

Université de Montréal

**TELEQ : Création et pré-validation d'un test de compréhension de lecture pour les
enfants québécois de niveau scolaire primaire**

par Mathilde Picotte-Lavoie

Département de psychologie, Faculté des arts et des sciences

Essai doctoral présenté en vue de l'obtention du grade de D.Psy, Doctorat en psychologie
clinique, option neuropsychologie clinique (D.Psy.)

Août 2020

© Mathilde Picotte-Lavoie, 2020

Table des matières

Résumé.....	3
Abstract	5
Liste des tableaux	7
Liste des graphiques	8
Liste des abréviations	9
Remerciements.....	11
Contexte théorique	15
Modèle de van Dijk et Kintsch (1983).....	18
Critique des outils de compréhension de lecture existants	23
Méthodologie.....	28
Développement du test <i>Une bouteille à la mer</i>	28
Participants (critères d'éligibilité et d'exclusion).....	31
Outils d'évaluation	33
Description du test <i>Une bouteille à la mer</i>	33
Outils utilisés afin de s'assurer de l'éligibilité des participants.	34
Tests utilisés afin de vérifier la fidélité et validité du test <i>Une bouteille à la mer</i>	36
Analyses statistiques prévues et prédictions	36
Fidélité test-retest.	36
Fidélité inter-juge.	37
Validité concordante	38
Validité critériée concomitante	38
Résultats	39
Description de l'échantillon final.....	39
Analyses préliminaires	41
Analyses de pré-validation.....	42
Fidélité test-retest.	42
Fidélité inter-juge.	43
Cohérence interne.....	44
Degré de difficulté des questions.	44
Validité concordante.	44
Validité critériée : niveau scolaire.....	45
Validité critériée : rendement scolaire – notes en français.....	46
Discussion	46
Conclusion	55

Résumé

Contexte. Afin d'évaluer la compréhension de lecture, les cliniciens n'ont d'autres choix que de se rabattre sur un nombre limité de tests dont la plupart ne sont pas adaptés à la réalité québécoise, étant d'origine européenne, sans compter que la majorité de ceux-ci ne répondent pas aux standards psychométriques, n'évaluent pas la compréhension de texte et ne se fondent pas sur un modèle théorique. **Objectifs.** Les objectifs de la présente étude consistent donc à développer et à pré-valider un test permettant de pallier à ces lacunes. Le test *Une bouteille à la mer* a été créé (épreuves de lecture, de rappel et de questions) et ses propriétés psychométriques ont été mesurées (fidélité test-retest, inter-juges, cohérence interne ; degré de difficulté des items ; validité concordante et critériée (niveau scolaire et rendement en français)). **Méthodologie.** Le test a été administré à deux reprises afin d'évaluer la fidélité test-retest. Deux correcteurs ont procédé à la correction du test afin de mesurer la fidélité inter-juges. Le degré de difficultés des items a été étudié. Un autre test de compréhension de lecture a été administré afin de mesurer la validité concordante et les performances des enfants ont été analysées en fonction de leur niveau scolaire et de leur note en français, mesurant ainsi la validité critériée. **Résultats.** Au total, 58 enfants franco-québécois de la 4^e à la 6^e année ont pris part à l'étude. En ce qui a trait aux propriétés psychométriques les coefficients de corrélation varient entre ,74 et ,86 pour la fidélité test-retest et entre ,96 et ,98 pour la fidélité inter-juges, ce qui s'avère être des coefficients acceptables, voir excellents. La cohérence interne surpasse quelque peu le seuil minimal escompté de ,70 et les questions varient quant à leur degré de difficulté, et ce, peu importe le niveau scolaire considéré. Le test *Une bouteille à la mer* et le test évaluant la compréhension de lecture corrélaient entre eux modérément et les performances des enfants tendent à augmenter en fonction du niveau scolaire et de leur rendement scolaire en français, ce qui dénote de bons

indices de validité concordante et critériée. **Conclusion.** En somme, le test *Une bouteille à la mer* est prometteur étant donné qu'il possède des qualités psychométriques globalement satisfaisantes. **Mots clés :** Compréhension de lecture, Psychométrie, Enfance, Modèle de van Dijk et Kintsch

Abstract

Context. In order to assess reading comprehension, clinicians have to rely on a limited number of tests, most of which are not adapted to the reality of Quebec, because most of them have been developed in Europe. Also, the majority of these tests only assess sentence comprehension and not text comprehension, not to mention that most of them do not meet psychometric standards and are not based on a theoretical model. **Objectives.** The aims of the present study were to develop and pre-validate a test that would address these shortcomings. The test *Une bouteille à la mer* was created (reading, recall, and question items) and its psychometric properties were measured (test-retest reliability, inter-rater reliability, internal consistency; degree of item difficulty; concurrent and criterion-related validity (grade level and performance in French)). **Methodology.** The test was administered twice to assess test-retest reliability. Two judges corrected the test to measure inter-rater reliability. The level of difficulty of each item was studied. Another reading comprehension test was administered to measure concurrent validity, and children's performance was analyzed according to their grade level and French score, thus measuring criterion-related validity. **Results.** A total of 58 Franco-Quebec children in grades 4 to 6 participated in the study. Regarding psychometric properties, correlation coefficients ranged from .74 to .86 for test-retest reliability and from .96 to .98 for inter-rater reliability, which are acceptable even excellent. Internal consistency was slightly above the expected threshold of .70. The questions varied in difficulty levels in all grade levels. The test *Une bouteille à la mer* and the other test assessing reading comprehension correlate with each other at a moderate degree. Children's performance tended to increase following their grade level and academic achievement in French, indicating good concurrent and criterion-related validity. **Conclusion.** In sum, *Une bouteille à la mer* test is promising given that it has globally satisfactory

psychometric qualities. **Key words:** Reading comprehension, Psychometrics, Childhood, van Dijk and Kintsch

Liste des tableaux

Tableau 1

Caractéristiques sociodémographiques de l'échantillon total ainsi que par niveaux

scolaires.....p.39

Tableau 2

Statistiques descriptives de l'échantillon total ainsi que par niveau

scolaire.....p.40

Tableau 3

Performances au test Une bouteille à la mer pour l'échantillon total ainsi que par niveau

scolaire.....p.45

Liste des graphiques

Graphique 1

Moyennes des scores aux questions en fonction du niveau scolaire (4^e, 5^e et 6^e année).....p.43

Liste des abréviations

ADEL UQAM : Apprenants en difficulté et littératie de l'Université du Québec à Montréal

APA : American Psychiatry Association

BALE : Batterie analytique du langage écrit

C : Processus non spécifiques de compréhension langagière

CCI : Coefficient de corrélation intraclasse

CELF CND-F : Évaluation Clinique des notions langagières fondamentales – version pour francophones du Canada

CÉREP : Comité d'éthique et de la recherche en arts et sciences

CEREP : Comité d'éthique de la recherche en éducation et en psychologie

CPO : Compréhension de paragraphe à l'oral

CS : Commission scolaire

CSDA : Commission scolaire des Affluents

CSDL : Commission scolaire de Laval

CSDS : Commission scolaire des Samares

CSPN : Commission scolaire Pierre-Neveu

CSRDN : Commission scolaire de la Rivière-du-Nord

CSSMI : Commission scolaire de la Seigneurie-des-Milles-Îles

D.OR.LEC : Disposition Orthographe Lecture

DSM : Manuel diagnostique et statistique des troubles mentaux

É.CO.S.SE : Épreuve de compréhension syntaxico-sémantique

Équipe LLL : Équipe des conseillères pédagogiques en orthopédagogie de Laval-Lanaudières-Larentides

I : Processus spécifiques d'identification des mots écrits

IC : Intervalle de confiance

L : Lecture

LÉNEA : Laboratoire d'études en neuropsychologie de l'enfant et de l'adolescent

MELS : Ministère de l'Éducation, du Loisir et du Sport

MSL : Modèle simple de la lecture

PDA : Progression des apprentissages

QI : Potentiel intellectuel

SDNV : Syndrome des dysfonctions non verbales

TRF : Test de rendement pour francophone

TSA : Trouble du spectre autistique

WIAT-II : Test individuel de rendement de Wechsler – seconde édition – version québécoise

WISC-V : Échelle d'intelligence de Wechsler pour enfants – 5^e édition – version pour francophones du Canada

Remerciements

Je tiens d'abord à remercier mon directeur de recherche, Dr Bruno Gauthier, pour son aide et son soutien tout au long des trois dernières années. Merci de m'avoir fait confiance en me laissant prendre part au projet TELEQ, projet que tu affectionnes particulièrement. J'ai beaucoup aimé travailler sur ce projet puisqu'en plus de s'avérer utile d'un point de vue clinique, il est le produit d'un travail d'équipe sans précédent. En effet, le TELEQ ne serait pas l'outil qu'il est sans l'apport d'un grand nombre de personnes. J'en profite pour remercier les auxiliaires de recherche avec qui j'ai eu la chance de travailler, soit Sophie Constant, Marion Cyr-Morin, Anabelle Guérard, Noémie Maheux-Racine et Evelyne Roy. Ce fût très agréable de vous côtoyer et de travailler avec vous.

Dans le cadre du présent projet, j'ai également eu la chance de collaborer avec une équipe de conseillères pédagogiques composée de femmes brillantes, drôles et travaillantes. Justine Bourgeois, Marie-Ève Brisette, Isabelle Brunet, Audray Grand-Maison, Marie-Josée Mongeau, Sylvie Nardone, Vanessa Savaria, Judy-Ann Valiquette et Genevière Turbide, merci du fond du cœur de m'avoir accueilli à bras ouverts au sein de votre équipe. Ce fût un réel plaisir d'échanger avec vous et de travailler à l'élaboration du présent outil avec vous. Merci également à Marie-France Bouchard pour la rédaction du texte *Une bouteille à la mer* ainsi qu'à Mélanie Bédard, Marie-Ève Côté et la Dre Line Laplante pour leurs recommandations vis-à-vis l'élaboration du test et leurs précieux conseils.

Aussi, mon expérience de recherche n'aurait pas été la même sans ma collègue, mon acolyte, mais surtout mon amie, Patricia Laniel. Merci pour les nombreux fous rires qui ont accompagné nos rencontres. Tu as été pour moi cette personne vers qui je pouvais me tourner

en cas de doutes et de questionnements et tu savais toujours quoi répondre pour me redonner confiance et raviver ma motivation. Mille mercis.

Je tiens également à remercier mon entourage sans qui je n'y serais simplement pas arrivée. Merci à mes parents pour leur amour inconditionnel et leur constant support qui m'a permis de mener à terme non seulement cet essai, mais l'entièreté de mes études doctorales. Merci à ma sœur pour son écoute et à mon frère pour ses conseils grammaticaux. Et que dire de mes amis...Charlie, Vincent, Michèle, Aurélie D.G., Flavie, Marie-Laurence, Amélie, Aurélie T., Laurence, merci de m'avoir épaulé dans les bons et moins bons moments. Je suis choyée de vous avoir. Et finalement, un merci tout particulier à mes consoeurs du D.Psy sur lesquelles j'ai pu compter dès le jour 1.

Introduction

L'essai présenté dans la section qui suit fût rédigé sous forme d'article scientifique. Le manuscrit sera soumis à la *Revue Canadienne d'orthophonie et d'audiologie* suite à l'acceptation dudit essai doctoral.

Article

TELEQ : Création et pré-validation d'un test de compréhension de lecture pour les enfants
québécois de niveau scolaire primaire

Manuscrit soumis à la *Revue canadienne d'orthophonie et d'audiologie (RCOA)*

Mathilde Picotte-Lavoie, Patricia Laniel et Bruno Gauthier¹

Université de Montréal

Note sur les auteurs

Mathilde Picotte-Lavoie, Département de Psychologie, Université de Montréal

Patricia Laniel, Département de Psychologie, Université de Montréal

Bruno Gauthier, Ph.D., Département de Psychologie, Université de Montréal

Le contenu de l'article se retrouve également dans l'essai doctoral de Mathilde Picotte-Lavoie

La correspondance concernant l'article peut être adressée à Mathilde Picotte-Lavoie

Département de psychologie

Université de Montréal (Campus Laval)

1700, rue Jacques-Tétreault

Laval (Québec) H7N 0B6 Canada

Courriel : mathilde.picotte-lavoie@umontreal.ca

¹ Le test qui fait l'objet de cette étude a été conceptualisé par la 1^{re} auteure et les milieux scolaires collaborateurs. Les deux premières autrices de l'article ont participé à la collecte de données, mais les analyses statistiques ainsi que la rédaction de l'article ont été entièrement réalisées par la 1^{ère} autrice, soit Mathilde Picotte-Lavoie, sous la supervision de Bruno Gauthier. Merci à Line Laplante et Marie-France Côté pour leur commentaires suite à la relecture d'une version initiale du projet.

