0.09(0.30)	0.00(0.00)	0.00(0.00)
Interactive DA	Word-picture matching	Nouns	2.53(0.64)	2.09(0.70)	1.77(0.59)	2.30(0.82)
Interactive DA	Word-picture matching	Verbs	2.53(0.51)	2.36(0.92)	2.15(0.89)	2.20(0.63)
Interactive DA	Word-picture matching	Sentences	4.47(1.30)	5.27(0.90)	3.08(1.38)	3.10(1.59)
Interactive DA	Recall	Nouns	1.20(1.14)	1.36(0.80)	0.46(0.66)	0.40(0.51)
Interactive DA	Recall	Verbs	1.13(0.64)	1.18(0.60)	0.38(0.50)	0.10(0.31)
Interactive DA	Recall	Sentences	0.73(1.10)	0.45(0.68)	0.00(0.00)	0.00(0.00)
Interactive DA	Naming	Nouns	1.20(1.20)	0.64(0.80)	0.46(0.66)	0.40(0.51)
Interactive DA	Naming	Verbs	1.13(0.83)	1.09(0.53)	0.38(0.50)	0.30(0.48)
Interactive DA	Naming	Sentences	0.53(0.91)	0.36(0.50)	0.00(0.00)	0.00(0.00)
Interactive DA	Acceptability judgment	Nouns	7.67(1.34)	8.36(0.50)	5.69(1.65)	7.60(0.96)
Interactive DA	Acceptability judgment	Verbs	6.80(1.01)	7.55(0.93)	5.38(1.66)	4.70(1.05)
Interactive DA	Acceptability judgment	Sentences	9.60(2.44)	10.82(1.94)	8.00(1.68)	8.20(1.81)
Interactive DA	Delayed recall	Nouns	0.60(1.54)	0.27(0.64)	0.08(0.27)	0.00(0.00)
Interactive DA	Delayed recall	Verbs	0.40(0.82)	0.09(0.30)	0.08(0.27)	0.10(0.31)

Situations	Tasks	Items	TD MON	TD BIL	DLD MON	DLD BIL
Interactive DA improvement	Word-picture matching	Nouns	1.00(0.63)	0.88(0.64)	1.33(0.49)	1.40(0.54)
Interactive DA improvement	Word-picture matching	Verbs	1.00(0.00)	1.50(0.57)	1.43(0.53)	1.14(0.37)
Interactive DA improvement	Word-picture matching	Sentences	2.00(1.09)	1.60(0.54)	2.62(1.26)	2.50(1.50)
Interactive DA improvement	Naming	Nouns	0.00(0.00)	0.09(0.30)	0.08(0.27)	0.00(0.00)
Interactive DA improvement	Naming	Verbs	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
Interactive DA improvement	Naming	Sentences	0.13(0.51)	0.18(0.40)	0.00(0.00)	0.00(0.00)
Number of cues	Word-picture matching	Nouns	0.47(0.64)	0.82(0.60)	1.31(0.75)	0.70(0.82)
Number of cues	Word-picture matching	Verbs	0.60(0.82)	0.64(0.92)	1.77(2.24)	1.00(0.81)
Number of cues	Word-picture matching	Sentences	3.33(3.45)	1.45(2.10)	7.77(5.77)	7.50(6.07)

Situations	Tasks	Items	TD MON	TD BIL	DLD MON	DLD BIL
Number of cues	Naming	Nouns	5.80(4.14)	7.82(4.19)	8.77(2.58)	8.00(1.94)
Number of cues	Naming	Verbs	5.13(2.77)	5.27(1.90)	6.62(2.43)	7.30(1.56)
Number of cues	Naming	Sentences	23.40(7.10)	24.27(6.84)	30.62(4.97)	29.80(6.84)

Note. Abbreviations: DA = dynamic assessment; Autonomous DA = autonomous computer game situation; Interactive DA = interactive story reading situation; Interactive DA improvement = improvement with cues in interactive story reading situation.

$N = 49$ ($n = 23$ children with a DLD and 26 children with a TD / $n = 21$ bilingual children and 28 monolingual children).

We chose to present the means and standard deviations by group because the medians by group did not allow us to appreciate the differences between the groups since all the medians were between 0 and 1 with few distinctions.

Discussion

We found that both DA situations showed an effect of clinical status with better performance of TD than DLD children, similar to previous research (e.g., Kapantzoglou et al., 2011). In the autonomous computer game situation, we only obtain differences between these two groups on three tasks with medium or large effect size: the word-picture matching involving nouns and sentences and in the acceptability judgment of sentences. In the interactive story reading situation, we obtain an effect of clinical status in more tasks and also involving different item types (e.g., recall task involving nouns, verbs and sentences) with mostly medium and large effect sizes. We found some rare influences of linguistic status with an advantage for monolingual children in the productive tasks (e.g., recall verb and noun and verb naming) and an advantage for bilingual children in the noun acceptability judgment task. It is possible that bilingual children are more comfortable in this task since they seem to have better metalinguistic abilities (Adescope et al., 2010), but our effect sizes associated with language status are small or medium. Overall, our results so far seem to favor the interactive story reading situation, since we find few and small language status effects in both directions even in the autonomous computer game (suggesting that the latter situation is not superior in terms of linguistic biases), but more and larger differences according to clinical status in the interactive story reading situation.

We find in both situations, that few words were produced by the children, as already described by Skoruppa (2019). The productive assessment tasks have already been characterized as very complex by different authors in the literature (e.g., Gray, 2003), and even the production training (additional repetition) and the easier free recall we devised here (since in this task, the child does not need to have made the link between the semantic referent and the linguistic item) does not seem to help. Furthermore, we also note that the productive tasks are less subject to the effect of clinical status, especially in the autonomous computer game

situation, but be more subject to the effect of linguistic status and age, with an advantage for monolingual children and older children. For example, in the noun naming of the interactive story reading situation, the bilingual children with DLD had the most difficulties. Alt et al. (2013) have also found this result but using items that respected the phonology of only one of the children's languages. We purposely chose items that respected the phonotactic rules of both languages of our bilingual children so that they would not be disadvantaged in the production tasks of our research. Nevertheless, an effect of language status is still noticeable in our results on the productive tasks. Children seem to perform better in the receptive tasks in our study, and we note greater differences with respect to clinical status, and less to language status, which is consistent with the literature (e.g., Kan & Windsor 2010). Thus, it seems that receptive tasks are particularly interesting to avoid differences between monolingual and bilingual children as much as possible and to associate task success with the clinical status of the participants.

With respect to the linguistic materials, we observe that verbs and sentences also seem to be more difficult to produce with lower averages per group. We note differences in the tasks including verbs only in the interactive story reading situation. However, overall, we do not find any clear advantage to privilege one or the other of the materials. Rather, it would be a combination of several tasks including different items that could be effective for diagnostic accuracy. We will discuss this topic later when we talk about the discriminative value.

Scaffolding

When graduated scaffolding is provided in the interactive story reading situation, we note improvements in performance for children in both groups (see graphics in Supplementary Material 7). In the word-picture matching tasks, we note an effect of clinical status by again running generalized linear models. Indeed, significant differences between TD vs. DLD in the task including nouns ($W^2 = 5.70, p = .017$) and a large effect size. In this task, DLD children

made more progress following cues. In the word-picture matching for verbs, we find an effect of linguistic status ($W^2 = 5.25, p = .022$), age ($W^2 = 5.08, p = .024$) and non-verbal IQ ($W^2 = 5.14, p = .023$). We also note an interaction between linguistic and clinical status in the word-picture matching task for verbs ($W^2 = 9.60, p = .002$), bilingual children with TD made more progress in this task.

Regarding the number of cues provided by the experimenter, we observe a tendency that more cues are given to DLD children in all tasks (see violin plots in Supplementary Material 8 and statistics in Supplementary Material 9). We observe an effect of the clinical status in almost all tasks of our generalized linear model. Specifically, children with DLD require more support than TD children in the noun ($W^2 = 5.25, p = .022$), verb ($W^2 = 4.17, p = .041$) and sentence word-picture matchings tasks ($W^2 = 10.85, p = .001$) as well as in verb ($W^2 = 6.68, p = .010$) and sentence naming tasks ($W^2 = 11.55, p < .001$). The effect sizes are only large for tasks that include sentences (see Supplementary Material 9). Production tasks, in this case naming, are subject to linguistic status. We note that bilingual children require more support than monolingual children to produce nouns ($W^2 = 6.99, p = .008$). In this production task, we also observe an effect of age, the necessary support reduces with age ($W^2 = 18.44, p < .001$). When we look at the type of support provided, we note an effect of clinical status in the sentence word-picture matching tasks ($W^2 = 8.15, p = .004$) and in the noun naming task ($W^2 = 4.99, p = .025$). The type of support provided was more complex for DLD children than for TD children (the level of cues is described in the Method). We observe an effect of linguistic status in the noun naming task ($W^2 = 4.73, p = .030$) and an interaction between clinical status and linguistic status on the word-picture matching task for nouns ($W^2 = 7.53, p = .006$). Monolingual children with DLD require a higher level of assistance than other children. We observe an effect of age in the verb-picture matching ($W^2 = 5.23, p = .022$), with older children requiring less cues.

In sum, we observe that with cues provided in the interactive story reading situation, children can make progress. We note that children with DLD often make more progress because they had lower scores without helps and are used to aids in the SLT context (de Weck & Marro, 2010). We found also it was clinically useful to observe the amount and type of cues provided to children in the situation of interactive story reading. We have seen that DLD children require a greater amount of assistance and more complex aids than TD children. Several research studies have also attested to the greater amount of scaffolding needed from DLD children (e.g., Hasson et al., 2013; Maragkaki & Hessels, 2017). It is also clinically interesting to note that bilingual children require more cues in productive tasks.

Discriminative value

Going beyond group differences, we examine the discriminative values of the situations and tasks on which TD and DLD children showed significant differences (see above). We generated Roc curves (Dunn, 2014; see Supplementary Material 10) and calculated Youden's indices for all different tasks. We combined the score obtained by the children (number of correct new words) per situation (using z-scores for three tasks for the autonomous computer game situation and ten for the interactive story reading situation). With this combined score, we obtained good percentages by referring to Plante & Vance (1994) (91 % of sensitivity and 85 % of specificity, see Table 4) for the tasks from interactive story reading situation. Like previous studies (e.g., Orellana et al., 2019), our situation and tasks show adequate classification between TD and DLD children. However, the computer game misclassifies eleven children with a TD (only 57 % of specificity), probably due to the overall difficulty of the activity, which is conducted with less interaction and without graduated helps. We also try to test the diagnostic accuracy of the number of cues gives in the interaction situation. We find that these percentages do not exceed the thresholds required by Plante & Vance (1994) either.

Interestingly, a hypothetical shortened version of the interactive DA with the two most promising tasks (recall and acceptability judgment) obtains very good values and the best Youden index. With this option, we note that only three children with DLD (87 % of sensitivity) and two children with TD (92 % of specificity) would be misclassified.

Table 4

Discriminative values

Situations	Sensitivity	Specificity	Threshold	Youden Indices
Autonomous DA	78 %	57 %	1.15	.36
Interactive DA	91 %	85 %	0.01	.76
Shortened Interactive DA	87 %	92 %	-0.22	.79
Number of cues	87 %	61 %	7.50	.49

Note. Abbreviations: DA = dynamic assessment, Autonomous DA = autonomous computer game situation, Interactive DA = interactive story reading situation.

$N = 49$ ($n = 23$ children with a DLD and 26 children with a TD / $n = 21$ bilingual children and 28 monolingual children).

General discussion and conclusion

We were able to show that whatever the proposed learning situation, TD and DLD children can learn new things, which encourages a diversity of learning contexts in educational structures or at home. We were also able to highlight that the interaction and the cues gives by the adult is facilitating for both groups of children, which encourages the active participation of the experimenter during the assessment phases to know the learning potential of the children. We were also able to show many significant differences between TD vs. DLD children in our DA situations. DLD children have more difficulty learning and require more

help from the experimenter. Differences appear according to the learning modality (receptive or productive), with production being more difficult for all children, especially young children with DLD and bilinguals. This encourages the use of receptive tasks in dynamic assessment. Few items were retained by the children, especially when it came to verbs and sentences, which is consistent with the existing literature that nouns seem to be more easily learned (e.g., Alt et al., 2004). However, the analysis of one linguistic domain (lexicon, morphosyntax) does not seem to be sufficient, it is their combination that proved discriminating.

Language status effects were infrequent in our data, confirming that DA clearly seems less prone to language bias than static tests (Maragkaki & Hessels, 2017). This is an important argument for its use in the SLT clinic. Furthermore, what is different with static tests is that when influences of language status were found in the interactive story reading situation, they sometimes highlighted an advantage of bilingual over monolingual children (e.g., acceptability judgment). Productive tasks, on the other hand, are more similar to static tests and sometimes benefit monolingual children. The use of non-words (as opposed to real French words) certainly allowed us to minimize potential linguistic biases but did not allow us to completely eradicate them. A limitation of our work is that we used standardized pre-tests with monolingual norms in both languages of the child to form our initial diagnostic groups. Of course, further research will be necessary to confirm our results, ideally in a larger prospective study. It would also be necessary to replicate this study with other language combinations to see if the results obtained here would be identical.

We obtained good to excellent percentages of sensitivity and specificity (Plante & Vance, 1994) for one of our situations. In our initial diagnostic goal, we can therefore affirm that our interactive reading situation with graduated live cues can be a clinically interesting option to identify monolingual and bilingual children with DLD. Furthermore, since there was no need for pre-tests, the time required to perform our assessment would not be excessive (about

one 45-minute session which could be shortened to 30 minutes when running only the most discriminant tasks) and thus encourages its use as well (Hasson & Joffe, 2007). We assume that this DA situation may be of interest to language test developers. Apart from, the pre-standardized DAPPLE test (Hasson, 2017; Hasson et al., 2013), there are, to our knowledge, no tests or batteries addressing dynamic assessment in the SLT clinic, leaving room for new tools to develop. It might also be interesting to combine our interaction DA situation with other promising tasks such as a NWRT (e.g., Schwob et al., 2021).

Acknowledgements

We would like to thank all participants and their families for taking part in this research, as well as colleagues, students and members of institutions who allowed us to contact the families. Parts of a prior version of this text were translated from French by the free version of the software Deep-L.

References

- Adescope, O., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, 80(2), 207-245. <https://10.3102/0034654310368803>
- Alt., M., Arizmendi, G. D., Gray, S., Hogan, T., Green, S., & Cowan, N. (2019). Novel Word Learning in children who are bilingual: comparison to monolingual peers. *Journal of speech, language, and hearing research*, 62(7), 2332-2360. https://doi.org/10.1044/2019_JSLHR-L-18-0009
- Alt, M., Meyers, C., & Figueroa, C. (2013). Factors that influence fast mapping in children exposed to Spanish and English. *Journal of Speech, Language, and Hearing Research*, 56(4), 1237-1248. [https://doi:10.1044/1092-4388\(2012/11-0092\)](https://doi:10.1044/1092-4388(2012/11-0092))
- Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 49(5), 941–954. [https://doi:10.1044/1092-4388\(2006/068\)](https://doi:10.1044/1092-4388(2006/068))
- Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research*, 47(2), 407–420. [https://doi:10.1044/1092-4388\(2004/033\)](https://doi:10.1044/1092-4388(2004/033))
- Armon-Lotem, S. (2012). Introduction: Bilingual children with SLI – the nature of the problem*. *Bilingualism: Language and Cognition*, 15(1), 1-4. <https://doi.org/10.1017/S1366728911000599>
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068-1080. <https://doi.org/10.1111/jcpp.12721>

- Brackenbury, T., & Fey, M. (2003). Quick incidental verb learning in 4. Year.Olds: Identification and Generalization. *Journal of Speech, Language, and Hearing Research*, 46(2), 313-327. [https://doi:10.1044/1092-4388\(2003/025\)](https://doi:10.1044/1092-4388(2003/025))
- Budoff, M. (1987). Measures for assessing learning potential. In C. S. Lidz (Ed.), *Dynamic assessment. An interactional approach to evaluating learning potential* (pp. 173-195). Guilford Press.
- Camilleri, B., & Law, J. (2007). Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology*, 9(4), 312-322. <https://doi.org/10.1080/14417040701624474>
- Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82–115). Guilford Press.
- Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R., & Coltheart, M. (2007). *Provas de avaliação da linguagem e da afasia em português*. CEGOC.
- Chiat, S. (2015). Non-Word Repetition. In S. Armon-Lotem, K. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children* (pp. 95-122). Multilingual Matters.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Publishers.
- De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, 40(1), 119-151. <https://10.1017/S0272263116000486>
- Denman, D., Speyer, R., Munro, N., Pearce, W., Chen, Y. W., & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4-12 years: A systematic review. *Frontiers in Psychology*, 1515(8), 1-28. <https://doi:10.3389/fpsyg.2017.01515>

- De Weck, G., & Marro, P. (2010). *Les troubles du langage chez l'enfant. Description et évaluation*. Elsevier Masson.
- Donders, A., Van der Heijden, G., Stijnen, T., & Moons, K. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 59(10), 1087-1091. <https://doi.org/10.1016/j.jclinepi.2006.01.014>
- Estes, K. G., Evans, J., & Else-Quest, N. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177-195. <https://doi:1092-4388/07/5001-0177>
- European Commission. (2012). Europeans and their languages. *Special Eurobarometer 386*. Brussels.
- Federal Statistical Office (2020, February 27). *Most frequent main languages*. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/langues-religions/langues.assetdetail.11947533.html>
- Feuerstein, R., Rand, Y., & Hoffman, M.B. (1979). *The dynamic assessment of retarded performers. The Learning Potential Assessment Device. Theory, instruments, and techniques*. University Park Press.
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed). Sage Edge.
- Glutting, J., Adams, W., & Sheslow, D. (2000). *Wide range intelligence test*. Wide Range.
- Gray, S. (2003). Word-Learning by Preschoolers with Specific Language Impairment: What Predicts Success? *Journal of Speech, Language, and Hearing Research*, 46(1), 56-57. [https://doi:10.1044/1092-4388\(2004/083\)](https://doi:10.1044/1092-4388(2004/083))
- Hasson, N. (2017). *The dynamic assessment of language learning*. Routledge.
- 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:10.1177/0265659007072142>

- Hasson, N., Camilleri, B., Jones, C., Smith, J., & Dodd, B. (2013). Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Language Teaching and Therapy*, 29(1), 57–75. <https://doi:10.1177/0265659012459526>
- Helloin, M. C., & Thibault, M. P. (2006). *Exalang 3-6*. Happyneuron.
- Hessels, M. G. P. (1997). Low IQ but high learning potential: Why Zeyneb and Moussa do not belong in special education. *Educational and Child Psychology*, 14, 121-136.
- Hessels-Schlatter, C., & Hessels, M. G. P. (2009). Clarifying some issues in dynamic assessment: Comments on Karpov and Tzuriel. *Journal of Cognitive Education and Psychology*, 8(3), 246-251. <https://doi.org/10.1891/1945-8959.8.3.246>
- Hessels, M. G. P., & Hessels-Schlatter, C. (2010). L'apport des tests d'apprentissage dans l'évaluation des capacités intellectuelles des élèves en difficultés scolaires. In M. G. P. Hessels & C. Hessels- Schlatter (Eds.). *Évaluation et intervention auprès d'élèves en difficultés* (pp. 5-35). Peter Lang.
- IBM Corp. (2020). *IBM SPSS Statistics for Macintosh* (Version 27.0) [Computer software]. IBM Corp.
- Kan, P., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 53(3), 739–756. [https://doi:10.1044/1092-4388\(2009/08-0248\)](https://doi:10.1044/1092-4388(2009/08-0248))
- Kapantzoglou, M., Restrepo, A., & Thompson, M. (2012). Dynamic Assessment of Word Learning Skills: Identifying Language Impairment in Bilingual Children. *Language, Speech, and Hearing Services in School*, 43(1), 81-96. [https://doi:10.1044/0161-1461\(2011/10-0095\)](https://doi:10.1044/0161-1461(2011/10-0095))

- Kaushanskaya, M., Gross, M., Sheena, E., & Roman R. (2017). Novel morpheme learning in monolingual and bilingual children. *American Journal of Speech and Language Pathology*, 26(2), 301-315. [https://doi:10.1044/2016_AJSLP-16-0011](https://doi.org/10.1044/2016_AJSLP-16-0011)
- Keller, K., Troesch, L., & Grob, A. (2015). A large receptive-expressive gap in bilingual children. *Frontiers in Psychology*, 6 (1284), 1-9. <https://doi.org/10.3389/fpsyg.2015.01284>
- Khomsi, A. (2001). Évaluation du langage oral. ECPA
- Kohnert K., & Danahy K. (2007). Young L2 learners' performance on a novel morpheme task. *Clinical Linguistics & Phonetics*, 21(7), 557–569. [https://doi:10.1080/02699200701374231](https://doi.org/10.1080/02699200701374231)
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), 155-163.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Lidz, C. S. (1987). Historical perspectives. In C. S. Lidz (Ed.), *Dynamic assessment. An interactional approach to evaluating learning potential* (pp. 3-35). Guilford Press.
- Lytle, S. R., Garcia-Sierra, A., & Kuhl, P. K (2018). Two are better than one: Infant language learning from video improves in the presence of peers. *Proceedings of the National Academy of Sciences of the United States of America*, 115(40), 9859-9866. www.pnas.org/cgi/doi/10.1073/pnas.1611621115
- Maragkaki, I., & Hessels, M. G. P. (2017). A pilot study of dynamic assessment of vocabulary in German for bilingual preschoolers in Switzerland. *Journal of Studies in Education*, 7(1), 32-49. [https://doi:10.5296/jse.v7i1.10392](https://doi.org/10.5296/jse.v7i1.10392)

- Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*, 48(2), 439–458. [https://doi::10.1044/1092-4388\(2005/030\)](https://doi::10.1044/1092-4388(2005/030))
- Nicoladis, E., & Genesee, F. (1997). Language development in preschool bilingual children. *Journal of speech-language pathology and audiology*, 21(4). 258-270. <https://doi.org/10.7939/R3348GW0N>
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, 57(11), 1247-1257. <https://doi.org/10.1111/jcpp.12573>
- Orellana, C., Wada, R., & Gillam, R. (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.
- Paradis, J. (2019). English second language acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors. In M. Brown et B. Dailey (ed.), *Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 11-26). Cascadilla Press. <http://www.lingref.com/buclid/43/BUCLD43-02.pdf>
- Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development & disorders: A handbook on bilingualism and second language learning*. Brookes.
- Peña, E. D., Reséndiz, M., & Gillam R. (2007). The role of clinical judgements of modifiability in the diagnosis of language impairment. *Advances in Speech-Language Pathology*, 9(4), 332-345. <https://doi:10.1080/14417040701413738>

- Peña, E. D., Gillam, R., & Bedore, L. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *Journal of Speech, Language, and Hearing Research*, 57(6), 2208-3330. https://doi:10.1044/2014_JSLHR-L-13-0151
- Peña, E. D., Bedore, L., & Kester, E. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Communication disorders*, 51(2), 192-202. <https://doi:10.1111/1460-6984.12199>
- Petersen, D. B, Tonn, P., Spencer, D., Foster, M. E. (2020). The classification accuracy of a dynamic assessment of inferential word learning for bilingual English/Spanish-speaking school-age children. *Language, Speech, and Hearing Services in Schools*, 51(1), 144-164. https://doi:10.1044/2019_LSHSS-18-0129
- Plante, E., & Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools*, 25(1), 15-24. <https://doi.org/10.1044/0161-1461.2501.15>
- Power-Point (2018). *Power-point for Macintosh*. [Computer software]. Microsoft.
- R Development Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Version 4.0.3 for Macintosh. [Computer software]. Vienna, Austria. <https://www.R-project.org/>
- Rey, A. (1934). D'un procédé pour évaluer l'éducabilité. *Archives de Psychologie*, 24, 297-337.
- Roseberry, C. A., & Connell, P. J. (1991). The use of an invented language rule in the differentiation of normal and language-impaired Spanish-speaking children. *Journal of Speech & Hearing Research*, 34(3), 596-603. <https://doi.org/10.1044/jshr.3403.596>
- Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P., & Skoruppa, K. (2021). Using nonword repetition to identify developmental language

- disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, ePub Ahead of Print
https://doi.org/10.1044/2021_JSLHR-20-00552
- Schwob, S., & Skoruppa (in press). L'identification d'un trouble développemental du langage oral chez les enfants bilingues français-portugais : les ressources à disposition des orthophonistes. *Revue canadienne d'orthophonie et d'audiologie*.
- Schwob, S., & Skoruppa, K. (submitted). Detecting developmental language disorder in monolingual and bilingual children: Comparison of language-specific and crosslinguistic non-word repetition tasks in French and Portuguese. *Journal of Speech, Language, and Hearing Research*.
- Sim-Sim, I. (2006). *Avaliação da linguagem oral - um contributo para o conhecimento do desenvolvimento linguístico das crianças portuguesas* (4^e ed.), Fundação Calouste Gulbenkian.
- Skoruppa, K. (2019). Novel noun and verb learning in mono- and multilingual children. *Travaux neuchâtelois de linguistique*, 71(1), 109-123.
http://www.unine.ch/files/live/sites/tranel/files/Tranel/71/109-123_Skorupa_def.pdf
- Thibault, M. P., & Helloin, M. C. (2010). *Exalang* 5-8. Happyneuron.
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426-445.
<https://doi.org/10.1177/1367006911403202>
- Thordardottir, E. (2015). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology*, 17(2), 97-114.
<https://doi.org/10.3109/17549507.2014.923509>
- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and

- processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, 54(2), 580-597. [https://doi.org/10.1044/1092-4388\(2010/09-0196\)](https://doi.org/10.1044/1092-4388(2010/09-0196))
- Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, K. de Jong & N. Meir (Eds.), *Assessing multilingual children* (pp. 301-331). Multilingual Matters.
- Van Horn, D., & Kan P. F. (2015). Fast mapping by bilingual children: Storybooks and cartoons. *Child Language Teaching and Therapy*, 32(1), 1-13. <https://doi:10.1177/0265659015584975>
- Volpin, L., Rezzonico, S., & de Weck, G. (2020). Enfants bilingues et prise en charge logopédique : panorama des pratiques déclarées. *Glossa*, 129(1) 16-31. <https://www.glossa.fr/index.php/glossa/article/view/651>
- Waxman, S., Fu., X., Arunachalam, S., Leddon, E., Geraghty, K., & Song., H. – J. (2013). Are nouns learned before verbs? Infants provide insight into a longstanding debate. *Child Development Perspectives*, 7(3), 10.1111/cdep.12032. <https://doi:10.1111/cdep.12032>
- Weismer, S. E., & Hesketh, L. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research*, 36(5), 1013-1025. <https://doi:10.1044/jshr.3605.1013>
- Woodrow, H. (1946). The ability to learn. *Psychological Review*, 53(3), 147-158. <https://doi.org/10.1037/h0053639>
- Vygotski, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Yoshida, H., Tran, D., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontier in psychology*, 210(2), 1-14. <https://doi:10.3389/fpsyg.2011.00210>

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32-35.

