The Development of Comprehension Monitoring Among Emerging English – French Bilingual Children

Klaudia Krenca[12], Kate Cain[3], Stefka Marinova-Todd[4], and Xi Chen[1]

[1] Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto

[2] Department of Psychology and Neuroscience, Dalhousie University

[3] Department of Psychology, Lancaster University

[4] School of Audiology and Speech Sciences, University of British Columbia

This is version submitted to the journal. Accepted for publication in Scientific Studies of Reading 21/03/2023. Please go to journal website for accepted version.

Author Note.

This research was funded by a Social Sciences and Humanities Research Council of Canada Grant # 435-2013-1741 (to X.C.). Opinions reflect those of the author(s) and do not necessarily reflect those of the funding agency. The author(s) have no conflicts of interest to disclose. This research would not have been possible without the support of the teachers, parents, and children at our partner schools, in addition to all of the research assistants of the Multilingualism and Literacy Lab. Portions of these findings were presented as a presentation at the 26th Annual Meeting of the Society for the Scientific Study of Reading, Toronto, Canada in July 2019.

Correspondence concerning this article should be addressed to Klaudia Krenca, Department of Applied Psychology and Human Development, Ontario Institute for Studies in

Education, University of Toronto, 252 Bloor St. W., Toronto, Ontario, M5S 1V6, Canada. ORCID 0000-0002-5845-967X. Email: k.krenca@mail.utoronto.ca

Abstract

Purpose. This one-year longitudinal study investigated the extent to which comprehension monitoring in children's first (L1) and second (L2) language predicts reading comprehension.

Method. Children's ability to detect inconsistencies in orally presented stories was measured by children's response to a judgement question about whether the story made sense, after each one, and by the identification of the inconsistency within the story. The participants included 115 children ($M_{ageGrade2}$ = 7.8 years, SD = 0.3) recruited from an early French immersion program in Canada.

Results. We computed a series of linear regression models, two in English and two in French. The concurrent results revealed that, in Grade 3, children's comprehension monitoring was a unique predictor of reading comprehension in English and French, over and above the contribution of word reading and vocabulary. This relationship was not observed in Grade 2. Notably, the longitudinal analyses indicated that Grade 2 children's comprehension monitoring in English made a significant contribution to English reading comprehension in Grade 3, over and above key control variables. However, this relationship was not established in French.

Conclusions. These results promote a call to include support for higher-level oral language skills during the early stages of bilingual reading instruction.

Keywords: comprehension monitoring; French immersion; inconsistency detection; oral language; reading comprehension

The Role of Comprehension Monitoring in Predicting Reading Comprehension Among Children in French Immersion

Children need to be able to understand what they read in order to succeed in school and participate in society. According to the Simple View of Reading framework (SVR: Gough & Tunmer, 1986), reading comprehension is the product of an individual's decoding and linguistic comprehension skills. Decoding is the ability to decipher printed words, whereas linguistic comprehension refers to the ability to comprehend language at the discourse level, including conversations, narratives, and informational oral texts (Kim & Pilcher, 2016). The relative relationship of these components changes over time with linguistic comprehension becoming increasingly more important as children's word reading proficiency develops (Keenan & Meenan, 2014). In this study, we explored a specific aspect of children's linguistic (or listening) comprehension, namely, comprehension monitoring, which is one's ability to think about and evaluate one's own comprehension of written or spoken text (Language and Reading Research Consortium (LARRC) & Yeomans-Maldonado, 2017), in relation to reading comprehension development in second language (L2) learners.

Successful text comprehension involves the construction of a coherent mental representation of the text's meaning called the "situation model" (Kintsch, 1988). To achieve a coherent situation model, readers (and listeners) must integrate propositions in the text into the situation model, thus updating their situation model of a text as each new piece of information is processed (Kintsch, 1988; van Dijk & Kintsch, 1983). Prior work has revealed that comprehension monitoring is an important skill that contributes to the construction of a coherent situation model for both written and spoken texts (Kintsch, 1988). Skilled comprehenders evaluate the adequacy of their comprehension as they process the text. If they have failed to

achieve a sufficient standard of coherence, they can apply repair strategies to restore comprehension (e.g., look up the meaning of an unknown word, or generate an inference to aid text integration).

To study comprehension monitoring for both written and spoken texts, many researchers have used an error detection task (e.g., Oakhill et al., 2005; Strasser et al., 2014). Individuals are presented with short texts that may or may not contain inconsistencies such as nonwords, violations with prior knowledge, or two contradictory segments (e.g., Oakhill et al., 2005; Oakhill & Cain, 2012). The task is typically to judge whether or not the text makes sense, and sometimes participants are asked also to identify the inconsistency. If an inconsistency is not detected, it is assumed that participants have failed to monitor their comprehension effectively. In our study, we used an inconsistency detection task that contained two pieces of information in separate sentences that were contradictory, so they could not both be integrated into the situation model for the text as a whole. This paradigm enabled us to assess the critical comprehension monitoring processes involved in the construction of the text's situation model.

For monolingual children, the ability to monitor comprehension of orally presented texts is related to general reading and listening comprehension skills (e.g., Kim, 2015; LARRC & Yeomans-Maldonado, 2017; Strasser et al., 2014). Despite its importance, the subcomponents that underlie listening comprehension have not received much attention, particularly among young bilingual children. Building on these studies, we investigated whether comprehension monitoring explains unique variance in bilingual readers' reading comprehension, and whether this relationship changes from Grade 2 to Grade 3. We examined these relationships among children acquiring French as a second language (L2) within an early immersion context in Canada. Since English is the home language and often also the dominant societal language,

children in this context become additive bilinguals in English and French (Johnson & Swain, 1997).