Contexte théorique

Position du problème

La compréhension de lecture² représente le but ultime de l'apprentissage de la lecture (Hawken, 2009; Nation, 2005, 2019). Au début du cursus scolaire, les enfants apprennent à lire pour finalement lire pour apprendre (Chall, 1983; St-Pierre et al., 2000). Autrement dit, la compréhension de lecture revêt un caractère essentiel puisqu'elle sous-tend les apprentissages en favorisant l'acquisition de connaissances. Si plusieurs enfants sont en mesure d'extraire le sens d'un texte, ce n'est pas le cas de tous (Clarke et al., 2010). Au Québec, près du quart des élèves de 6^e année échoue l'épreuve ministérielle en lecture, les prédisposant ainsi à vivre des difficultés au secondaire puisque la lecture n'y est que peu enseignée de façon formelle et explicite (Desrosiers et Tétreault, 2012; Van Grunderbeeck et al., 2004). Un dépistage précoce des difficultés en compréhension de lecture est donc nécessaire puisqu'un faible niveau de compréhension nuit à l'acquisition des connaissances et engendre des difficultés scolaires d'ordre général (Desrosiers et Tétreault, 2012; Potocki et al., 2014) compromettant le développement de l'enfant. Les outils actuellement disponibles afin d'évaluer cette compétence sont peu nombreux et ne sont pas adaptés à la réalité linguistique et culturelle québécoise. Par ailleurs, la majorité de ceux-ci ne sont pas fondés sur des théories cognitives et ne répondent pas aux standards psychométriques en matière de fidélité et de validité.

L'objectif de la présente étude consiste donc à développer et à pré-valider un nouveau test d'évaluation de la compréhension de lecture chez les enfants franco-québécois de la 4^e à la 6^e année. Ce test s'apparentera aux épreuves de compréhension de lecture auxquelles les élèves

² Dans le présent article, le terme « compréhension de lecture » renvoie à l'habileté à comprendre un texte écrit. Ce faisant, les termes « compréhension de lecture », « compréhension écrite » et « compréhension de l'écrit », sont des termes utilisés de manière interchangeable.

sont exposés en classe et permettra de mieux desservir la population pédiatrique franco-québécoise en ce qui a trait à l'identification des élèves présentant des difficultés en compréhension de lecture.

La compréhension de lecture

La compréhension de lecture peut être définie comme étant « un processus simultané d'extraction et de construction de sens chez le lecteur par le biais d'une interaction constante entre ses connaissances et les informations contenues dans le texte » (St-Pierre et al., 2000, p.36). Cette définition met en évidence la complexité de la compréhension de l'écrit puisqu'elle montre que, pour comprendre un texte, le lecteur doit non seulement lire les mots et les phrases qui composent le texte, mais il doit également les comprendre, les mettre en relation et les contextualiser afin de se créer une représentation mentale intégrée de la situation décrite dans le texte. Pour ce faire, de nombreux processus langagiers et cognitifs doivent être mis à contribution simultanément et séquentiellement, ce pour quoi la compréhension de lecture est considérée comme étant une compétence multidimensionnelle (Turcotte et Talbot, 2017) et de haut niveau (St-Pierre et al., 2000).

D'un point de vue théorique, le cadre de référence le plus souvent cité afin de rendre compte de la compréhension de lecture est le modèle simple de la lecture (MSL) (Gough et Tunmer, 1986; Hoover et Gough, 1990). Selon celui-ci, la compréhension de lecture (L) résulte de l'interaction entre les habiletés de décodage (identification des mots écrits : D) et les habiletés de compréhension linguistique (compréhension orale : C) ($L = D \times C$) (Gough et Tunmer, 1986). À elles seules, ces habiletés ne permettent pas de comprendre le langage écrit, toutes deux doivent être employées (Castles et al., 2018; Nation, 2019). Un déficit au niveau de l'une de ces

habiletés, ou de ces deux habiletés (*D* et/ou *C*), compromet la compréhension écrite (Chapleau et al., 2014; Gough et Tunmer, 1986; Hoover et Gough, 1990; Israel et Duffy, 2014).

Le cadre général d'apprentissage de la lecture-écriture proposé par la Dre Line Laplante (2011), reprend le MSL, mais une terminologie légèrement différente y est employée. Les habiletés de décodage portent le nom de processus spécifiques d'identification des mots écrits, étant spécifiques au langage écrit, alors que les habiletés de compréhension linguistique portent le nom de processus non spécifiques de compréhension langagière, étant non spécifiques au langage écrit ; ces processus permettant de comprendre tout type de discours (écrit et oral) (Laplante, 2011). Cette façon de conceptualiser les processus non spécifiques tient compte de la contribution relative de ces processus à travers le temps (Hoover et Gough, 1990). En effet, au début de l'apprentissage de la lecture, les pauvres habiletés de décodage (*D*) limitent la capacité de tout lecteur à comprendre un texte. À mesure que ces habiletés se perfectionnent et s'automatisent, la compréhension écrite (*L*) dépend davantage des habiletés de compréhension linguistique (Gough et Tunmer, 1986; Lervåg et al., 2018; Perfetti et al., 2005; Verhoeven et Van Leeuwe, 2008). D'ailleurs, la corrélation entre la compréhension de lecture et la compréhension linguistique augmente avec les années, pour finalement atteindre ,90 chez les adultes devenus « experts » de la lecture (Gernsbacher et al., 1990). Aussi, lorsque les habiletés de décodage sont pleinement maîtrisées ($D = 0$), le lecteur en vient à comprendre aussi bien un discours écrit qu'oral ($L = C$); l'écrit pouvant être oralisé (Majerus et al., 2020). Des processus similaires sous-tendent donc la compréhension écrite et orale (Gough et Tunmer, 1986; Hoover et Gough, 1990; Majerus et al., 2020; Nation, 2019).

Une critique a néanmoins été formulée à propos du MSL : celui-ci ne tient pas compte des caractéristiques propres au langage écrit (Majerus et al., 2020). En effet, à l'écrit, le

vocabulaire employé est plus soutenu, et ce, même dans les livres destinés aux enfants d'âge préscolaires (Montag et al., 2015) et les structures phrastiques employées sont plus complexes (Montag, 2019; Montag et MacDonald, 2015). Ce faisant, un lecteur novice est confronté à un langage écrit dont la complexité surpasse celle du langage oral (Majerus et al., 2020).

Au Québec, le cursus scolaire est tel que lors des premières années du primaire, l'accent est surtout mis sur l'apprentissage de la lecture et l'amélioration de la fluidité de lecture; travaillant ainsi les processus spécifiques (Ministère de l'Éducation du Loisir et du Sport [MELS]³, 2009a). Néanmoins, en se référant à la progression des apprentissages du MELS, il est possible de constater que des stratégies de compréhension de lecture sont enseignées dès la 1^{ère} année du primaire. Cela dit, il est attendu que la vaste majorité de celles-ci soient maîtrisées qu'à compter de la 4^e année (MELS, 2009a). Le test développé dans le cadre de la présente étude s'adressera donc aux élèves de la 4^e à la 6^e chez qui les stratégies de compréhension sont acquises, ou en voie de l'être, et pour qui la compréhension écrite dépend davantage de processus non spécifiques. Le test prendra assise sur le modèle cognitif le plus influent en matière de processus non spécifiques, soit le modèle de van Dijk et Kintsch (Kintsch et Van Dijk, 1978; Laplante, 2011; Van Dijk et Kintsch, 1983), tel que décrit dans le cadre général d'apprentissage de la lecture-écriture du Dre Line Laplante (2011).

Modèle de van Dijk et Kintsch (1983)

Ce modèle fait état des différents niveaux de traitement nécessaires à la construction d'une représentation mentale de la situation décrite dans un discours écrit ou oral (Bernard,

³ Le Ministère de l'Éducation du Loisir et du Sport (MELS) porte désormais le nom du Ministère de l'Éducation et de l'Enseignement Supérieur (MEES) suite à la fusion du MELS et du ministère de l'Enseignement supérieur, de la Recherche et de la Science (MESRS). Cela dit, la progression des apprentissages a été rédigée alors que le ministère portait toujours le nom du MELS, d'où la présente référence.

2017; Laplante, 2011; Pecatte, 2015; Van Dijk et Kintsch, 1983). La mise en œuvre simultanée de ces niveaux de traitement, combinée à leurs constantes interactions (mécanismes de rétroactions), permettent la construction de cette représentation mentale.

Le premier niveau de traitement, soit l'analyse syntaxique, consiste à traiter séquentiellement les mots et les phrases contenus dans le texte. Pour ce faire, les mots sont d'abord décodés via des processus perceptuels, pour ensuite se voir attribuer un sens ainsi qu'une fonction au sein de la phrase (déterminants, noms, verbes, groupes nominaux, groupes du verbe, etc.) (Bernard, 2017; Kintsch et Rawson, 2005; Laplante, 2011; Van Dijk et Kintsch, 1983). Ce traitement se limite à l'analyse de la structure syntaxique d'une succession de mots et ne tient pas compte de la sémantique globale d'un groupe de mots ou d'une phrase. Les phrases traitées peuvent donc respecter les règles syntaxiques sans avoir de sens pour autant. C'est lors des prochains niveaux de traitement qu'une analyse sémantique plus holistique sera effectuée par le lecteur.

Le deuxième niveau de traitement mène à l'élaboration d'une représentation sémantique du discours, dénommée base de texte (Bernard, 2017; Laplante, 2011; Van Dijk et Kintsch, 1983b). Celle-ci est constituée de propositions, soit des unités sémantiques issues de la combinaison du sens de plusieurs mots contenus dans le discours. Afin de comprendre le discours, les propositions doivent être liées entre elles de manière à former un réseau complexe d'informations (Laplante, 2011). Pour ce faire, les propositions sont traitées localement et globalement, formant ainsi deux niveaux hiérarchiques : la microstructure (traitement local) et la macrostructure (traitement global).

La microstructure est composée de micropropositions traitées de manière séquentielle. Afin d'assurer la cohérence de la microstructure, divers processus (microprocessus) sont

sollicités dont l'interprétation des anaphores (comprendre la chaîne anaphorique) et l'identification des marqueurs de relation (comprendre leur sens et leur fonction au sein de la phrase) (Laplante, 2011). Les anaphores sont des mots utilisés pour en remplacer d'autres (p.ex. : « Jean et Marie aiment le sport. *Il* pratique le ski et *elle*, la natation ») (Giasson, 1990, p.55). Il s'agit généralement d'un pronom qui vient remplacer un nom, ou encore, une proposition. Le lecteur compétent est en mesure d'identifier correctement l'entité auquel renvoie l'anaphore, soit le référent. Quant aux marqueurs de relation (ou connecteurs), ils sont utilisés afin de relier des événements entre eux (Giasson, 1990) (p.ex. : « J'ai mal au ventre *parce que* j'ai trop mangé de chocolat ») (Laplante, 2011, p.148). Ils permettent d'établir une relation de continuité et/ou de causalité entre deux micropropositions. La microstructure servira à l'élaboration d'une représentation sémantique plus globale, soit la macrostructure.

La macrostructure se compose des idées essentielles du texte et est influencée par les connaissances du lecteur, ses croyances et ses objectifs (Van Dijk et Kintsch, 1983). Plusieurs caractéristiques inhérentes au texte guident la construction de la macrostructure, telles que le genre de discours, les thèmes abordés et l'organisation des idées (Biber, 1986; Van Dijk et Kintsch, 1983). Elle se compose de macropropositions qui résultent de l'application de trois macrorègles : 1) Suppression : Éliminer les micropropositions non pertinentes ; 2) Généralisation : Substituer une série de micropropositions par une macroproposition qui reprend l'idée véhiculée par ces micropropositions (Kintsch et Van Dijk, 1978; Van Dijk et Kintsch, 1983a) et 3) Construction : Remplacer une série de micropropositions par une macroproposition qui tient compte du contenu de chacune de ces micropropositions (Van Dijk et Kintsch, 1983). Contrairement à la généralisation, les micropropositions sur lesquelles s'applique la

construction renvoient à un même fait et non à une même idée (Pecatte, 2015; Van Dijk et Kintsch, 1983).

L'application de ces règles permet d'aller à l'essentiel, tout en préservant la signification du discours (Kintsch et Van Dijk, 1978). Les macropropositions, issues de ces macrorègles, doivent être organisées. Pour ce faire, le lecteur peut tirer profit de la structure du discours. Les discours sont généralement dotés d'une structure, dont certaines sont dites conventionnelles. C'est le cas de la structure de récit dont les principales composantes sont : la situation initiale, l'élément perturbateur, les péripéties, la résolution et la situation finale (Giasson, 1990; Laplante, 2011). Lorsqu'un lecteur réalise qu'un discours suit une structure conventionnelle, ce dernier récupère en mémoire à long terme les connaissances qu'il possède sur ladite structure afin de l'aider à organiser l'information et ainsi mieux la retenir (Giasson, 1990; Laplante, 2011). L'ensemble des connaissances relatives à la structure d'un discours constitue un schéma. Il en existe plusieurs types dont le schéma de récit, propre à la structure de récit, qui se définit comme une « représentation interne idéalisée des parties d'un récit typique » (Giasson, 1990, p. 99). L'enfant d'âge scolaire, étant souvent confronté à des structures de récit, se sert du schéma de récit afin de faire des prédictions concernant le déroulement de l'histoire, ainsi que pour juger de la pertinence de l'information (Giasson, 1990; Kintsch et Van Dijk, 1978). Le schéma oriente donc l'application des macrorègles ainsi que l'organisation de la macrostructure.

Par ailleurs, certaines informations contenues dans le discours sont implicites et doivent être rendues explicites par le lecteur afin que la cohérence de la base de texte soit préservée. Pour y arriver, le lecteur doit faire des inférences, c'est-à-dire qu'il doit, à partir du texte et de ses connaissances, faire des liens de manière à déceler les informations implicites du texte (Laplante, 2011). Les inférences permettent d'outrepasser une compréhension littérale

(Giasson, 1990) et s'avèrent donc indispensables à la compréhension écrite. Il s'agit également également d'un processus sollicité lors de la construction de la microstructure et de la macrostructure permettant ainsi d'assurer la cohésion de la base de texte d'un point de vue local et global. Bien qu'il existe différents systèmes de classification des inférences, le cadre de référence du Dre Laplante (2011) ainsi que le modèle de van Dijk et Kintsch (1983) ne proposent ni ne mettent de l'avant, un système de classification des inférences.