[https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)

Supplementary Material 1

Acceptability judgment distractors

Item	List	Distractor	Type
/fli/	A	/bli/	Phonological
/lati/	A	/ati/	Phonological
/plifu/	A	/plifi/	Phonological
/lati/	A	/fli/	Semantic
/plifu/	A	/lati/	Semantic
/fli/	A	/lati/	Semantic
/upit/	A	/upi/	Phonological
/lut/	A	/upit/	Semantic
/lut/	A	/lup/	Phonological
/upit/	A	/lut/	Semantic
/flinat/	A	/linat/	Phonological
/lut/	A	/flinat/	Semantic
/plifu upim/	A	/plifu upi/	Inflection
/plifu lum/	A	/plif lum/	Phonological
/lati upit/	A	/upit lati/	Order
/lati lut/	A	/lut lati/	Order
/lati lut/	A	/lati upit/	Semantic
/plifu flinam/	A	/plifu finam/	Phonological
/lati flinat/	A	/lati flinam/	Inflection
/lati upit/	A	/plifu upim/	Semantic
/otu/	B	/oti/	Phonological
/stabi/	B	/blu/	Semantic

Item	List	Distractor	Type
/blu/	B	/flu/	Phonological
/stabi/	B	/otu/	Semantic
/stabi/	B	/tabi/	Phonological
/blu/	B	/stabi/	Semantic
/fakat/	B	/faka/	Phonological
/fakat/	B	/vat/	Semantic
/vat/	B	/vap/	Phonological
/vat/	B	/fakat/	Semantic
/aplut/	B	/plut/	Phonological
/vat/	B	/aplut/	Semantic
/otu fakam/	B	/otu faka/	Inflection
/otu vam/	B	/ot vam/	Phonological
/stabi fakat/	B	/fakat stabi/	Order
/stabi vat/	B	/vat stabi/	Order
/stabi vat/	B	/stabi fakat/	Semantic
/otu aplum/	B	/otu apum/	Phonological
/stabi aplut/	B	/stabi aplum/	Inflection
/stabi fakat/	B	/otu fakam/	Semantic

Supplementary Material 2

Example of a story for noun learning

Original version :

Dans un monde lointain extraterrestre, une chasse aux trésors s'organise chaque année par le grand chef. Le gagnant de cette chasse aux trésors part dans le monde voisin jaune pour le découvrir.

Cette année, pour participer à la chasse aux trésors, le grand chef choisit une */plifu/* et un */lati/*.

Une */plifu/* est une créature bizarre verte et petite.

Un */lati/* est un personnage bizarre bleu et grand.

Le grand chef cache un objet. Le premier des extraterrestres qui le trouve, gagne la chasse et peut partir dans le monde voisin jaune pour un voyage.

L'objet à trouver est un */fli/*

Un */fli/* est un objet de décoration brun

La */plifu/* et le */lati/* choisis par le grand chef partent à la chasse au */fli/*.

Le */lati/* cherche activement dans les maisons.

Le */lati/* veut absolument partir en voyage dans le monde voisin. Le */fli/* se cache effectivement tout près, derrière une maison.

La */plifu/* cherche dans la forêt. La */plifu/* ne le sait encore pas mais le */fli/* ne se cache malheureusement pas dans la forêt.

Tout à coup, derrière une maison, le */lati/* trouve le */fli/*.

Malgré tous ses efforts, la */plifu/* ne la pas trouvé.

Il gagne donc la chasse aux trésors et pourra partir sur le monde voisin jaune pour le découvrir.

Translated version:

In a distant extraterrestrial world, a treasure hunt is organized every year by the great chief. The winner of this treasure hunt goes to the neighboring yellow world to discover it.

This year, to participate in the treasure hunt, the great chief chooses a */plifu/* and a */lati/*.

A */plifu/* is a small and green strange creature.

A */lati/* is a strange, blue and big character.

The big chief hides one object. The first of the aliens who finds it, wins the hunt and can go to the neighboring yellow world for a trip.

The object to be found is a */fli/*.

A */fli/* is a brown decorative object.

The */plifu/* and the */lati/* chosen by the great chief go hunting for the */fli/*

The */lati/* is actively searching in houses.

The */lati/* absolutely wants to go on a trip to the neighboring world. The */fli/* actually hides nearby, behind a house.

The */plifu/* searches in the forest. The */plifu/* doesn't know it yet, but the */fli/* is unfortunately not hiding in the forest.

Suddenly, behind a house, the */lati/* finds the */fli/*.

Despite all his efforts, the */plifu/* did not find it.

So, he wins the treasure hunt and will be able to go to the neighboring yellow world to discover it.

Supplementary Material 3

Hierarchy of mediation

Word-picture	Stimuli	Cue 1	Cue 2	Cue 3
matching	example			
Noun	/lati/	/lati/	Examples: yes, Strange alien	blue big
Verb	he /upit/	he /upit/	good, Figure easy	Needs himself Roll to the left
Sentence	the /lati upit/	the /lati upit/ / do you congratulations, remember the continue character or what he does?	Strange blue alien	big Figure Needs Roll to easy himself the lef

Naming	Stimuli	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5	Cue 6	Cue 7
	example							
Noun	/lati/	a /lati/	Examples: yes, Strange alien	Blue	Big	A	A l	
Verb	he /upit/	he /upit/	good, congratulations,	Figure easy	Needs	Roll to the lef	He	He u
				himself				
Sentence	a /lati upit/	the /lati upit/ / do you remember the character or what he does?	continue	Strange alien	blue big	Figure easy	Needs himself	Roll to the lef
							A	A l
								A lati
								A upi

Note. Abbreviations: cue 1 = repetition or/and question; cue 2 = encouragements; cue 3 = semantics clues; cue 4 = determiner of pronoun; cue 5 = 1st phoneme; cue 6 = noun + 1st phoneme of the verb; cue 7 = all except inflection.

Supplementary Material 4

Factors explaining the number of correct words in tasks

Situations	Tasks	Items	Disorder	<i>d</i>	Bilingualism	<i>d</i>	Disorder *	Age	Nonverbal
							Bilingualism		IQ
Autonomous DA	Word-picture	Nouns	$W^2 = 4.59, p$	$d =$	$W^2 = 6.34, p$	$d = -$	$W^2 = 0.42, p$	$W^2 = 2.82, p$	$W^2 = 1.48 p$
	matching		$= .032^*$	.641	$= .012^*$	.647	$= .837$	$= .093$	$= .223$
Autonomous DA	Word-picture	Verbs	$W^2 = 1.14, p$	$d = -$	$W^2 = 2.39, p$	$d = -$	$W^2 = 0.47, p$	$W^2 = 1.89, p$	$W^2 = 0.33, p$
	matching		$= .286$	.939	$= .122$	1.025	$= .828$	$= .169$	$= .561$
Autonomous DA	Word-picture	Sentences	$W^2 = 4.97, p$	$d =$	$W^2 = 0.57, p$	$d = .139$	$W^2 = 0.93, p$	$W^2 = 0.99, p$	$W^2 = 0.01, p$
	matching		$= .026^*$	1.076	$= .450$		$= .334$	$= .318$	$= .916$
Autonomous DA	Recall	Nouns	$W^2 = 1.01, p$	$d =$	$W^2 = 0.00, p$	$d = -$	$W^2 = 0.39, p$	$W^2 = 6.21, p$	$W^2 = 0.56, p$
			$= .314$	.244	$= .993$	.365	$= .529$	$= .013^*$	$= .454$
Autonomous DA	Recall	Verbs	$W^2 = 0.08, p$	$d = -$	$W^2 = 9.17, p$	$d = -$	$W^2 = 3.53, p$	$W^2 = 27.63,$	$W^2 = 11.13,$
			$= .768$	.614	$= .002^{**}$	.367	$= .060$	$p < .001^{***}$	$p = .001^{***}$

Situations	Tasks	Items	Disorder	<i>d</i>	Bilingualism	<i>d</i>	Disorder *	Age	Nonverbal
							Bilingualism		IQ
Autonomous DA	Recall	Sentences	$W^2 = 0.01, p$	$d = -$	$W^2 = 0.00, p$	$d = -$	$W^2 = 3.57, p$	$W^2 = 0.00, p$	$W^2 = 0.67, p$
			= .891	.454	= .961	.412	= .059	= .944	= .410
Autonomous DA	Naming	Nouns	$W^2 = 2.26, p$	$d =$	$W^2 = 2.20, p$	$d = -$	$W^2 = 0.09, p$	$W^2 = 7.11, p$	$W^2 = 0.39, p$
			= .133	.568	= .138	.900	= .756	= .008**	= .531
Autonomous DA	Naming	Verbs	$W^2 = 1.17, p$	$d = -$	$W^2 = 0.50, p$	$d = -$	$W^2 = 1.60, p$	$W^2 = 3.81, p$	$W^2 = 0.89, p$
			= .277	.154	= .477	.038	= .205	= .051	= .343
Autonomous DA	Naming	Sentences	$W^2 = 1.26, p$	$d = -$	$W^2 = 1.63, p$	$d = -$	$W^2 = 1.50, p$	$W^2 = 0.11, p$	$W^2 = 0.00, p$
			= .260	.006	= .201	.016	= .219	= .735	= .979
Autonomous DA	Acceptability judgment	Nouns	$W^2 = 1.79, p$	$d =$	$W^2 = 3.72, p$	<i>Na</i>	$W^2 = 0.23, p$	$W^2 = 1.91, p$	$W^2 = 0.08, p$
			= .180	1.051	= .054		= .630	= .166	= .770
Autonomous DA	Acceptability judgment	verbs	$W^2 = 0.84, p$	$d = -$	$W^2 = 4.04, p$	$d = -$	$W^2 = 0.29, p$	$W^2 = 3.67, p$	$W^2 = 0.48, p$
			= .773	.207	= .044*	.704	= .590	= .055	= .484

Situations	Tasks	Items	Disorder	d	Bilingualism	d	Disorder *	Age	Nonverbal
							Bilingualism		IQ
Autonomous DA	Acceptability judgment	Sentences	$W^2 = 5.19, p = .023^*$	$d = 0.688$	$W^2 = 0.28, p = .593$	$d = -.717$	$W^2 = 0.00, p = .941$	$W^2 = 0.79, p = .373$	$W^2 = 0.19, p = .659$
Autonomous DA	Delayed recall	Nouns	$W^2 = 0.34, p = .556$	$d = .452$	$W^2 = 2.50, p = .113$	$d = -.267$	$W^2 = 0.67, p = .411$	$W^2 = 0.05, p = .821$	$W^2 = 0.16, p = .687$
Autonomous DA	Delayed recall	Verbs	$W^2 = 3.25, p = .071$	$d = .060$	$W^2 = 1.09, p = .296$	$d = .311$	$W^2 = 3.41, p = .065$	$W^2 = 0.03, p = .856$	$W^2 = 0.01, p = .911$
Interactive DA	Word-picture matching	Nouns	$W^2 = 2.28, p = .131$	$d = -.377$	$W^2 = 0.31, p = .576$	$d = -1.569$	$W^2 = 6.35, p = .012^*$	$W^2 = 0.02, p = .884$	$W^2 = 4.62, p = .031^*$
Interactive DA	Word-picture matching	Verbs	$W^2 = 5.01, p = .025^*$	$d = .427$	$W^2 = 5.72, p = .017^*$	$d = .479$	$W^2 = 0.42, p = .513$	$W^2 = 8.46, p = .004^{**}$	$W^2 = 8.07, p = .004^{**}$
Interactive DA	Word-picture matching	Sentences	$W^2 = 12.47, p < .001^{***}$	$d = .927$	$W^2 = 0.46, p = .497$	$d = -.393$	$W^2 = 0.02, p = .871$	$W^2 = 8.16, p = .004^{**}$	$W^2 = 0.00, p = .967$

Situations	Tasks	Items	Disorder	d	Bilingualism	d	Disorder *	Age	Nonverbal
							Bilingualism		IQ
Interactive DA	Recall	Nouns	$W^2 = 9.15, p$ $= .002^{**}$	$d =$.726	$W^2 = 1.58, p$ $= .208$	$d = .541$	$W^2 = 0.00, p$ $= .958$	$W^2 = 28.15,$ $p < .001^{***}$	$W^2 = 2.81, p$ $= .094$
Interactive DA	Recall	Verbs	$W^2 = 32.66,$ $p < .001^{***}$	$d =$ 1.581	$W^2 = 4.41, p$ $= .036^*$	$d = .819$	$W^2 = 0.39, p$ $= .529$	$W^2 = 25.59,$ $p < .001^{***}$	$W^2 = 0.23, p$ $= .630$
Interactive DA	Recall	Sentences	$W^2 = 7.98, p$ $= .005^{**}$	$d =$.575	$W^2 = 2.42, p$ $= .119$	$d = .331$	$W^2 = 0.65, p$ $= .417$	$W^2 = 8.39, p$ $= .004^{**}$	$W^2 = 0.41, p$ $= .517$
Interactive DA	Naming	Nouns	$W^2 = 1.87, p$ $= .171$	$d = -$.246	$W^2 = 8.30, p$ $= .004^{**}$	$d = .270$	$W^2 = 3.90, p$ $= .048^*$	$W^2 = 17.73,$ $p < .001^{***}$	$W^2 = 1.38, p$ $= .240$
Interactive DA	Naming	Verbs	$W^2 = 18.71,$ $p < .001^{***}$	$d =$ 1.081	$W^2 = 3.90, p$ $= .048^*$	$d = .385$	$W^2 = 0.24, p$ $= .624$	$W^2 = 9.45, p$ $= .002^{***}$	$W^2 = 1.83, p$ $= .176$

Situations	Tasks	Items	Disorder	d	Bilingualism	d	Disorder *	Age	Nonverbal
							Bilingualism		IQ
Interactive DA	Naming	Sentences	$W^2 = 4.79, p$ $= .029^*$	$d =$.291	$W^2 = 0.99, p$ $= .319$	$d = -$.055	$W^2 = 1.50, p$ $= .219$	$W^2 = 2.25, p$ $= .133$	$W^2 = 0.38, p$ $= .536$
Interactive DA	Acceptability judgment	Nouns	$W^2 = 13.10,$ $p < .001^{***}$	$d =$.342	$W^2 = 6.90, p$ $= .009^{**}$	$d = -$ 1.387	$W^2 = 5.24, p$ $= .022^*$	$W^2 = 4.72, p$ $= .030^*$	$W^2 = 0.13, p$ $= .715$
Interactive DA	Acceptability judgment	Verbs	$W^2 = 31.25,$ $p < .001^{***}$	$d =$ 2.903	$W^2 = 0.60, p$ $= .438$	$d =$ 3.598	$W^2 = 2.73, p$ $= .098$	$W^2 = 1.81, p$ $= .178$	$W^2 = 2.65, p$ $= .104$
Interactive DA	Acceptability judgment	Sentences	$W^2 = 9.07, p$ $= .003^{**}$	$d =$ 1.071	$W^2 = 0.14, p$ $= .708$	$d = .124$	$W^2 = 0.25, p$ $= .614$	$W^2 = 3.48, p$ $= .062$	$W^2 = 2.42, p$ $= .120$
Interactive DA	Delayed recall	Nouns	$W^2 = 1.08, p$ $= .297$	$d =$.242	$W^2 = 1.32, p$ $= .251$	$d = .314$	$W^2 = 0.41, p$ $= .520$	$W^2 = 4.31, p$ $= .038^*$	$W^2 = 0.01, p$ $= .892$
Interactive DA	Delayed recall	Verbs	$W^2 = 0.63, p$ $= .426$	$d = -$.066	$W^2 = 0.77, p$ $= .377$	$d = -$.036	$W^2 = 1.11, p$ $= .291$	$W^2 = 1.91, p$ $= .166$	$W^2 = 0.80, p$ $= .371$

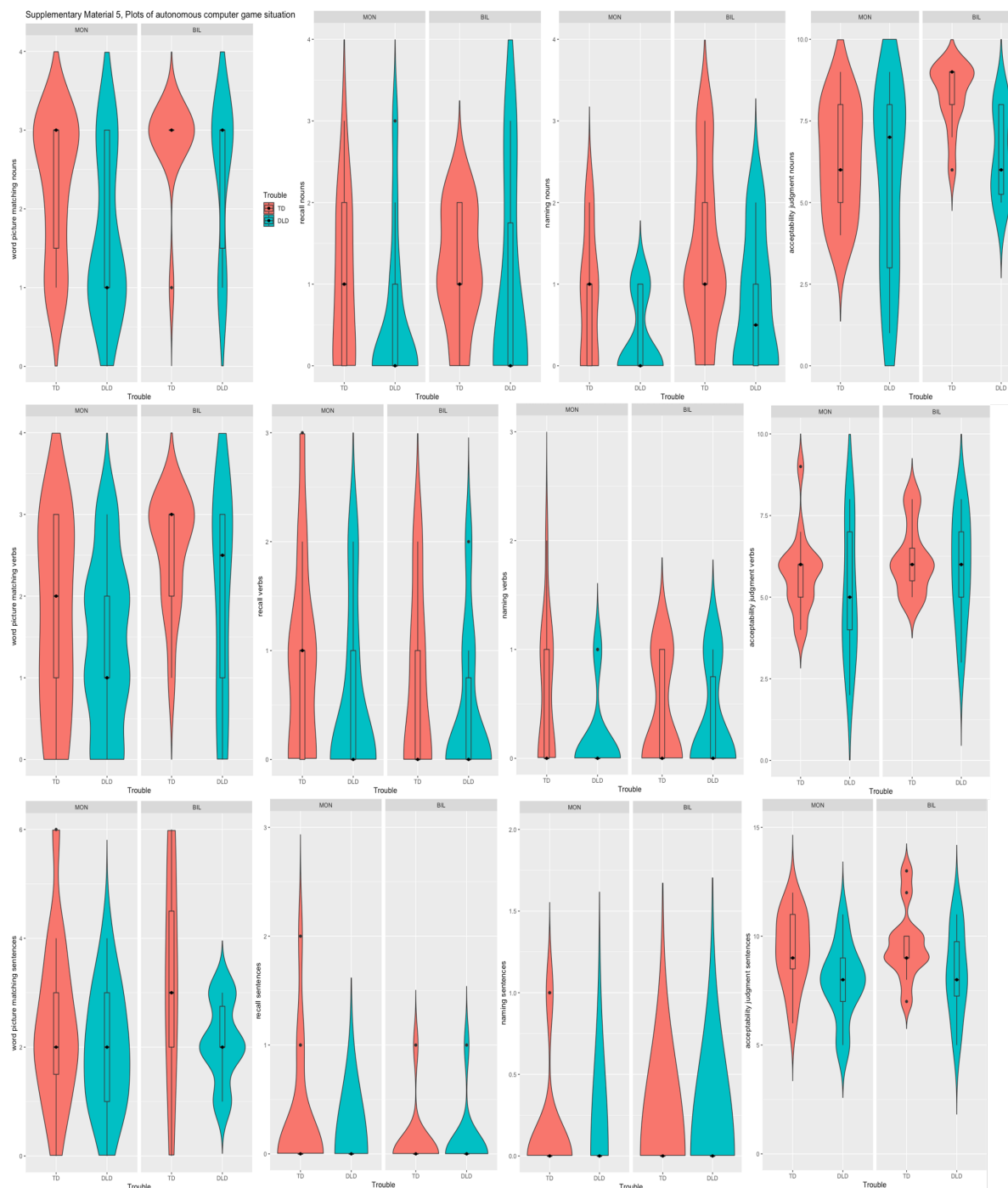
Situations	Tasks	Items	Disorder	d	Bilingualism	d	Disorder *	Age	Nonverbal
							Bilingualism		IQ
Interactive DA improvement	Word-picture matching	Nouns	$W^2 = 5.70, p = .017^*$	$d = -1.503$	$W^2 = 0.13, p = .713$	$d = .010$	$W^2 = 0.17, p = .672$	$W^2 = 0.44, p = .504$	$W^2 = 1.14, p = .284$
Interactive DA improvement	Word-picture matching	Verbs	$W^2 = 1.43, p = .232$	$d = .545$	$W^2 = 5.25, p = .022^*$	$d = .190$	$W^2 = 9.60, p = .002^{**}$	$W^2 = 5.08, p = .024^*$	$W^2 = 5.14, p = .023^*$
Interactive DA improvement	Word-picture matching	Sentences	$W^2 = 2.41, p = .120$	$d = -1.965$	$W^2 = 0.98, p = .322$	$d = .822$	$W^2 = 0.05, p = .823$	$W^2 = 1.02, p = .311$	$W^2 = 0.29, p = .589$
Interactive DA improvement	Naming	Nouns	$W^2 = 0.37, p = .540$	$d = .056$	$W^2 = 0.21, p = .640$	$d = .073$	$W^2 = 1.14, p = .285$	$W^2 = 0.02, p = .887$	$W^2 = 0.00, p = .974$
Interactive DA improvement	Naming	Verbs	/		/		/	/	/
Interactive DA improvement	Naming	Sentences	$W^2 = 3.06, p = .080$	$d = .328$	$W^2 = 0.00, p = .985$	$d = -.028$	$W^2 = 0.02, p = .870$	$W^2 = 0.57, p = .447$	$W^2 = 4.70, p = .030^*$

Note. Abbreviations: DA = dynamic assessment; Autonomous DA = autonomous computer game situation; Interactive DA = interactive story reading situation; Interactive DA improvement = improvement with cues in interactive story reading situation; d = effect size from Cohen's d .

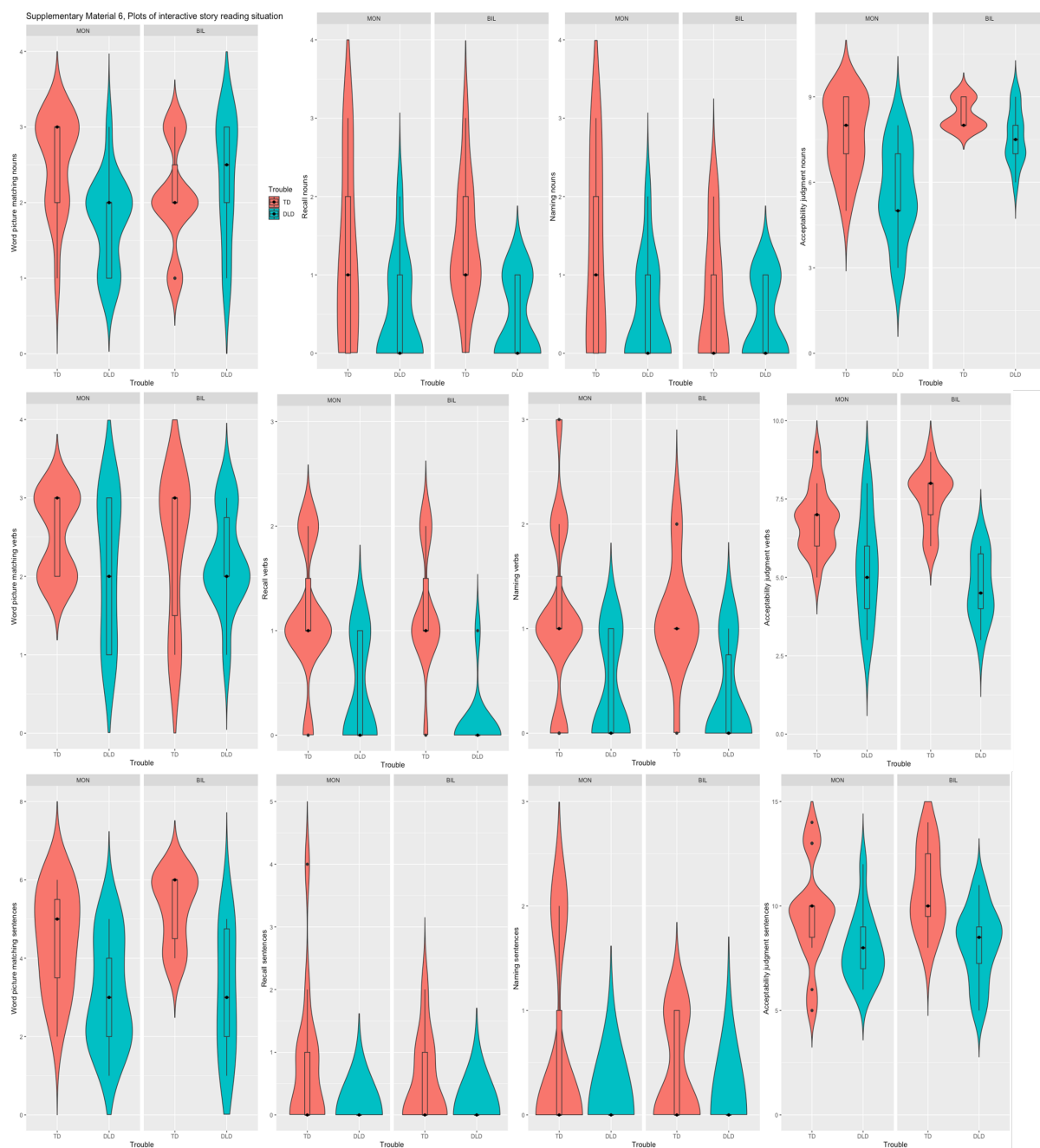
$N = 49$ ($n = 23$ children with a DLD and 26 children with a TD / $n = 21$ bilingual children and 28 monolingual children).