The Relationship Between Comprehension Monitoring and Listening Comprehension

Listening comprehension involves the processes of extracting and constructing meaning via lower-level oral language skills (e.g., vocabulary, grammatical knowledge) as well as higher order processes at the discourse level (e.g., comprehension monitoring and inference, see Kim & Pilcher, 2016). As already noted, comprehension monitoring refers to the processes by which readers evaluate their understanding of a text (LARRC & Yeomans-Maldonado, 2017). Skilled readers (and listeners) are more likely to recognize a lapse in comprehension (e.g., resolving apparent inconsistencies with their understanding so far or statements conflicting with background knowledge) and employ various strategies in order to restore coherence in their understanding (e.g., re-read sections of the text to check for meaning). Individuals with poor comprehension monitoring skills may not be aware that their comprehension has failed, hence, they may not take appropriate remedial action (e.g., ask questions). Therefore, a lack of comprehension monitoring would result in an inaccurate situation model. We modelled our tasks and study design on those used in previous work on monolingual readers' comprehension monitoring for narrative passages (e.g., Oakhill & Cain, 2012); we used an inconsistency detection task and assessed performance longitudinally, using statistical controls for word reading and lower-level oral language skills, which are both strongly related to reading comprehension. This enabled us to examine the specific relationship between comprehension monitoring and reading comprehension, to test the reproducibility of earlier study findings, and to assess the generalizability of those findings and multicomponent accounts of reading comprehension for a different reader group.

The Relationship Between Comprehension Monitoring and Literacy Among Young Readers

There is evidence to suggest that comprehension monitoring is observed in early language development (Skarakis-Doyle, 2002), but still developing after several years of formal literacy instruction (Baker, 1984). Several studies have demonstrated a unique relation between 5- to 11-year-olds comprehension monitoring and their reading or listening comprehension after controlling for other key variables, such as vocabulary and working memory (Cain et al., 2001; Cain et al., 2004; Kim, 2016). Another body of research has demonstrated that children with reading comprehension difficulties struggle with critical comprehension monitoring processes, such as detecting internal inconsistencies in text (Oakhill et al., 2005). Related to this, van der Schoot et al. (2012) found that poor comprehenders were able to detect an inconsistency when the two pieces of conflicting information were near to each other, but were less likely to detect inconsistent information that was further apart, and the inconsistency could only be detected by comparison with the current situation model. Together, these studies suggest that younger and poorer readers' reading and listening comprehension is constrained by their comprehension monitoring processes.

The literature examining the development of comprehension monitoring using a longitudinal design is scarce. Oakhill and Cain (2012) found that comprehension monitoring (evaluated via an inconsistency detection task) of children aged 7 to 8 years was a unique predictor of reading comprehension 4 years later, over and above general verbal ability, vocabulary, and the autoregressive effect of reading comprehension. LARRC and Yeomans-Maldonado (2017) studied the development of comprehension monitoring in a sample of English-speaking children from Grade 1 to Grade 3 (assessed with an inconsistency detection

task based on the work of Oakhill & Cain, 2012) and its relation to reading comprehension. Children's ability to monitor comprehension significantly increased at each time point demonstrating age-related development. Of note, comprehension monitoring in Grade 1 was a significant predictor of reading comprehension in Grade 3, even after controlling for decoding, vocabulary, and working memory.

The critical role of comprehension monitoring in successful reading comprehension is further demonstrated by intervention research. A 4-week reading strategy intervention that involved explicit instruction in teaching comprehension monitoring strategies (e.g., identifying main ideas, summarizing, generating questions, and clarifying) resulted in improved comprehension monitoring behaviours (reading an inconsistent piece of text more slowly, indicating that a problem was detected) for Dutch fourth-grade children (Wassenburg, Bos et al., 2015). The researchers did not observe the same differences in the reading behaviours of younger Grade 3 children who had completed the intervention. However, critically, children in the intervention condition improved their general reading comprehension performance and reading motivation scores compared to controls, supporting other research that finds a strong relation between comprehension monitoring performance and reading comprehension.

To the best of our knowledge, there are only two studies that have empirically examined comprehension monitoring among bilingual learners. The first, by Morrison (2004), studied participants who were older than the young learners in our research; they were undergraduates who had completed a French immersion program. Morrison (2004) examined the role of comprehension monitoring, evaluated with an inconsistency detection task, in reading comprehension in both English (L1) and French (L2). The passages contained two types of errors: micro-level errors involve homonyms or words in which one letter has changed (e.g.,

expanse vs. expense); macro-level errors (e.g., words or sentences that contradict information presented in the previous sentences) which are more likely to be detected if the reader has a coherent representation of the passage. Macro-level error detection in English and French did not predict French reading proficiency. While the researchers acknowledged the small sample size, they did not assess known predictors of reading comprehension such as participants' word reading and vocabulary skills (e.g., LARRC & Yeomans-Maldonado, 2017). The other study by Currie et al. (2021) contrasted comprehension monitoring of narrative and expository text in Grade 6 children who were either monolingual or Spanish-English dual language learners (DLLs) when they had started pre-kindergarten. The results indicated that readers adopted different standards of monitoring for different genres, with lower levels of detection when processing expository text. Of relevance to our study, the Spanish-English DLLs in Currie et al.'s (2021) study were less likely to detect inconsistencies when reading compared to monolinguals, possibly attributed to unobserved effects of socioeconomic status and language exposure at home.