Le troisième niveau de traitement mène à l'élaboration du modèle de situation qui correspond à une représentation mentale signifiante de la situation (Kintsch et Van Dijk, 1978; Laplante, 2011; Van Dijk et Kintsch, 1983). Elle émane de l'intégration de la base de texte aux connaissances antérieures du lecteur (Laplante, 2011). Ce n'est que lorsque le lecteur établit un tel lien qu'il peut avoir une compréhension approfondie de tout type de discours et qu'il peut acquérir de nouvelles connaissances (Bianco, 2010).

Le test développé dans le cadre de la présente étude évaluera donc l'ensemble des processus associés aux différents niveaux de traitement, soit : l'analyse syntaxique, les microprocessus, les macroprocessus, les inférences ainsi que la capacité du lecteur à se créer une représentation mentale signifiante de la situation décrite dans un texte.

Les enfants ayant de la difficulté à se construire une représentation mentale de la situation décrite dans un texte sont à risque de recevoir un diagnostic de trouble spécifique des apprentissages dans le domaine de la compréhension de lecture, tel que défini par le Manuel diagnostique et statistique des troubles mentaux, 5^e édition (DSM-5) (American Psychiatric Association (APA), 2013). Ce trouble se caractérise par une « difficulté à comprendre ce qui est lu (p.ex. peut lire un texte correctement, mais ne pas comprendre l'ordre, les relations, les déductions ou les significations plus profondes de ce qui est lu) » (APA, 2013, p.76). Ce

diagnostic est posé lorsque les habiletés de compréhension de lecture sont en deçà du niveau escompté pour l'âge, tel qu'objectivé par l'utilisation d'outils standardisés administrés individuellement, ainsi que lorsque les difficultés persistent au-delà de six mois malgré la mise en place d'interventions ciblées. De plus, ces difficultés doivent se manifester au cours du développement et ne doivent pas être imputables à un enseignement inadéquat ou un trouble perceptuel, neurologique, mental ou médical.

Critique des outils de compréhension de lecture existants

Afin de statuer sur la présence ou l'absence de ce trouble spécifique des apprentissages, des outils standardisés sont mis à la disposition des cliniciens. Pour juger de la qualité de ces outils, différents éléments seront examinés dont : les caractéristiques inhérentes à l'outil (le type de lecture impliquée (orale ou silencieuse), l'unité de texte auquel renvoient les items (phrases ou texte) et la forme des items (rappel, réponses à court ou long développement, réponses à choix multiples, sélection d'une image analogue et complémentation de phrases)), leurs fondements théoriques, leurs propriétés psychométriques ainsi que la représentativité de l'échantillon normatif.

Les tests faisant partie de batteries d'évaluation du langage écrit seront d'abord analysés, suivis des outils dédiés uniquement à la compréhension de l'écrit. Seuls les outils fréquemment utilisés au Québec seront abordés.

En ce qui a trait aux batteries d'évaluation du langage écrit, la majorité de celles-ci ne détiennent pas de sous-test dédié à la compréhension de lecture. La *Batterie analytique du langage écrit* (BALE ; Jacquier-Roux et al., 2010) est l'une des rares batteries à posséder une épreuve de compréhension écrite. Ce sous-test est composé de 12 phrases devant être lues à voix haute par l'enfant. Par la suite, quatre images sont présentées à l'enfant et celui-ci doit choisir

laquelle correspond à la phrase lue préalablement. Ce sous-test est en fait une version abrégée de l'*Épreuve de compréhension syntaxico-sémantique* (É.CO.S.SE ; Lecocq, 1996). La BALE et l'É.CO.S.SE comportent d'importantes limites lorsqu'utilisés auprès des élèves franco-québécois. Aucune information n'est prodiguée en ce qui a trait au modèle théorique sous-tendant ces sous-tests, hormis une augmentation de la complexité des items en matière de structure syntaxique. Ce faisant, seules des difficultés sur le plan syntaxique peuvent être mises en exergue, limitant ainsi l'analyse des difficultés en matière de compréhension écrite. De plus, ces outils se limitent à la compréhension de phrases; or la compréhension de lecture est une compétence qui s'étend bien au-delà de la simple compréhension de phrases (Potocki et al., 2014). Il a aussi été démontré que de tels outils font davantage appel aux habiletés de décodage qu'aux habiletés de compréhension (Keenan et al., 2008), omettant ainsi d'évaluer d'importants aspects de la compréhension de lecture. Les mêmes limites en matière de compréhension de phrases s'appliquent à l'épreuve L3 (issue de la batterie *Disposition orthographe lecture* : D.OR.LEC ; Lobrot, 1973; Lobrot et Adnet-Piat, 1976) et au *Woodstock-Johnson* (traduit de l'anglais) (Woodcock et al., 1990) évaluant tous deux la compréhension en demandant aux enfants de compléter des phrases en évoquant ou sélectionnant les mots appropriés. Il s'agit également d'épreuves ayant été développées il y a de cela plusieurs années et n'ayant pas été mises à jour depuis (L3 : 1973, D.OR.LEC : 1973). En ce qui concerne les qualités psychométriques de la BALE et de l'É.CO.S.SE, il n'existe pas de données à ce sujet. Par conséquent, il est difficile de statuer sur la validité et la fidélité de ces tests (Bouchard et al., 2009). Enfin, la BALE et l'É.CO.S.SE sont des outils français dont les normes ont été développées en Europe. Les données normatives ne sont donc pas représentatives de la population pédiatrique franco-québécoise.

Une autre batterie, soit le *Test de rendement individuel de Wechsler* (WIAT-II CND-F ; Wechsler, 2008) est utilisée afin d'évaluer les apprentissages scolaires des individus âgés entre 6 et 29 ans. Cette batterie contient un sous-test consacré à la compréhension de lecture où l'enfant doit lire à voix haute ou silencieusement des phrases ou de courts paragraphes avant de répondre oralement à des questions les concernant. Bien que cette batterie présente d'excellentes qualités psychométriques ainsi que des normes adaptées à la population pédiatrique franco-québécoise (Wechsler, 2005, 2008), le sous-test *Compréhension de lecture* ne repose sur aucune théorie cognitive. Les concepteurs avancent tout de même l'idée qu'il évaluerait la compréhension littérale et les inférences logiques (Turcotte et Talbot, 2017). Très peu d'informations peuvent donc être recueillies par rapport aux processus lacunaires, ce qui complique la prise en charge des enfants ayant des difficultés de compréhension. Il est tout de même fréquemment utilisé en clinique (Turcotte et Talbot, 2017) puisqu'il permet de statuer sur le rendement de l'enfant et ainsi savoir si ce dernier a accumulé un retard ou présente un déficit en compréhension de lecture comparativement aux élèves du même âge ou du même niveau scolaire que lui.

Quant au *Test de rendement pour francophone* (TRF ; Sarrazin, 1995), il s'agit d'une batterie de tests normalisés au Québec. Cette batterie fût d'abord conçue pour les adultes, puis adaptée aux élèves québécois suivant des programmes scolaires réguliers. Le TRF évalue les habiletés scolaires de base et se compose de six sous-tests dont l'un évalue exclusivement la compréhension de l'écrit. Lors de ce sous-test, il est demandé à l'enfant de lire des textes variés (affiches, publicités, textes éducatifs, poèmes, etc.) avant de devoir répondre à des questions. Tout comme la WIAT-II, le TRF ne se fonde sur aucune théorie cognitive, en plus de dater.

En ce qui concerne les outils d'évaluation se consacrant uniquement à l'évaluation de la compréhension de texte chez l'enfant d'âge primaire, un seul outil semble actuellement disponible, soit *La forme noire* (Maeder et Charlois, 2010); outil dédié aux enfants de 9 à 12 ans. Lors de la passation de cet outil, l'enfant doit lire un récit, pour ensuite compléter sept épreuves : rappel de récit, choix de titres, questions, jugement d'ordre d'importance, détection d'erreurs, résolution d'anaphores et tri et chronologie d'images. Chaque épreuve évalue de manière prédominante une habileté impliquée dans la compréhension de texte. Malgré le fait qu'il s'agisse d'un test relativement complet (Potocki et al., 2014), il présente certains inconvénients. En effet, le test n'offre aucune information concernant la fidélité, seule la validité a fait l'objet d'une étude de validation (Maeder et Charlois, 2010). Cette étude montre une progression du score en fonction de l'âge et met en lumière l'existence de patrons d'évolution propres aux différentes compétences évaluées. Ceci dit, seules des interprétations qualitatives des résultats quantitatifs sont fournies. Aucune analyse quantitative ne semble avoir été réalisée, remettant ainsi en question la véracité des conclusions avancées. Enfin, les normes ont été développées en France et ne sont pas adaptées aux enfants franco-québécois.

Afin de pallier à la pénurie d'outils en compréhension de texte, l'équipe de recherche des *Apprenants en difficulté et littératie de l'Université du Québec à Montréal* (ADEL UQAM, 2020) a conçu des épreuves de compréhension de lecture s'adressant aux élèves du primaire (4^e à 6^e année) et du secondaire (1^{ère} et 2^e année), scolarisés au Québec. Pour chaque niveau scolaire, deux textes accompagnés de questions sont mis à la disposition de tout professionnel désirant évaluer les habiletés de compréhension écrite (ADEL UQAM, 2020). Cependant, la majorité de ces épreuves n'ont pas fait l'objet d'études de validation, hormis un test s'adressant aux élèves de 6^e année (Turcotte et Talbot, 2017). L'administration de ce test se fait en grand groupe et

nécessite que les élèves lisent un texte informatif avant de répondre à 13 questions portant sur le texte. Il s'agit d'un test qui a été validé sur le plan théorique (modèle d'Irwin (2007) et modèle de classification des inférences de Bianco et Coda (2002) adapté par Dupin de Saint-André (2008), ainsi que sur le plan psychométrique. Aucune information concernant la fidélité inter-juges et test-retest n'est prodiguée et aucune norme n'a été développée.

Force est de constater que les outils actuellement mis à la disposition des cliniciens afin d'évaluer la compréhension de lecture présentent de nombreuses lacunes. En plus d'être peu nombreux, la majorité de ceux-ci ne sont pas adaptés à la clientèle pédiatrique québécoise puisqu'ils ont été développés, validés et normés en Europe (Bouchard et al., 2009; Garcia et Desrochers, 1997), s'appuyant donc sur des normes culturelles et linguistiques qui ne représentent pas la francophonie québécoise (Bouchard et al., 2009; St-Pierre et al., 2000). La plupart ne répondent pas aux standards psychométriques; les critères de validité et de fidélité étant défaillants, voire même inexistant (Bouchard et al., 2009). La majorité de ces outils renseignent peu sur la nature des difficultés rencontrées en compréhension de lecture puisqu'ils ne s'appuient pas sur des modèles cognitifs. La pénurie d'outils adaptés aux enfants québécois ainsi que les pauvres qualités psychométriques et le peu d'assises théoriques de ces outils soulignent le besoin criant de mettre sur pied un outil d'évaluation standardisé et adapté à la clientèle pédiatrique franco-québécoise. C'est pourquoi le test *Une bouteille à la mer* a été développé. Ce dernier s'ajoutera au *Test d'évaluation du langage écrit québécois* (TELEQ), une batterie d'évaluation du langage écrit développé au *Laboratoire d'études en neuropsychologie de l'enfant et de l'adolescent* (LÉNEA).

Méthodologie

Développement du test *Une bouteille à la mer*

Le test *Une bouteille à la mer* a été développé en collaboration avec l'équipe des conseillères pédagogiques en orthopédagogie des commissions scolaires (CS) des régions administratives de Laval-Laurentides-Lanaudière (équipe LLL), composée de Mesdames Justine Bourgeois (CS des Samares : CSDS), Marie-Ève Brissette (CS de la Rivière-du-Nord : CSRDN), Isabelle Brunet (CSRDN), Audray Grand-Maison (CS de Laval : CSDL), Sylvie Nardone (CSDL), Julie Brunelle (CSDL), Marie-Josée Mongeau (CS des Affluents : CSDA), Geneviève Turbide (CS des Laurentides : CSL), Vanessa Savaria (CS de la Seigneurie-des-Mille-Îles : CSSMI) et Judy-Ann Valiquette (CS Pierre-Neveu : CSPN). L'autrice jeunesse et orthopédagogue Marie-France Bouchard a également participé à l'élaboration du test en composant le texte sur lequel il repose. Pour ce faire, elle devait respecter certaines directives prenant assises sur les échelles de niveaux de compétences établis par le *Ministère de l'Éducation, du Loisir et du Sport* (MELS, 2009b) : le texte devait être de type narratif (récit), puisque les enfants d'âge primaire sont régulièrement exposés à de telles structures de texte. La longueur de celui-ci ne devait pas dépasser les 600 mots et les phrases composants la surface de texte devaient être majoritairement des phrases simples (2 à 3 propositions), bien que des phrases inversées et négatives devaient y être incluses. Des anaphores (pronoms) et des marqueurs de relations (et, ou, puis, ensuite, d'abord, enfin, après, premièrement, finalement, hier, demain, le lendemain, lorsque, quand, avant de, pendant que, au cours de) devaient être intégrés par l'autrice. Pour ce qui est du contenu du texte, celui-ci devait être adapté aux élèves de la 4^e à la 6^e année dans la mesure où il devait susciter leur intérêt tout en demeurant accessible (ni trop facile, ni trop difficile). Une brève description physique et psychologique des

protagonistes était exigée et le prénom de l'un d'eux devait être ambigu, pouvant être associé au sexe féminin ou masculin.