* $p < .05$. ** $p < .01$. *** $p < .001$

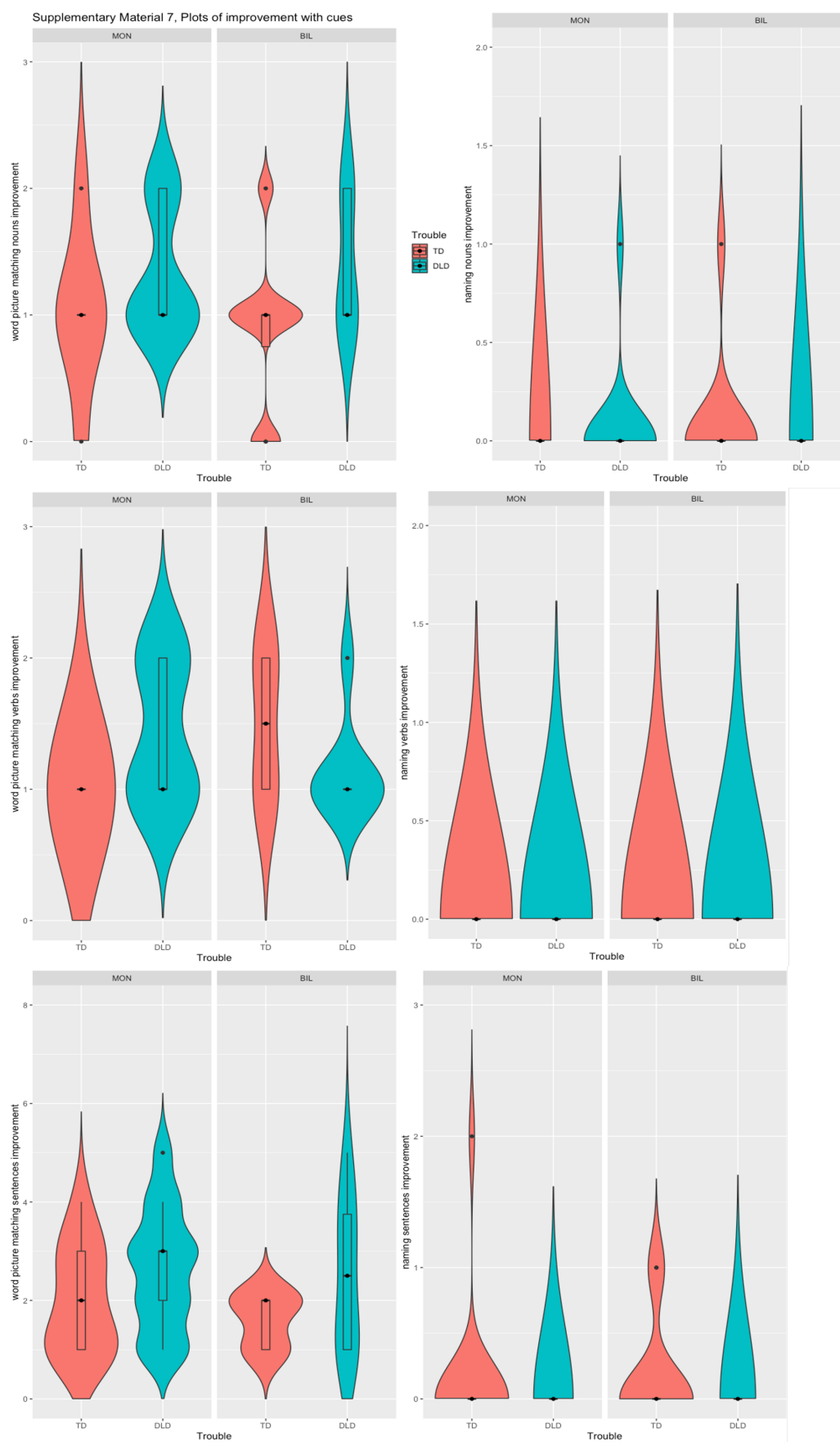
Supplementary Material 5



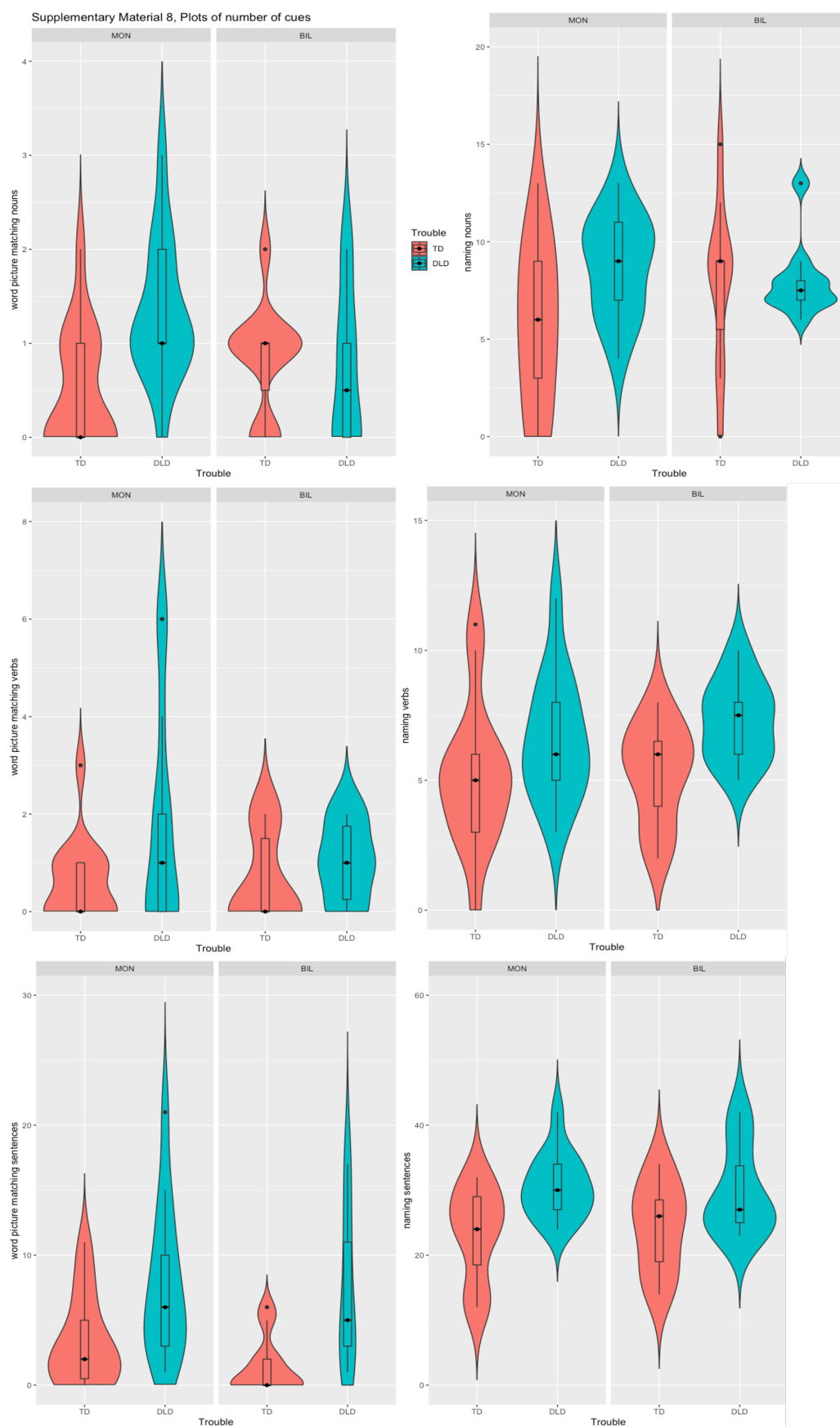
Supplementary Material 6



Supplementary Material 7



Supplementary Material 8



Supplementary Material 9

Statistic of number of scaffoldings

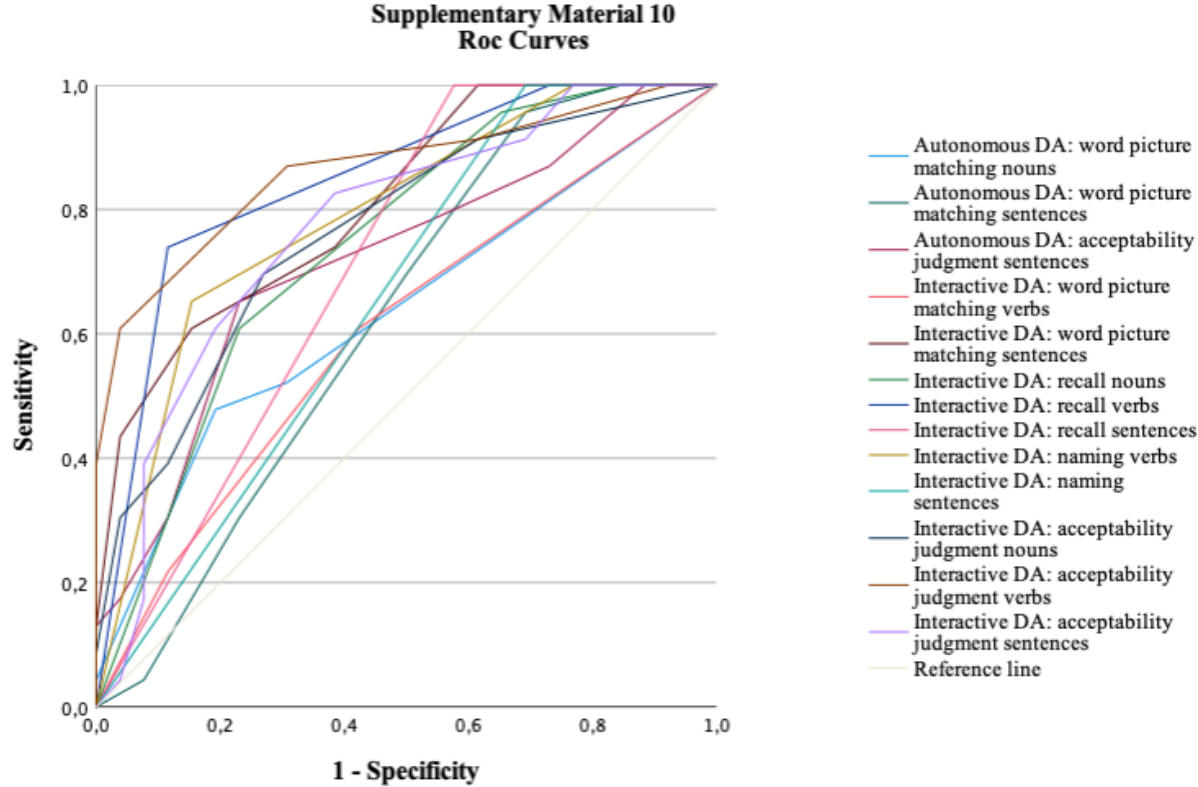
Tasks	Items	Disorder	<i>d</i>	Bilingualism	<i>d</i>	Disorder *	Age	Non-verbal
						Bilingualism		IQ
Word- picture matching	Nouns	$W^2 = 5.25, p = .022^*$	$d = -.032$	$W^2 = 0.54, p = .461$	$d = 1.536$	$W^2 = 5.25, p = .022^*$	$W^2 = 0.19, p = .890$	$W^2 = 3.73, p = .053$
Word- picture matching	verbs	$W^2 = 4.17, p = .041^*$	$d = -.181$	$W^2 = 0.11, p = .739$	$d = 1.105$	$W^2 = 2.71, p = .096$	$W^2 = 3.02, p = .082$	$W^2 = 1.39, p = .237$
Word- picture matching	Sentences	$W^2 = 10.85, p = .001$ ***	$d = -1.001$	$W^2 = 0.11, p = .735$	$d = -.002$	$W^2 = 0.14, p = .705$	$W^2 = 7.14, p = .008^{**}$	$W^2 = 0.59, p = .439$
Naming	Nouns	$W^2 = 2.62, p = .106$	$d = 2.118$	$W^2 = 6.99, p = .008^{**}$	$d = -.645$	$W^2 = 6.65, p = .010^{**}$	$W^2 = 18.44, p < .001^{***}$	$W^2 = 1.19, p = .274$

Tasks	Items	Disorder	<i>d</i>	Bilingualism	<i>d</i>	Disorder *	Age	Non-verbal
						Bilingualism		IQ
Naming	Verbs	$W^2= 6.68, p =$	$d = - .725$	$W^2= 2.91, p =$	<i>Na</i>	$W^2= 0.01, p =$	$W^2= 2.32, p =$	$W^2= 2.42, p =$
		.010**		.088		.911	.128	.117
Naming	Sentences	$W^2= 11.55, p$	$d = - .867$	$W^2= 0.05, p =$	$d = 8.200$	$W^2= 0.08, p =$	$W^2= 2.26, p =$	$W^2= 3.48, p =$
		= .001***		.811		.766	.101	.062

Note. $n = 49$ ($n = 23$ children with a DLD and 26 children with a TD / $n = 21$ bilingual children and 28 monolingual children).

* $p < .05$. ** $p < .01$. *** $p < .001$.

Supplementary Material 10



7.1 Analyses additionnelles

Dans l'article présent au chapitre 7, nous avons renoncé, par manque de place, à présenter les corrélations statistiques entre les résultats obtenus par les enfants à notre situation interactive de lecture d'histoires et leurs performances aux sous-tests statiques standardisés. Nous proposons ainsi ces informations ci-dessous.

Correlations with standardized tasks

To assess whether the difficulties children show in our DA situations are in line with their actual language profile, we now test whether there were non-parametric correlations between each time two variables: the performance in the different interactive story reading tasks that show significant differences between DLD and TD children and their performances on some of the standardized subtests in French (used as inclusion criteria in the study). Recall that DLD children in our research have various difficulties and failed different subtests in the initial assessment (see Methods section). Indeed, we found significant correlations between children's performance on the sentence-picture matching and sentence acceptability judgment tasks with the standardized sentence comprehension task ($p < .05$). We can also report a correlation between verb naming and verb recall and the standardized naming subtest (verb naming: $\tau_b = 0.250, p < .05$; verb recall: $\tau_b = 0.251, p < .05$).

Thus, our tasks seem to tap into similar skills as static linguistic assessments in one language. In addition to phonological memory deficits evidenced by their failure in the NWRT, DLD children in our research had a variety of language difficulties on standardized tests in French (naming and sentence comprehension) that seem to be closely related to their failure on certain tasks in our interactive story reading situation.

7.2 Informations additionnelles

Nous élaborons actuellement avec une mémorante de Master en Logopédie, une revue systématique de littérature sur l'usage de l'évaluation dynamique dans un contexte monolingue et bilingue. Une revue systématique récente publiée par Orellana et al. (2019) s'est notamment intéressée au bilinguisme mais nous avons envie de venir compléter l'absence de données théoriques portant sur l'évaluation dynamique dans un contexte monolingue et d'inclure les études s'étant attelées à l'analyses des résultats des deux populations.

Partie VI : Discussion et conclusion

8. Discussion

Nous débuterons cette partie de discussion en effectuant une synthèse générale des résultats obtenus au fil de ce travail. Nous les mettrons également en résonance avec les données de la littérature et nous évoquerons les implications cliniques de l'analyse de celles-ci en reprenant nos sous-questions de recherche permettant de traiter notre questionnement général (8.1). Nous finaliserons la partie de discussion en évoquant les limites de ce travail ainsi que les ouvertures possibles (8.2). Nous clôturerons cette thèse de doctorat par une conclusion au chapitre 9.

8.1 Synthèse de la thèse et implications cliniques

La Partie I de cette thèse de doctorat s'est intéressée à décrire théoriquement le TDL, le contexte de bilinguisme ainsi que les difficultés des logopédistes dans l'évaluation des enfants bilingues. Après avoir énoncé ces fondements, nous avons présenté, dans la Partie II, notre problématique de recherche, nos questions de recherche et hypothèses théoriques. Nous allons reprendre nos questionnements et y apporter des éléments de réponses grâce aux données analysées dans les Parties III, IV et V (chapitre 4, 5, 6 et 7) de cette thèse de doctorat.

Quelles sont les tâches langagières et/ou modes d'évaluation permettant une identification précise de la présence d'un TDL chez une population monolingue et bilingue français-portugais d'âge scolaire ?

8.1.1 Sous-questions 1

Parmi nos sous-tests statiques langagiers, quelles sont les tâches langagières (sous-tests standardisés- dénomination, compréhension syntaxique, répétitions de non-mots - ; questionnaire parental) et/ou modes d'évaluation (dans une ou deux langues, utilisant des non-

mots langues-spécifiques ou *cross-linguistiques*) qui révèlent une différence de performance significative entre les enfants avec un DT et avec un TDL dans un contexte monolingue et bilingue ?

a Quelle tâche langagière et/ou mode d'évaluation fournit la meilleure précision diagnostique (pourcentage de sensibilité et spécificité) ? Permet-elle d'améliorer la classification initiale de nos sous-groupes ?

Nous nous sommes attelées, dans la Partie III, à l'évaluation des enfants bilingues au moyen de tests standardisés provenant de batteries francophones et lusophones avec des normes monolingues en testant plusieurs modes d'évaluation (dans une langue, en français et dans la langue dominante, puis dans les deux langues parlées par l'enfant, et évaluation nécessitant l'usage de non-mots). Nous avons pu dégager le pouvoir discriminant de la tâche de RNM en français dans l'identification d'un TDL au sein de notre population bilingue français-portugais. Nos résultats vont dans le sens de nos hypothèses qui présumaient qu'un mode d'évaluation recourant aux non-mots permettraient de mettre en lumière des effets du statut clinique vs. statut linguistique (nous revenons sur les différences des RNM dans ce qui suit). Nous avons également pu observer la qualité en termes de précision diagnostique de plusieurs questions adressées aux parents via notre questionnaire parental, à savoir l'âge de la combinaison de deux mots ainsi que leur inquiétude quant au développement langagier de leur enfant. Les pourcentages de précision diagnostique des questions étaient parfois meilleurs que certaines tâches standardisées disponibles sur le marché. Toutefois, comme formulé dans l'article présenté au point 4, les parents d'enfants avec un TDL se sont montrés plus inquiets que les parents d'enfants avec un DT, cependant tous les enfants avec un TDL de notre recherche sont suivis en logopédie. Il y a donc un biais non négligeable concernant les réponses parentales au sujet de leur inquiétude. Il s'agit, en effet, d'une post-prédiction vu que les parents s'expriment

d'avantage sur ce qu'ils ressentent et connaissent maintenant au sujet de leur enfant (ex. mon enfant possède un TDL) plutôt que sur ce qu'ils ressentaient dans le passé. Si une étude longitudinale avait été menée, il aurait été intéressant de savoir si ces mêmes parents auraient signalé leur inquiétude lorsque leur enfant avait trois ans par exemple. Malgré ce biais, nous pouvons confirmer notre hypothèse qui était de pouvoir utiliser les données indirectes de façon complémentaire dans le bilan des enfants bilingues et nos résultats nous rappellent que les modes d'évaluation indirectes ne doivent pas être sous-estimés. Sur le terrain, il convient ainsi d'accorder une place importante aux questions que l'on pose aux parents en recourant à des questionnaires ou lors d'une anamnèse détaillée (Kehoe, 2009). Bien que nous ayons évoqué les limites inhérentes aux interprètes dans les bilans logopédiques, notamment impliquant des ressources financières et du personnel qualifié (Boerma et Blom, 2017 ; voir article au point 1.2), il semble indispensable de pouvoir les mandater lorsque c'est nécessaire pour la séance d'anamnèse et de restitution. Les questions à poser aux parents pour identifier des facteurs de risque de TDL ainsi que la transmission des informations liées au diagnostic sont importantes et doivent être à la portée de la compréhension des parents de diverses cultures et parlant différentes langues (Rosenbaum, 2019).

La Partie III a également mis en lumière le faible pouvoir discriminant d'autres sous-tâches en français et en portugais auprès de notre échantillon ayant des profils linguistiques hétérogènes. Les tâches de dénomination et de compréhension orale en français et en portugais ne nous ont que peu renseignées sur le statut clinique de nos participants bilingues comme nous avons pu le supposer. Bien que nous ayons observé une légère amélioration de la précision diagnostique en appliquant des adaptations aux seuils de pathologie définis par le test en fonction du taux d'exposition langagier des enfants à la langue de la tâche (Thordardottir, 2015a), cette amélioration n'avait de sens uniquement pour la tâche de RNM en français puisque les autres tâches en français et en portugais n'obtenaient que de très faibles

pourcentages de précision diagnostique avant et après cette adaptation. À noter que nous avons utilisé une adaptation des seuils de pathologie en fonction du taux d'exposition langagier en prenant en considération le taux d'exposition actuelle des langues de l'enfant, choix méthodologique qui diffère de Thordardottir (2015a) s'étant intéressée au bilinguisme cumulatif. Dans cette Partie, notre choix ne semble pas avoir eu beaucoup d'impact puisque nous avons testé de nombreux modes et outils d'évaluation et que les deux langues de l'enfant bilingue ont été évaluées. De plus, pour rappel, l'intérêt premier de notre recherche étant l'évaluation dynamique au moyen d'apprentissages langagiers indépendants aux langues parlées par les enfants, nous n'avons donc pas choisi de consacrer une partie plus importante de la thèse à ces éléments. Nous n'avons, dans cette Partie, pas trouvé d'avantages diagnostics à utiliser un mode d'évaluation dans une (français et langue dominante) ou les deux langues (français et portugais) de l'enfant, certainement car ces outils possédaient des normes monolingues et ont été conçus pour des enfants parlant une seule langue. Nos résultats vont dans le sens de la littérature existante sur le sujet ; par exemple, Scheidnes et Tuller (2016) avaient également utilisé des tests avec normes monolingues pour évaluer les deux langues des enfants bilingues anglais-français de leur recherche et ils avaient trouvé des résultats dépendants du taux d'exposition des enfants bilingues, créant ainsi des erreurs de diagnostic. Dans une visée clinique, il est nécessaire de reconnaître que de proposer des tâches évaluatives dans les deux langues de l'enfant bilingue, peut nous fournir des informations qualitatives intéressantes pour son évaluation et sa potentielle prise en charge (ex. si l'enfant peine à comprendre l'usage des monèmes interrogatifs dans ses deux langues, l'idée serait de proposer une intervention ayant pour but de renforcer cet aspect).

Malgré les constats précédents, nous avons été en mesure d'améliorer la classification initiale des participants, notamment grâce à la RNM en français qui s'est montrée la tâche la plus sensible et spécifique de cette Partie. Nous notons que 12 enfants auraient pu être assignés

à d'autres groupes mais ont été exclus des analyses portant sur l'évaluation dynamique (voir suite de la section). Il est important de mentionner brièvement que la RNM en français s'est montrée précise et que cela n'a pas été le cas pour la RNM en portugais, ce qui peut s'expliquer de différentes façons. Puisque la tâche a été conçue pour des enfants monolingues et qu'elle ait été standardisée et normée auprès d'un petit échantillon (de Almeida et al., 2016), il est possible qu'elle ne se montre pas très sensible. Nous avons également relevé que les deux RNM variaient en termes de longueur et complexité phonologique, toutefois, nous avons également observé que la majorité des items en portugais possèdent une syllabe et peuvent donc être plus faciles pour les enfants et ne pas impliquer suffisamment la mémoire à court-terme verbale (voir article au point 4).