The Current Study

The present study was designed to understand the development of comprehension monitoring in the L1 (English) and the L2 (French) among children enrolled in a French immersion program. The observation of poor comprehension monitoring in bilinguals and second language learners in previous studies (Currie et al., 2021; Morrison, 2004) is not necessarily predicted. Comprehension monitoring is a metacognitive skill, involving reflecting one's own task performance (Flavell, 1979). There is a wealth of literature reporting a bilingual advantage on cognitive skills such as executive function, memory updating, and inhibition (e.g., Bialystok, 2009). However, that advantage may not necessarily transfer to language-dependent

tasks, given the weaker performance on language and literacy tasks often reported for dual language learners (Lesaux et al., 2010). Furthermore, a meta-analytic review contrasting performance of monolingual and bilingual adults reported a small disadvantage for bilinguals on language tasks (Lehtonen et al., 2018). Thus, the study of comprehension monitoring in bilingual children is an important line of enquiry to add to this scant body of knowledge, that has the potential to provide theoretical insights into bilingual reading development and guidelines for classroom practice.

The following research objectives guided our study. First, we investigated to what extent comprehension monitoring (for aurally presented texts) predicts unique variance in reading comprehension, over and above key variables (e.g., word reading and vocabulary), and whether this relationship changes among children in French immersion from second to third grade. In order to address our first research objective, data collected in Grade 2 and 3 was analyzed concurrently with language and literacy variables at each time point predicting within-language reading comprehension outcomes. Studies of monolinguals (e.g., Cain et al., 2004) find that children's comprehension monitoring explains unique variance in reading comprehension across this age group. However, using the Simple View of Reading as a framework, which theorizes a relative change in the predictive power of word reading and language comprehension skills over time, we expected a stronger relationship between comprehension monitoring and reading comprehension at Grade 3 relative to Grade 2.

The second objective was to determine whether and to what extent comprehension monitoring in Grade 2 serves as a unique predictor of reading comprehension when children are in Grade 3. We adopted a conservative approach by including the respective variables measured during Grade 2 as an autoregressor when estimating their potential contribution one year later.

Without accounting for the autoregressor, relations among other predictors can be artificially inflated (Kenny, 1975). As a result, the inclusion of the autoregressive effect of reading comprehension rules out the possibility that the relations among the predictors with reading comprehension in third grade were due to their association with reading comprehension in the second grade, supporting a causal relation between those variables and the outcome measure (de Jong & van der Leij, 2002).

In relation to the second objective, we first hypothesized that children's comprehension monitoring would show significant improvement in English and French as children progressed from Grade 2 to Grade 3. In relation to the prediction of reading comprehension, we expected that comprehension monitoring would explain unique variance in the development of children's reading comprehension, in accordance with the only reported study to examine this relationship longitudinally among English-monolingual children (LARRC & Yeomans-Maldonado, 2017). To date, there are no studies that examine the extent to which comprehension monitoring serves as a potential foundational skill in the prediction of future reading comprehension in bilingual readers, and not much is known about reading comprehension in a French immersion setting. However, previous research with monolinguals, shows that comprehension monitoring in the early grades predicts later reading comprehension, even after controlling for lexical-level skills (e.g., LARRC & Yeomans-Maldonado, 2017; Oakhill & Cain, 2012), and there is evidence that comprehension monitoring is malleable so could be a suitable target for intervention in the early grades (Language and Reading Research Consortium (LARRC), Jiang, & Logan, 2019; Wassenburg, Bos, et al., 2015).

The current study was addressed within the context of an early French immersion program in two predominantly English-speaking regions in Canada. This immersion

environment presents a unique context for language learning, which leads to additive English-French bilingualism (Swain & Lapkin, 2005). In other words, English is the children's stronger language as well as the language of the community, but instruction in English is not provided in the early grades (Genesee & Jared, 2008). French is the second or additional language and all the school instruction is provided in French. However, the L2 vocabulary size of children who begin French immersion is limited because these children live in predominantly English-speaking environments with minimal opportunity to hear and speak French outside of the classroom. This learning environment differs from many second language users (e.g., Spanish – English bilinguals) where their L1 is a minority language in the community and there are no formal literacy opportunities in this language.

Method

Participants

A total of 166 Grade 2 children participated in the present study as part of a larger project. They were tested in Grade 2 and Grade 3. Thus, the study design yielded both cross-sectional and longitudinal data. Participants were recruited from early French immersion schools in two predominantly English-speaking regions in Canada. The children received French instruction beginning in kindergarten. The participants were, therefore, in the early stages of acquiring French as an additional language. Attrition analyses conducted on demographic variables (level of education, place of birth, gender, first language status) as well as on the variables used in the main analyses for this study (see Table 1) showed no significant differences between children who dropped out of the study (n = 51) and those who remained in the study between Grade 2 to Grade 3 (p values > .05). For these reasons, we were able to rule out selective attrition in the sample.