Suite à la rédaction d'une première version du texte, l'équipe LLL a travaillé à l'élaboration de deux épreuves relatives à celui-ci. La première épreuve en était une de rappel où l'enfant était amené à rapporter l'histoire dans ses propres mots sans avoir recours au texte. Afin d'évaluer la qualité du rappel, les conseillères pédagogiques avaient mis sur pied une grille de consignation dans laquelle le texte était divisé en propositions, regroupées selon la structure de récit (situation de départ, élément déclencheur, déroulement, péripéties, dénouement et situation finale). Chaque proposition était numérotée de 1 à 3 selon leur degré d'importance (1 étant les propositions de première importance et 3 les propositions superflues, non nécessaires à la compréhension du texte). Au cours de l'épreuve, l'examineur identifiait les propositions rappelées par l'enfant lors de l'administration de cette épreuve, évaluant ainsi l'intégrité de la macrostructure. La deuxième épreuve, quant à elle, en était une de questions. Dix-sept questions avaient été formulées, chacune évaluant de façon prépondérante un processus issu du modèle de van Dijk et Kintsch (1983), tel que décrit dans le cadre général d'apprentissage de la lecture-écriture de Laplante (2011). Selon la réponse de l'enfant, celui-ci obtenait 0 ou 1 point, pour un maximum de 17 points. Le solutionnaire se trouvait à même le protocole de passation. Enfin, un guide de passation avait été créé explicitant, pour chacune des épreuves, les consignes, les modalités de passation et les processus évalués.

Cette version pilote du test fut ensuite distribuée par les conseillères pédagogiques dans leurs réseaux scolaires respectifs. Plus d'une centaine d'enfants de la 4^e à la 6^e année ont répondu au test administré par des orthopédagogues formés à cette fin. Suite à cette diffusion, les rappels et réponses des élèves aux questions ont été compilés et analysés qualitativement et des

modifications furent apportées au texte et aux épreuves y étant rattachées. Le texte fut simplifié (abandon du récit à la première personne, clarification du genre associé au prénom Gab et diminution des dialogues) et un préambule, de même qu'un personnage (le chien Fripouille) furent ajoutés. De plus, étant donné les difficultés rencontrées par les orthopédagogues lors du rappel en raison de l'exhaustivité de la grille de consignation, celle-ci fut abandonnée et une nouvelle grille fut créée. Pour ce faire, les rappels des enfants furent analysés qualitativement et seuls les éléments centraux de l'histoire furent retenus afin de faire partie intégrante de la grille de consignation. Certains enfants commettaient des erreurs en rappelant l'histoire ; erreurs qui pouvaient être mineures – n'altérant pas le sens de l'histoire – ou majeures – altérant le sens de l'histoire. Un tableau a donc été mis à la disposition des examinateurs afin que ceux-ci puissent compiler ces erreurs et ainsi nuancer leurs propos lorsque viendra le temps de rapporter les résultats. Pour ce qui est des questions, certaines d'entre elles laissaient place à des réponses qui s'éloignaient du processus visé, ce pour quoi plusieurs furent reformulées. Par ailleurs, suite à la simplification du texte, des questions furent supprimées et d'autres créées, portant le nombre total de questions à quatorze. Un tableau regroupant l'ensemble de celles-ci a été créé afin d'indiquer quel processus du modèle de van Dijk et Kintsch (1983) est évalué par chaque question (voir Annexe 1). Ce tableau a été révisé en entier par une collaboratrice experte, soit une étudiante au doctorat en éducation. La Dre Line Laplante a également été consultée lorsqu'un soutien était de mise en lien avec des questions plus complexes. Finalement, le protocole ainsi que le manuel technique furent revus afin de tenir compte de l'ensemble des modifications apportées.

Participants (critères d'éligibilité et d'exclusion)

Afin d'être éligibles à la présente étude, les enfants devaient fréquenter un établissement francophone au moment de l'étude et avoir le français comme langue maternelle. Les enfants ayant grandi dans un environnement bilingue étaient éligibles, pourvu que le français ait été l'une des langues maternelles. Les enfants ayant subi un échec en français (c.-à-d. obtenus en deçà de 60% en français, toutes compétences confondues : lire des textes variés, écrire des textes variés, communiquer oralement), ayant redoublé une année scolaire, ou encore, ayant reçu ou recevant des interventions intensives en orthophonie ou en orthopédagogie en lien avec le langage écrit étaient exclus de l'étude. Il en va de même pour les enfants ayant reçu un diagnostic de trouble neurodéveloppemental, neurologique ou de santé mentale. Les enfants ayant obtenu des scores situés à plus de deux écarts-types sous la moyenne pour leur âge aux tests mesurant les habiletés langagières et le potentiel intellectuel (QI), ont également été exclus de l'échantillon. Les tests utilisés afin de mesurer les habiletés langagières sont l'*Alouette-R* (Lefavrais, 2005), évaluant la fluidité de lecture, soit la précision et la vitesse de lecture, et le sous-test *Compréhension de paragraphe à l'oral* de la batterie : Évaluation Clinique des notions langagières fondamentales, version pour francophone du Canada (CELF CDN-F ; Wiig et al., 2009), évaluant les processus non spécifiques de compréhension langagière (voir section **Outils d'évaluation**). Ces exclusions ont été réalisées dans le but de s'assurer que les enfants n'aient pas de difficultés majeures au niveau du décodage et de la compréhension du langage oral. Quant au QI, celui-ci est estimé via trois sous-tests : *Similitudes*, *Matrices* et *Séquences de Chiffres* de l'Échelle d'intelligence de Wechsler pour enfants, 5^e édition, version pour francophones du Canada (WISC-V ; Wechsler, 2016) (voir section **Outils d'évaluation**).

Procédure

Le projet TELEQ a été approuvé le 29 juin 2015 par le Comité d'éthique et de la recherche en arts et sciences (CÉRAS) (n° de certificat CERAS-2015-16-080-P). Un amendement au projet fût demandé afin de tenir compte de l'ajout du test *Une bouteille à la mer* au TELEQ. Une réémission du certificat d'éthique fut obtenue le 2 avril 2019 par le Comité d'éthique de la recherche en éducation et en psychologie (CÉREP).

Les participants ont été recrutés dans des écoles de la CSDL et de la CSPN sur une base volontaire. Un formulaire de sollicitation était distribué dans les classes de 4^e, 5^e et 6^e année. Les parents intéressés par l'étude devaient remplir le formulaire en y inscrivant leur numéro de téléphone et leur courriel. Les formulaires complétés étaient ensuite acheminés à la chercheuse qui se chargeait de contacter ces derniers par téléphone. Les critères d'inclusion et d'exclusion étaient vérifiés et une première rencontre était fixée pour les enfants éligibles. Un courriel était ensuite envoyé leur rappelant la date de la rencontre et comprenant le formulaire de consentement. La première rencontre durait approximativement deux heures trente et se déroulait à l'Université de Montréal (Campus Laval). Au début de la rencontre, le formulaire de consentement était revu avec les parents et expliqué à l'enfant. Ceux-ci devaient le signer, sans quoi la rencontre ne pouvait avoir lieu. La première rencontre consistait en une évaluation sommaire du QI ainsi qu'en une évaluation des habiletés langagières des enfants. Le test *Une bouteille à la mer* était également administré, de même que le sous-test *Compréhension de lecture* du Test individuel de rendement de Wechsler, seconde édition, version québécoise (WIAT-II ; Wechsler, 2008) (voir section **Outils d'évaluation**). Ceux-ci ont été administrés afin de comparer leurs résultats entre eux et ainsi mesurer la validité concordante. Parallèlement à cette évaluation, les parents étaient invités à remplir le questionnaire sociodémographique, partiellement complété lors de l'entretien téléphonique. À la fin de cette rencontre, une seconde

évaluation était fixée de deux à six semaines plus tard. Celle-ci ne durait que 30 minutes puisque seul le test *Une bouteille à la mer* était administré, permettant ainsi de s'assurer de la fidélité du test à travers le temps (fidélité test-retest). Cette seconde rencontre pouvait se dérouler à l'école de l'enfant auprès d'un orthopédagogue formé à cette fin, ou à l'Université de Montréal (Campus Laval) auprès d'un membre de l'équipe de recherche. Les enfants ayant participé à l'étude recevaient 10\$ suite à la première ou deuxième rencontre. Les parents étaient également dédommés pour leurs frais de déplacement ou de stationnement, jusqu'à concurrence de 20\$. En ce qui a trait aux participants ayant été recrutés à la CSPN, la procédure fut quelque peu différente puisque la première rencontre prenait place dans des locaux prêtés par la CS et la deuxième rencontre n'avait pas lieu. La compensation financière était donc remise directement après la première rencontre avec l'enfant.

Outils d'évaluation

Description du test *Une bouteille à la mer*. Le test *Une bouteille à la mer* se compose de trois épreuves : lecture, rappel et questions. Avant même d'entamer la tâche, trois enveloppes sont présentées à l'enfant. Celui-ci doit en choisir une. Chacune de ces enveloppes contient le même texte, mais les enfants ne sont pas au courant de ce subterfuge. Cette façon de faire reprend l'idée de Maeder et Charlois (2010), autrices ayant développé le test de compréhension de lecture *La forme noire*. De ce fait, les enfants pensent que l'examineur ne sait pas quel texte sera lu et auront tendance à rapporter plus d'éléments afin de faciliter la compréhension de l'histoire par l'examineur. Une fois l'enveloppe sélectionnée, l'enfant retire le texte qui s'y trouve et doit le lire. Il peut le faire dans la modalité de son choix, c'est-à-dire silencieusement, à voix haute ou de manière subvocalisée. Peu importe la modalité de lecture choisie, le temps de lecture est mesuré. Subséquemment, le texte est retiré de la vue de l'enfant et celui-ci doit

rappeler l'histoire en donnant le plus grand nombre de détails possibles, permettant ainsi d'évaluer la macrostructure. Le rappel de l'enfant est corrigé à l'aide de la grille prévue à cet effet. Le score maximal pouvant être obtenu est de 19. Le texte est ensuite remis à l'enfant sans que celui-ci ne soit encouragé à l'utiliser. S'ensuit la période de questions, durant laquelle quatorze questions sont posées oralement à l'enfant. Chaque question évalue de manière prépondérante un processus du modèle de van Dijk et Kintsch (1983) : analyse syntaxique (1 question), anaphores (5 questions), marqueurs de relation (3 questions), inférences (6 questions) et élaboration du modèle de situation (2 questions). Cela dit, bien que les questions ciblent un processus spécifique, les processus issus dudit modèle sont interdépendants et reposent l'un sur l'autre. À titre d'exemple, la macrostructure prend assise sur la microstructure qui nécessite une analyse syntaxique de prime abord. Ce faisant, en évaluant la macrostructure, on évalue indirectement les éléments de la microstructure et de l'analyse syntaxique. Pour chacune des questions, l'examineur est amené à spécifier si l'enfant a répondu de mémoire ou si celui-ci s'est plutôt référé au texte afin d'y répondre. La dernière question diffère des autres puisque l'enfant est amené à recréer l'histoire de manière chronologique via l'utilisation de languettes de papier sur lesquelles sont inscrits les éléments principaux de l'histoire. Un maximum de 2 points est alloué par question, portant le total de cette épreuve à 28. La correction des questions se fait via l'utilisation du solutionnaire se trouvant dans le guide de passation. Un score total de compréhension (/47) est obtenu en faisant la somme des scores du rappel (/19) et des questions (/28). Comme *Une bouteille à la mer* évalue l'ensemble des processus du modèle de van Dijk et Kintsch (1983), des hypothèses peuvent être émises quant aux processus potentiellement déficitaires et pour lesquels une prise en charge serait nécessaire.

Outils utilisés afin de s'assurer de l'éligibilité des participants.

Échelle d'intelligence de Wechsler pour enfants, 5e édition, version pour francophones du Canada (WISC-V ; Wechsler, 2016). La WISC-V est utilisée dans le cadre de cet essai afin d'obtenir un estimé du QI et de la mémoire de travail verbale des participants. Les sous-tests *Similitudes*, *Matrices* et *Séquences de chiffres* ont été administrés puisque ceux-ci sont fortement corrélés au QI global ($r = ,73$; $r = ,69$ et $r = ,69$ respectivement) (Wechsler, 2016). Le sous-test *Similitudes* évalue les capacités d'abstraction de l'enfant en lui demandant de créer des liens sémantiques entre deux concepts verbaux, alors que le sous-test *Matrices* évalue plutôt les habiletés de raisonnement de l'enfant en l'amenant à choisir, parmi une série de figures, celle complétant une suite logique. Quant au sous-test *Séquence de chiffres*, il permet d'évaluer la mémoire de travail, soit l'habileté à maintenir et à manipuler de l'information sur un court laps de temps, en demandant à l'enfant de rapporter en ordre direct, inverse et croissant une série de chiffres énoncés oralement.