En regard de la qualité diagnostique de la RNM, nous avons consacré notre Partie IV à celle-ci. Nous nous sommes d'abord intéressées théoriquement à cette tâche en réalisant une revue systématique de littérature et méta-analyse sur son utilité dans l'identification d'un TDL dans un contexte monolingue et bilingue, sujet totalement novateur puisque la dernière méta-analyse sur le sujet (à notre connaissance, Estes et al., 2007) s'est uniquement penchée sur des études réalisées auprès d'enfants monolingues. Nous avons ainsi pu mettre en évidence une large taille d'effet de la RNM dans un contexte monolingue et bilingue. Auparavant, différents auteurs avaient déjà mis en avant le pouvoir discriminant des RNM auprès de différentes langues (ex. Bortolini et al, 2006) et différentes combinaisons de langues (ex. Armon-Lotem et Meir, 2016), mais une méta-analyse regroupant les différents contextes linguistiques n'avait encore jamais vu le jour. Grâce à nos analyses quantitatives de sous-groupes, nous avons pu mettre en lumière une influence de la taille d'effet en fonction de la population étudiée (avantage pour les enfants bilingues vs. monolingues) et de la spécificité langagière de la tâche (avantage de la tâche *cross-linguistique* vs. langue spécifique). Toutefois, notre méta-régression

incluant trois modérateurs (statut linguistique, spécificité langagière des tâches et âge des enfants) ne montre aucune différence entre ces conditions, ce qui peut s'expliquer par le nombre restreint et inégal d'études incluses dans les analyses quantitatives. À noter que les tâches de RNM incluses dans notre article n'évaluaient pas toutes les mêmes éléments. Les RNM langues-spécifiques avaient l'avantage d'avoir des items qui variaient en termes de complexité phonologique (structure syllabique) et de faire ainsi appel aux capacités phonologiques des enfants (ex. précision et représentations notamment). À contrario, les items de la répétition *cross-linguistique*, souhaitant être utilisable pour la majorité des langues, variaient uniquement en longueur et impliquaient ainsi la mémoire à court terme verbale. La large taille d'effet moyenne trouvée dans cet article atteste donc des difficultés de précision phonologique et de mémoire à court-terme verbale des enfants monolingues et bilingues avec un TDL. Toutefois, il est intéressant d'observer que cette taille d'effet variait en fonction des types de tâche (langue-spécifique ou *cross-linguistique*) et donc des domaines qu'elles impactent. À noter que nous aurions également pu effectuer d'autres distinctions plus fines que celles proposées dans cet article. En effet, au sein des RNM langues-spécifiques certains auteurs ont essayé de neutraliser la présence de complexité phonologique pour ne faire varier que la longueur syllabique (ex. Thordardottir et Brandeker, 2013), dans le but de ne pas recourir aux capacités phonologiques des enfants (ex. précision, représentations). À contrario, d'autres auteurs ne souhaitaient pas d'influence de la longueur syllabique et ont ainsi proposé uniquement des items courts pour ne pas mobiliser la mémoire à court-terme verbale et faire varier seulement la complexité phonologique (ex. Dos Santos et Ferré, 2018). Toutefois, pour effectuer des analyses statistiques, il aurait été nécessaire d'avoir un minimum d'études par sous-groupes (environ dix) ce qui a limité nos distinctions. Nous aurions également été intéressées à analyser l'influence du statut socio-économique sur les performances des enfants aux RNM mais, à

nouveau, trop peu d'études ont mentionné ces indications et ces éléments n'ont donc pas pu être exploités comme des modérateurs dans notre méta-analyse.

Toujours dans la Partie IV, nous avons cherché à analyser la qualité diagnostique des RNM en français, en portugais et *cross-linguistique* auprès de notre population monolingue et bilingue en prenant comme référence notre classification initiale. Nous avons supposé que des différences inhérentes à la construction des tâches (langues-spécifiques, *cross-linguistique*) puissent apparaître dans nos résultats. Nous nous sommes donc attelées à leur comparaison et nous nous sommes intéressées également à leur précision diagnostique. Nous avons relevé que les trois tâches permettent de mettre en évidence des différences entre les enfants avec un DT et un TDL qu'ils soient monolingues ou bilingues. Face aux différentes RNM, les enfants avec un TDL de notre recherche ont toujours éprouvé plus de difficultés que les enfants avec un DT. Toutefois, la précision diagnostique a été la meilleure dans le cas de la RNM en français tant pour les enfants monolingues que bilingues. Comme relaté déjà auparavant, ce n'est pas le cas de la RNM lusophone également issue d'un test standardisé. Une des raisons possibles est que la standardisation et normalisation ont eues lieu sur un nombre restreint d'enfants (de Almeida et al., 2016). À noter que divers auteurs (ex. Betz et al., 2013) font état que bien des tests disponibles sur le terrain ne disposent pas de précision diagnostique jugée satisfaisante ou ne rapportent pas ces résultats dans les manuels. Nous notons également d'importantes différences de résultats obtenus en fonction de notre choix d'utiliser un seuil en percentiles selon les normes disponibles dans le test (voir article au point 4) ou un seuil basé sur le score par item ou phonème erroné (article au point 6) obtenu par les enfants. Pour la RNM en portugais, l'usage d'un score par item erroné s'est montré clairement plus informatif pour dissocier les enfants avec un DT et un TDL que les normes disponibles dans le manuel de test. Alors que nous aurions pu supposer que l'analyse par phonèmes erronés soit plus précise, nous avons pu attester que c'est le score par items erronés qui a fourni la meilleure précision diagnostique pour les trois RNM

de manière générale (détails dans l'article au chapitre 6). Nous avons également noté cette conclusion dans la revue systématique (présente au point 5), ce constat rendant la cotation des tâches de RNM plus aisée en clinique (Pham et Ebert, 2020). Toutefois, la prise en considération de l'effet de longueur rend nécessaire l'usage d'une cotation en nombre de phonèmes erronés afin d'avoir davantage de renseignements sur les stratégies employées par les enfants (ex. les enfants avec un DT ont tenté de produire davantage alors que les enfants avec un TDL se sont interrompus lorsque les items étaient longs). Nous nous sommes également intéressées à l'influence de l'âge sur les performances en RNM. Nous relevons un effet de l'âge dans nos données présentées au point 6. Effectivement, les enfants plus âgés auraient moins de difficultés à répéter adéquatement les non-mots présentés. Il est possible qu'en fonction de la nature des items (longueur et complexité) de nos RNM, nous ayons été confrontées à des effets plafond des enfants les plus âgés de la recherche. Toutefois, dans notre méta-analyse présente au point 5, nous n'avons pas vu d'influence significative du modérateur âge sur les tailles d'effet. À noter encore que d'autres auteurs s'étant intéressés à des enfants plus âgés, par exemple Girbau et Schwartz (2007), relèvent que leur population avec TDL âgée entre 8 ;3 et 10 ;3 ans a échoué significativement plus à la RNM que les enfants avec un DT. Leur RNM variant en termes de longueur (jusqu'à cinq syllabes) et de complexité phonologique. Ainsi, il semble exister un réel intérêt d'utiliser une telle tâche jusqu'à environ dix ans, pour autant que les items utilisés soient plus complexes que pour des enfants plus jeunes. Finalement, nous avons également observé des proportions d'erreurs statistiquement différentes entre les enfants avec un DT et un TDL dans les tâches langues-spécifiques. La proportion d'erreur étant également importante dans la tâche *cross-linguistique* mais ne variant pas selon le statut clinique.

Bien que la RNM *cross-linguistique* ait également montré son potentiel (effet du statut clinique), celle-ci semble plutôt refléter des difficultés en mémoire à court-terme verbale plutôt que de traitement phonologique. Pour rappel, la RNM *cross-linguistique* ne varie qu'en termes

de longueur vs. complexité phonologique et, selon plusieurs auteurs (ex. Marini et al., 2017), les enfants avec un DT et un TDL auraient du mal avec la croissance en longueur des items. À contrario, seuls les enfants avec un TDL vs. DT présentent des difficultés pour répéter des non-mots jugés complexes phonologiquement (Dos Santos et Ferré, 2018 ; Coady et Evans, 2008 ; Estes et al., 2007). C’est notamment ce que nous avons trouvé dans notre article présent au point 6 avec une interaction significative entre complexité et présence de trouble uniquement pour les RNM basées sur les langues-spécifiques. Dans notre recherche, il est également possible que nous ayons trouvé des différences significatives plus importantes entre les enfants bilingues avec un DT et avec un TDL ainsi que des meilleures précisions diagnostiques aux RNM langues-spécifiques vs. *cross-linguistique* car les enfants maîtrisent deux langues proches phonologiquement. Nous évoquerons cette piste plus en détails dans la section d’ouverture (voir point 8.2). Au niveau de la recherche, l’intérêt de la RNM *cross-linguistique* a été démontré notamment puisqu’elle est utilisable par la plupart des langues (Antoniejevic-Elliott et al., 2019), ce qui la rend particulièrement intéressante lorsqu’il n’existe peu voire pas d’outils standardisés pour évaluer les enfants monolingues ou bilingues dans certaines langues. Des collectes de données sont en cours et une normalisation verra peut-être le jour suite au projet international de mise en commun des résultats obtenus en fonction de différents contextes linguistiques (voir informations additionnelles au point 6.1). Ainsi, l’intérêt de cette tâche *cross-linguistique* se verra également étendu au contexte clinique.

8.1.2 Sous-questions 2

Est-ce que nos deux situations d’évaluation dynamique permettent de révéler une différence de performance significative entre les enfants avec un DT et un TDL dans un contexte monolingue et bilingue ?

a Parmi nos deux situations d'évaluation dynamique, laquelle fournit la meilleure précision diagnostique ? Sera-t-elle en mesure d'améliorer la classification intermédiaire de nos sous-groupes ?

La Partie V de cette thèse s'est focalisée sur l'évaluation dynamique. Cette tâche, bien que très prometteuse, n'a pas encore été suffisamment explorée dans le contexte de l'identification d'un TDL. En effet, la méta-analyse récente d'Orellana et al. (2019) inclut seulement sept études portant sur le diagnostic de TDL au sein d'une population bilingue. Nous remarquons ainsi que les études à ce sujet sont moindres par rapport à la répétition de non-mots par exemple. De plus, dans leur méta-analyse, Orellana et al. (2019) ne recensent qu'une seule étude s'étant attelée à l'analyse de la morphosyntaxe, domaine fréquemment impacté dans les TDL (Bishop et al., 2017). Enfin, au sujet des tests d'apprentissage, seul le test pré-standardisé de Hasson et al. (Hasson, 2017 ; Hasson et al., 2013) semble présent dans le domaine de la logopédie. Ce sont ces différentes raisons qui nous ont poussé à nous y intéresser et à y consacrer un chapitre dans ce travail.

Nous avons mis sur pieds deux situations d'évaluation dynamique variant en termes d'activité (jeu autonome sur ordinateur et lecture d'histoires en interaction) et de procédure d'aide de la part de l'expérimentateur. Au sein de celles-ci, les enfants, assignés par groupes en référence à notre classification intermédiaire (voir Figure 4), étaient amenés à apprendre de nouveaux substantifs, verbes et structures syntaxiques (flexion verbale) et ils étaient ensuite évalués via plusieurs tâches (réceptives et productives). À travers nos deux situations d'évaluation dynamique, nous notons de nombreuses différences entre les enfants avec un DT et un TDL monolingues et bilingues de notre recherche avec des tailles d'effet moyennes et larges. Nous notons que les performances sont influencées en grande partie par la présence d'un TDL et d'une façon moindre par d'autres facteurs. Plus précisément, le statut linguistique

semble avoir moins d'impact que dans des tests statiques, ce qui confirme notre hypothèse à ce sujet et qui semble prometteur pour favoriser l'utilisation de l'évaluation dynamique dans le contexte clinique. Nous avons obtenu des pourcentages de précision diagnostique non convenables (Plante et Vance, 1994) pour notre situation de jeu sur ordinateur, ce qui diffère de notre hypothèse de départ. Plus précisément, les pourcentages de sensibilité sont adéquats mais la situation n'est que peu spécifique (peu d'enfants avec un DT sont rejetés par le test). Cette situation d'apprentissage plus autonome et avec moins de médiation semble avoir été plus complexe même pour les enfants avec un DT. À contrario, notre situation de lecture d'histoires en interaction et avec indices graduels s'est montrée suffisamment discriminante atteignant des seuils proches de 80 % de précision diagnostique (Plante et Vance, 1994). Dans un souci d'utilisation clinique, nous avons également tenté de calculer le pouvoir discriminant d'une version raccourcie de notre situation de lecture d'histoires en interaction (incluant seulement certaines tâches évaluatives). Ainsi, nous supposons que si cette tâche peut être intéressante pour compléter une batterie de tests par exemple, il ne sera pas possible de proposer dix tâches d'évaluation. De ce fait, nous avons analysé les résultats d'une version plus courte incluant les tâches d'évaluation d'évocation et de jugement d'acceptabilité toutes ayant montré des différences significatives entre les enfants avec un DT et un TDL. Les pourcentages de sensibilité et de spécificité sont également très intéressants puisque supérieurs à 80 % (Plante et Vance, 1994). Nous avons choisi les tâches d'évocation car elles évaluent la production langagière mais de façon plus souple que la dénomination (pas nécessaire d'associer le référent sémantique à l'item linguistique). Les tâches de jugement d'acceptabilité ont aussi été sélectionnées car elles ciblent les capacités réceptives et la précision des représentations phonologiques et sémantiques des enfants. Bien entendu, avec ces résultats prometteurs, il semblerait intéressant d'exécuter la situation dans cette version abrégée, mais nous ne l'avons

actuellement pas testée puisque tous les enfants de notre recherche ont accompli toutes les sous-tâches.

Dans notre article au chapitre 7, nous avons également relevé que les enfants de notre recherche ont appris peu d'éléments dans nos deux situations, notamment quand il s'agissait de produire les items (ex. tâches de dénomination et d'évocation). Nous pouvons supposer que l'utilisation des phonèmes communs aux deux langues (ex. uniquement trois voyelles) n'a pas facilité la tâche aux enfants et qu'ils puissent avoir eu de la difficulté dans la distinction des items et de leur mémorisation. En outre, nous pouvons noter des différences en fonction des procédures utilisées : lorsque de l'aide graduelle est fournie aux enfants (situation interactive de lectures d'histoires), les enfants apprennent et produisent davantage (voir chapitre 7). Quelques rares effets de statut linguistique ont été perceptibles dans les tâches de production avec un avantage des enfants monolingues, comme l'avait aussi trouvé certains auteurs (ex. Alt et al., 2013). Des tâches réceptives au sein d'une situation d'évaluation dynamique s'avèrent donc indispensables pour neutraliser ces effets. Nous avons supposé que les enfants bilingues auraient pu être avantagés dans les situations d'apprentissages en raison de leur supériorité dans les tâches nécessitant leurs fonctions exécutives (ex. Adescope et al., 2010 ; Blom et al., 2014) ; nos données confirment partiellement cette hypothèse avec quelques rares avantages des enfants bilingues dans nos situations d'évaluation dynamique. Cependant, nous pouvons confirmer que les résultats de nos deux situations d'évaluation dynamique sont en majorité influencés par le statut clinique des participants vs. leur statut linguistique même lorsque les non-mots sont intégrés dans un contexte de lecture d'histoires en français (alors que nous avions initialement supposé moins d'influence du statut linguistique dans la tâche de jeu sur ordinateur).

Dans nos analyses, nous nous sommes également penchées sur l'aide fournie par l'expérimentateur à l'enfant testé. Lorsque des indices graduels sont donnés dans notre situation

de lecture d'histoires, nous notons des différences entre les enfants avec un DT et un TDL, ces derniers nécessitant davantage d'aide dans les tâches d'évaluation de désignation et de dénomination, ce qui confirme notre hypothèse de départ. Les aides sont plus nombreuses mais également plus complexes que pour les enfants avec un DT. Par ailleurs, nous notons que malgré la quantité d'aide plus importante, les enfants avec un TDL s'améliorent également, voire plus parfois, grâce aux indices de l'expérimentateur. En effet, l'étayage fourni a été parfois plus profitable aux enfants avec un TDL car ils semblent habitués à celui-ci dans leurs différents contextes de vie, en clinique logopédique (de Weck et Marro, 2010 ; Weismer et Hesketh, 1993) et avec leur entourage familial (Ingold et al., 2008). De plus, puisqu'ils partent de scores inférieurs aux enfants avec un DT sans étayage, il est possible qu'ils s'améliorent davantage lorsque l'expérimentateur les aide. L'aide proposée par l'adulte se situait dans la zone proximale de développement des enfants (Vygotski, 1934/1985) et ils ont ainsi pu apprendre plus que quand ils étaient seuls face à la tâche (avant l'étayage ou dans la situation de jeu sur ordinateur). Au même titre, plusieurs recherches avaient attesté le recours à davantage d'aide de la part des enfants avec un TDL (ex. Camilleri et Law, 2007 ; Maragkaki et Hessels, 2017).

Les enfants avec un TDL de notre recherche présentaient des profils langagiers différents. Ils n'ont pas tous échoués de manière identique aux sous-tests statiques utilisés pour l'inclusion à notre étude, certains rencontrant des difficultés plus importantes en réception qu'en production, ou vice versa et dans un ou plusieurs domaines langagiers. Ces diverses difficultés se sont aussi reflétées dans nos situations d'évaluation dynamique qui avaient l'avantage de proposer plusieurs tâches évaluant plusieurs domaines (lexical et morphosyntaxique) et modalités langagières (production et compréhension). Au point 7.2, nous avons d'ailleurs relevé des corrélations entre les mesures obtenues aux sous-tests standardisés et les performances des enfants à la situation interactive de lecture d'histoires. Dans notre article

présent au chapitre 7, nous avons pu affirmer, que dans nos situations d'évaluation dynamique, c'est l'analyse de plusieurs domaines du langage en expression et réception qui a permis la meilleure précision diagnostique. L'utilisation de plusieurs tâches évaluatives a aussi permis de neutraliser les quelques rares influences du statut linguistique de nos participants, comme noté précédemment dans les tâches de production par exemple. À noter encore, que nous n'avons pratiquement pas trouvé d'influence du QI non-verbal des enfants sur leurs performances aux situations d'évaluation dynamique, au contraire d'autres études ou méta-analyses (ex. Kan et Windsor, 2010).

L'évaluation dynamique semble ainsi être une épreuve à utiliser davantage dans le contexte d'évaluation logopédique, notamment celle ayant recours à l'utilisation de non-mots dans le but d'éviter les biais linguistiques (Kapantzoglou et al., 2012 ; Petersen et al., 2020). Nos deux situations d'évaluation dynamique ont poursuivi un objectif de diagnostic (Hessels et Hessels-Schlatter, 2010), toutefois, même dans celles-ci, le logopédiste est en mesure de pouvoir observer les stratégies d'apprentissage de l'apprenant (Poehner, 2008), ses forces et ses faiblesses, le nombre et le type d'aide nécessaire pour mettre en lumière son potentiel d'apprentissage, ce qui constitue des éléments précieux également pour le traitement. En effet, il semble encore aujourd'hui que l'écart entre l'évaluation et l'intervention dans les contextes spécialisés (dont la logopédie) est conséquent (Tiekstra et al., 2016). L'évaluation dynamique se veut donc être une proposition pour rapprocher ces deux phases fondamentales du travail clinique (Hasson et Joffe, 2007 ; Tiekstra et al., 2016).

8.1.3 Synthèse au questionnement

Après nos analyses d'évaluation dynamique, nous avons pu mettre en lumière que cinq enfants pouvaient être classifiés différemment que notre regroupement intermédiaire l'ait proposé. En effet, trois enfants assignés au groupe TDL semblent mieux performer que

l'ensemble du groupe (deux enfants bilingues et un enfant monolingue) et deux enfants monolingues avec un DT éprouvent des difficultés. Il semble donc qu'au terme de toutes nos analyses (sous-tests statiques et situations d'évaluation dynamique), des doutes subsistent quant à nos sous-groupes d'enfants (avec ou sans TDL). Le regroupement établi à la fin de cette thèse a été nommé *complémentaire* (voir Figure 4 dans la section méthodologique) mais il s'agit de se questionner si l'évaluation dynamique doit être considérée comme une évaluation supplémentaire après la passation de sous-tests statiques ou si celle-ci peut venir remplacer la passation de ceux-ci. Il convient aussi de s'interroger sur la nature même du TDL puisque des difficultés de classification existent bel et bien pour les enfants bilingues et monolingues de notre recherche. Bien que des efforts aient été réalisés pour définir celui-ci (ex. Bishop et al., 2017) et pour le diagnostiquer (propos de cette thèse), il semble encore complexe de pouvoir affirmer clairement ce qu'il regroupe et comment l'identifier.

Afin d'obtenir le profil langagier le plus précis possible de l'enfant bilingue (Peña et al., 2016) et préciser encore notre classification, il aurait été cliniquement intéressant de combiner plusieurs tâches utilisées dans cette thèse (ex. l'étude de Boerma et Blom, 2017, combine plusieurs de ces tâches). En regard de la qualité diagnostique de certaines questions adressées aux parents, nous avons, au point 4, tenté une combinaison incluant la RNM en français et certaines questions adressées aux parents (âge de la phase combinatoire et inquiétude parentale) mais celle-ci ne s'est pas montrée convaincante dans notre recherche, puisque le pouvoir discriminant des deux tâches associées n'améliorait pas celui des tâches prises isolément. Nous aurions été tentées de tester une seconde combinaison avec les deux outils ayant la meilleure précision diagnostique de notre thèse, nous fournissant des pourcentages proches de l'excellence (Plante et Vance, 1994) : la RNM en français et la tâche d'évaluation dynamique de lecture d'histoires en interaction. Ces deux tâches proviennent toutes les deux d'un mode d'évaluation ne recourant pas ou peu à l'une ou l'autre des langues de l'enfant. L'intérêt de la

RNM et de la situation d'évaluation dynamique avec interaction est également qu'elles couvrent différents domaines langagiers pouvant être affectés par la présence d'un TDL. En effet, les enfants rencontrant des fragilités phonologiques (précision et/ou mémoire à court-terme verbale) ont éprouvé des difficultés dans la RNM en français. Toutefois, l'évaluation d'un seul domaine langagier n'est pas suffisante puisque certains enfants avec un TDL ne présentent pas de difficultés phonologiques. Notre situation d'évaluation dynamique de lecture d'histoires a permis quant à elle de mettre en avant des difficultés dans différentes modalités (réceptives et productives) et dans différents domaines (lexicales et morphosyntaxiques) des enfants avec un TDL de notre recherche. Toutefois, puisque les nouveaux apprentissages langagiers sont fortement influencés par de multiples facteurs tels que la motivation ou encore les ressources attentionnelles et intellectuelles (voir Kan et Windsor, 2010 ; Kucker et al., 2015), nous aurions privilégié la passation des deux tâches. À notre connaissance, nous aurions été les premières à tenter une combinaison de ces deux outils. Toutefois, puisque la RNM en français a été utilisée comme sous-test pour classer les enfants de notre thèse, nous avons renoncé à combiner la RNM à notre situation d'évaluation dynamique avec interaction pour dégager leurs pourcentages de précision diagnostique cumulés. Cependant, au terme de notre travail, nous pouvons mentionner que l'utilisation de la RNM en français et l'utilisation d'une situation d'évaluation dynamique (lecture d'histoires en interaction) peuvent, à ce jour, constituer des outils efficaces pour identifier un TDL chez une population monolingue et bilingue français-portugais puisque leurs pourcentages de précision diagnostique sont élevés (Plante et Vance, 1994). Pour les bilans d'enfants bilingues français-portugais, une de nos suggestions, à la fin de cette thèse, serait donc de proposer une évaluation dynamique de lectures d'histoires et une tâche de RNM pour identifier la présence d'un TDL. Il s'agirait donc de combiner sous-test statique (ex. RNM) et évaluation dynamique. Il serait ensuite cliniquement intéressant d'effectuer des sous-tests dans les deux langues de l'enfant afin

d'affiner les connaissances que l'enfant possède dans ses deux langues dans une visée plus d'intervention.

Au terme de cette thèse de doctorat et pour fournir encore quelques pistes de réponses à notre questionnement général, nous pouvons mentionner que les spécificités des enfants bilingues doivent être nécessairement prises en considération pour leur évaluation langagière. Effectivement, plusieurs modes d'évaluation ont été testés au cours de ce travail et seul un mode d'évaluation incluant à minima, voire aucune des langues de l'enfant bilingue et faisant usage de non-mots permet de neutraliser les biais linguistiques et d'associer l'échec aux tâches au statut clinique des enfants bilingues de notre recherche. Toutefois, nous notons également que les tâches à préconiser dans le bilan de ces enfants ne sont pas complètement différentes de celles proposées aux enfants monolingues. En effet, les tâches de RNM ont également montré leur intérêt pour évaluer les enfants monolingues de différentes langues (Coady et Evans, 2008). D'ailleurs, la RNM en français utilisée dans ce travail est issue d'un test monolingue français qui est disponible sur le marché francophone depuis plusieurs années. Les questions posées aux parents sont également toujours présentes dans le cas de l'anamnèse des enfants monolingues. Enfin, l'évaluation dynamique a également montré son intérêt dans une évaluation monolingue (ex. Camilleri et Law, 2007). Les données obtenues à la fin de ce travail sont donc très prometteuses pour les bilans logopédiques des enfants parlant une ou plusieurs langues.

8.2 Limites et ouvertures

Une des limites importantes de notre travail concerne le choix de nos tâches. Nous avons essayé d'explorer le plus d'outils et méthodes de cotation possibles pour évaluer les enfants bilingues. Toutefois, nous avons laissé de côté certaines tâches, bien qu'elles paraissent également très prometteuses. Par exemple, l'utilisation d'une tâche narrative (ex. Boerma et

Blom, 2017) aurait également pu faire partie de notre inventaire de tâche à utiliser et tester auprès de notre population. Cette tâche aurait pu être proposée dans l'une (ex. Rezzonico et al., 2015) ou les deux langues (ex. Altman et al., 2016) de l'enfant bilingue. Une combinaison incluant une tâche narrative et une RNM a notamment montré son potentiel d'identification des enfants bilingues avec un TDL dans l'étude de Boerma et Blom (2017). Ainsi, une piste d'ouverture pourrait être d'observer si l'ajout d'une tâche narrative à notre évaluation dynamique de lecture d'histoires améliorerait nos pourcentages de sensibilité et spécificité dans l'identification d'un TDL.