The final sample consisted of 115 children (52 males, $M_{ageGrade2}$ = 7.8 years, *SD* = 0.3, age range = 7.2 – 8.4 years) from 11 elementary French immersion schools. With respect to the children's place of birth, 94% were born in Canada (*n* = 105).[1] Approximately 73% (*n* = 84) of the sample was only exposed to English at home. Of the 115 participants, 12% (*n* = 14) had at least one guardian who spoke an unofficial language to the child at least 50% of the time. Similar to Authors (2020), we consider these children as English Language Learners. This group was linguistically diverse (*n* = 6 for Mandarin, *n* = 2 for Turkish, and *n* = 1 for Dutch, German, Gujarati, Spanish, Tamil, and Vietnamese). The children's overall exposure to French was generally limited to input received at school. Based on the demographic data, 15% of children never spoke French at home, while 79% rarely spoke French.[2] With respect to the highest level of education obtained, 90% (*n* = 103) of guardians had a university-level bachelor's degree or higher.

Measures

Parental Demographic Questionnaire. Parents completed a demographic questionnaire adapted from the Alberta Language Environment Questionnaire (ALEQ, Paradis et al., 2010). They responded to questions such as the highest level of education attained, self-rated fluency in English, French, and/or another language [5-point scale from (0) *not fluent* to (4) *very fluent*], current language use by family members in the home, and the richness of the children's language environment, and whether the children experienced any learning challenges (e.g., speech, hearing, behavioural). In terms of the current study, socioeconomic status was estimated based

[1] Place of birth is missing for three participants.

[2] Data is based on 79 participants who responded to this question using a 5-point Likert scale from (0) = *never* to (4) = *almost always*.

on the parent or guardian's highest level of education [6-point scale from (0) *primary* to (5) *University – PhD*). The information obtained from this questionnaire was used to describe the sample.

Non-Verbal Reasoning. During the second year of the longitudinal study (first grade), participants were assessed on the reasoning by analogy subtest of the Matrix Analogies Test in English (Naglieri, 1985) to measure non-verbal reasoning. The children were presented with an incomplete visuospatial matrix and asked to select one of six options that accurately completed each picture. There were 16 items and testing was discontinued when four consecutive items were answered incorrectly. The raw score represents the sum of correct responses.

Phonological Short-Term Memory. During the second year of the longitudinal study (first grade), the Memory for Digits subtest of the Comprehensive Test of Phonological Processing – Second Edition (CTOPP-2; Wagner et al., 2013) was used to measure participants' ability to code information phonologically for temporary storage in short-term memory. For this task, children listened to orally presented sets of digits and were then asked to repeat the numbers aloud in the same order. The number of digits increased across trials. The subtest consists of 28 items and was administered in English. Feedback was provided for the first four items and the task was discontinued when three consecutive items were answered incorrectly. The raw score represents the sum of correct responses.

Receptive Vocabulary. English receptive vocabulary was measured using the fourth edition of the Peabody Picture Vocabulary Test (PPVT-IV, Form A; Dunn & Dunn, 2007). Administration followed standardized procedures for establishing a basal and ceiling. Children identified one of four pictures that best illustrated a target word spoken by the tester. Three practice items preceded test items. Testing was discontinued on the last item of a set on which

the child made eight errors or more. The raw scores refers to the total number of words identified correctly (maximum 228 items).

Form A of the Échelle de Vocabulaire en Images Peabody (EVIP; Dunn et al., 1993), normed on a sample of Canadian children who were native speakers of French, was used to assess students' French receptive vocabulary. A target word was spoken, and children identified one of four pictures that best illustrated the item. Three practice items preceded test items. Testing was discontinued when the child made six errors within a set of eight items. The raw scores refers to the total number of words identified correctly (maximum 170 items).

Word Reading Accuracy. The letter-word identification subtest of the Woodcock-Johnson III Test of Achievement (Form B; Woodcock et al., 2001) was administered as an untimed measure of non-contextual single word reading skills. The raw scores refers to the total number of words read correctly (maximum 76 items). The basal and discontinuation rule of the standardized procedure were followed. A raw score was calculated as the total correct items including credit for items below the basal.

The Lecture de Mots (word reading) subtest of the Wechsler Individual Achievement Test-Second Edition (WIAT-II; Wechsler, 2005) was used to assess French word reading.[3] This measure has been standardized on Canadian children who are native speakers of French. In this task, the children were asked to read aloud words that gradually increased in difficulty. The raw scores refers to the total number of words read correctly (maximum 84 items). Testing was discontinued after having produced seven incorrect responses on the word items.

[3] The administration of this test differed from standardized procedure. Only the "La Carte des Mots" (Word Reading Card) was used to assess children's word reading accuracy from this subtest.

Reading Comprehension. The Gates-MacGinitie Reading Comprehension Test, Fourth Edition (GMRT-IV; MacGinitie et al., 2000), Level 2 and Level 3 – Form S; was used to assess children's English reading comprehension. In Grade 2, the written passages were presented in units of one or more sentences and children selected one of three pictures that best represented the meaning of what had been read. In Grade 3, children read short passages as a whole (no pictures) and answered corresponding multiple-choice questions (one of four choices) after each passage. The Level 2 test consists of 39 items, while the Level 3 measure contains 48 items. The total score refers to the number of correct answers completed within a 35-minute time limit. The children were permitted to refer to the passages during question answering.