L'Alouette, version révisée (Alouette-R ; Lefavrais, 2005). Il s'agit d'un test de lecture utilisé afin d'évaluer la fluidité de lecture des participants. Ce test est fréquemment utilisé (Écalle, 2010) en recherche afin d'évaluer le niveau de lecture des enfants, ainsi qu'en milieu clinique afin de statuer sur la présence ou l'absence d'une dyslexie (Bertrand et al., 2010). Lors de ce test, l'enfant dispose de trois minutes pour lire à voix haute un texte de 265 mots dénué de sens (Lefavrais, 2005). Le temps de lecture est mesuré, le nombre de mots lus et le nombre d'erreurs sont dénombrés, permettant ainsi de calculer le nombre de mots lus correctement et d'aboutir à un score de vitesse et un score de précision de lecture.

Évaluation clinique des notions langagières fondamentales, version pour francophones du Canada (CELF CDN-F ; Wiig et al., 2009). Le sous-test *Compréhension de paragraphes à l'oral* de la CELF a été administré, puisque celui-ci permet d'évaluer la

compréhension orale. Lors de ce sous-test, trois paragraphes sont lus oralement à l'enfant. Des questions relatives à ces paragraphes lui sont ensuite posées afin d'évaluer sa compréhension.

Tests utilisés afin de vérifier la fidélité et validité du test *Une bouteille à la mer*.

Le Test individuel de rendement de Wechsler, seconde édition, version québécoise (WIAT-II ; Wechsler, 2008). Le test *Compréhension de lecture* a été administré afin de mesurer la validité concordante avec le test *Une bouteille à la mer*. Bien que ce sous-test présente des limites, il fut choisi afin d'établir la validité concordante puisqu'il est fréquemment utilisé par les praticiens du Québec afin d'évaluer la compréhension de lecture. Tel que mentionné, lors de ce sous-test l'enfant est amené à lire des phrases et de courts paragraphes pour ensuite répondre à des questions les concernant.

Analyses statistiques prévues et prédictions

Fidélité test-retest. Ce type de fidélité consiste à comparer les résultats d'un même test lors de temps de mesure distincts; les résultats devant être similaires d'une passation à l'autre. Si tel est le cas, il est alors possible de conclure que les résultats obtenus ne sont pas le fruit du hasard, mais qu'ils représentent de manière fidèle la performance de l'individu (Bouchard et al., 2009). La fidélité test-retest sera mesurée en comparant les résultats obtenus au test *Une bouteille à la mer* lors de la première et de la deuxième passation. Les corrélations de Pearson obtenues devraient être fortes et significatives (Bouchard et al., 2009). Plus spécifiquement, McCauley et Swisher (1984) avance qu'un coefficient de corrélation minimal de ,90 devrait être obtenu afin que le test soit considéré comme fidèle. Cela dit, d'autres balises moins conservatrices existent et avancent plutôt que le seuil minimal devant être atteint est de ,70 (Souza et al., 2017). Il est d'usage qu'un coefficient de ,70 soit considéré comme acceptable, un coefficient de ,80 comme étant satisfaisant et un coefficient de ,90 comme étant excellent (Le

Corff et Yergeau, 2016). Le coefficient obtenu en matière de fidélité test-retest devrait donc atteindre le seuil minimal de ,70, bien qu'un coefficient proche de ,90 soit espéré. Par ailleurs, comme les participants se sont créés un modèle de situation en lisant le texte *Une bouteille à la mer* lors de la première lecture du texte, ce modèle sera réactivé en mémoire à long terme lors de la seconde lecture du texte, permettant ainsi à l'enfant de l'étoffer davantage. Ce faisant, un effet d'apprentissage est attendu entre les deux passations.

Fidélité inter-juge. Cette fidélité renvoie à l'idée que deux évaluateurs (juges) indépendants devraient quantifier la performance d'un même individu de la même façon. Ce faisant, le score obtenu devrait être représentatif de la performance de l'évalué et non le reflet des impressions des juges (Bouchard et al., 2009). Ce type de fidélité sera mesuré en comparant les résultats au test *Une bouteille à la mer* évalué par deux juges distincts pour les mêmes participants lors de la première passation. Des coefficients de corrélations intra-classes (CCI) fortes et significatives devraient être retrouvées (Bouchard et al., 2009). En fait, des corrélations surpassant ,75 sont attendues, tel que le suggère les critères émis par Koo et son équipe (2016), la fidélité étant considérée comme pauvre en deçà de ,50, modérée entre ,50 et ,75, bonne entre ,75 et ,90 et excellente au-delà de ,90.

Cohérence interne (ou homogénéité). Celle-ci nous renseigne sur le degré de cohésion des items formant un test et repose sur la prémisse que les items censés mesurer un même construit devraient mener à des résultats similaires (Bouchard et al., 2009). Afin de mesurer la cohérence interne, l'alpha de Cronbach sera calculé à partir des items (c.-à-d. les 14 questions) pour l'ensemble de l'échantillon ainsi que par niveau scolaire. Pour que le test soit considéré comme ayant une bonne cohérence interne, l'alpha obtenu doit être de plus de ,70 (Nunnally 1978, 1994).

Degré de difficulté des items. Afin d'avoir une idée du degré de difficulté des items, la moyenne des scores pour chacune des questions a été calculée, et ce, en tenant compte du niveau scolaire, permettant ainsi de voir la tendance des performances aux items. Comme le test a été conçu de manière à s'apparenter aux épreuves de compréhension de lecture effectuées dans le contexte de la classe, les questions devraient varier entre elles quant à leur degré de difficulté.

Validité concordante. Cette validité renseigne sur le degré de correspondance entre un test en cours de développement et un autre test préalablement validé et évaluant le même construit (Bouchard et al., 2009). Afin d'évaluer cette validité, les résultats au test *Une bouteille à la mer* seront comparés aux résultats obtenus au sous-test *Compréhension de lecture* de la WIAT-II. Bien que ces tests évaluent les habiletés de compréhension de lecture chez les enfants, la nature et la forme de ceux-ci diffèrent, si bien que l'on s'attend à des corrélations de Pearson significatives et positives, mais de force moyenne. Afin d'être considéré de force moyenne, le coefficient de corrélation doit se situer autour de ,30, selon les critères émis par Cohen en 1988. En revanche, un coefficient est considéré comme faible lorsqu'il se situe autour de ,10 et fort lorsqu'il se situe autour de ,50 (Cohen, 1988). Ces critères s'appliquent à tous les indices de validité.

Validité critériée concomitante. De manière générale, la validité critériée consiste à définir le degré d'adéquation entre les résultats obtenus à un test et un critère externe étant lié au construit d'intérêt. La validité critériée qualifiée de concomitante consiste à déterminer l'accord entre un résultat à un test et un critère préalablement déterminé (Stephenson et Parent, 2012). Dans la présente étude, deux critères ont été sélectionnés, soient le niveau scolaire (4^e, 5^e ou 6^e année du primaire) et la note en français (compétence *Lire des textes variés*). Ainsi, on prévoit que le score global du test *Une bouteille à la mer* augmente en fonction du niveau

scolaire. Des analyses de variance (ANOVA simple) seront effectuées ainsi que des tests de Tukey afin de vérifier cette prédiction. Dans un deuxième temps, il est attendu que les résultats au test *Une bouteille à la mer* corréleront avec les notes en français (compétence « *Lire des textes variés* ») obtenues dans le dernier bulletin du participant. Des corrélations de Spearman seront effectuées et devraient s'avérer significatives, positives et de force moyenne. Cette analyse non paramétrique a été préférée à l'analyse corrélationnelle de Pearson puisque les relations entre les variables n'étaient pas toujours linéaires, mais étaient monotones, répondant ainsi au postulat de la corrélation de Spearman.

Résultats

Les analyses statistiques ont été conduites avec le logiciel SPSS (versions 25 et 26).

Description de l'échantillon final

Sur les 72 participants, seulement 58 participants constituent l'échantillon final puisque 14 participants répondaient aux critères d'exclusion : deux participants se sont avérés être en situation d'échec en français; un participant a été suivi de façon intensive en orthopédagogie et découvert suite à sa participation à l'étude et 11 participants ont obtenu un score se situant à plus de deux écarts-types sous la moyenne pour leur âge à l'*Alouette-R* ($n = 8$), au sous-test *Compréhension de paragraphe à l'oral* ($n = 2$) et *Séquence de chiffres* ($n = 1$). L'échantillon final se compose donc de 58 participants âgés entre 105 et 150 mois ($M = 128,74$, $É.T. = 11,20$). La majorité de ceux-ci ont comme langue maternelle le français ($n = 47$), bien que certains ont évolué dans un environnement bilingue où, en plus du français, l'arabe était parlé ($n = 5$), le chinois ($n = 1$), l'espagnol ($n = 1$), le libanais ($n = 1$), le russe ($n = 2$), le perse ($n = 1$) et le créole ($n = 1$). En se référant au Tableau 1, présentant les caractéristiques sociodémographiques de l'échantillon final, il est possible de constater que les participants se répartissent assez

équitablement entre la 4^e, 5^e et 6^e année ; qu'il se compose principalement de filles; que la plupart des participants ont été recrutés au trimestre d'hiver (sauf pour les 5^e année) ; que ceux-ci proviennent majoritairement de la région de Laval; qu'en majorité leurs mères ont complété des études universitaires et que les revenus familiaux annuels de près moitié des participants surpassent 100 000\$. Il s'agit également d'enfants réussissant majoritairement bien en français.

Tableau 1

Caractéristiques sociodémographiques de l'échantillon total ainsi que par niveaux scolaires

Variables	Total (n = 58)	4^e année (n = 24)	5^e année (n = 19)	6^e année (n = 15)
Genre				
Filles (%)	39 (67,24)	16 (66,67)	13 (68,42)	10 (66,67)
Garçons (%)	19 (32,76)	8 (33,33)	6 (31,8)	5 (33,33)
Trimestre				
Automne (%)	11(18,97)	4 (16,67)	6 (31,58)	1 (6,67)
Hiver (%)	32 (55,17)	16 (66,67)	6 (31,58)	10 (66,67)
Printemps (%)	15 (25,86)	4 (16,67)	7 (36,84)	4 (26,67)
Régions				
Laval (%)	40 (68,97)	14 (58,33)	13 (68,42)	13 (86,67)
Laurentides (%)	18 (31,03)	10 (41,67)	6 (31,58)	2 (13,33)
Revenu familial				
Moins de 100 000\$ (%)	28 (49,12)	14 (58,33)	7 (38,89)	7 (4,67)
100 000\$ et plus (%)	29 (50,89)	10 (41,67)	11 (61,11)	8 (53,33)
Niveau d'éducation de la mère				
Aucune étude universitaire	19 (33,93)	9 (37,50)	10 (35,71)	2 (14,29)
Études universitaires et plus	37 (66,07)	15 (62,50)	18 (64,29)	12 (85,71)
Niveau en français ^a				
Forts (au-delà de 80%)	39 (67,24)	14 (58,33)	11 (61,11)	14 (93,33)
Moyens-Faibles (en deçà de 80%)	18 (31,58)	10 (41,67)	7 (38,89)	1 (6,67)

^a Notes : Une donnée manquante pour un élève de 5^e année.

Quant aux Tableau 2, celui-ci rassemble toutes les informations relatives aux statistiques descriptives de l'échantillon final.

Tableau 2

Statistiques descriptives de l'échantillon total ainsi que par niveau scolaire

Variables	Total			4 ^e année			5 ^e année			6 ^e année		
	<i>M</i> (É.T.)	Min	Max	<i>M</i> (É.T.)	Min	Max	<i>M</i> (É.T.)	Min	Max	<i>M</i> (É.T.)	Min	Max
Âge (mois)	128,74 (11,20)	105	150	118,29 (4,75)	105	126	129,68 (3,73)	121	134	144,27 (3,01)	140	150
Scores z												
AL - Précision	-0,25 (0,64)	-1,90	0,84	-0,31 (0,64)	-1,89	0,84	-0,07 (0,57)	-1,39	0,79	-0,36 (0,70)	-1,52	0,74
AL - Vitesse	0,23 (0,73)	-1,06	2,59	0,21 (0,77)	-0,88	2,59	0,25 (0,78)	-1,06	1,66	0,22 (0,64)	-0,71	1,37
Scores standardisés												
CELF – CPO	11,02 (2,42)	6	15	10,26 (2,03)	6	15	10,95 (2,37)	7	14	12,27 (2,69)	7	14
WISC-5 – SI	11,36 (2,20)	7	17	10,83 (2,16)	7	17	11,47 (2,12)	8	16	12,07 (2,82)	8	17
WISC-5 – MA	11,16 (3,32)	5	18	10,67 (3,70)	5	18	10,37 (2,65)	6	16	12,93 (2,94)	6	16
WISC-5 – SC	11,22 (2,41)	5	17	10,75 (2,49)	5	15	10,89 (2,05)	7	15	12,40 (2,44)	7	17

Notes. AL = Alouette-R; CELF = Évaluation clinique des notions langagières fondamentales; CPO = Compréhension de paragraphe à l'oral; WISC-5 = Échelle d'intelligence de Wechsler pour enfants 5e édition version pour francophones du Canada; SI = Similitudes; MA = Matrices; SC = Séquences de chiffres.