Parmi les tâches issues de tests standardisés, nous avons initialement choisi d'intégrer une répétition de phrases, dont l'échec semble également attribuable au statut clinique des enfants vs. statut linguistique (ex. Fleckstein et al., 2018). Toutefois, après notre étude pilote, nous avons renoncé à celle-ci puisque la seule tâche de répétition de phrases lusophone normée était très longue (plus de 40 phrases à répéter) et nous avons ainsi obtenu des effets planchers pour les enfants avec un DT et un TDL. Nous aurions également pu tester l'utilisation de la répétition de phrases en français dont les phénomènes linguistiques spécifiquement difficiles aux enfants avec un TDL de différentes langues ont été contrôlés (outil LITMUS, Marinis et Armon-Lotem, 2015). Des précédentes études (de Almeida et al., 2016 ; Fleckstein et al., 2017) ont montré son intérêt dans l'utilisation des évaluations bilingues. Comme pour la tâche narrative, il serait possible dans un futur travail de pouvoir tester une combinaison incluant une tâche de répétition de phrases et notre évaluation dynamique de lecture d'histoires.

Nous avons également fait le choix d'utiliser une version raccourcie et à l'écrit du questionnaire PabiQ (Tuller, 2015). Nous avons laissé de côté certains éléments présents dans le questionnaire initial tel que la qualité d'exposition et la quantité d'exposition de façon cumulative. Notre but était d'avoir des informations au sujet de certains facteurs de risque (ex. présence de difficultés dans l'entourage familial ou encore inquiétude parentale) et de connaître

le taux d'exposition actuel aux langues de l'enfant. En effet, ces informations étaient intéressantes pour nous uniquement dans la Partie III de ce travail, dans laquelle, nous avons, de toute manière, procédé à une évaluation dans les deux langues des enfants et au moyen de plusieurs tâches recourant à des non-mots dont le taux d'exposition a peu d'influence. Toutefois, nous sommes conscientes que les informations que nous avons omises auraient pu nous guider dans l'interprétation de certaines données. Par exemple, en sachant si l'enfant bénéficie fréquemment de moments de lecture partagée, nous aurions pu supposer qu'il soit plus familier à la situation d'évaluation dynamique de lecture d'histoires (Partie V). De plus, nous savons également que la dominance linguistique est dynamique et qu'elle varie en fonction des moments, des contextes de vie ainsi que des domaines langagiers (Bedore et al., 2012), toutefois, nous avons fait le choix de nous intéresser à la dominance actuelle puisque nous avons fait usage de plusieurs modes et tâches d'évaluation et que notre intérêt principal était l'évaluation dynamique. Nous supposons donc que nos décisions méthodologiques n'ont pas préterité les performances des enfants bilingues.

Plusieurs facteurs n'ont pas été pris en considération dans nos analyses de RNM (chapitre 6) alors qu'ils peuvent être source d'influence pour les enfants, c'est le cas de l'effet de ressemblance des non-mots de la tâche de RNM en français à des mots existants (*neighborhood density*, McKean et al., 2013). Ce risque existe bel et bien lorsque les items sont créés sur la phonotactique d'une langue. Toutefois, nous ne supposons pas que les enfants de notre recherche aient été influencés par cette ressemblance puisque la tâche ne s'est pas montrée facile pour eux quel que soit leur groupe d'appartenance (DT ou TDL ; monolingue bilingue) et que la tâche s'est révélée très sensible et spécifique. Nous avons privilégié une analyse de l'effet de longueur et de la complexité phonologique comme cela est fréquemment le cas dans les études portant sur la RNM (notre revue systématique au chapitre 5 atteste du nombre élevé d'études s'y étant intéressé).

Nos situations d'évaluation dynamique ont été conçues pour proposer de nouveaux apprentissages dans le domaine du lexique (substantifs et verbes) et de la (morpho)syntaxe (règle de flexion verbale). Nous sommes conscientes que la phonologie a eu un impact dans nos situations d'apprentissage. En effet, la mémoire à court-terme verbale a été fortement mobilisée chez les enfants, de même que nous nous sommes intéressées à la précision de leurs représentations phonologiques (ex. dans la tâche de jugement d'acceptabilité) et que nous avons intégré une règle morphologique pour les aider dans leur apprentissage (voir Alt et al., 2004 ; Gray, 2003). Nous avons décidé de ne pas tester l'apprentissage dans le domaine phonologique car les tâches d'apprentissages de nouveaux sons sont complexes pour des enfants de l'âge de notre étude (ex. la population de l'étude de Fuhrmeister et al., 2020, est composée d'enfants dès 10 ans et d'adultes monolingues) et que nous avons voué un intérêt aux RNM (Partie IV) s'attellant à tester la mémoire à court-terme verbale et la précision phonologique.

Nous ne nous sommes pas intéressées aux implications culturelles de la combinaison de langues choisie dans la thèse. Les enfants bilingues de notre recherche étaient pour une partie bilingues français-portugais européens et pour les autres français-portugais brésiliens. La culture de ces enfants peut jouer un rôle dans leurs habitudes et manières de prendre part aux activités. La culture est une variable complexe tout comme le caractère des enfants, qu'ils soient monolingues ou bilingues. En effet, certains enfants de notre recherche se sont montrés plus expressifs lors des situations d'évaluation dynamique en produisant plus de mots même s'ils n'en n'étaient pas sûrs. Alors que d'autres enfants, plus analytiques, ont préféré peu voire pas produire d'items par peur de faire faux (Bates, 1988). L'apprentissage de nouveaux mots semble ainsi multifactoriel (ex. items, tâches, âge, motivation et caractère de l'enfant, voir Kucker et al., 2015). L'influence du caractère et du statut linguistique l'enfant (voir article

présenté au point 7) sur ses performances de production, encourage également le recours à des évaluations réceptives (ex. via une désignation) qui semblent moins affectées par ces biais potentiels. Nous notons également que les performances aux situations d'évaluation dynamique ont parfois été influencées par les ressources attentionnelles des enfants. Cela a notamment été le cas dans les tâches de jugement d'acceptabilité dans lesquelles les enfants devaient juger si la réponse était correcte ou incorrecte. Il est parfois arrivé que les enfants fassent usage de stratégies qui peuvent être basées sur le hasard. La situation de jeu sur ordinateur a été plus propice à la mise en place de telles stratégies puisque l'enfant apprenait de façon autonome. La forme de l'enfant le jour de la passation a aussi exercé une influence et nous supposons également que l'interaction avec l'expérimentateur (dans notre situation de lectures d'histoires) a permis d'enrôler l'enfant dans la tâche, ce qui n'a pas été possible lorsque l'enfant était plus autonome (jeu sur ordinateur). Nous pouvons supposer que ce sont pour ces différentes raisons que nous avons trouvé un avantage de la situation interactive de lecture d'histoires vs. jeu autonome sur ordinateur dans notre article présent au chapitre 7.

Notre tâche d'évaluation dynamique avec interaction a montré son potentiel dans l'identification des enfants monolingues et bilingues avec un TDL. L'expérimentateur faisant passer la tâche dans notre recherche était le même pour tous les enfants. Alors bien qu'un script des histoires à raconter ainsi que de l'étayage à fournir aient été rédigés, il serait intéressant de répliquer l'étude avec différents expérimentateurs et voir si les pourcentages de sensibilité et de spécificité restent autant élevés, étape nécessaire pour s'assurer de la fidélité inter-juges dans la conception d'un test par exemple. En effet, nous venons de mentionner que le style d'apprentissage des enfants influe sur les résultats aux tâches mais il existe également le style des logopédistes dans la façon de faire passer les évaluations et de se rendre actif dans celles-ci qui constitue un facteur d'influence à ne pas négliger. Il est également important de noter qu'une autre limite de ce travail est que l'auteur de cette thèse n'ait pas été aveugle au statut

des participants. En effet, elle était avertie de leur groupe diagnostic et elle était parfois la logopédiste clinicienne des enfants avec un TDL. Il est, par exemple, tout à fait possible que ses connaissances aient exercé une influence sur le nombre ou type d'étayage dans la situation d'évaluation dynamique de lecture d'histoires.

Une limite ou paradoxe concerne le fait que nous nous sommes basées sur une classification des enfants avec un TDL imparfaite mais nécessaire au début de notre recherche et que nous ayons utilisé certaines tâches jugées comme peu discriminantes, dans notre étude, pour inclure ou exclure les enfants de nos deux groupes expérimentaux à notre recherche portant sur l'évaluation dynamique. Plus précisément, bien que les enfants avec un TDL aient été identifiés comme tel par les logopédistes praticiens et qu'ils soient suivis depuis au moins trois mois en logopédie, nous avons mis en lumière les difficultés pour se positionner sur ce type de trouble chez une population bilingue. Nous avons donc ensuite cherché à confirmer le TDL en utilisant des tâches standardisées provenant de tests en français et en portugais avec des normes monolingues. Cependant, voici ici tout le paradoxe de notre démarche puisque nous avons, tout au long de ce travail, mentionné que ces mesures n'étaient que peu discriminantes. Nous avons donc décidé de suivre une ligne de conduite stricte en excluant tous les enfants n'ayant pas répondu comme attendu aux sous-tests langagiers pour nos situations d'évaluation dynamique. Plus précisément, plusieurs enfants ($n = 12$) ont été exclus pour nos analyses portant sur l'évaluation dynamique (voir article au point 7) puisqu'ils ne répondaient pas comme attendus aux sous-tests statiques. À la fin de ce travail, nous avons donc tenté de proposer une meilleure classification des enfants que celle initialement utilisée qui s'est avérée imparfaite. Toutefois, au terme de ce travail, des doutes subsistent quant à la classification de nos sous-groupes (voir aussi 8.1.3). Nous ne sommes pas les seules à être confrontées à ce paradoxe puisque la plupart des recherches sur l'identification des TDL reposent sur ce type de

démarche méthodologique. À plus grande échelle que cette thèse, il serait possible de choisir de s'intéresser à des cohortes d'enfants sans avoir de renseignements sur leurs difficultés, ainsi nous pourrions, chez certains enfants, identifier des difficultés langagières et éventuellement suspecter la présence d'un TDL (comme Archibald et Joanisse, 2009, avec la RNM). Il serait également possible de mener une étude longitudinale afin de réaliser une analyse prospective et déterminer si les enfants possèdent véritablement des difficultés persistantes s'apparentant à un TDL.

Une piste d'ouverture serait de reproduire cette étude en analysant une combinaison de langues moins proches phonologiquement que celle du français et du portugais. Nous pourrions ainsi voir si les résultats obtenus, notamment en ce qui concernent la RNM et l'évaluation dynamique, sont toujours aussi prometteurs. Pour la RNM en français, nous pourrions nous attendre à des résultats différents si les enfants bilingues partageaient d'autres langues. Nous pouvons, par exemple, supposer que des enfants avec un DT et un TDL maîtrisant le français et l'albanais (langue balkanique fréquemment parlée en Suisse, Office Fédéral de la Statistique, 2020) puissent avoir du mal à répéter les non-mots de la RNM en français. Durant ce travail de doctorat, nous avons laissé un peu de côté les débats portant sur la RNM et son évaluation de la mémoire à court-terme verbale et/ou de traitement phonologique (voir Coady et Evans, 2008) puisque la combinaison de langues partagée par notre population était proche phonologiquement. Toutefois, dans d'autres cas de figure, comme celui du français-albanais, il semble que des tâches évaluant la mémoire à court-terme verbale puissent être plus discriminantes que celles impliquant le traitement phonologique et ayant été conçues sur des langues spécifiques. En effet, les caractéristiques phonologiques des non-mots créés sur la base d'une langue spécifique sont susceptibles d'influencer les performances à la RNM pour des enfants parlant des langues moins proches phonologiquement. Dans ce cas de figure, les

variations en termes de longueur qui semblent universelles, constituent un argument pour penser que la RNM *cross-linguistique* (Chiat, 2015) s'avérerait peut être plus intéressante cliniquement à exploiter qu'elle l'a été pour les enfants bilingues français-portugais de notre recherche. Sur la base de notre conception des stimuli de nos situations d'évaluation dynamique, des enfants bilingues français-albanais pourraient également rencontrer des difficultés. Il faudrait alors faire des adaptations, en essayant de s'assurer que les items soient composés de phonèmes communs aux langues et respectent les règles phonotactiques de celles-ci. Il serait également nécessaire de vérifier si la règle de flexion verbale que nous avons créée s'avérerait pertinente pour ces enfants bilingues, c'est-à-dire en s'assurant que ce type de règle n'existe pas déjà en albanais.

Nous souhaitons également formuler quelques perspectives de notre travail, en particulier au sujet de l'évaluation dynamique. Puisque nous avons obtenu des pourcentages de précision diagnostique élevés, nous supposons que nos situations d'évaluation dynamique puissent intéresser des concepteurs de tests logopédiques. Nous serions disposées à nous plonger dans le long processus de construction d'un test. Hormis, le test pré-standardisé DAPPLE (Hasson, 2017 ; Hasson et al., 2013), il n'existe, à notre connaissance, aucun test ou batterie portant sur l'évaluation dynamique en clinique logopédique, ce qui laisse la place à de nouveaux outils pour se développer.

9. Conclusion

Au travers de cette thèse de doctorat, nous avons mis en lumière le travail d'évaluation rendu complexe pour les logopédistes confrontés à une population bilingue de plus en plus hétérogène partageant des combinaisons de langues peu étudiées dans la littérature scientifique. Nous nous sommes intéressées à des tâches fréquemment employées dans le contexte clinique et disponibles dans différentes batteries de tests (ex. RNM) et nous avons également voué un intérêt particulier à des tâches encore moins utilisées telles que l'évaluation dynamique. Nous avons testé et comparé leur efficacité auprès d'une population monolingue et bilingue-français portugais, fréquente en Suisse romande (Office Fédéral de la Statistique, 2020). Finalement, nous avons tenté de dresser une liste des outils et méthodes ayant montré leur intérêt dans l'identification des TDL auprès de cette population. Nous espérons ainsi pouvoir aider les professionnels à appréhender de manière plus sereine les bilans de ces enfants.

Le choix de réaliser une thèse par articles s'est opéré en espérant que nos papiers publiés puissent avoir clairement un impact sur les pratiques des professionnels et notamment des logopédistes recevant ces enfants en bilan logopédique. Nous pensons également que nos recherches serviront de bases pour des études plus larges portant sur l'évaluation langagière des enfants bilingues. En effet, nos recherches ont mis en lumière la nécessité de la poursuite des recherches scientifiques sur ce sujet et concernant des combinaisons de langues moins étudiées.

Bibliographie

- Adescope, O., Thompson, T. et Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, 80(2), 207-245. <https://10.3102/0034654310368803>
- Alt, M. et Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 49(5), 941–954. [https://doi:10.1044/1092-4388\(2006/068\)](https://doi:10.1044/1092-4388(2006/068))
- Alt, M., Meyers, C. et Figueroa, C. (2013). Factors that influence fast mapping in children exposed to Spanish and English. *Journal of Speech, Language, and Hearing Research*, 56(4), 1237-1248. [https://doi:10.1044/1092-4388\(2012/11-0092\)](https://doi:10.1044/1092-4388(2012/11-0092))
- Altman, C., Armon-Lotem, S., Fichman, S. et Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English-Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, 37(1), 165-193. <https://doi:10.1017/S0142716415000466>
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)* (5^e éd.). American Psychiatric Association Publishing.
- American Speech-Language-Hearing Association (s. d). *Bilingual Service Delivery*. Repéré le 18 mars 2020 à, <https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589935225§ion=Overview>
- Animatron (2018). *Animatron studio*. <https://www.animatron.com/studio>
- Antonišević-Elliott, S., Lyons, R., O' Malley, M. P., Meir, N., Haman, E., Banasik, N., Carroll, C., McMenamin, R., Rodden, M. et Fitzmaurice, Y. (2019). Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics & Phonetics*, 34(4), 293-311. <https://doi:10.1080/02690940.2019.1644444>

- Archibald, L. et Joanisse, M. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, 52(4), 899-914. [https://doi.org/10.1044/1092-4388\(2009/08-0099\)](https://doi.org/10.1044/1092-4388(2009/08-0099))
- Armon-Lotem, S. (2012). Introduction: Bilingual children with SLI – the nature of the problem*. *Bilingualism: Language and Cognition*, 15(1), 1-4. <https://doi.org/10.1017/S1366728911000599>
- Armon-Lotem, S. et de Jong, J. (2015). Introduction. Dans S. Armon-Lotem, K. de Jong, et N. Meir (éd.), *Assessing Multilingual Children* (p. 1-25). Multilingual Matters.
- Armon-Lotem, S. et Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, 51(6), 715-731. <https://doi.org/10.1111/1460-6984.12242>
- Avenet, S., Lemaître, M. -P. et Vallée, L. (2016). DSM5 : quels changements pour les troubles spécifiques du langage oral. *Neuropsychiatrie de l'enfance et de l'adolescence*, 64(2), 81-92. <https://doi.org/10.1016/j.neurenf.2015.12.002>
- Bates, E., Bretherton, I. et Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press.
- Bedore, L., Peña, E., Summers, C., Boerger, K., Resendiz, M., Greene, J., Bohman, T. et Gillam, R. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism*, 15(3), 616-629. <https://doi.org/10.1017/S1366728912000090>

- Betz, S., Eickhoff, J. et Sullivan S. (2013). Factors influencing the Selection of Standardized Tests for the Diagnosis of Specific Language Impairment. *Language, Speech, and Hearing Services in Schools*, 44, 133-146.
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A. et Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068-1080. <https://doi.org/10.1111/jcpp.12721>
- Blom, E., Küntay, A. C., Messer, M., Verhagen, J. et Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, 128(1), 105-119. <http://dx.doi.org/10.1016/j.jecp.2014.06.007>
- Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F. et Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International journal of Language & Communication Disorders*, 51(6), 626-638. <https://doi:10.1111/1460-6984.12234>
- Boerma, T. et Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, 66(1), 65-76. <https://doi.org/10.1016/j.jcomdis.2017.04.001>
- Bortolini, U., Arfé, B., Caselli, C., Degasperi, L., Deevy, P. et Leonard, L. (2006). Clinical markers for specific language impairment in Italian: The contribution of clitics and non- word repetition. *International journal of Language & Communication Disorders*, 41(6), 695- 712. <https://doi:10.1080/13682820600570831>
- Breault, C., Béliveau, M. -J., Labelle, F., Valade, F. et Trudeau. (2019). Le trouble développemental du langage (TDL) : Mise à jour interdisciplinaire. *Neuropsychologie*

Clinique et Appliquée, 3, 46-63. https://oaaqvousinforme.com/wp-content/uploads/2019/10/NCACN-Volume3_final-coup-de-chapeau.pdf

- Camilleri, B. et Law, J. (2007). Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology*, 9(4), 312-322. <https://doi.org/10.1080/14417040701624474>
- Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. Dans C. S. Lidz (dir.), *Dynamic assessment: An interactional approach to evaluating learning potential* (p. 82–115). Guilford Press.
- Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R. et Coltheart, M. (2007). *Provas de avaliação da linguagem e da afasia em português*. CEGOC.
- Chiat, S. (2015). Non-Word Repetition. Dans S. Armon-Lotem, K. de Jong, et N. Meir (dir.), *Assessing Multilingual Children* (p. 125-150). Multilingual Matters.
- Chiat, S. et Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research*, 59(1), 1179-1189. https://doi:10.1044/2016_JSLHR-L-15-0293
- Coady, J. et Evans, J. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *Journal of Language & Communication Disorders*, 43(1), 1-40. <https://doi:10.1080/13682820601116485>
- Conti-Ramsden, G. et Botting, N. (1999). Classification of children with specific impairment: Longitudinal considerations. *Journal of Speech, Language, and Hearing Research*, 42(5). <https://doi:10.1044/jslhr.4205.1195>
- De Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L. et Zebib, R. (2016). L'identification d'enfants bilingues avec trouble spécifique du langage en France. *SHS Web of Conferences*, 27(10005), 1-16. <https://doi:10.1051/shsconf/20162710005>

- De Cat, C., Gusnanto, A. et Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, 40(1), 119-151. <https://doi.org/10.1017/S0272263116000486>
- De Montfort Supple, M. et Söderpalm E. (2010). Child Language disability: A historical review. *Topics in Language Disorders*, 30(1). <https://doi.org/10.1097/TLD.0b013e3181d0a13e>
- Denman, D., Speyer, R., Munro, N., Pearce, W., Chen, Y. W. et Cordier, R. (2017). Psychometric properties of language assessments for children aged 4-12 years: a systematic review. *Frontiers in Psychology*, 1515(8), 1-28. <https://doi.org/10.3389/fpsyg.2017.01515>
- De Weck, G. et Marro, P. (2010). *Les troubles du langage chez l'enfant. Description et évaluation*. Elsevier Masson.
- Dos Santos, C. et Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition*, 25(1), 58-71. <https://doi.org/10.1080/10489223.2016.1243692>
- Dubreuil-Piché, S., Mayer-Crittenden, C. et Lachance, J. (2017). Nonword repetition and sentence imitation as clinical markers for primary language impairment in bilingual French-English- and English-French speaking children in Northern Ontario: A narrative review. *Diversity of research in health journal* 1(1), 61-69. <https://pubs.biblio.laurentian.ca/index.php/drhj-rdrs/article/view/19>
- Estes, K. G., Evans, J. et Else-Quest, N. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177-195. <https://doi.org/10.1092-4388/07/5001-0177>
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5^e éd). Sage Edge.

- Filmora Screen (s.d). *Wondershare Filmora for Macintosh*, [Computer software].
<https://filmora.wondershare.net>
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E. et Zebib, R. (2018). How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition*, 25(1), 85-101. <https://doi.org/10.1080/10489223.2016.1192635>
- Fuhrmeister, P., Schlemmer, B. et Myers, E. B. (2020). Adults show initial advantages over children in learning difficult nonnative speech sounds. *Journal of Speech, Language, and Hearing Research*, 63(8), 2667-2679. https://doi.org/10.1044/2020_JSLHR-19-00358
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balcuiniene, I., Bohnacker, U. et Walters, J. (2015). Assessment of Narrative Abilities in Bilingual Children. Dans S. Armon-Lotem, K. de Jong et N. Meir (dir.), *Assessing Multilingual Children*. (p. 243-277). Multilingual Matters.
- Girbau, D. et Schwartz, R. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, 42(1), 59-75. <https://doi:10.1080/13682820600783210>
- Glutting, J., Adams, W. et Sheslow, D. (2000). *Wide range intelligence test*. Wide Range.
- Gray, S. (2003). Word-Learning by Preschoolers with Specific Language Impairment: What Predicts Success? *Journal of Speech, Language, and Hearing Research*, 46(1), 56-57.
[https://doi:10.1044/1092-4388\(2004/083\)](https://doi:10.1044/1092-4388(2004/083))
- Grosjean, F. (2015). *Parler plusieurs langues. Le monde des bilingues*. Albin Michel.
- Hambly, H., Wren, Y., McLeod, S. et Roulstone, S. (2013). The influence of bilingualism on speech production: A systematic review. *International Journal of Communication Disorders*, 48(1), 1-24. <https://doi:10.1111/j.1460-6984.2012.00178.x>

- Hadley, P. A. et Rice, M. L. (1991). Conversational responsiveness in speech and language-impaired children. *Journal of Speech and Hearing Research*, 34(6), 1308-1317.
<https://doi.org/10.1044/jshr.3406.1308>
- Hasson, N. (2017). *The dynamic assessment of language learning*. Routledge.
- Hasson, N. et Joffe, V. (2007). The case for Dynamic Assessment in speech and language therapy. *Child Language Teaching and Therapy*, 23(1), 9–25.
<https://doi:10.1177/0265659007072142>
- Hasson, N. et Botting, N. (2010). Assessment of children with language impairments: A pilot study. *Child Language Teaching and Therapy*, 26(3), 249-272.
<https://doi:10.1177/0265659009349982>
- Hasson, N., Camilleri, B., Jones, C., Smith, J. et Dodd, B. (2013). Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Language Teaching and Therapy*, 29(1), 57–75. <https://doi:10.1177/0265659012459526>
- Helloin, M. C. et Thibault, M. P. (2006). *Exalang 3-6*. Happyneuron.
- Hessels, M. G. P. et Hessels-Schlatter, C. (2010). L'apport des tests d'apprentissage dans l'évaluation des capacités intellectuelles des élèves en difficultés scolaires. Dans M. G. P. Hessels & C. Hessels-Schlatter (dir.). *Évaluation et intervention auprès d'élèves en difficultés* (p. 5-35). Peter Lang.
- IBM Corp. (2020). *IBM SPSS Statistics for Macintosh* (Version 27.0) [Computer software]. IBM Corp.
- Ingold, J., Gendre, S., Rezzonico, S., Corlateanu, C. et Da Silva, C. (2008). Diversité des étayages des mères d'enfants tout-venant et dysphasiques dans deux situations d'interaction. *Travaux neuchâtelois de linguistique*, 49(1), 69-82.
http://www.unine.ch/files/live/sites/tranel/files/Tranel/49/Ingoldetal_69-82_corr.pdf