The Gates-MacGinitie Reading Tests, Second Canadian Edition (GMRT-II; MacGinitie & MacGinitie, 1992) was translated into French (Level B, Form 4 for Grade 2; Level C, Form 4 for Grade 3) and administered in the same way as the English task. The Grade 2 test consists of 46 items, while the Grade 3 measure contains 47 items. This task has been used in previous research (e.g., D'Angelo et al., 2020, Lee & Chen, 2019).

Comprehension Monitoring. An experimental measure of comprehension monitoring that assessed children's ability to detect inconsistencies in orally presented stories was developed based on previous literature (see LARRC & Yeomans-Maldonado, 2017). The same test stories were used in Grade 2 and 3. There were 12 test stories that were either entirely consistent (4 stories) or included inconsistent information (8 stories). Only the scores for the inconsistent stories were used for analyses. Identification of inconsistent information (two propositions in the text that conflict in meaning) required integration of two elements across two separate sentences. An example of an inconsistent story is: *"Last night Jill walked home through the park. There was no moonlight, so Jill could hardly see her way. Jill often takes this route home because it is*

a good shortcut to her house. She walked along a narrow path. The moon was so bright that it lit the way. Jill lives on the other side of the park." For the 8 inconsistent stories presented, children were asked (a) if the story made sense, and if they answered correctly to this question, they were then asked (b) what was wrong with the story (i.e., "There was no moonlight so Jill could hardly see her way" and "The moon was so bright that it lit the way.") Part (b) was scored as either incorrect (score = 0) or correct (score = 1), and it was this part which was used to compute the total score for this task which ranged from 0 to 8 points. Before administering the experimental items, three practice stories were provided to ensure that children understood the task. A parallel measure was designed in French and was reviewed by a doctoral student who is a native speaker of French. In terms of the scoring, we did not award a correct response for children who may have known how to describe the inconsistency in English but did not have adequate vocabulary in French to detect the discrepancy.[4]

Procedure

Parents of the participating children completed a consent form (at each wave of the study) and a questionnaire (at the beginning of Grade 1) to collect basic demographic and background information on children's home language environment. All children were assessed during the spring term of each timepoint at the student's school during regular instructional hours and took approximately 45-60 minutes (over 3 – 4 sessions). English measures were administered before French measures for all participants. English instructions were used to explain the English measures, and both English and French instructions were provided for the French measures to

[4] Given the experimental nature of the task, Cronbach's alpha reliability is in line with previous research among English monolingual children. Internal consistency for the same experimental comprehension monitoring task ranged from .73 to .84 across Grades 1-3 (LARRC & Yeomans-Maldonado, 2017).

ensure that students understood task directions. Comprehension monitoring, word reading, and vocabulary were assessed in an individual format whereas, reading comprehension was assessed in a small group format. The participants were assessed individually outside of their classroom by carefully trained research assistants who were proficient in both English and French.

Results

Descriptive Statistics

Table 1 summarizes the descriptive statistics for all English and French measures. Both raw and standard scores are reported for all the standardized measures, whereas raw scores were used for the statistical analyses. Norms for standardized tests are based on English monolingual samples, however, the majority of the children were English L1 speakers and research has shown that French immersion children perform similarly on English measures to children in English stream programs (e.g., Au-Yeung et al., 2015). The standard scores for English word reading accuracy, English and French receptive vocabulary are based on a distribution with a mean of 100 and a standard deviation of 15. At both time points, the standard scores for English vocabulary were one standard deviation above average, whereas children's performance in French vocabulary was below average. In terms of English word reading accuracy, the standard scores were within average range relative to age-level peers at both grades. The scaled (standard) scores for nonverbal reasoning and phonological short-term memory, assessed when the children were in Grade 1, indicated that the children in our sample fell within the average range. Finally, with respect to English reading comprehension, the stanine score represents a statistical (normalized) transformation of percentile ranks in which the range of reading achievement is divided into nine equal units with a mean of five and a standard deviation of two. A stanine score of six represents that the raw score is above average. In our sample, children's reading

comprehension performance was within the average range relative to age-level peers at both grades.

***** Insert Table 1 here *****

Table 2 presents bivariate correlations among the measures across both time points with significant correlations among most variables. There was a significant correlation between comprehension monitoring and reading comprehension, the primary variables of interest in this study, across both grades in English (i.e., Grade 2, $r = .18$, Grade 3, $r = .47$) and in French (i.e., Grade 2, $r = .41$, Grade 3, $r = .32$). Comprehension monitoring was significantly correlated within English (Grade 2, $r = .49$) and French (Grade 3, $r = .29$), and between English (Grade 2, $r = .19$) and French (Grade 3, $r = .32$). In terms of the remaining predictors, there was a significant correlation between comprehension monitoring and word reading across both grades in English (i.e., Grade 2, $r = .19$, Grade 3, $r = .28$) and in French (i.e., Grade 2, $r = .38$, Grade 3, $r = .50$). Moreover, there was a significant correlation between comprehension monitoring and vocabulary across both grades in English (i.e., Grade 2, $r = .37$, Grade 3, $r = .33$) and in French (i.e., Grade 2, $r = .26$, Grade 3, $r = .54$).

***** Insert Table 2 here *****

There were three extreme univariate outliers ($z < \pm 3.29$; Tabachnick & Fidell, 2013) within each grade based on the z scores of all the measures used in the analysis.[5] Overall, the rate of missing data was acceptable with percentages ranging from 0% to 11% -- listwise deletion was selected before running the analyses. Little's Missing Completely at Random Test (Little, 1988) indicated that the data were missing completely at random in Grade 2, χ^2 (113, $N = 115$) =

[5] The same pattern of findings emerged when analyses were performed without outliers; therefore, we report the analyses based on the entire sample.