Analyses préliminaires

Avant d'entamer les analyses de pré-validation, l'association entre diverses variables sociodémographiques et le score global de compréhension de lecture a été mesurée. L'objectif

étant de vérifier si ces variables ont un impact sur le score global. Pour ce faire, les scores globaux du test ont été convertis en score z, calculés à partir de la moyenne et de l'écart-type des participants selon leur niveau scolaire. Par la suite, des analyses de variances (ANOVA simple) ont été menées. Ces analyses indiquent que les scores globaux des participants ne diffèrent pas selon leur genre (fille ($M = 0,11$, $\acute{E}.-T. = 0,95$) ou garçon ($M = -0,21$, $\acute{E}.-T. = 1,04$) : $F(1,56) = 1,40$, $p = ,241$), leur trimestre scolaire (automne ($M = -0,16$, $\acute{E}.-T. = 0,96$), hiver ($M = -0,03$, $\acute{E}.-T. = 0,98$), printemps ($M = 0,19$, $\acute{E}.-T. = 1,04$) : $F(2,55) = 0,43$, $p = ,655$), leur région d'appartenance (Laval ($M = -0,09$, $\acute{E}.-T. = 1,01$) ou Laurentides ($M = 0,21$, $\acute{E}.-T. = 0,91$) : $F(1,56) = 1,15$, $p = ,287$), leur revenu familial (moins de 100 000\$ ($M = -0,12$, $\acute{E}.-T. = 1,03$) ou plus de 100 000\$ par année ($M = 0,17$, $\acute{E}.-T. = 0,89$) : $F(1,55) = 1,31$, $p = ,257$) ou le niveau de scolarité de leur mère (pas d'études universitaires ($M = -0,12$, $\acute{E}.-T. = 0,90$) ou études universitaires ($M = 0,07$, $\acute{E}.-T. = 1,04$) : $F(1,54) = 0,45$, $p = ,504$). Ceci dit, une différence significative subsiste entre les participants ayant obtenus 80% ou plus en français ($M = 0,25$, $\acute{E}.-T. = 0,89$) et ceux ayant obtenus en deçà de 80% ($M = -0,44$, $\acute{E}.-T. = 0,98$), $F(1,55) = 6,87$, $p = ,01$. Seule la note en français est associée au score global au test *Une bouteille à la mer*.

Analyses de pré-validation

Fidélité test-retest. Les analyses corrélationnelles menées afin de statuer sur la fidélité test-retest du test ont été effectuées auprès de 39 participants. Cette diminution de l'échantillon s'explique par le fait que les participants en provenance des Laurentides n'ont pas participé à la seconde rencontre d'évaluation, tout comme trois autres participants non joignables. Les scores aux épreuves de rappel, de questions ainsi que les scores globaux de compréhension de lecture des deux temps de mesure ont été corrélés. Ces corrélations (de Pearson) se sont avérées fortes (Cohen, 1988), positives et significatives, tel qu'attendu (score global : $r = ,86$, $p < ,001$; rappel :

$r = ,74, p < ,001$ et questions : $r = ,83, p < ,001$). Le coefficient associé au rappel est acceptable, alors que les coefficients associés aux scores globaux et aux questions sont satisfaisants (Le Corff et Yergeau, 2016; Souza et al., 2017).

Des test t appariés ont par la suite été effectués sur l'échantillon total afin de déterminer si les participants s'amélioraient ou non entre les deux temps de mesure. Les résultats révèlent la présence de différences significatives, que ce soit pour le score global de compréhension de lecture (test : $M = 26,63, \acute{E}.-T. = 8,32$; retest : $M = 31,91, \acute{E}.-T. = 7,31 : t(38) = 7,77, p < ,001, d = 0,77$), le score au rappel (test : $M = 9,96, \acute{E}.-T. = 3,23$; retest : $M = 12,47, \acute{E}.-T. = 2,89 : t(38) = 7, p < ,001, d = 0,47$) et le score aux questions (test : $M = 16,67, \acute{E}.-T. = 5,89$; retest : $M = 19,44, \acute{E}.-T. = 4,77 : t(38) = 5,29, p < ,001, d = 0,64$). En moyenne, les participants obtiennent 5,28 points de plus ($\acute{E}.-T. = 4,25$) au score global de compréhension, soit 2,51 ($\acute{E}.-T. = 2,24$) points de plus au score de rappel et 2,77 ($\acute{E}.-T. = 3,27$) points de plus au score de questions, lorsque réévalués de deux à quatre semaines plus tard. Les d de Cohen ont été calculés en utilisant les écarts-types du premier temps de mesure. Le d de Cohen pour le score global révèle la présence d'un large effet, alors que l'effet est moyen pour les scores du rappel et des questions.

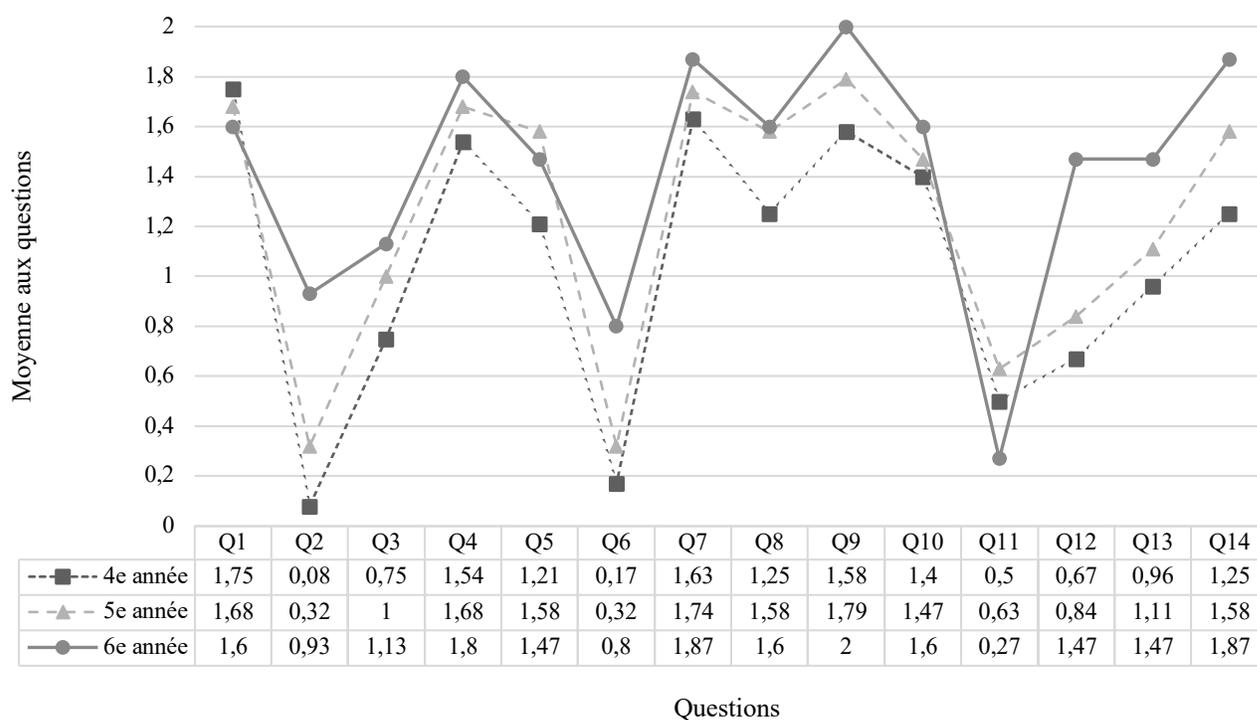
Fidélité inter-juge. Afin de savoir si le test *Une bouteille à la mer* possède une bonne fidélité inter-juges ou non, des corrélations intra-classe ont été effectuées suivant un modèle aléatoire à deux facteurs, pour des mesures uniques dont l'intérêt est l'hétérogénéité de ces mesures. Les scores globaux, les scores au rappel ainsi que les scores aux questions, tels qu'évalués par deux juges distincts, ont donc été comparés. Les CCI qui en résultent sont forts et positifs, que ce soit pour les scores globaux ($,98$; intervalle de confiance (IC) 95% [$,97 - ,99$]), les scores de rappel ($,96$; IC 95% [$,93 - ,98$]) et les scores des questions ($,98$; IC 95% [$,97 - ,99$]), témoignant d'une excellente fidélité inter-juge (Koo et Li, 2016).

Cohérence interne. L'alpha de Cronbach mesuré en considérant l'ensemble de l'échantillon est de ,74, atteignant ainsi le critère de Nunnally (1978, 1994) fixé à ,70.

Degré de difficulté des questions. Le Graphique 1 regroupe l'ensemble des moyennes des questions par niveau scolaire. Tel qu'il est possible de le constater, le degré de difficultés des questions varie : certaines questions étant plus faciles, d'autres plus difficiles. Cette variabilité est présente pour tout niveau scolaire confondu.

Graphique 1

Moyennes des scores aux questions en fonction du niveau scolaire (4^e, 5^e et 6^e année)



Validité concordante. En comparant le score global brut de compréhension de lecture du test *Une bouteille à la mer* avec le score brut obtenu au sous-test de *Compréhension de lecture* de la WIAT-II, une corrélation moyenne, positive et significative est obtenue pour l'ensemble de l'échantillon ($r = ,37, p = ,004$), tel que prévu.

Validité critériée : niveau scolaire. Le Tableau 3 répertorie les performances des participants au test *Une bouteille à la mer* pour l'échantillon entier ainsi que par niveau scolaire. En s'y référant, il est possible de constater que les moyennes augmentent d'un niveau scolaire à l'autre, et ce, peu importe le score d'intérêt. Des analyses de variance (ANOVA simple) ont été menées afin de déterminer si les différences de moyennes observées sont statistiquement significatives.

En ce qui a trait au score global de compréhension de lecture, le seuil de significativité est atteint ($F(2,55) = 4,28, p = ,019$) indiquant la présence d'une ou de plusieurs différences entre les groupes. La taille d'effet (R^2) est de 0,1346, l'effet explique donc 13,46% de la variance totale. Il s'agit d'une grande taille d'effet selon Cohen (1988). Des tests de Tukey ont par la suite été effectués et révèlent la présence d'une différence significative entre les élèves de 4^e et de 6^e année ($p = ,014$) et non entre les élèves de 4^e et 5^e année et les élèves de 5^e et 6^e année. Un contraste polynomial orthogonal montre toutefois que les moyennes suivent une progression linéaire ($p = ,005$).

En ce qui concerne le score du rappel, le seuil de significativité n'est pas atteint ($F(2,55) = 1,81, p = ,173$). Aucune différence significative n'existe donc entre les groupes, et ce, malgré une augmentation des moyennes à travers les années. Cette progression ne peut cependant être qualifiée de linéaire, bien qu'elle s'en approche (contraste polynomial orthogonal : $p = ,063$).

Pour ce qui est du score des questions, le seuil de significativité est atteint ($F(2,55) = 4,83, p = ,012$). L'effet est de grande taille ($R^2 = 0,1494$) (Cohen, 1988) et explique 14,94% de la variance totale. Les tests de Tukey montrent que seuls les élèves de 4^e et 6^e année se distinguent significativement les uns des autres ($p = ,009$). Néanmoins, une progression linéaire

subsiste entre les trois années scolaires telle que le suggère un contraste polynomial orthogonal ($p = ,003$).

Tableau 3

Performances au test Une bouteille à la mer pour l'échantillon total ainsi que par niveau scolaire (scores bruts)

Niveaux scolaires	Total			4 ^e année			5 ^e année			6 ^e année		
	<i>M</i> (É.T.)	Min	Max	<i>M</i> (É.T.)	Min	Max	<i>M</i> (É.T.)	Min	Max	<i>M</i> (É.T.)	Min	Max
Score global	26,98 (7,82)	9,5	39,5	24,04 (7,05)	9,5	38	27,42 (7,53)	15	39	31,13 (7,81)	15,5	39,5
Rappel	10,10 (3,21)	3	17	9,29 (2,75)	4,5	14	10,21 (3,38)	3	15	11,27 (3,5)	5,5	17
Questions	16,90 (5,31)	5	6	14,79 (4,66)	5	24	17,21 (4,98)	8	24	19,78 (5,49)	8	26

Validité critériée : rendement scolaire – notes en français. Afin de déterminer si les notes en français (compétence « *Lire des textes variés* ») corrélaient avec le score global de compréhension de lecture, des corrélations de Spearman ont été effectuées. Lorsque l'ensemble de l'échantillon est considéré, les corrélations sont fortes (Cohen, 1988), positives et significatives pour le score global ($r_s = ,53, p < ,005$) et de questions ($r_s = ,51, p < ,05$) et d'intensité modérée pour le score de rappel ($r_s = ,45, p < ,005$).

Discussion

L'objectif de la présente étude était de développer un test évaluant la compréhension de lecture et de le pré-valider. Le test *Une bouteille à la mer* créé dans le cadre de la présente étude s'adresse aux élèves franco-québécois de la 4^e à la 6^e année du primaire. Ce test s'ajoutera au

TELEQ, une batterie d'évaluation en cours de développement, qui permettra de pallier au manque d'outils cliniques adaptés, validés et dédiés à l'évaluation du langage écrit. Le TELEQ se compose actuellement de deux sous-tests, l'un évaluant les habiletés de décodage et l'autre évaluant les habiletés d'orthographe. Ceux-ci permettent de statuer sur l'intégrité des processus spécifiques d'identification des mots écrits. Le test *Une bouteille à la mer* permettra quant à lui de statuer sur l'intégrité des processus non spécifiques de compréhension langagière (modèle de van Dijk et Kintsch (1978, 1983)). Lors de ce test, l'enfant est amené à lire un texte à voix haute, silencieusement ou de manière subvocalisée selon sa préférence. Il lui est ensuite demandé de rappeler l'histoire dans ses propres mots en évoquant le plus de détails possible. Des questions lui sont par la suite posées oralement. Au total, 58 élèves de la 4^e à la 6^e année ont pris part à l'étude.