- Kambanaros, M. et Grohmann, K. (2015). More general all-purpose verbs in children with specific language impairment? Evidence from Greek for not fully lexical verbs in language development. *Applied Psycholinguistics*, 36(5), 1-29.
<https://doi:10.1017/S0142716414000034>
- Kan, P. et Kohnert, K. (2008). Fast mapping by bilingual preschool children. *Journal of Child Language*, 35(3), 495-514. <https://doi:10.1017/S0305000907008604>
- Kan, P. et Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 53(3), 739–756.
[https://doi:10.1044/1092-4388\(2009/08-0248\)](https://doi:10.1044/1092-4388(2009/08-0248))
- Kapalková, S., Polisenská, K. et Vicenová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: novel scoring methods. *International Journal of Language and Communication Disorders*, 48(1), 78-89.
<https://doi:10.1111/j.1460-6984.2012.00189.x>
- Kapantzoglou, M., Restrepo, A. et Thompson, M. (2012). Dynamic Assessment of Word Learning Skills: Identifying Language Impairment in Bilingual Children. *Language, Speech, and Hearing Services in School*, 43(1), 81-96. [https://doi:10.1044/0161-1461\(2011/10-0095\)](https://doi:10.1044/0161-1461(2011/10-0095))
- Kehoe, M. (2009). Plurilinguisme et logopédie: le défi de l'évaluation, *Langage et pratiques*, 44, 69-82
https://arld.ch/fileadmin/user_upload/Documents/ARLD/WWW/Editeurs/Logopediste_s/Langages_pratiques/44-textes.pdf
- Khamsi, A. (2001). *Évaluation du langage oral*. ECPA
- Kihm, A. (2016). *Fiche langue le portugais européen*. Site du CNRS.
https://lgidf.cnrs.fr/sites/lgidf.cnrs.fr/files/images/PORTUGAIS_24_06_19_A4.pdf
- Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence,

- and implications for clinical actions. *Journal of communication disorders*, 43(6), 456-473. <https://doi.org/10.1016/j.jcomdis.2010.02.002>
- Kohnert K. et Danahy K. (2007). Young L2 learners' performance on a novel morpheme task. *Clinical Linguistics & Phonetics*, 21(7), 557–569. <https://doi:10.1080/02699200701374231>
- Koo, T. K. et Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), 155-163.
- Kucker, S., McMurray, B. et Samuelson, L. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child development perspectives*, 9(2), 74-78. <https://doi:10.1111/cdep.12110>
- Leonard, L. (2014). *Children with specific language impairment* (2^e éd.). A Bradford book.
- Maragkaki, I. et Hessels, M. (2017). A pilot study of dynamic assessment of vocabulary in German for bilingual preschoolers in Switzerland. *Journal of Studies in Education*, 7(1), 32-49. <https://doi:10.5296/jse.v7i1.10392>
- Marini, A., Ruffino, M., Sali, M. E. et Molteni, M. (2017). The role of phonological working memory and environmental factors in lexical development in Italian-speaking late talkers: A one-year follow-up study. *Journal of Speech, Language, and Hearing Research*, 60(12), 3462-3473. https://doi.org/10.1044/2017_JSLHR-L-15-0415
- Marinis, T. et Armon-Lotem, S. (2015). Sentence Repetition. Dans S. Armon-Lotem, K. de Jong, et N. Meir (dir.), *Assessing Multilingual Children* (p. 95-122). Multilingual Matters.
- McKean, C., Letts, C. et Howard, D. (2013). Developmental change is key to understanding primary language impairment: The case of phonotactic probability and nonword repetition. *Journal of Speech, Language, and Hearing Research*, 56(5), 1579–1594. [https://doi.org/10.1044/1092-4388\(2013/12-0066\)](https://doi.org/10.1044/1092-4388(2013/12-0066))

- McGregor, K., Oleson, J., Bahnsen, A. et Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Communication Disorders*, 48(3), 307-319. <https://doi:10.1111/1460-6984.12008>
- Nounproject (2018). *Nounproject*. <https://thenounproject.com>
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G. et Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, 57(11), 1247-1257. <https://doi.org/10.1111/jcpp.12573>
- Oetting, J., Rice, M. et Swank, L. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research*, 38(2), 434-445. <https://doi:10.1044/jshr.3802.434>
- Office Fédéral de la Statistique. (2020, 27 février). *Langues principales les plus fréquentes*. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/langues-religions/langues.assetdetail.11947533.html>
- Orellana, C., Wada, R. et Gillam, R. (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
- O'Toole, C., Gatt, D., Hickey, T. M., Miękisz, A., Haman, E., Armon-Lotem, S., Rinker, T., Ohana, O., Dos Santos, C. et Kern, S. (2017). Parent report of early lexical production in bilingual children: A cross-linguistic CDI comparison. *International Journal of Bilingual Education and Bilingualism*, 20(2), 124-145. <https://doi:10.1080/13670050.2016.1179258>
- Paint (2018). *Paint for Macintosh*. [Computer software]. Microsoft.

- Paradis, J., Genesee, F. et Crago, M. B. (2011). *Dual language development & disorders: A handbook on bilingualism and second language learning* (2^e éd.). Brookes.
- Paul, R. et Norbury, C. (2012). *Language Disorders from infancy through adolescence*. Elsevier.
- Pearson, Z. B. (2013). Distinguishing the bilingual as a late talker from the late talker who is bilingual. Dans L. Rescorla et P. Dale (dir.), *Late talkers* (p. 67-91). Paul Brookes.
- Peña, E., Iglesias, A. et Lidz, C. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology*, 10(2), 138-154. [https://doi.org/10.1044/1058-0360\(2001/014\)](https://doi.org/10.1044/1058-0360(2001/014))
- Peña, E., Reséndiz, M., & Gillam R. (2007). The role of clinical judgements of modifiability in the diagnosis of language impairment. *Advances in Speech-Language Pathology*, 9(4), 332-345. <https://doi.org/10.1080/14417040701413738>
- Peña, E. D., Gillam, R. et Bedore, L. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *Journal of Speech, Language, and Hearing Research*, 57(6), 2208-3330. https://doi.org/10.1044/2014_JSLHR-L-13-0151
- Peña, E. D., Bedore, L. M. et Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders*, 51(2), 192-202 <https://doi.org/10.1111/1460-6984.12199>
- Pérez, M. (2000). Les enjeux de la lusophonie. Le portugais, langue de communication internationale. *Lusotopie*, 7(1), 610-620. https://www.persee.fr/doc/luso_1257-0273_2000_num_7_1_1402
- Petersen, D. B, Tonn, P., Spencer, D. et Foster, M. E (2020). The classification accuracy of a dynamic assessment of inferential word learning for bilingual English/Spanish-speaking

- school-age children. *Language, Speech, and Hearing Services in Schools*, 51(1), 144-164. https://doi.org/10.1044/2019_LSHSS-18-0129
- Pham, G. et Ebert, K. D. (2020). Diagnostic accuracy of sentence repetition and nonword repetition for developmental language disorder in Vietnamese. *Journal of Speech, Language, and Hearing Research*, 6(5),1521-1536. https://doi.org/10.1044/2020_JSLHR-19-00366
- Plante, E. et Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools*, 25(1), 15-24. <https://doi.org/10.1044/0161-1461.2501.15>
- Poehner, M. (2008). *Dynamic assessment. A Vygotskian approach to understanding and promoting L2 development*. Springer Science +Business Media.
- Power-Point (2018). *Power-point for Macintosh*. [Computer software]. Microsoft.
- R Development Core Team (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Version 4.0.3 for Macintosh*. [Computer software]. Vienna, Austria. <https://www.R-project.org/>
- Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews*, 17(2), 141-150. <https://doi.org/10.1002/ddrr.1108>
- Rezzonico, S., Chen, X., Cleave, P., Greenberg, J., Hipfner-Boucher, K., Johnson, C., Milburn, T., Pelletier, J., Weitzman, E. et Girolametto, L. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI. *International Journal of Language & Communication disorders*, 50(6), 830-841. <https://doi.org/10.1111/1460-6984.12179>
- Rezzonico, S., Goldberg, A., Mak Katy, K.-Y., Yap, S., Milburn, T., Belletti, A. et Girolametto, L. (2016). Narratives in Two Languages: Storytelling of Bilingual Cantonese–English Preschoolers. *Journal of Speech, Language, and Hearing Research*, 59(3), 521-532. <https://doi.org/10.1111/1460-6984.12179>

- Rice, M. L. (2013). Language growth and genetics of specific language impairment. *International Journal of Speech Language Pathology*, 15(3), 223-233. <https://doi:10.3109/17549507.2013.783113>
- Roseberry, C. A. et Connell, P. J. (1991). The use of an invented language rule in the differentiation of normal and language-impaired Spanish-speaking children. *Journal of Speech & Hearing Research*, 34(3), 596-603. <https://doi:10.1044/jshr.3403.596>
- Rosenbaum, F. (2019). *Les mots pour les maux de l'exil : chemins narratifs de reconnaissance*. Fabert.
- Rudolph, J. M. (2017). Case history risk factors for specific language impairment: A systematic review and meta-analysis. *American journal of speech-language pathology*, 26(3), 991-1010. https://doi:10.1044/2016_AJSLP-15-0181
- Scheidnes, M. et Tuller, L. (2016). Assessing successive bilinguals in two languages: A longitudinal look at English-speaking children in France. *Journal of communication disorders*, 64(1), 45-61. <https://doi:10.1016/j.jcomdis.2016.10.001>
- Schulz, P. (2015). Comprehension of exhaustive Wh-questions. Dans S. Armon-Lotem, K. de Jong, et N. Meir (dir.), *Assessing Multilingual Children* (p. 76-95). Multilingual Matters.
- Schwob, S. (2020). Le défi de l'évaluation langagière des enfants bilingues : état des lieux des ressources disponibles en logopédie. *Travaux neuchâtelois de linguistique*, 72(2), 63-87. http://www.unine.ch/files/live/sites/tranel/files/Tranel/72/63-87_Schwob.pdf
- Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P. et Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, ePub Ahead of Print https://doi.org/10.1044/2021_JSLHR-20-00552

- Sim-Sim, I. (2006). *Avaliação da linguagem oral - um contributo para o conhecimento do desenvolvimento linguístico das crianças portuguesas* (4^e ed.), Fundação Calouste Gulbenkian.
- Skoruppa, K. (2019). Novel noun and verb learning in mono- and multilingual children. *Travaux neuchâtelois de linguistique*, 71(1), 109-123.
http://www.unine.ch/files/live/sites/tranel/files/Tranel/71/109-123_Skorupa_def.pdf
- Thibault, M. P. et Helloin, M. C. (2010). *Exalang* 5-8. Happyneuron.
- Thomas, S., Schulz, J. et Ryder, N. (2019). Assessment and diagnosis of Developmental Language Disorder: The experiences of speech and language therapists. *Autism & Developmental Language Impairments*, 4, 1-12.
<https://doi.org/10.1177/2396941519842812>
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426-445.
<https://doi.org/10.1177/1367006911403202>
- Thordardottir, E. (2015a). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. Dans S. Armon-Lotem, K. de Jong et N. Meir (dir.), *Assessing multilingual children* (p. 331-359). Multilingual Matters.
- Thordardottir, E. (2015b). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology*, 17(2), 97-114.
<https://doi.org/10.3109/17549507.2014.923509>
- Thordardottir, E., Rothenberg, A., Rivard, M. E., et Naves. R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, 4(1), 1-21.
<https://doi.org/10.1080/14769670500215647>

- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N. et Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, 54(2), 580-597. [https://doi.org/10.1044/1092-4388\(2010/09-0196\)](https://doi.org/10.1044/1092-4388(2010/09-0196))
- Thordardottir, E. et 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://dx.doi.org/10.1016/j.jcomdis.2012.08.002>
- Tiekstra, M., Minnaert, A. et Hessels, M. (2016). A review scrutinizing the consequential validity of dynamic assessment. *Educational Psychology*, 36(1), 112-137. <https://doi:10.1080/01443410.2014.915930>
- Tomblin, J. B., Records, N., Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*, 39(6), 1284-1294. <https://doi:10.1044/jshr.3906.1284>
- Tomblin, J. B., Zhang, Y., Buckwalter, P. et O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research*, 46(6), 1283-1296. [https://doi:10.1044/1092-4388\(2003/100\)](https://doi:10.1044/1092-4388(2003/100))
- Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. Dans S. Armon-Lotem, K. de Jong et N. Meir (dir.), *Assessing multilingual children* (p. 301-331). Multilingual Matters.
- Vance, M. (2008). Short-term memory in children with developmental language disorder. Dans C. F. Norbury, J. B. Tomblin et D. V. M. Bishop (dir.), *Understanding developmental language disorders* (p. 23-39). Psychology Press.

- Van der Lely, H. (2005). Domain-specific cognitive systems: Insight from grammatical-SLI. *Trends in Cognitive Sciences*, 9(2), 53-59. <https://doi:10.1016/j.tics.2004.12.002>
- Van Horn, D. et Kan P. F. (2015). Fast mapping by bilingual children: Storybooks and cartoons. *Child Language Teaching and Therapy*, 32(1), 1-13. <https://doi:10.1177/0265659015584975>
- Volpin, L., Rezzonico, S. et de Weck, G. (2020). Enfants bilingues et prise en charge logopédique : panorama des pratiques déclarées. *Glossa*, 129(1) 16-31. <https://www.glossa.fr/index.php/glossa/article/view/651>
- Weismer, S. E. et Hesketh, L. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research*, 36(5), 1013-1025. <https://doi:10.1044/jshr.3605.1013>
- Vygotski, L. S. (1934/1985). *Pensée et Langage*. Éditions sociales.

Annexes

Annexe A : Création des items des tâches d'évaluation dynamique

1. Voyelles et consonnes utilisées communes aux deux langues :

- /a, i, u/
- /p, b, t, d, k, g, s, z, l, m, n, f, v/
- Les phonèmes ch-j, s-z, k-g pouvant poser des difficultés aux jeunes enfants ont été au maximum évité dans notre recherche.

2. Constitution des items :

Les nouveaux mots disposent de règles phonotactiques, phonologiques et morpho phonologique. Ils doivent, en outre (Kapantzoglou et al., 2012 ; Chiat, 2015 ; Alt et Plante, 2006) :

- Être des non-mots afin d'éviter les biais linguistiques
- Être des non-mots bien distincts les uns des autres
- Être composés de séquences phonotactiques fréquentes dans les langues utilisées par les enfants
- Être composés de structures syllabiques communes aux langues utilisées
- Varier en fonction de leur longueur (ex. 1-2 syllabes)
 - Concernant les non-noms, les extraterrestres sont systématiquement composés de deux syllabes. A l'inverse, les trésors sont eux composés d'une seule syllabe
- Ne pas ressembler à de réels voisins phonologiques
- Respecter une règle morphophonologique (ex. les substantifs se terminent par une syllabe ouverte mais diffère si le mot est féminin ou masculin). En effet, les indices phonologiques semblent pouvoir aider les enfants à retenir et apprendre les mots (ex. Gray, 2003)

- Les substantifs doivent être équivalents en nombre en ce qui concerne la catégorie du genre (féminin et masculin)
- Les substantifs doivent être équivalents en nombre en ce qui concerne leur caractère animé et inanimé
 - Les extraterrestres sont de caractère animé et les trésors de caractère inanimé
- Les verbes varient selon le caractère transitif et intransitif
 - Par histoire, deux verbes sont transitifs et un est intransitif (n'agit pas sur un objet)
 - Les verbes représentent des actions qui ne peuvent être dénommées par un verbe en français (ex. tourner sur soi-même en sautant ; chanter et rire simultanément).

3. Liste des items

- Substantifs :

Histoire A : /lati/, /plifu/, /fli/

Histoire B : /otu/, /stabi/, /blu/

- Verbes :

Histoire A : /upit/, /flinat/, /lut/

Histoire B : /fakat/, /vat/, /aplut/

- Phrases :

Combinaison des substantifs et des verbes.

Annexe B : Script et captures d'écran du jeu sur ordinateur

« Salut, mon nom est Fap, je viens de la planète Glip et aujourd'hui je t'emmène avec mon vaisseau spatial sur ma planète pour te faire découvrir de nouvelles créatures qui font des choses extraordinaires.

Tiens-toi bien on décolle.

Nous voilà dans le vaisseau spatial. Durant tout le voyage, Madame Flamap, la professeure va te faire connaître et t'apprendre le nom des créatures et des objets de notre planète et ce qu'elles font.

Quand ce signal apparaîtra tu devras parler ou répéter. Par exemple, peux-tu me dire ton prénom ?

SUPER !

Quand c'est ce signal qui apparaît tu devras me montrer sur l'écran de contrôle le mot demandé.

Par exemple, peux-tu presser sur Madame Flamap la professeure ?

SUPER !

Enfin, quand ce signal apparaîtra tu devras bien écouter.

Plus tard, nous allons aussi aider ma cousine Blach qui est aussi en train d'apprendre le nom des créatures de ma planète. Elle fait parfois des erreurs. Tu devras alors lui dire si c'est juste ou si c'est faux en cliquant sur le bon bonhomme.

Pendant tout le voyage, il est possible que tu ne te souviennes plus des noms des créatures. Ce n'est pas grave, essayes de faire de ton mieux.

Nous continuons notre voyage pour arriver sur la planète Glip.

Nous voici arrivés sur Glip. Madame Flamap va te présenter quelques créatures et objets bizarres présents sur Glip. Tout au long de l'aventure, elle te donnera des étoiles pour te féliciter. Dès que tu en auras reçu 5, tu pourras connaître et apprendre ce que font les créatures.

Pour commencer, écoutes-la bien et regardes les images.

Ça va ? C'est super, tu as bien écouté, Madame Flamap te donne ta première étoile !

Nous allons voir maintenant si tu reconnais déjà les créatures et les objets. Ecoutes bien Madame Flamap, elle va te dire le nom d'un objet ou créature et tu dois monter quelle est la bonne image.

Super !

Tu connais déjà bien les créatures de ma planète ! Madame Flamap te donne la deuxième étoile !

Pour continuer à t'entraîner à connaître le nom des créatures et objet de la planète Glip, tu dois maintenant répéter les mots après Madame Flamap. Tu es prêt ? C'est parti !

Très bien, Madame Flamap te donne la troisième étoile. Plus qu'une seule et tu pourras connaître et apprendre ce que font les créatures.

Madame Flamap te montre les images et tu dois dire comment s'appellent les créatures ou les objets.

Bravo ! Tu reçois ta 4^{ème} étoile !

Nous allons retrouver ma cousine Blach qui est aussi en train d'apprendre le nom des créatures et objets de la planète Glip. Regarde la vidéo et écoute ce qu'il dit, s'il dit le mot correctement tu peux presser sur le bonhomme content. Par contre, si Blach se trompe tu peux presser sur le bonhomme triste.

Pour les premiers mots, l'ordinateur de bord va te dire si ce que tu penses est juste ou faux.
Ensuite, tu n'auras plus d'aide.

Tu n'auras maintenant plus d'aide. Fait de ton mieux et tu obtiendras la 5^{ème} étoile.

Très bien ! Tu as fini de corriger ma cousine Blach et Mme Flamap te donne la 5^{ème} étoile.

Tu vas maintenant pouvoir observer et apprendre ce que font ces créatures sur ma planète.

Mais avant de continuer, nous allons faire une petite pause et retourner sur terre. Prépare-toi on décolle !

J'espère que tu t'es bien reposé pendant la pause. Nous retournons sur Glip. Attention, on décolle.

Nous voilà de retour sur la planète Glip pour observer ce que font les créatures. Tu commences par bien écouter Madame Flamap qui va te dire le nom des actions.

Super, tu as très bien écouté ! Madame Flamap te donne la première étoile.

On poursuit, tu dois maintenant montrer la bonne action. Attention, d'abord toutes les actions sont montrées en même temps puis chacune t'est remontrée une fois. Prends le temps de réfléchir puis quand tu es sûr clique sur la bonne image qui correspond à l'action.

Bravo ! Tu connais déjà bien les choses que peuvent faire les créatures de Glip. Madame Blach te donne la deuxième étoile.

Voici l'étape de la répétition. Tu écoutes bien ce que Madame Flamap dit et tu répètes pour encore mieux apprendre les actions.

Super, voici ta troisième étoile. Nous passons à la dernière phase d'apprentissage pour les actions. Tu verras une créature de la planète Glip faire des choses et tu dois dire ce que tu vois.

Tu connais déjà bien ces nouveaux mots ! Bravo. Tu obtiens ta 4^{ème} étoile.

Nous retrouvons encore une fois ma cousine Blach, et nous allons voir comment elle connaît les actions des créatures. Regarde la vidéo et écoute ce qu'elle dit, si elle dit le mot correctement tu peux presser sur le bonhomme content. Par contre, si Blach se trompe tu peux presser sur le bonhomme triste.

Pour les premiers mots, l'ordinateur de bord va te dire si ce que tu penses est juste ou faux. Ensuite, tu n'auras plus d'aide.

Tu n'auras maintenant plus d'aide. Fait de ton mieux et tu obtiendras la 5^{ème} étoile.

Super ! Tu as bien corrigé Blach. Madame Flamap te donne la 5^{ème} étoile.

Avant de continuer à voir comment les autres créatures font ces actions nous allons faire une pause et retourner sur terre. Accroche-toi, on décolle.

J'espère que tu as bien profité de ta pause. Nous allons retourner sur Glip pour une dernière fois. On y va !

On est arrivé ! Nous allons faire une dernière session d'apprentissage. Une fois que Madame Flamap t'aura donné les 5 étoiles, nous aurons fini ce jeu et tu connaîtras toutes les actions et toutes les créatures et objets de la planète Glip.

Tu écoutes Mme Flamap qui va te dire le nom des actions réalisées par les différentes créatures.

SUPER ! Tu as bien écouté. Tu reçois ta première étoile.

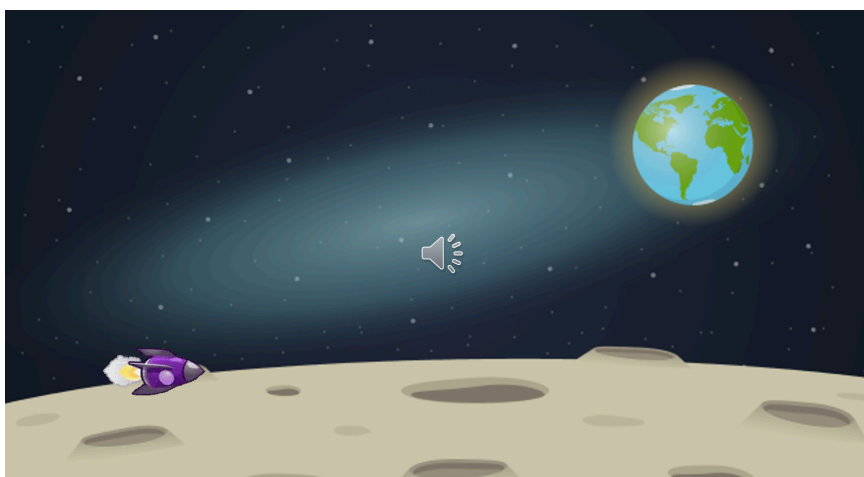
[...] *identique à précédemment.*

Tu as maintenant fini d'apprendre et de connaître les créatures et ce qu'elles font sur Glip. Je te félicite ! Madame Blach est aussi très fière de toi !

Nous allons retourner sur terre. On y va.

Salut à la prochaine ».

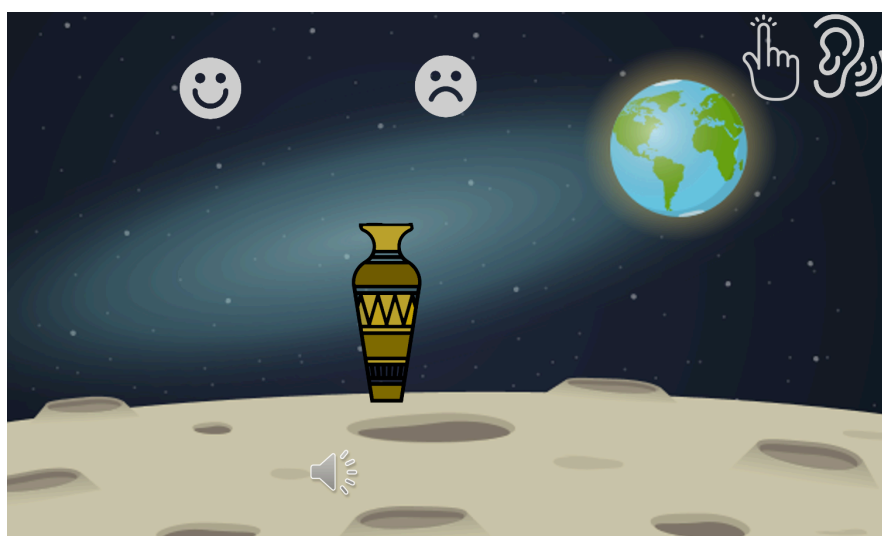
Exemple contexte du jeu :



Exemple désignation :



Exemple jugement d'acceptabilité :



Annexe C : Histoires évaluation dynamique

Histoire 1A

Dans un monde lointain extraterrestre, une chasse aux trésors s'organise chaque année par le grand chef. Le gagnant de cette chasse aux trésors part dans le monde voisin jaune pour le découvrir.

Cette année, pour participer à la chasse aux trésors, le grand chef choisit une */plifu/* et un */lati/*

Une */plifu/* est une créature bizarre verte et petite.

Un */lati/* est un personnage bizarre bleu et grand.

Le grand chef cache un objet. Le premier des extraterrestres qui le trouve, gagne la chasse et peut partir dans le monde voisin jaune pour un voyage.

L'objet à trouver est un */fli/*.

Un */fli/* est un objet de décoration brun.

La */plifu/* et le */lati/* choisis par le grand chef partent à la chasse au */fli/*.

Le */lati/* cherche activement dans les maisons.