128.32, $p = .154$, and Grade 3, χ^2 (100, $N = 115$) = 117.05, $p = .117$. Skewness and kurtosis values fell within the acceptable range (i.e., statistic/SE ± 3.29; Tabachnick & Fidell, 2013) with the exception of Grade 2 English comprehension monitoring (negative skew, -5.77), Grade 2 English reading comprehension (negative skew, -5.27), Grade 3 English comprehension monitoring (negative skew, -5.52) and Grade 3 English word reading (negative skew, -4.43). These variables were corrected using x^2 or antilogarithm transformations (Tabachnick & Fidell, 2013).[6] Using a $p < .001$ criterion for Mahalanobis distance (Tabachnick & Fidell, 2013), no multivariate outliers were detected within each grade. The assumptions of heteroscedasticity and independent errors were met for all regression models upon the examination of scatterplots with the standardized residual scores plotted against the standardized predicted values. The degree of collinearity for all regression analyses was found to be acceptable (tolerance > .460; variance inflation factor < .217) and all correlations were below .90, reducing the likelihood of multicollinearity (Tabachnick & Fidell, 2013).

Finally, to prepare the data for analyses, we examined whether there was statistical support for merging the samples of children who spoke English as a primary language at home and those who were exposed to additional home languages into one sample. A chi-square test of independence indicated a non-significant relationship between language group and testing site, χ^2 (1, $N = 115$) = .10, $p = .752$. A Box's M using the Grade 2 and 3 measures, indicated no significant difference in variance-covariance patterns between the two language groups on English (Grade 2, Box's $M = 34.36$, $p = .090$, Grade 3, Box's $M = 26.23$, $p = .347$) and French

[6] Unless indicated, the same pattern of findings emerged when analyses were performed with raw scores, transformed scores, and bootstrapped scores; therefore, we report the analyses with raw scores for ease of interpretability.

(Grade 2, Box's $M = 14.77$, $p = .899$, Grade 3, Box's $M = 25.71$, $p = .335$) reading and language measures. Based on these results and given that the children exposed to additional languages met the inclusion criteria (non-native speakers of French), the two groups were combined to create one sample.

The Concurrent Role of Comprehension Monitoring in Reading Comprehension

The first aim of this study was to examine whether and to what extent there were grade related changes in comprehension monitoring among children in French immersion. Our first research objective examined the unique contribution of comprehension monitoring for aurally presented texts to reading comprehension concurrently in Grade 2 and Grade 3. To address this objective, we computed a series of linear regressions models, two in English and two in French (Table 3). In each regression, age, nonverbal reasoning, phonological short-term memory, and parental education were entered in the first step to take into account children's general cognitive ability and demographic background. In particular, nonverbal intelligence (Catts et al., 2001) and the highest level of parental education obtained (Jeong et al., 2017) have been associated with reading comprehension. Word reading accuracy was entered as the second step, receptive vocabulary was included in the third step, and comprehension monitoring was entered in the final step. All standardized regression coefficients reflect values obtained from the final model.

The first two linear regression analyses examined whether comprehension monitoring skills were related to reading comprehension in Grade 2. Beyond the 51% of the variance taken up by the control variables, English comprehension monitoring in Grade 2 made no independent contribution to English reading comprehension, indicating a non-significant within-language relationship. We then examined the within language contribution of comprehension monitoring to French reading comprehension in Grade 2. Beyond the 52% of the variance taken up by the

control variables, French comprehension monitoring in Grade 2 explained 2% of unique variance in French reading comprehension in Grade 2, illustrating a marginally non-significant within-language relationship. Collectively, the variables considered in the regression model explained over 54% of the variance in Grade 2 French reading comprehension.

The next set of analyses examined the concurrent contribution of comprehension monitoring to reading comprehension in Grade 3. Beyond the 41% of the variance explained by the control variables, English comprehension monitoring in Grade 3 accounted for a statistically significant portion of unique variance (8%) in English reading comprehension, representing a within-language contribution. All of the variables entered in the model explained about 49% of the variance in children's English reading comprehension in Grade 3. We then examined the within language contribution of comprehension monitoring to French reading comprehension in Grade 3. Beyond the 49% of the variance taken up by the control variables, French comprehension monitoring in Grade 3 explained 4% of unique variance in French reading comprehension in Grade 3, revealing a within-language relationship. Altogether, the variables considered in the regression model explained over 53% of the variance in Grade 3 French reading comprehension.

***** Insert Table 3 here *****

The Longitudinal Role of Comprehension Monitoring in Reading Comprehension

In line with our second research objective, we examined the contribution of comprehension monitoring to reading comprehension from Grade 2 to Grade 3 in English and French using a series of cross-lag regression analyses with autoregressive controls. We adopted a conservative approach by including the respective variables measured during Grade 2 as an autoregressor when estimating their potential contribution one year later, in Grade 3. Without

accounting for the autoregressor, relations among other predictors can be artificially inflated (Kenny, 1975). As a result, the inclusion of the autoregressive effect of reading comprehension ruled out the possibility that the relations among the predictors with reading comprehension in Grade 3 were due to their association with reading comprehension in Grade 2, supporting a causal relation between those variables and the outcome measure (de Jong & van der Leij, 2002).