Suite au développement dudit test, les qualités psychométriques de ce dernier ont été vérifiées. Pour ce qui est de la fidélité, celle-ci est globalement adéquate que ce soit pour la fidélité test-retest, la fidélité inter-juge et la cohérence interne. En ce qui a trait à la fidélité test-retest, des coefficients de corrélations acceptables (rappel) et satisfaisants (questions, score global) ont été trouvés entre les deux temps de mesure. Ces résultats indiquent que les performances des enfants sont constantes à travers le temps, c'est-à-dire que les enfants qui réussissent bien au test, réussissent tout aussi bien lors du retest et inversement. Or, une amélioration notable est constatée au retest, ce qui indique la présence d'un effet de pratique. Ceci était attendu puisque les enfants, en lisant le texte pour la première fois, se créent une représentation mentale signifiante de l'histoire (modèle de situation) favorisant ainsi son encodage. Lorsqu'ils lisent à nouveau le texte, cette représentation mentale s'active et s'actualise, permettant d'atteindre une compréhension plus approfondie du texte. Bref, le test

Une bouteille à la mer présente une bonne fidélité test-retest, mais un effet de pratique est présent. Il serait donc préférable que ce test, une fois utilisé, ne soit pas réutilisé à court terme. Cela dit, s'il l'est, il importe que les cliniciens tiennent compte de l'effet de pratique dans leur interprétation de la performance au test.

Pour ce qui est de la fidélité inter-juge, de très fortes corrélations sont obtenues lorsque les scores au test *Une bouteille à la mer* (score global, de rappel et de questions) de deux juges sont comparés. La fidélité inter-juge est donc excellente (Koo et Li, 2016), ce qui signifie qu'un même participant est corrigé sensiblement de la même façon, peu importe l'évaluateur qui procède à l'évaluation. Ce faisant, la grille de correction du rappel ainsi que le solutionnaire mis à la disposition des cliniciens afin de corriger les réponses aux questions sont suffisamment clairs et étoffés pour permettre une correction standard entre cliniciens.

En ce qui concerne la cohérence interne, le seuil de ,70 fixé par Nunnally (1978, 1994) est respecté lorsque l'ensemble des participants est considéré. Ces résultats appuient l'idée que le test évalue un seul et même construit, soit celui de la compréhension de lecture.

En s'attardant aux moyennes des scores aux questions, il est possible de constater que les questions 1, 4, 7 et 9 s'avèrent faciles, et ce, peu importe le niveau scolaire considéré, les moyennes étant toutes très élevées. Or, ces questions pourraient s'avérer utiles lors du dépistage des élèves aux prises avec des difficultés de compréhension de lecture, pouvant potentiellement bien discriminer les bons des mauvais compreneurs.

Par ailleurs, lorsque l'on considère les niveaux scolaires indépendamment, des interrogations émergent quant à l'apport de certaines questions. En 4^e année, les moyennes aux questions 2 et 6 sont très basses, indiquant que ces questions sont rarement réussies par les élèves de 4^e année. Comme les élèves recrutés n'ont pas de difficultés sur le plan de la compréhension

de lecture, il est possible de conclure en la présence d'un effet plancher. Ces questions, lorsqu'utilisées en 4^e année, seraient probablement moins efficaces pour discriminer les bons des mauvais compreneurs. Il en va de même pour les élèves de 5^e année pour qui ces questions sont également moins bien réussies. Il faut dire que celles-ci évaluent l'inférence et l'analyse syntaxique (compréhension de l'accord adjectival et de l'accord d'un participe passé avec le verbe « être »), des compétences qui, selon la PDA, sont apprises en 5^e année, mais acquises en 6^e année (MELS, 2009a). De tels résultats ne sont donc pas surprenants.

De surcroît, certaines questions s'avèrent moins bien réussies par les élèves de 6^e année que par les élèves de 5^e année, voire même de 4^e année. C'est le cas notamment des questions 5 et 11 qui visent toutes deux l'évaluation de l'inférence. Cette baisse de performance est étonnante, car, telle que mentionnée, cette compétence devrait être maîtrisée en 6^e année. Ce résultat est d'autant plus étonnant considérant que la majorité des élèves de 6^e année de l'échantillon ont d'excellentes notes en français (> 80% pour 14 des 15 élèves). En se référant aux questions, il est possible d'émettre l'hypothèse que les réponses aux questions semblent peut-être évidentes, et de ce fait, les élèves n'exposent pas l'entièreté des éléments leur ayant permis de faire ladite inférence. Il aurait été sans doute préférable de permettre aux évaluateurs de poser une question supplémentaire de manière à encourager les élèves à étoffer ou préciser leurs réponses. Afin de confirmer ou infirmer cette hypothèse, une analyse qualitative des réponses aurait pu être menée.

Il est également possible de constater que peu de questions sont de difficulté modérée, la majorité d'entre elles ayant été plutôt bien réussies, et ce, surtout en 6^e année. Cela est probablement dû au fait que le test a été administré à des élèves normolecteurs dont les notes en

français sont élevées. La majorité de ceux-ci ayant au-delà de 75% (voir au-dessus de 80% pour les élèves de 6^e année).

À la lumière de ces constats, il est possible d'affirmer que certaines questions pourraient être révisées, voire enlevées dépendamment du niveau scolaire d'intérêt. Or, il serait hâtif de le faire étant donné la petite taille de l'échantillon, le manque de variabilité par rapport aux notes en français ainsi qu'en raison de l'absence d'un groupe clinique.

Pour ce qui est de la validité concordante, les résultats sont tels qu'attendus. En effet, lorsque le score global du test *Une bouteille à la mer* est comparé au résultat du sous-test *Compréhension de lecture* de la WIAT-II, une corrélation moyenne est obtenue, corroborant notre prédiction de départ voulant que les élèves qui réussissent bien à l'un des tests, réussissent tout aussi bien à l'autre test, et inversement. Comme ces tests diffèrent quant à leur forme (texte continu versus courts paragraphes) et leur utilisation (test d'évaluation versus test de rendement) une corrélation de force moyenne était attendue.

Pour ce qui est de la validité critériée concomitante, deux critères avaient préalablement été sélectionnés, soit le niveau scolaire et le rendement scolaire (illustré par la note en français dans la compétence « *Lire des textes variés* »).

Pour ce qui est du niveau scolaire, une tendance développementale est observée. D'une année scolaire à l'autre, les moyennes augmentent, qu'il s'agisse du score global, du score au rappel ou du score aux questions. Une telle progression n'est pas surprenante puisque les élèves, de la 4^e à la 6^e année, développent une foule de compétences liées à la compréhension écrite. Néanmoins, il importe de rappeler que les élèves ont, pour la plupart, de bonnes notes en français, particulièrement les élèves de 6^e année. Cette caractéristique propre à l'échantillon

n'est pas anodine étant donné que la tendance développementale observée pourrait être fortement influencée par cette caractéristique.

Aussi, d'un point de vue statistique, ces améliorations ne sont pas toutes significatives. Les moyennes aux scores globaux et aux scores des questions diffèrent seulement entre la 4^e et la 6^e année. L'absence de différence significative entre la 4^e et la 5^e année, de même qu'entre la 5^e et 6^e année, pourrait s'expliquer par d'importants écarts-types. Néanmoins, l'étendue de ces écarts-types est encourageante puisqu'elle montre la présence de différences de performances à même un niveau scolaire, ce que le test vise à détecter.

Aussi, aucune différence de moyennes n'est constatée pour le score de rappel. Le score de rappel semble donc moins discriminant que le score aux questions et le score global chez des enfants normolecteurs. Toutefois, auprès d'une population clinique, cette épreuve pourrait s'avérer discriminante. Carlisle (1999) a d'ailleurs démontré que les élèves aux prises avec des troubles d'apprentissage performant généralement moins bien aux tâches de rappel. Par ailleurs, la petite taille d'échantillon pourrait expliquer pourquoi certains seuils de significativité ne sont pas atteints. Tel que le mentionne Cohen (1988), afin de détecter la présence d'un effet moyen avec 80% d'exactitude, 52 participants doivent constituer chacun des trois groupes. Or, 24 participants forment le groupe des 4^e année, 19 le groupe des 5^e année et 15 le groupe des 6^e année, ce qui ne serait pas suffisant.

En ce qui a trait au rendement, les coefficients de corrélation dont la force varie entre modérée (score au rappel) et élevée (score global et score aux questions), indiquent que la performance au test *Une bouteille à la mer* s'accroît à mesure que la note en français augmente. Les scores obtenus refléteraient donc bien le rendement des élèves en français. D'ailleurs, dans les analyses préliminaires, il était question d'une différence de moyennes entre les enfants ayant

obtenu plus de 80% en français (compétence « *Lire des textes variés* ») et ceux ayant obtenu moins de 80% en français, et ce, en contrôlant pour le niveau scolaire. Effectivement, les enfants ayant obtenus plus de 80% en français réussissaient en moyenne mieux au test *Une bouteille à la mer* (score global) que les enfants ayant obtenus en deçà de 80% en français.

Limites de l'étude

Malgré les résultats prometteurs obtenus, l'étude présente certaines limites qu'il importe de considérer lors de l'utilisation du test. Certaines de ces limites concernent les caractéristiques propres à l'échantillon, alors que d'autres portent sur le développement du test ainsi que sur le test en soi.

La première limite en lien avec les caractéristiques de l'échantillon concerne le nombre de participants. En effet, le fait d'avoir peu de participants par niveau scolaire diminue la puissance des tests statistiques. Il est donc parfois difficile de trouver un effet même si un effet est réellement présent. Par ailleurs, le manque de participants limite le choix d'analyse statistique pouvant être menée. À titre d'exemple, il aurait été intéressant de faire une analyse factorielle exploratoire afin de rendre compte de l'aspect multidimensionnel de la compréhension de lecture.

Une autre limite concerne la représentativité de l'échantillon. Bien que la majorité des variables sociodémographiques ne semblent pas avoir d'impact sur le score global du test, il n'en demeure pas moins que l'échantillon n'est pas représentatif de la population des enfants francophones québécois de la 4^e à la 6^e année. La majorité des participants sont des filles ayant été recrutées en hiver à Laval, provenant d'un milieu aisé dont les mères ont fait des études universitaires. La plupart de ceux-ci avaient également des notes élevées en français (compétence « *Lire des textes variés* »). Une plus grande variabilité serait à privilégier.

En ce qui a trait au développement du test, celui-ci n'a pas été élaboré en suivant des étapes prédéfinies de conceptualisation, telle que celle suggérée par De Vellis (2016) par exemple. De plus, le test ne se fonde pas sur un modèle psychométrique. Comme celui-ci a été développé de manière à ressembler aux tests de compréhension de lecture auxquels les enfants sont exposés au primaire (test écologique), il est attendu que les questions le composant varient en matière de degré de difficulté. Or, le postulat de tout modèle psychométrique est que les items (c.-à-d. les questions) sont censés avoir le même pouvoir discriminatif. Il n'est donc pas possible de procéder à des analyses psychométriques classiques. Cela dit, les qualités psychométriques du test peuvent tout de même être mesurées, tel que démontré jusqu'à présent.

En ce qui a trait au test en soi, différentes limites sont à considérer. La première renvoie à l'épreuve de rappel. Bien que celle-ci permette d'évaluer efficacement la capacité de l'enfant à se créer une représentation globale de la situation décrite dans un texte (macrostructure), cette épreuve fait également appel à des habiletés langagières orales (Reed et Vaughn, 2012). Les difficultés orales de certains élèves pourraient donc affecter négativement leur score au rappel.

La deuxième limite renvoie plutôt au genre de texte employé. En effet, les textes narratifs (récits) seraient plus facilement compris par les élèves du primaire que les textes explicatifs (informatifs) puisque le récit décliné en 5 temps est le genre de texte auquel les élèves sont le plus souvent exposés (Best et al., 2008). Il importe de nommer cette limite, car on ne peut déduire qu'en réussissant ce test, un élève a de bonnes habiletés à comprendre tout genre de texte.

Enfin, peu de questions sont associées à chacun des processus issus du modèle de van Dijk et Kintsch (1983). Le test permet donc d'avoir une idée globale de la compétence qu'est la

compréhension écrite, mais seules des hypothèses peuvent être émises quant aux processus lacunaires.

Piste de recherche future

Afin d'assurer la pérennité du test, l'étude de validation doit se poursuivre auprès d'un échantillon plus grand dont la représentativité (surtout en matière de notes en français) est meilleure de manière à confirmer ou infirmer les résultats obtenus dans la présente étude.

Celle-ci devrait également inclure l'établissement de normes afin de comparer les performances des enfants franco-québécois entre eux et ainsi formellement statuer sur la présence ou l'absence d'un déficit en compréhension de lecture.

Il pourrait être pertinent de poursuivre le développement du test en y ajoutant de nouvelles épreuves évaluant spécifiquement un processus issu du modèle de van Dijk et Kintsch (1983), étoffant ainsi le score global de compréhension de lecture.

Aussi, comme le test *Une bouteille à la mer* s'adresse aux élèves de la 4^e à la 6^e année, une autre étude pourrait avoir comme objectif de développer et de valider un test de compréhension de lecture pour les élèves plus jeunes, allant de la 1^{ère} à la 3^e année du primaire.