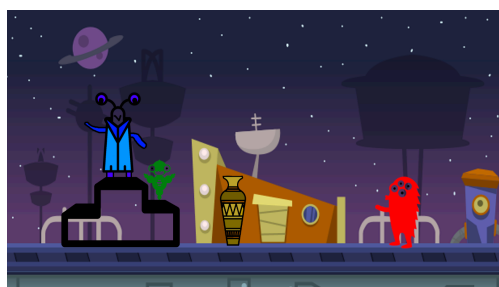
Le */lati/* veut absolument partir en voyage dans le monde voisin. Le */fli/* se cache effectivement tout près, derrière une maison.

La */plifu/* cherche dans la forêt. La */plifu/* ne le sait encore pas mais le */fli/* ne se cache malheureusement pas dans la forêt

Tout à coup, derrière une maison, le */lati/* trouve le */fli/*.

Malgré tous ses efforts, la */plifu/* ne le pas trouvé.

Il gagne donc la chasse aux trésors et pourra partir sur le monde voisin jaune pour le découvrir.



Histoire 1B

Dans un monde lointain extraterrestre, une chasse aux trésors s'organise chaque année par le grand chef. Le gagnant de cette chasse aux trésors part dans le monde voisin jaune pour le découvrir.

Cette année, pour participer à la chasse aux trésors, le grand chef choisit une */otu/* un */stabi/*.

Une */otu/* est une créature bizarre rouge qui a trois yeux.

Un */stabi/* est un personnage bizarre jaune qui a des cheveux.

Le grand chef cache un objet. Le premier des extraterrestres qui le trouve, gagne la chasse et peut partir dans le monde voisin jaune pour un voyage.

L'objets à trouver est une */blu/*.

Une */blu/* est un objet violet avec des hanches.

La */otu/* et le */stabi/* choisis par le grand chef partent à la chasse à la */blu/*.

Le */stabi/* cherche activement dans les maisons.

Le */stabi/* veut absolument partir en voyage dans le monde voisin. La */blu/* se cache effectivement tout près, derrière une maison.

La */otu/* cherche dans la forêt. La */otu/* ne le sait encore pas mais la */blu/* ne se cache malheureusement pas dans la forêt.

Tout à coup, derrière une maison, le */stabi/* trouve la */blu/*.

Malgré tous ses efforts, la */otu/* ne la pas trouvé.

Il gagne donc la chasse aux trésors et pourra partir sur le monde voisin jaune pour le découvrir.



Histoire 2A

Comme promis, le grand gagnant de la chasse aux trésors part à la découverte du monde voisin jaune qui se situe à quelques kilomètres. Pour s’y rendre, il monte dans la soucoupe volante.

Les habitants du monde jaune sont des animaux extraordinaires qui font tous de la gymnastique magique.

Tous les habitants pratiquent cette gymnastique magique. Malheureusement, les animaux du monde jaune sont en entraînement intensif pour les prochains jeux olympiques et ne sont donc pas là.

Seul le grand chef, un peu vieux et plus très en forme, est présent pour accueillir le jeune extraterrestre. Il lui donne la possibilité de regarder un livre qui explique les différentes figures magiques.

Il s’installe et commence par regarder les images des figures magiques. Après un long moment, il essaye de réaliser les acrobaties.

Il commence son entraînement par une figure qui semble facile et pour laquelle il a besoin que de lui-même : il doit faire une roulade sur la gauche.

Il */upit/* une fois. Il */upit/* deux fois. En fait, il */upit/* assez facilement.

Il veut passer à la figure suivante.

Il continue avec une autre figure. Il */lut/* une baguette magique.

Cette figure est différente, il n’a pas besoin de force physique mais il a besoin de beaucoup de concentration pour faire approcher une baguette magique. Il */lut/* une baguette magique mais celle-ci tombe.

Le grand chef propose de réessayer encore une fois. Finalement, il réussit : il */lut/* la baguette magique.

Avant de continuer avec les autres figures, il enchaîne les deux figures à la suite pour s’en souvenir.

Il */upit/* et il */lut/* une baguette magique.

Il veut pouvoir les remonter à ses amis à son retour.

Le temps presse, il termine par une dernière figure très appréciée dans le monde jaune.

Il */flinat/* un ballon.

Cette figure est très amusante. Elle consiste à agrandir un ballon à distance. Il */flinat/* le ballon mais celui-ci rétrécit.

Le grand chef propose de recommencer et d'être encore plus concentré. Il */flinat/* le ballon et celui-ci reste intact.

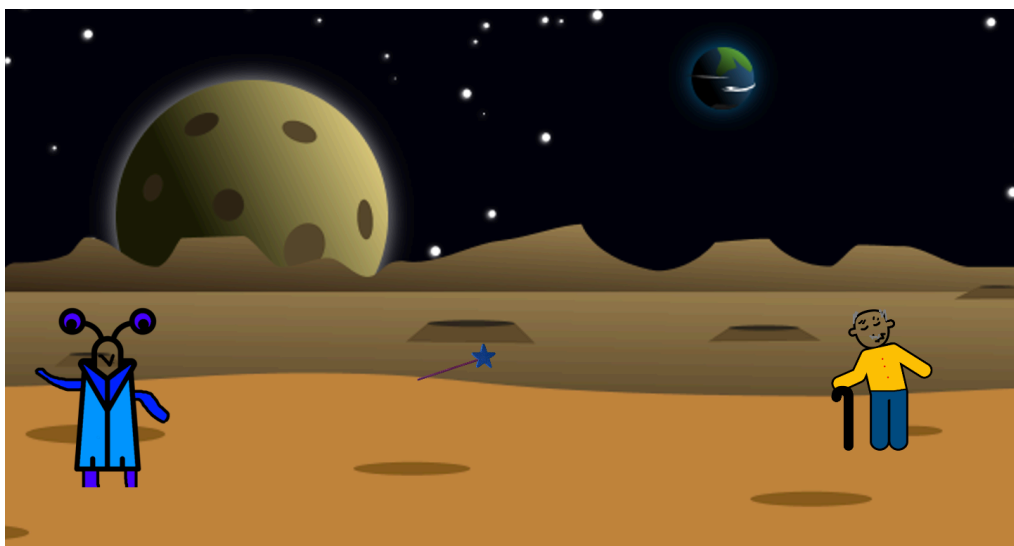
Avant de repartir et pour être sûr, il */flinat/* le ballon une dernière fois.

C'est le moment de rentrer, il remercie et dit au revoir au grand chef.

Dans la soucoupe volante, pour rentrer, il essaye de se souvenir des figures, il */upit/*, il */lut/* une baguette magique et il */flinat/* un ballon dans ses pensées.

Une fois arrivé dans son monde, tous ses amis l'attendent et veulent voir les figures de la gymnastique magique.

Il monte sur scène et fait la démonstration de toutes les figures apprises. Il */lut/* une baguette, il */upit/* et il souhaite terminer par la figure la plus amusante. Il */flinat/* un ballon pour chacun de ses amis.



Histoire 2B

Comme promis, le grand gagnant de la chasse aux trésors part à la découverte du monde voisin jaune qui se situe à quelques kilomètres. Pour s’y rendre, il monte dans la soucoupe volante.

Les habitants du monde jaune sont des animaux extraordinaires qui font tous de la gymnastique magique.

Tous les habitants pratiquent cette gymnastique magique. Malheureusement, les animaux du monde jaune sont en entraînement intensif pour les prochains jeux olympiques et ne sont donc pas là.

Seul le grand chef, un peu vieux et plus très en forme, est présent pour accueillir le jeune extraterrestre. Il lui donne la possibilité de regarder un livre qui explique les différentes figures magiques.

Il s’installe et commence par regarder les images des figures magiques. Après un long moment, il essaye de réaliser les acrobaties.

Il commence son entraînement par une autre figure qui semble facile et pour laquelle il a besoin que de lui-même : il doit tourner sur lui-même en descendant.

Il */fakat/* une fois.

Il */fakat/* deux fois.

En fait, il */fakat/* assez facilement.

Il veut passer à la figure suivante.

Il continue avec une autre figure qui est la plus amusante et qui est très appréciée dans le monde jaune. Il */vat/* une feuille.

Cette figure est différente, il n’a pas besoin de force physique mais il a besoin de beaucoup de concentration pour faire avancer par terre une feuille. Il */vat/* une feuille mais il n’arrive pas à la faire avancer jusqu’à lui.

Le grand chef propose de réessayer encore une fois. Finalement, il réussit : il */vat/* une feuille.

Avant de continuer avec les autres figures, il enchaîne les deux figures à la suite pour s'en souvenir.

Il */fakat/* et il */vat/* une feuille.

Il veut pouvoir les remontrer à ses amis à son retour.

Le temps presse, il termine par la dernière figure.

Il */aplut/* une maison.

Cette figure consiste à faire disparaître une maison et il a donc besoin de sa tête pour réfléchir.

Il */aplut/* une maison mais celle-ci réapparaît.

Le grand chef propose de recommencer et d'être encore plus concentré. Il */aplut/* une maison et elle disparaît réellement.

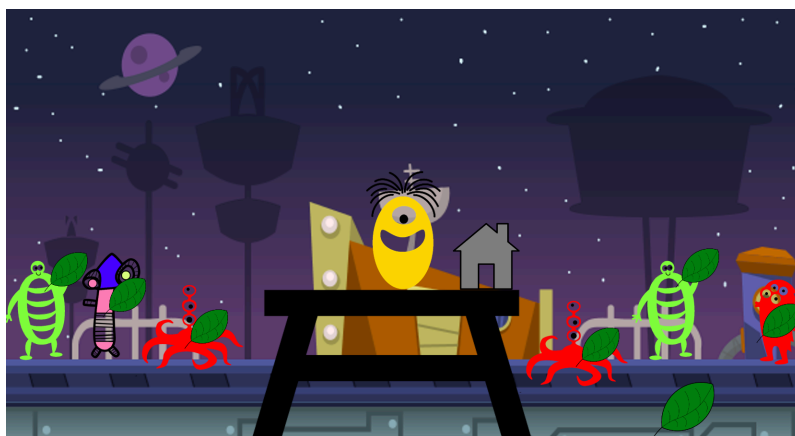
Avant de repartir et pour être sûr, il */aplut/* une maison une dernière fois.

C'est le moment de rentrer, il remercie et dit au revoir au grand chef.

Dans la soucoupe volante, pour rentrer, il essaye de se souvenir des figures, il */fakat/*, il */vat/* une feuille et il */aplut/* une maison dans ses pensées.

Une fois arrivé dans son monde, tous ses amis l'attendent et veulent voir les figures de la gymnastique magique.

Il monte sur scène et fait la démonstration de toutes les figures apprises. Il */aplut/* une maison, il */fakat/* et il souhaite terminer par la figure la plus amusante. Il */vat/* une feuille pour chacun de ses amis.



Histoire 3A

Un nouveau défi attend le */lati/* et la */plifu/*.

Après être rentré du monde jaune, il s'est beaucoup entraîné et il souhaite maintenant participer et remporter les jeux olympiques de gymnastique magique avec son amie.

Pour cela, ils se rendent dans le monde jaune et s'entraînent plusieurs semaines pour essayer de gagner la compétition.

Le */lati/* a plus d'expérience que la */plifu/* pour réaliser les figures de la gymnastique magique. En effet, il */lut/* la baguette, il */flinat/* le ballon et il */upit/* depuis plusieurs mois maintenant. Depuis son premier voyage dans le monde jaune, le */lati lut/* la baguette et il */upit/* mieux que personne.

Le */lati upit/* encore à une reprise. Elle est débutante et observe d'abord les figures. La */plifu upim/* facilement après avoir regardé le */lati/* réalisé les figures.

Par contre, les autres figures sont plus difficiles. La */plifu lum/* la baguette de manière correcte mais celle-ci tombe parfois. Alors il lui remontre une fois comment il faut faire : le */lati lut/* la baguette. Pour finir, la */plifu flinam/* le ballon très difficilement. Celui-ci n'arrête pas de se dégonfler. Elle */flinam/* le ballon encore une fois pour essayer.

Pour mettre leurs chances de leurs côtés et essayer de remporter les jeux olympiques, ils ont une idée : puisque la */plifu flinam/* le ballon difficilement, elle ne fera pas cette figure. Ils décident ainsi que le */lati flinat/* le ballon et la */plifu upim/* et elle */lum/* la baguette.

Après plusieurs semaines d'entraînement, la compétition des jeux olympiques commence. Il y a beaucoup de concurrents dans le stade olympique !

Le */lati/* et son amie sont stressés. Ils ont beaucoup répété et souhaitent gagner !

La démonstration commence. La */plifu lum/* la baguette. Elle ne tombe pas et il y a beaucoup d'applaudissements du public. Elle continue le spectacle et elle */upim/*.

Arrive le moment de la dernière figure : le */lati flinat/* le ballon superbement.

Mais surprise, la */plifu/* finit également par cette figure, Superbe ! Le public est conquis. Les amis sont fiers et attendent les résultats. Miracle ! Ils gagnent les jeux olympiques !



Histoire 3B

Un nouveau défi attend le */stabi/* et la */otu/*.

Après être rentré du monde jaune, il s'est beaucoup entraîné et il souhaite maintenant participer et remporter les jeux olympiques de gymnastique magique avec son amie.

Pour cela, ils se rendent dans le monde jaune et s'entraînent plusieurs semaines pour essayer de gagner la compétition.

Le */stabi/* a plus d'expérience que la */otu/* pour réaliser les figures de la gymnastique magique. En effet, il */vat/* une feuille, il */fakat/*, il */aplut/* une maison depuis plusieurs mois maintenant. Depuis son premier voyage dans le monde jaune, le */stabi vat/* une feuille et il */fakat/* mieux que personne.

Le */stabi fakat/* encore à une reprise. Elle est débutante et observe d'abord les figures. La */otu fakam/* facilement après avoir regardé le */stabi/* réalisé les figures.

Par contre, les autres figures sont plus difficiles. La */otu vam/* la feuille de manière correcte mais celle-ci n'arrive pas jusqu'à elle. Alors il lui remontre une fois comment il faut faire : le */stabi vat/* la feuille. Pour finir, la */otu aplum/* une maison très difficilement. Celle-ci réapparaît à chaque fois. Elle */aplum/* une maison encore une fois pour essayer.

Pour mettre leurs chances de leurs côtés et essayer de remporter les jeux olympiques, ils ont une idée : puisque la */otu aplum/* la maison difficilement, elle ne fera pas cette figure. Ils décident ainsi que le */stabi aplut/* la maison et la */otu fakam/* et elle */vam/* la feuille.

Après plusieurs semaines d'entraînement, la compétition des jeux olympiques commence. Il y beaucoup de concurrents dans le stade olympique !

Le */stabi/* et son amie sont stressés. Ils ont beaucoup répété et souhaitent gagner !

La démonstration commence. La */otu vam/* la feuille. Elle arrive à la faire avancer jusqu'à elle et il y a beaucoup d'applaudissements du public. Elle continue le spectacle et elle */fakam/*.

Arrive le moment de la dernière figure : le */stabi aplut/* une maison superbement !

Mais surprise, la */otu/* finit également par cette figure, Superbe ! Le public est conquis.

Les amis sont fiers et attendent les résultats. Miracle ! Ils gagnent les jeux olympiques !



Annexe D : Création des tâches d'évaluation

- Désignation

Consigne : ex. « montre-moi le /lati/ »

Distracteurs :

- Substantifs et verbes : les deux autres items sont présentés comme des distracteurs
- Phrases : trois autres éléments de distracteur : image avec bon extraterrestre mais pas la bonne action ; image avec la bonne action mais pas le bon extraterrestre ; image avec aucun des bons éléments.

- Évocation

Consigne : ex. « est-ce que tu te rappelles du nom des extraterrestres et des objets ? »

- Dénomination

Consigne : ex. « comment s'appelle cet extraterrestre ? »

- Jugement d'acceptabilité

Consigne : ex : « Regardes la vidéo et écoute ce qu'il dit, s'il dit le mot correctement tu peux presser sur le bonhomme content »

Distracteurs :

- Phonologiques : ex. omission du phonème initial de l'item cible (ex. ati pour lati), omission du phonème final (de la règle morphophonologique ou de la flexion ; ex. upi pour upit), substitution d'un phonème (ex. bli pour blu), réduction d'un groupe consonantique (ex. finat pour flinat)

- Sémantiques : ex. substitution par un item proche (ex. autre extraterrestre, plifu pour lati), substitution par un item plus éloigné (ex. objet, fli pour lati)
- Syntaxiques : ex. ordre des mots dans la phrase (ex. upit lati pour lati upit).

Annexe E : Questionnaire parental



UNIVERSITÉ DE
NEUCHÂTEL

Institut des sciences
du langage et de la
communication

Centre de logopédie

Rue de la Pierre-à-Mazel 7
CH-2000 Neuchâtel

Secrétariat ;
Rue Pierre-à-Mazel 7
CH-2000 Neuchâtel
Tél : +41 (0)32 718 18 29
secreariat.orthophonie@unine.ch

Salomé Schwob
Assistance-doctorante
Chaire de Logopédie I
Salome.schwob@unine.ch
076/498.34.06

Questionnaire aux parents des enfants bilingues

Chers parents,

Nous vous remercions pour l'intérêt porté à notre étude. Veuillez prendre le temps de remplir le questionnaire ci-dessous. Il porte sur votre famille, votre utilisation des langues et le développement de votre enfant¹.

Prénom de l'enfant :

Date de naissance :

Lieu et pays de naissance :

A. Développement

1. A quel âge votre enfant a-t-il commencé ...

a. À produire ses premiers mots : ____ ans ____ mois

b. À combiner deux mots (ex. maman partie, lapin cassé) : ____ ans ____ mois

2. Avez-vous été inquiet concernant le développement langagier de votre enfant ?

Oui ☐

Non ☐

3. Votre enfant a-t-il déjà présenté des difficultés dans les domaines suivants?

a. Langage/parole (ex. trouble du langage, dysphasie, zézaïement, schlintement)

Oui ☐

Non ☐

Si oui, à quel âge ?

¹ Nous avons essayé de formuler toutes les questions de façon aussi neutre et complète que possible afin de refléter différentes situations familiales. Si une question ne correspond pas à votre situation, veuillez écrire vos commentaires à la fin de ce document (partie D). Si vous ne pouvez pas fournir des informations spécifiques (ex. âge exact), veuillez faire une estimation plausible. Vous pouvez également nous contacter pour toute question ou ambiguïté.

b. Ouïe (ex. surdit , otites fr quentes)

Oui ☐

Non ☐

Si oui,   quel  ge ?

c. Autres (ex. autisme, trouble de l'attention, hyperactivit )

Oui ☐

Non ☐

Si oui,   quel  ge ?

4. Votre enfant a-t-il d j  re u un soutien ou de la th rapie en lien avec le langage ?

Oui ☐

Non ☐

Si oui,

- o  (ex.  cole/orthophoniste)?
-   quel  ge ?
- pendant combien de temps ?   quelle fr quence ?

5. Existe-t-il des ant c dents familiaux (ex. (grands)-parents, fr res/s urs, oncles/tantes) de difficult s langagi res (ex. difficult s de compr hension, trouble du langage, dyslexie-dysorthographe).

Si oui, Lesquelles ? Chez qui ?

B. Langues et famille

6. A part le fran ais, votre enfant est-il en contact avec d'autres langues ?

Oui ☐

Non ☐

Si oui, Lesquelles ?

7. Veuillez s.v.p. remplir le tableau suivant pour chaque langue parl e par l'enfant.

Langue (ex. fran�ais, portugais)	Depuis quel �ge ? (_____ an _____ mois)	O� ? (ex. maison, �cole)	Avec qui ? (ex : p�re, m�re, belle-m�re, fille au pair)

8. Pour chacune des langues, veuillez s.v.p. indiquer leur fréquence d'utilisation via les tableaux suivants (à la maison/ à l'école).

Langue	<u>À la maison</u>				
	Fréquence d'utilisation				
	<i>Presque toujours (90-100%)</i>	<i>Souvent (65-90%)</i>	<i>La moitié du temps (35-65%)</i>	<i>Rarement (10-35%)</i>	<i>Jamais (0- 10%)</i>
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

Langue	<u>À l'école</u>				
	Fréquence d'utilisation				
	<i>Presque toujours (90-100%)</i>	<i>Souvent (65-90%)</i>	<i>La moitié du temps (35-65%)</i>	<i>Rarement (10-35%)</i>	<i>Jamais (0- 10%)</i>
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

C. Contexte familial

1. Quelle sont les formations obtenues par les adultes qui vivent avec l'enfant ? (ex. père, mère, belle-mère). Les formations à l'étranger comptent aussi. Veuillez s.v.p. remplir le tableau suivant.

Adulte	Formations				
	<i>Ecole primaire (4-12 ans/1H à 8H)</i>	<i>Ecole secondaire (12-15 ans/ 9H à 11 H)</i>	<i>Apprentissage, CAP, école professionnelle</i>	<i>Maturité gymnasiale (lycée, gymnase)</i>	<i>Haute Ecole, Université, EPFL</i>
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

D. Commentaires

(voir note de bas de page 1)

E. Cordonnées

Vos coordonnées nous sont nécessaires afin que nous puissions vous contacter en cas de questions.

La date à laquelle vous remplissez le questionnaire :

Votre nom et prénom :

Votre adresse :

Votre numéro de téléphone :

Votre email :

Annexe F : Analyse des répétitions de non-mots

Comptage :

- Nombre de mots corrects : cette catégorie contient le score (0/1) de l'enfant : 1 en cas de réussite de l'item par l'enfant ; 0 si l'item est répété incorrectement
- Nombre de phonèmes erronés : cette catégorie contient le nombre de phonèmes erronés par rapport à l'item cible. Le positionnement du phonème dans le mot est pris en considération dans ce calcul.

En nous référant à Dollaghan et Campbell (1997), il est possible d'aligner les syllabes du mot produit et de l'item cible pour effectuer le comptage. Toutefois, il est toujours important de s'intéresser aux productions et de se demander comment à procéder l'enfant.

- Ex : l'enfant répète plu^{fi} à la place de plu^{fu} : 2 phonèmes erronés, 3 phonèmes corrects et 1 erreur de type métathèse

Types d'erreurs :

Les erreurs doivent être détaillées en genre et en nombre avec exemple.

Le positionnement des phonèmes par rapport à l'item cible est pris en considération. Les erreurs sont détaillées en fonction de 3 niveaux (phonème, syllabe, mot). Les erreurs sont dissociées en fonction des consonnes et des voyelles. Les erreurs portant sur les semi-voyelles sont comptabilisées parmi les erreurs portant sur les voyelles.

En fonctionnement de l'alignement, il n'est pas toujours facile de dire quel phonème ou syllabe a été modifiée. Nous décidons ainsi de toujours aligner les syllabes et analyser le plus finement possible.

Exemple:

/Li ta pi mu di/

/Li pa ___ mu di/

Ainsi, le t est substitué par p et le pi est omis.

Classification basée sur Kapalkova et al. (2013) (tableau 8 de l'article).

- Phonèmes

- Substitution de consonne
- Substitution de voyelle
- Ajout consonne
- Ajout voyelle

- Syllabes

- Omission consonne

Parfois l'enfant s'est interrompu, ex. gri-, les items manquants sont notés comme des omissions

Parfois l'enfant n'a pas produit le mot, ex. /, les erreurs sont donc comptabilisées comme des omissions

- Réduction de groupe consonantique
- Omission voyelle

- Mots

- Métathèse : déplacement de phonèmes ou de syllabes ou d'une partie plus grande que la syllabe. Le phonème ou autre segment est alors inversé ou

interverti avec un autre segment de l'item initial.

Ex. « kreki » devient « ikrek », nous avons 1 métathèse du segment « krek ».

- Assimilation. Les assimilations peuvent être progressives ou régressives mais nous ne détaillons pas cela dans la grille d'analyse.

- Lexicalisation : l'enfant produit un mot proche du non-mot entendu

Exception : « maki » n'est pas considéré comme une lexicalisation de la part de l'enfant dans cette recherche. Il est parfois difficile de dire s'il s'agit de lexicalisations, cela dépend souvent des langues parlées par le juge (ex. français et portugais, vs. français).

- Création : lorsque plus de 4 phonèmes erronés (à partir de 5 du coup) sont présents dans une production. L'erreur est comptabilisée comme création sans détailler les autres types d'erreur

Ex. l'enfant produit /silabulalimalilala/ pour /sibumakila/ ; 8 erreurs de 2 types (1 substitution et 8 ajouts) = 1 création.

Si les erreurs sont nombreuses mais concernent des omissions, ex. l'enfant dit seulement le début de l'item, les erreurs sont comptabilisées comme des omissions et non comme des créations.

Annexe G : Analyse évaluation dynamique

1. Évaluation spontanée

De manière générale :

- L'évaluation spontanée doit être notée dans toutes les tâches : 1 point = juste / 0 point = faux
- Si le non-mot n'est pas exactement correct, compter 0 point dans évaluation spontanée
 - Grâce aux autres catégories d'analyse (nb de syllabe, nb de phonèmes correct et erroné) nous pourrions analyser plus finement les productions de l'enfant
 - Lorsque l'enfant fait une liaison avec le pronom précédent le non-mots (ex. il upit), nous comptons l'item comme correct

2. Types d'erreurs

De manière générale :

- Le type d'erreur doit être précisé dans les tâches de rappel différé, désignation, dénomination et évocation.
- Dans la tâche d'évocation, il n'est pas toujours aisé de déterminer quel item a été produit par l'enfant (à quel item l'enfant se réfère), nous essayons de chercher l'item le plus ressemblant à celui produit
- Dans la tâche de rappel différé, nous acceptons les réponses sous forme de phrases et de verbes. Seul le verbe est comptabilisé au niveau du nombre de phonème correctement produit.