Beyond the 48% of the variance explained by the control variables, English comprehension monitoring in Grade 2 accounted for a statistically significant portion of unique variance (7%) in English reading comprehension in Grade 3, representing a within-language contribution. All of the variables entered in the model explained about 55% of the variance in children's English reading comprehension in Grade 3. A supplementary analysis was conducted with the autoregressive variable entered in the first step of the regression model ($\Delta R^2 = .37$, $p < .001$). English comprehension monitoring in Grade 2 was a significant predictor of English reading comprehension in Grade 3 ($\Delta R^2 = .07$, $p = .001$, $\beta = .29$, $p = .001$). We then examined the within language longitudinal contribution of comprehension monitoring to French reading comprehension in Grade 3. Beyond the 56% of the variance taken up by the control variables, French comprehension monitoring in Grade 2 made no independent contribution to French reading comprehension in Grade 3, indicating a non-significant relationship.

***** Insert Table 4 here *****

Discussion

The objectives of the present study were twofold. First, we investigated whether comprehension monitoring contributes to concurrent reading comprehension performance among children in French immersion, and whether this relationship changes across the developmental course from Grade 2 to Grade 3. With regards to the first objective, the results partially

supported our hypothesis. In Grade 2, comprehension monitoring was not a significant predictor of reading comprehension in English and French. By contrast, in Grade 3, the ability to monitor one's comprehension explained unique variance in reading comprehension in both languages. The second objective of this study was to explore longitudinally whether comprehension monitoring in Grade 2 serves as a unique predictor of reading comprehension in Grade 3. The results were again in partial support of our initial hypothesis. After controlling for within-language variables in Grade 2 (including the autoregressor), English comprehension monitoring explained unique variance in English reading comprehension one year later. In contrast, the same pattern of findings was not observed between French comprehension monitoring in Grade 2 and French reading comprehension in Grade 3.

Overall, our findings are broadly in line with those of studies with L1 learners (e.g., Cain et al., 2004; Oakhill & Cain, 2012) and importantly and uniquely extend those findings to second language learners. First, our results align with the SVR (Gough & Tunmer, 1986), as we observed a strong contribution of decoding skills in explaining the variance in Grade 2 reading comprehension. As children progressed from Grade 2 to Grade 3, the influence of decoding skills on reading comprehension decreased, while the influence of comprehension monitoring increased. Interestingly, in Grade 2, comprehension monitoring did not explain concurrent unique variance in reading comprehension in either language, but it was a significant predictor in Grade 3. Notably, Grade 3 comprehension monitoring explained unique variance in reading comprehension in both languages, even after controlling for established predictors associated with reading.

We interpret our findings to support other work that shows that children's oral language skills make less of a contribution to reading comprehension for younger readers, because

decoding is the critical determinant of reading comprehension in the beginning stages of reading development (Chall, 1983; LARRC, 2015). As children get older, decoding skills become more efficient enabling language comprehension skills to exert a greater influence on the prediction of reading comprehension. Furthermore, the texts they encounter become increasingly complex, placing greater demands on oral language. As a result, the strength of the relationship between word reading skills and reading comprehension decreases, and that between language skills and reading comprehension increases (García & Cain, 2014; LARRC, 2015). Our concurrent findings align with this research base.

An alternative (and not mutually exclusive) interpretation is that our reading comprehension assessment for Grade 2 was more heavily dependent on decoding than language comprehension skills, as has been demonstrated for other assessments (see, e.g., Keenan & Meenan, 2014). The format for Grade 2 required children to select a picture that matched the meaning of each successive sentence in the text; the format for Grade 3 involved answering multiple-choice questions after reading (longer) passages. We do not think the difference in format can fully explain our grade differences here because even the Grade 2 reading comprehension items required text integration. For example, children could not select the correct picture (e.g., “boy catching a bus” vs “boy catching a ball”) to match the meaning of a final sentence (‘He was just in time to catch it’) if they interpreted that sentence in isolation; they could only select the correct picture if they understood this in the context of the previous two sentences (“Dan’s bus was coming. Dan started to hurry.”). Furthermore, other work demonstrates that children and adults construct situation models even for two-clause sentences (e.g., Blything et al., 2015; Münte et al., 1998; Pyykkonen & Jarvikivi, 2012). However, the longer passage format for Grade 3 may involve greater application of comprehension monitoring

skills; longer passages provide context to support interpretation for children to deduce the correct meaning of an unknown word, when the information in the text does not align with their background knowledge, and when two pieces of information are difficult to integrate (Hogan et al., 2011). Future work should test the reproducibility of these findings across a range of grades and reading comprehension assessments and might also include multiple measures of reading comprehension in order to examine the relationship between comprehension monitoring and a latent variable in order to minimize measurement error.