Finalement, il serait pertinent de mener une étude auprès d'une population clinique afin de pouvoir statuer sur la sensibilité et la spécificité du test. Cela dit, le diagnostic de trouble spécifique des apprentissages dans le domaine de la compréhension de lecture est rarement donné. D'ailleurs, la prévalence de ce trouble n'est pas établie au Québec ni au Canada. D'autres troubles neuropsychologiques s'accompagnent souvent de difficultés en compréhension écrite. C'est le cas notamment du syndrome des dysfonctions non verbales (SDNV), du trouble du spectre autistique (TSA) ainsi que du trouble primaire du langage (dysphasie) (Nation, 2005). Les enfants possédant l'un de ses diagnostics et étant connus pour avoir des lacunes en

compréhension écrite, pourraient former le groupe contrôle. Ce faisant, l'utilité clinique de ce dernier serait également démontrée.

Conclusion

En conclusion, le test *Une bouteille à la mer* s'avère prometteur, ses propriétés psychométriques étant somme toute bonnes, que ce soit en matière de fidélité ou de validité. Par ailleurs, il offre l'avantage d'avoir été conçu et développé au Québec, et ce, spécifiquement pour les élèves francophones. Il est donc adapté à la réalité linguistique et culturelle franco-québécoise, en plus de suivre la progression des apprentissages (PDA) proposés par le MELS (2009). Ce test prend également assise sur un modèle théorique, ce qui permet de mettre en exergue les processus non spécifiques de compréhension langagière potentiellement déficitaires, guidant ainsi les interventions qui seront prodiguées auprès de ces élèves. Il importe de souligner que d'une discipline à l'autre, les modèles théoriques de référence ne sont pas toujours les mêmes, limitant ainsi la communication entre professionnels et entravant, par le fait même, la prise en charge des élèves. Comme ce test est le fruit d'une collaboration entre neuropsychologues et orthopédagogues et que le modèle théorique choisi convient aux deux disciplines, ce test crée un pont entre celles-ci et rend la prise en charge d'autant plus efficace.

Références

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5e éd.). Arlington, VA : American Psychiatric Publishing.
- Apprenants en difficulté et littératie [ADEL UQAM]. (2020). *Épreuves de compréhension en lecture entre la 4^e année et la 2^e secondaire*. <https://adel.uqam.ca/docs-category/epreuves-comprehension-lecture-entre-4e-sec2/>
- Bernard, F. (2017). *Les mécanismes de la lecture: Développement normal et pathologique, enfance et vieillissement*. De Boeck Supérieur.
- Bertrand, D., Fluss, J., Billard, C. et Ziegler, J. C. (2010). Efficacité, sensibilité, spécificité : comparaison de différents tests de lecture. *L'Année psychologique*, 110(02). <https://doi.org/10.4074/s000350331000206x>
- Best, R. M., Floyd, R. G. et Mcnamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading psychology*, 29(2), 137-164.
- Bianco, M. et Coda, M. (2002). La compréhension en quelques points. *La compréhension*, 93-97.
- Biber, D. (1986). Strategies of discourse comprehension By Teun A. van Dijk and Walter Kintsch. *Language*, 62(3), 664-668.
- Bouchard, M.-E. G., Fitzpatrick, E. M. et Olds, J. (2009). Analyse psychométrique d'outils d'évaluation utilisés auprès des enfants francophones. *Canadian Journal of Speech-Language Pathology & Audiology*, 33(3).
- Carlisle, J. F. (1999). Free recall as a test of reading comprehension for students with learning disabilities. *Learning Disability Quarterly*, 22(1), 11-22.

- Castles, A., Rastle, K. et Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5-51.
- Chall, J. S. (1983). *Learning to read: The great debate*. McGraw-Hill.
- Chapleau, N., Laplante, L. et Brodeur, M. (2014). Enseigner la morphologie dérivationnelle pour apprendre l'orthographe lexicale. *Québec français*, (171), 83-84.
- Clarke, P. J., Snowling, M. J., Truelove, E. et Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological science*, 21(8), 1106-1116.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic.
- De Vellis, R. F. (2003). *Scale Development: Theory and Applications* (2nd ed). Sage Publications.
- Desrosiers, H. et Tétreault, K. (2012). Les facteurs liés à la réussite aux épreuves obligatoires de français en sixième année du primaire [ressource électronique]: un tour d'horizon.
- Dupin de Saint-André, M. (2008). Pistes pour travailler les inférences au primaire. *Québec français*, (150), 66-67.
- Écalle, J. (2010). L'évaluation de la lecture et des compétences associées. *Revue française de linguistique appliquée*, 15(1), 105-120.
- Garcia, L. J. et Desrochers, A. (1997). L'évaluation des troubles du langage et de la parole chez l'adulte francophone. *Revue d'orthophonie et d'audiologie*, 21(4), 271-284.
- Gernsbacher, M. A., Varner, K. R. et Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 430.
- Giasson, J. (1990). *La compréhension en lecture*. De Boeck Université.

- Gough, P. B. et Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, 7(1), 6-10.
- Hawken, J. (2009). Pour un enseignement efficace de la lecture et de l'écriture [ressource électronique]: une trousse d'intervention appuyée par la recherche.
- Hoover, W. A. et Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127-160.
- Irwin, J. W. (2007). *Teaching reading comprehension processes*. Pearson College Division.
- Israel, S. E. et Duffy, G. G. (2014). *Handbook of research on reading comprehension*. Routledge.
- Jacquier-Roux, M., Lequette, C., Pouget, G., Valdois, S. et Zorman, M. (2010). BALE: batterie analytique du langage écrit. *Grenoble: Laboratoire Cogni-Sciences*.
- Keenan, J. M., Betjemann, R. S. et Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific studies of reading*, 12(3), 281-300.
- Kintsch W. et Rawson, K.A. (2005). Comprehension. Dans M.J. Snowling et C. Hulme (dir.), *The Science of Reading: A Handbook* (p.209-226). Malden, États-Unis: Blackwell Publishing Ltd.
- Kintsch, W. et Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, 85(5), 363.
- 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.
- Laplante, L. (2011). L'évaluation diagnostique des difficultés d'apprentissage de la lecture. Dans *L'Évaluation de la littératie* (pp. 139-174). University of Ottawa Press.

- Lecocq, P. (1996). *L'É. co. s. se une épreuve de compréhension syntaxico-sémantique (manuel et épreuve): Deux volumes* (vol. 1). Presses Univ. Septentrion.
- Le Corff, Y. et Yergeau, É. (2016). *Définition du concept de fidélité*. Repéré à <http://psychometrie.espaceweb.usherbrooke.ca/la-methode-dequivalence/page-d-exemple>
- Lefavrais, P. (2005). *Test d'analyse de la lecture et de la dyslexie*: Alouette-R: Paris: Éditions du centre de psychologie appliquée.
- Lefebvre, P., Bruneau, J. et Desmarais, C. (2012). Analyse conceptuelle de la compréhension inférentielle en petite enfance à partir d'une recension des modèles théoriques. *Revue des sciences de l'éducation*, 38(3), 533-553.
- Lervåg, A., Hulme, C. et Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, 89(5), 1821-1838.
- Lobrot, M. (1973). D.OR.LEC, Disposition, ORthographe, LECture. *Paris : EAP*.
- Lobrot, M. et Adnet-Piat, E. (1976). *Lire: avec épreuves pour évaluer la capacité de lecture (D.OR.LEC)*.
- Maeder, C. (2010). *La forme noire (9-12 ans). Test de compréhension écrite de récits*. Isbergues, France : Ortho-Édition.
- Maeder, C. et Charlois, A.-L. (2010). Validation d'un test de compréhension écrite de récit:«La forme noire». *Glossa*, 108, 69-85.
- Majerus, S., Jambaqué, I., Mottron, L., Van der Linden, M. et Poncelet, M. (2020). *Traité de Neuropsychologie de l'Enfant—2ème édition*.

- McCauley, R. J. et Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders*, 49(1), 34-42.
- Ministère de l'Éducation du Loisir et des Sports. (2009a, 24 août). *Progression des apprentissages au primaire, Français langue d'enseignement*. http://www.education.gouv.qc.ca/fileadmin/site_web/documents/education/jeunes/pfeq/PDA_PFEQ_francais-langue-enseignement-primaire_2011.pdf
- Ministère de l'Éducation du Loisir et des Sports. (2009b). *Les échelles de niveaux de compétence au primaire*. <http://www.education.gouv.qc.ca/references/tx-solrtyperecherchepublicationtx-solrpublicationnouveaute/resultats-de-la-recherche/detail/article/les-echelles-des-niveaux-de-competence-au-primaire/pubLang/0/>
- Montag, J. L. (2019). Differences in sentence complexity in the text of children's picture books and child-directed speech. *First language*, 39(5), 527-546.
- Montag, J. L., Jones, M. N. et Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological science*, 26(9), 1489-1496.
- Montag, J. L. et MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8-and 12-year-old children and adults. *Journal of Experimental Psychology: General*, 144(2), 447.
- Nation, K. (2005). *Children's Reading Comprehension Difficulties*.
- Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties*, 24(1), 47-73.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd edit.). *New York*.
- Nunnally, J. C. (1994). *Psychometric theory 3E*. Tata McGraw-hill education.

- Pecatte, M. M. (2015). *Evaluation informatisée des habiletés de lecture et de compréhension de l'enfant: vers des profils de lecteurs.*
- Perfetti, C. A., Landi, N. et Oakhill, J. (2005). The acquisition of reading comprehension skill. *The science of reading: A handbook*, 227-247.
- Potocki, A., Bouchafa, H., Magnan, A. et Ecalle, J. (2014). Évaluation de la compréhension écrite de récits chez l'enfant de 7 à 10 ans: vers des profils de compreneurs. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 64(5), 229-239.
- Reed, D. K. et Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific studies of reading*, 16(3), 187-217.
- Sarrazin, G. (1995). Test de Rendement pour Francophones (TRF): Ottawa, Ontario, Canada: Harcourt Canada Ltd/Psychological Corporation.
- Souza, A. C. d., Alexandre, N. M. C. et Guirardello, E. d. B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde*, 26, 649-659.
- Stephenson, R. et Parent, N. (2012). *Introduction à la psychométrie* (1^{ère} éd.). Montréal, Canada : Chenelière Éducation.
- St-Pierre, M.-C., Dalpé, V., Lefebvre, P., Giroux, C., Dalpé, M.-C. S.-P. V. et Céline, P. L. G. (2000). *Difficultés de lecture et d'écriture*. PUQ.
- Turcotte, C. et Talbot, N. (2017). Élaboration d'une épreuve de compréhension en lecture en 6^e année du primaire favorisant l'articulation enseignement-apprentissage-évaluation. *Mesure et évaluation en éducation*, 40(3), 37-67.

- Van Dijk, T. A. et Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, États-Unis : Academic Press.
- Van Grunderbeeck, N., Théorêt, M., Chouinard, R. et Cartier, S. (2004). Suggestions de pratiques d'enseignement favorables au développement de la lecture chez les élèves du secondaire. *Montréal: Université de Montréal*.
- Verhoeven, L. et Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(3), 407-423.
- Wechsler, D. (2005). *Test de rendement individuel de Wechsler™, version pour francophones du Canada*. Toronto, Canada : Pearson.
- Wechsler, D. (2008). *Test de rendement individuel de Wechsler™, version pour francophones du Canada, manuel des normes québécoises (2^e édition)*. Toronto, Canada : Pearson.
- Wechsler, D. (2016). *Échelle d'intelligence de Wechsler pour enfants™, version pour francophone du Canada (5^e ed.)*. Toronto, Canada : Pearson.
- Wiig, E. H., Secord, W. A., & Semel, E. (2009). *Évaluation Clinique des notions langagières fondamentales- version pour francophones du Canada (CELF CDN-F)*. (L. Boulianne & M. Labelle, Trans.). Toronto, Ontario: Pearson Canada Assessment.
- Woodcock, R. W., Johnson, M. B. et Mather, N. (1990). *Woodcock-Johnson psycho-educational battery--Revised*. DLM Teaching Resources.

Annexe 1 – Processus cognitifs ciblé par les questions

QUESTIONS	Analyse syntaxique	Microprocessus Anaphores	Microprocessus Marques de relation	Macroprocessus	Inférences	Élaboration du modèle de situation
1) Dans l’histoire, qui a ramassé les coquillages ?	X	X				
2) Au début de l’histoire, pourquoi on dit que plus rien n’intéresse les enfants ?	X	X	X	X	X	
3) Dans le premier paragraphe, comment les enfants se sentent par rapport au vieux marin ?	X	X	X			
4) Qui découvre la bouteille et de quelle manière ?	X	X	X			
5) Qui est Fripouille et qu’est-ce qui te permet de le savoir ?	X	X	X	X	X	
6) Est-ce qu’on peut savoir si Fripouille est un mâle ou une femelle. Si oui, qu’est-ce qui te permet de le savoir ?	X					
7) Qui réussit à enlever la cire de la bouteille et comment ?	X	X	X			
8) La bouteille s’ouvre. Comment ont-ils réussi à retirer le papier de la bouteille ?	X	X	X	X	X	
9) Dans l’histoire, qu’est-ce que William étire doucement ?	X	X				
10) Qu’est-ce qui surprend les enfants en découvrant le bout de papier ?	X	X	X	X	X	
11) Comment les personnages découvrent-ils ce qui se cache sur le papier ?	X	X	X	X	X	
12) À la fin du texte les enfants doivent partir. Pourquoi ?	X	X	X	X	X	
13) En te référant aux événements de la fin de l’histoire, quelle pourrait être la suite selon toi?	X	X	X	X	X	X
14) Mise en ordre	X	X	X	X		
	1	5	3	1	6	1