Dans le cas où un enfant dit /aplum/ nous comptons 1 erreur de type substitution de flexion puisque c'est le personnage masculin, à savoir, le stabi, qui réalise l'action.

- Plusieurs erreurs peuvent se combiner à l'intérieur d'une même production.

- Lorsque les erreurs sont plurielles et doivent être notées en ajoutant un « + ».

Ex. /klitu/ est produit à la place de /plifu/ = substitution phonème + substitution phonème

Ex. /plufi/ est produit » à la place de /plifu/ = substitution phonème + substitution phonème

- Nous considérons la place des erreurs seulement pour les phonèmes finaux correspondant aux flexions.

ex. pour les substantifs : si l'enfant produit /flimu/ à la place de /plifu/, nous relevons une seule erreur de type substitution de phonème (m-p).

ex. pour les verbes : si l'enfant produit /lupit/ pour /aplut/, nous notons une seule erreur de type substitution de phonème (a-i).

ex. pour les phrases : si l'enfant produit /flifu flimat/ à la place de /plifu upim/. Nous relevons une erreur de substitution dans le substantif (f-p) ; et une création de verbe puisque – de 2 phonèmes sont corrects puisque le « m » ne peut être considéré comme intact étant donné qu'il correspond à la flexion dans la réponse attendue.

Types d'erreurs existants :

1. Substitution phonème
2. Substitution substantif : l'enfant substitue un substantif par un autre existant /lati/ à la place de /plifu/.

Ex. dans une phrase de type /lati lut/ à la place de /plifu lum/, on considère l'erreur comme une « substitution substantif + substitution flexion »

- Si l'enfant substitue un substantif par un autre existant mais n'étant pas présent dans sa liste d'apprentissage, ex. l'enfant dit /stabi/ pour /lati/, l'erreur est

également considérée comme « substitution substantif ». Par contre, si l'item n'est pas exact ex. /stab/ pour /lati/, l'erreur est de type création

3. Substitution verbe : ex. l'enfant produit /lut/ à la place de /upit/
 - Si l'enfant substitue un verbe par un autre existant mais n'étant pas présent dans sa liste d'apprentissage, ex. l'enfant dit /vat/ pour /lut/, l'erreur est également considérée comme « substitution verbe »
 - Si le verbe substitué n'est pas exactement l'équivalent de celui substitué, ex. l'enfant mentionne /otu fakat/ pour /otu vam/, l'erreur est de type « substitution verbe + substitution flexion »
4. Substitution item :
 - Lorsque dans les phrases, l'enfant produit/désigne /lati lut/ à la place de « plifu flinam » -> tout l'item est différent
 - Lorsque l'enfant mentionne un verbe à la place d'un substantif /upit/ pour /lati/ Si l'item n'est pas exact, l'enfant dit /upi/ (= upit) pour /stabi/ ,ou /vat/ pour /aplum/, l'erreur est de type création.
5. Substitution flexion : Ex. /upit/ pour /upim/ , /fakal/ pour /fakat/
 - Pour considérer l'erreur comme une substitution de flexion, la substitution doit concerner une consonne en position finale dans un item verbe.
6. Ajout phonème
7. Absence de réponse : l'enfant ne répond pas ou dit ne pas savoir
8. Création substantif : lorsque – de 2 phonèmes corrects et ne peut pas être mieux expliqué par un autre type d'erreur
9. Création verbe : lorsque – de 2 phonèmes corrects et ne peut pas être mieux expliqué par un autre type d'erreur

10. Création phrase : lorsque - de 2 phonèmes corrects dans chacun des items (2 dans verbe et 2 dans substantif) et ne peut pas être mieux expliqué par un autre type d'erreur. Ex. /saksak/ pour /otu fakam/. La segmentation entre le substantif et le verbe est souvent difficile à établir

- Lors de phrases, la création peut être une erreur parmi d'autres. Ex l'enfant produit /plifu tima/ pour /lati lut/, on voit ici une erreur de type « substitution substantif » + « création verbe »

11. Répétition ébauche substantif/ verbe: cette catégorie ne concerne que les productions après étayage.

- Lorsque l'enfant répète une ébauche de l'adulte correctement mais que malheureusement le contenu spontané n'a pas été produit. Le reste de la production de l'enfant (ex. ce qui a été produit spontanément ou ce qu'il a été omis après ébauche) est noté par d'autres types d'erreurs.

Ce qui n'a pas été répété est également inscrit par d'autres types d'erreurs (omission notamment). Attention les autres erreurs concernent ce qui a été produit spontanément par l'enfant (ex. si l'enfant répète « a » et qu'il ne donne pas d'autres phonèmes pour le verbe /aplut/ nous sommes dans 3 omissions de phonèmes + 1 omission de flexion, mais non pas dans une erreur de type création.

Lorsque l'enfant a produit un début d'item correctement mais que la fin est erronée (ex. /flipu/ pour /fli/), l'adulte va fournir de l'étayage sous forme d'ébauche. En cas d'erreur de l'enfant dans la partie d'étayage (ex. /fla/ pour /fli/), nous ne comptabiliserons pas l'erreur « répétition ébauche substantif » puisque l'enfant avait déjà, avant l'étayage, produit ce début d'item correctement. Nous noterons ainsi que l'erreur « substitution phonème »

12. Description substantif : ex. « c'est un extraterrestre bleu et grand »

13. Description verbe : ex. « il fait avancer une baguette magique »

- Lors de phrases, si l'enfant dit « Il se fait tourner soi-même » pour /stabi fakat/, nous considérons cela comme une « description de verbe »
- Les descriptions ne sont pas toujours pertinentes. Ex. pour /aplut/ l'enfant peut dire « il fait apparaître la maison » vs « disparaître la maison »
- Il arrive dans les cas de phrases, que la description porte uniquement sur le verbe dans ce cas les erreurs sont notées comme « omission substantif + description verbe ». Les phonèmes erronés sont notés à partir de la cible attendu (ex. l'enfant dit « il roule » pour /le lati aplut/, nous comptabilisons 9 phonèmes erronés et 0 correct).
- En cas de lexicalisation ou d'usage du français (ex. description) nous ne comptons considérons cela comme une absence de phonèmes/syllabes corrects.
Ex. il ramène l'objet pour /lut/

14. Omission phonème : /pit/ pour /upit/

15. Omission substantif /lut/ pour /lati lut/

16. Omission verbe : /lati/ pour /lati lut/

Nous sommes dans 2 erreurs de type « omission substantif » + « omission verbe » vs « absence de réponse » lorsque l'enfant dit ex. « il s'appelle Mr. Banane » pour /lati flinat/.

Lorsque l'enfant énonce seulement un déterminant ou un pronom nous sommes dans la catégorie « absence de réponse ».

Lorsque l'enfant énonce l'objet réalisant l'action « baguette » pour « lut la baguette », nous sommes dans une erreur de type « omission verbe »

17. Omission flexion (/upi/ pour /upit/)

18. Élimination : lorsque l'enfant fournit une réponse par élimination.

Ex. /pas upit/

19. Lexicalisation : l'enfant lexicalise le non-mot.

Ex. l'enfant produit « stabilo » pour /stabi/.

Les phonèmes sont comptabilisés comme pour les non-mots. C'est-à-dire que nous comptons le nombre de phonèmes corrects et erronés à partir d'une transcription phonétique de la lexicalisation.

Pour considérer l'erreur comme une lexicalisation, il est nécessaire de trouver une ressemblance entre le mot et le non-mots. Il est parfois difficile de faire la distinction entre les lexicalisations et les descriptions. Ex. l'enfant produit « tournevis » pour /fakat/, nous notons que le non-mots /fakat/ consiste à tourner sur soi-même et qu'il ne possède aucune ressemblance phonologique avec le mot tournevis, nous décidons ainsi de coder cette erreur comme une « description verbe »

20. Substitution morphologie française : l'enfant modifie le non-mot avec une morphologie française. Ex. « une plifeuse » pour /plifu/.

21. Commentaire sur la tâche : l'enfant dit par ex. « c'est dur » à la place de l'item

22. Commentaire à l'expérimentateur : par ex. « montre-moi toi la /blu/ »

23. Question : l'enfant pose une question pour obtenir de l'aide. Ex. « c'est lequel il est petit ou grand ? »

24. Inversion : lorsque dans une phrase, l'enfant inverse l'ordre des mots, soit le substantif et le verbe. Ex. /fakat le stabi/.

25. Inversion phonème : lorsque tous les phonèmes du non-mots sont présents et seul l'ordre est erroné. Ex. /lub/ pour /blu/.

26. Ambigu : toutes erreurs ne pouvant être classifiées dans les autres catégories

3. Nombre de syllabes correctes

De manière générale :

- 27. Le nombre de syllabe doit être précisé dans les tâches de rappel différé, de dénomination et d'évocation.
- 28. Compter le nombre de syllabe correctement énoncé
- 29. Lorsque les erreurs sont de type « ajout » et que la syllabe initiale est correcte. Nous comptabilisons la syllabe comme juste. Ex. /flipi/ pour /fli/ -> nous comptabilisons une syllabe correcte.

4. Nombre de phonèmes erronés

De manière générale :

- 30. Le nombre de phonème erroné doit être précisé dans les tâches de rappel différé, de dénomination et d'évocation.
- 31. Compter le nombre de phonème erroné par rapport à l'item cible.
- 32. Les phonèmes erronés peuvent être des ajouts, des substitutions, des omissions.
- 33. Toujours compter de façon favorable.

Ex : si l'enfant dit /upit/ pour /lut/. Nous comptabilisons 2 phonèmes corrects et 2 phonèmes erronés (1 étant considéré comme substitué et 1 comme ajouté)

Ex. dans les phrases : Ex. l'enfant produit /otu fakap/ vs /stabi aplut/, nous comptabilisons le nombre de phonèmes erronés entre les substantifs et entre les verbes.

Pour les substantifs, il s'agit de 4 phonèmes incorrects (2 substitutions o, u ; 2 omissions) et pour les verbes de 4 (3 substitutions f, k, a et une erreur de flexion p).

Nous avons donc en tout 8 phonèmes erronés dans cette phrase.
- 34. Même pour les substitutions de flexion, les phonèmes corrects et erronés sont comptés en faveur de l'enfant. Ex. l'enfant produit /fatak/ pour /fakat/ : nous considérons que

l'enfant des phonèmes sont correctement produits mais que des erreurs sont présentes :
substitution phonème + substitution flexion.

35. Pour la tâche avec étayage : nous comptabilisons après l'étayage, le nb de phonèmes erronés produits par l'enfant, pas ceux ayant été répété ou n'ayant pas été dit.

5. Nombre de phonèmes corrects

De manière générale :

36. Le nombre de syllabe doit être précisé dans les tâches de rappel différé, de dénomination et d'évocation.

37. Compter le nombre de phonèmes correctement énoncé ou ayant en commun avec l'item cible.

Ex. /plufi/ est produit » à la place de /plifu/. Le nombre de phonèmes correct est de 5 (totalité).

Ex. dans les phrases : Ex. l'enfant produit /otu fakap/ vs /stabi aplut/, nous comptabilisons le nombre de phonèmes corrects entre les substantifs et entre les verbes.

Pour les substantifs, il s'agit de 1 (t) et pour les verbes de 1 (p). Le phonème « p » n'est pas considéré comme correct car il se trouve à la place de la flexion. Nous avons donc en tout 3 phonèmes correctement produits dans cette phrase.

38. Dans les verbes, toujours faire attention à la flexion. Lorsque l'enfant produit /flipi/ à la place de /upit/, nous relevons 4 erreurs (omission flexion + substitution phonème + ajout phonème + ajout phonème). 2 erreurs se rapportent au même phonème (flexion : ajout phonème + omission flexion). Nous comptons donc favorablement le nombre de phonèmes erronés et corrects : 3 phonèmes erronés (alors que 4 erreurs) et 2 phonèmes corrects.

39. Pour les erreurs de type « création », nous comptabilisons le nombre de phonèmes produits par l'enfant, idem pour compter le nombre de phonèmes erronés et corrects. Cette comptabilisation est parfois difficile puisque la production se rapproche peu de l'item cible.
40. Pour la tâche avec étayage : nous comptabilisons après l'étayage, le nombre de phonèmes corrects produits par l'enfant, pas ceux ayant été répété ou n'ayant pas été dit.

6. Évaluation sans étayage

Score spontané (1) lorsque l'adulte n'a pas fourni d'étayage.

7. Étayage

En général :

41. Si lorsque sans étayage l'enfant produit une réponse erronée et qu'avec l'étayage il ne fournit pas de nouvelles réponses. Nous codons « absence de réponse ». Toutefois, les scores déjà accumulés malgré l'erreur sont reportés lors de la cotation avec étayage.
42. Il arrive parfois qu'un bout de l'item ne soit pas repris dans la partie étayage (notamment lorsque le substantif a déjà été produit de façon correcte ou erronée, il arrive que l'enfant ne le corrige pas). En ce qui concerne les types d'erreurs après étayage, nous notons que celles réalisées dans le contexte étayage.
43. Pour compter les scores après l'étayage, nous prenons en considération la plupart du temps, la dernière réponse donnée par l'enfant, c'est-à-dire la réponse donnée après le plus grand nombre d'étayage de la part de l'expérimentateur. En somme, nous prenons en considération la meilleure réponse donnée par l'enfant et il se peut que ça ne soit pas

sa dernière réponse. Lorsqu'il existe un doute sur la meilleure production de l'enfant, un commentaire est inséré dans la colonne AJ.

- Nous distinguons ensuite dans 2 colonnes différentes ce qui a été répété par l'enfant (comme une écholalie d'apprentissage) de ce qui a été produit spontanément suite à l'étayage de l'adulte. Cf. 10 et 11
- A noter, qu'à la fin de chaque étayage, si la bonne réponse n'est pas fournie par l'enfant, l'expérimentateur donne la bonne forme à l'enfant. Si à ce moment-là l'enfant répète nous ne codons pas cela comme une répétition.
- Lorsque l'ébauche phonémique a été donnée par l'expérimentateur, nous notons dans la production spontanée la production de l'enfant avec une () pour le phonème ébauché. Ex. /(o)tu/. Et dans la colonne de répétition ce qui a été répété dans (). Nous notons également en () ce qui a été répété par l'enfant suite à la production de l'adulte

44. Lorsque l'étayage est donné par un autre adulte (ex. parent), les codes d'étayage sont utilisés et sont répertoriés sous le codage de l'enfant. Les différents étayages sont insérés dans la colonne de commentaires AJ. Les étayages ne sont toutefois pas répertoriés dans ce qui est donné par l'expérimentateur.

8. Évaluation avec étayage

Score lorsque l'adulte a fourni une aide (0/1).

45. Ce score avec étayage concernant le score en production spontanée

46. En général, la production après étayage est différente et dissociée de la production spontanée sans étayage. Toutefois, dans le cadre des phrases, lorsque le substantif a été correctement produit en production spontanée il est tout de même comptabilisé après avec étayage.

Les seules productions qui changent sont celles qui ont subi un étayage.

9. Étayage (voir aussi guide étayage, Annexe G)

De manière générale :

47. A chaque fois, spécifier chaque aide. Si plusieurs aides ont été fournies les noter avec un « + ». « Ex. A1 + A2 ».
48. Il est possible que parfois l'expérimentateur ne fournisse pas un étayage (ex. a cru entendre le mot correctement). Nous notons alors absence dans les codes à insérer concernant l'étayage.
49. Lorsqu'un étayage est donné mais est chevauché par l'enfant, l'étayage (code) est noté avec le signe de transcription auto-interruption (-).
50. A la fin de chaque étayage, si la bonne réponse n'est pas fournie par l'enfant, l'expérimentateur peut lui donner la bonne forme. Cette information est insérée dans les commentaires.
51. Il n'est pas toujours aisé de déterminer le nombre d'étayage. Nous considérons généralement un type d'étayage par tour de parole.

Pour tous les nouveaux mots :

52. A1 : répétition de l'item/de la demande / questions (ex. tu te souviens comment il s'appelle ?
53. A2 : encouragements / feedbacks correctifs (les phatiques sont considérés comme des feedbacks correctifs. Ex. ouais). Les recadrages de comportement sont également codés par A2.

Substantifs et verbes :

54. Désignation :

- A3 : Indices sémantiques

55. Dénomination :

- A3 : indices sémantiques
- A4 : Déterminant/pronom
- A5 : Déterminant/pronom + ébauche phonémique (avec cluster)

Phrases :

56. Désignation :

- A3 : questions
- A4 traits sémantiques substantif
- A5 traits sémantiques verbe

57. Dénomination :

- A1 : répétition de la demande ou questions (ex. comment il s'appelle lui ?)
- A3 : déterminant
- A4 : déterminant + ébauche phonémique (cluster)
- A5 : déterminant + substantif + ébauche phonème verbe (cluster)
- A6 : déterminant + substantif + verbe sans flexion

10. Réaction à l'étayage

Insérer si plusieurs réactions à la suite.

58. Réponse

59. Absence de réponse

Le nombre de syllabes, phonèmes corrects et erronés est noté pour les productions spontanées de l'enfant.

11. Commentaires

Dans la colonne commentaires AJ, sont insérés tous les commentaires liés aux passations et notamment les feedbacks finaux correctifs parfois fournis par l'expérimentatrice. Les phatiques finaux sont également insérés dans la colonne « commentaires ». Ainsi que les cas particuliers avec commentaires et justifications.

Annexe H : Guide d'étayage

Histoire 1A

Désignation	Items	Aide 1	Aide 2	Aide 3		
		Répétition	Encouragements	Indices sémantiques		
	Le /lati/	Le /lati/		Créature	Grand	Bleu
	Le /fli/	Le /fli/		Objet	Décoration	Brun
	La /plifu/	La /plifu/		Créature	Petite	Verte

Dénomination	Items	Aide 1	Aide 2	Aide 3	Aide 4	Aide 5
		Répétition /questions	Feedback encouragements	Traits sémantiques	Déterminant	Déterminant + Ébauche phonémique (avec cluster)
Une	Comme appelle-t-on /plifu/ cette créature ?			Créature, petite, verte	Une	Une pl
Un /fli/	Comme appelle-t-on cet objet ?			Objet, décoration, brun	Un	Un fl
Un /lati/	Comme appelle-t-on cette créature ?			Créature, grand, bleu	Un	Un l

Histoire 1B

Désignation	Items	Aide 1	Aide 2	Aide 3		
		Répétition de l'item	Encouragements	Indices sémantiques		
	Le /stabi/	Le /stabi/		Créature	Cheveux	Jaune
	La /blu/	La /blu/		Objet	Hanses	Violet
	La /otu/	La /otu/		Créature	3 yeux	Rouge

Dénomination	Items	Aide 1	Aide 2	Aide 3	Aide 4	Aide 5
		Répétition /questions	Encouragements	Traits sémantiques	Déterminant	Déterminant + Ébauche phonémique (ou cluster)
	Une /otu/	Comme appelle-t-on cette créature ?		Créature, 3 yeux, rouge	Une	Une ho
	Une /blu/	Comme appelle-t-on cet objet ?		Objet, hanches, violet	Une	Un bl
	Un /stabi/	Comme appelle-t-on cette créature ?		Créature, cheveux, jaune	Un	Un st

Histoire 2A

Désignation	Items	Aide 1	Aide 2	Aide 3		
		Répétition de l'item	Encouragements	Indices sémantiques		
	Il /upit/	Il /upit/		Lui-même	Roulade	Gauche
	Il /lut/ qqch	Il /lut/ qqch		Concentration	Approcher	Baguette
	Il /flinat/ qqch	Il /flinat/ qqch		Agrandir	Amusant	Ballon

Dénomination	Items	Aide 1	Aide 2	Aide 3	Aide 4	Aide 5
		Répétition	Feedback	Traits sémantiques	Pronom	Pronom + Ébauche
		/questions	encouragements			phonémique
	Il /upit/	Qu'est-ce qu'il fait ?		Lui-même, roulade, gauche	Il	Il u
	Il /lut/	=		Concentration, approcher, baguette	Il	Il l
	Il /flinat/	=		Agrandir, amusant, ballon	Il	Il fl

Note. Le signe = indique que les éléments sont identiques à la ligne précédente.

Histoire 2B

Désignation	Items	Aide 1	Aide 2	Aide 3		
		Répétition	Encouragements	Indices sémantiques		
	Il /fakat/	Il /fakat/		Lui-même	Descendre	Tourner
	Il /vat/ qqch	Il /vat/ qqch		Concentration	Faire avancer parterre	Feuille
	Il /aplut/ qqch	Il /aplut/ qqch		Besoin de sa tête pour réfléchir	Disparaître	Maison

Dénomination	Items	Aide 1	Aide 2	Aide 3	Aide 4	Aide 5
		Répétition / questions	Encouragements	Traits sémantiques	Pronom	Pronom + Ébauche phonémique
	Il /fakat/	Qu'est-ce qu'il fait ?		Lui-même, descendre, tourner	Il	Il f
	Il /vat/	=		Concentration, faire avancer par terre, feuille	Il	Il v
	Il /aplut/	=		Besoin de sa tête pour réfléchir, disparaître, maison	Il	Il a

Note. Le signe = indique que les éléments sont identiques à la ligne précédente.

Histoire 3A

Désignation	Items	Aide 1	Aide 2	Aide 3	Aide 4			Aide 5		
		Répétition	Encouragements	Question	Traits sémantiques substantif			Traits sémantiques verbe		
	Le /lati upit/	Le /lati upit/		Tu te souviens du personnage ? de l'action ?	Créature	Grand	Bleu	Lui-même	Roulade	Gauche
	Le /lati lut/ qqch	Le /lati lut/ qqch		=	=	=	=	Concentration	Approcher	Baguette
	La /plifu flinam/ qqch	La /plifu flinam/ qqch		=	Créature	Petite	Verte	Agrandir	Amusant	Ballon
	La /plifu upim/	La /plifu upim/		=	=	=	=	Lui-même	Roulade	Gauche

La /plifu	La /plifu	=	=	=	Concentration	Approcher	Baguette
lum/	lum/ qqch						
qqch							
Le /lati	Le /lati	=	Créature	Grand	Bleu	Agrandir	Amusant
flinat/	flinat/ qqch						Ballon
qqch							

Note. Le signe = indique que les éléments sont identiques à la ligne précédente.

Dénomination	Items	Aide 1	Aide 2	Aide 3	Aide 4	Aide 5	Aide 6
		Répétition /questions	Encouragements, feedbacks	Déterminant	Déterminant + 1 phonème substantif	Déterminant + substantif + 1 ^{er} phonème verbe	Substantif + verbe sans flexion
	La /plifu upim/	Qu'est-ce qui se passe ?		Une	Une pl	Une plifu u	Une plifu upi
	Le /lati lut/	=		Un	Un l	Un lati l	Un lati lu
	La /plifu flinam/	=		Une	Une pl	Une plifu fl	Une plifu flina
	Le /lati upit/	=		Un	Un l	Un lati u	Un lati upi
	La /plifu lum/	=		Une	Une pl	Une plifu l	Une plifu lu
	Le /lati flinat/	=		Un	Un l	Un lati fl	Un lati flina

Note. Le signe = indique que les éléments sont identiques à la ligne précédente.

Histoire 3B

Désignation	Items	Aide 1	Aide 2	Aide 3	Aide 4			Aide 5		
		Répétition	Encouragements	Questions	Traits sémantiques substantif			Traits sémantiques verbe		
Le	Le /stabi			Tu te souviens	Créature	Cheveux	Jaune	Lui-même	Descendre	Tourner
/stabi	fakat/			du						
fakat/				personnage ?						
				de l'action ?						
Le	Le /stabi			=	=	=	=	Concentration	Faire	Feuille
/stabi	vat/ qqch								avancer	
vat/									parterre	
qqch										
La /otu	La /otu			=	=	3 yeux	Rouge	Besoin de sa	Disparaître	Maison
aplum/	aplum/							tête pour		
qqch	qqch							réfléchir		

La /otu	La /otu	=	=	=	=	Elle-même	Descendre	Tourner
fakam/	fakam/							
La /otu	La /otu	=	=	=	=	Concentration	Faire	Feuille
vam/	vam/ qqch						avancer	
qqch							parterre	
Le	Le /stabi	=	=	Cheveux	Jaune	Besoin de sa	Disparaître	Maison
/stabi	aplut/					tête pour		
aplut/	qqch					réfléchir		
qqch								

Note. Le signe = indique que les éléments sont identiques à la ligne précédente.

Dénomination	Items	Aide 1	Aide 2	Aide 3	Aide 4	Aide 5	Aide 6
		Répétition / questions	Encouragements/feedbacks	Déterminant	Déterminant + 1 ^{er} phonème substantif	Déterminant + substantif + 1 ^{er} phonème verbe	Substantif + verbe sans flexion
	La /otu fakam/	Qu'est-ce qui se passe ?		Une	Une o	Une otu f	Une otu faka
	Le /stabi vat/	=		Un	Un st	Un stabi v	Un stabi va
	La /otu aplum/	=		Une	Une o	Une otu a	Une otu aplu
	Le /stabi fakat/	=		Un	Un st	Un stabi f	Un stabi faka
	La /otu vam/	=		Une	Une o	Une otu v	Une otu va
	Le /stabi aplut/	=		Un	Un st	Un stabi a	Un stabi aplu

Note. Le signe = indique que les éléments sont identiques à la ligne précédente.