Our second research objective concerned the longitudinal prediction of reading comprehension, and our findings add to our understanding of the longitudinal development of comprehension monitoring in relation to reading comprehension among bilingual learners. In terms of children's first language, English comprehension monitoring explained unique variance in English reading comprehension one year later, after controlling for vocabulary and word reading, including the autoregressor, in Grade 2. It could be argued that children's efficient decoding skills in their L1 (i.e., English) enable the expression of influence of higher-level oral language skills, such as comprehension monitoring, in line with the predictions of the simple view of reading. There may also be developments in the application of comprehension monitoring skills across development. Comprehension monitoring is evident in preschoolers (Skarakis-Doyle, 2002) and our study indicates that comprehension monitoring is still developing after several years of formal literacy instruction among bilingual readers, as has been shown for monolingual readers (LARRC & Yeomans-Maldonado, 2017). Furthermore, Grade 2 comprehension monitoring had a significant effect on Grade 3 reading comprehension beyond the autoregressor and control variables. This finding suggests that students' comprehension

monitoring in Grade 2 predicted unique gains in their reading comprehension from Grade 2 to Grade 3.

In terms of children's second language, no significant gains were observed in French comprehension monitoring from Grade 2 to Grade 3. French decoding skills continued to explain unique variance in French reading comprehension one year later, indicating that children's reading comprehension was still highly dependent on lower-level skills in the L2. It is somewhat surprising that French comprehension monitoring explained unique variance in children's reading comprehension in the concurrent analyses (i.e., Grade 3), but not longitudinally (i.e., from Grade 2 to Grade 3). A possible explanation for these results is that vocabulary plays a much stronger role in explaining reading comprehension in children's L2, presumably because there is more variability in the range of scores (see Table 1). Children who attend early French immersion programs are situated within a unique educational context since English is usually the home and societal language, but not the language of literacy instruction, with French input limited to the school environment. Since vocabulary acquisition plays a strong role in second language development, we infer that children in Grade 2 were not able to draw on higher-level oral language skills, such as the ability to monitor one's comprehension. Indeed, the correlation between French comprehension monitoring and French vocabulary increased from Grade 2 to Grade 3, suggesting that children had more capacity to activate higher-level oral language skills such as comprehension monitoring in Grade 3. Taken together, the longitudinal results suggest that, in the beginning stages of reading, children rely on oral language strategies developed in their stronger language (i.e., English). The results indicate that once children attain a certain level of proficiency in L2 vocabulary and decoding skills, they will be equipped to process language more deeply by involving higher-level skills, such as comprehension monitoring.

In addition to the limitations of the current study and directions for future research that have already been noted, we consider some additional critical methodological issues here. Based on our observations of skewness, the English comprehension monitoring task was relatively easy for our sample, and a notable proportion of children hit ceiling for the measure in Grade 2 (23% of the sample) and Grade 3 (50% of the sample). Such limited variability could have constrained the observed relationships between variables. Consequently, future work should consider developing texts across a greater range of difficulty, including inconsistencies that are less obvious and/or separated by several lines of filler text, as has been done in previous research (e.g., Oakhill et al., 2005). Likewise, future studies may consider developing a more comprehensive assessment that includes different passages with inconsistencies that vary in terms of emotional, causal, temporal or spatial information (see Wassenburg, Beker et al., 2015), and which include different text genres (Currie et al., 2021) among bilingual readers across diverse L1 and L2 combinations. Moreover, capturing real time comprehension monitoring with eye tracking methodology could be used to capture moment-to-moment processes in a natural reading context, without drawing the child's attention to the presence of inconsistencies in the text through instructions to evaluate the text.

With respect to educational implications, these findings highlight the importance of comprehension monitoring in early reading comprehension for L2 learners. Previous research has shown that comprehension monitoring is malleable and can be taught in the oral language context for monolingual prekindergartners from low socioeconomic backgrounds (Kim & Phillips, 2016; LARRC, Jiang & Logan, 2019), and is also effective for typically developing (National Institute of Child Health and Human Development, 2000) and struggling readers (Gersten et al., 2001). Support for comprehension monitoring of spoken texts may be particularly

useful for bilingual learners who have not developed good decoding skills in the L2. Comprehension monitoring can be supported through shared storybook reading and the modelling of monitoring through think-alouds, in which the parent or teacher can demonstrate how good readers are aware of the mental processes they use while listening and reading (Oster, 2001), and how to employ practical reading strategies upon encountering a breakdown in their understanding. Finally, one clear drawback to our study was that the measurement of reading comprehension changed from Grade 2 to Grade 3. To this end, researchers and psychologists should develop standardized reading measures and create norms for bilingual children in order to enhance educational assessments for this population.

To summarize, the present study adds to our understanding of the development of comprehension monitoring among English – French bilingual readers from Grade 2 to Grade 3. The concurrent results revealed that in Grade 3, children's comprehension monitoring served as a unique predictor of reading comprehension within English and French. Notably, the longitudinal analyses indicated that Grade 2 children's comprehension monitoring in English made a significant contribution to English reading comprehension in Grade 3. Taken together, the results from this study are essential to expanding theoretical models of reading comprehension development among bilingual learners and to foster the development of higher-level oral language skills in the early stages of reading.

References

Authors (2020)

Au-Yeung, K., Hipfner-Boucher, K., Chen, X., Pasquarella, A., D'Angelo, N., & Deacon, S.H. (2015). Development of English and French language and literacy skills in EL1 and EL French immersion students in the early grades. *Reading Research Quarterly*, *50*(2), 233-254. https://doi.org/10.1016/0022-0965(84)90127-9

Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology*, *38*(2), 289-311. https://doi.org/10.1016/0022-0965(84)90127-9

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(1), 3-11. https://doi.org/10.1017/S1366728908003477

Blything, L. P., Davies, R., & Cain, K. (2015). Young children's comprehension of temporal relations in complex sentences: The influence of memory on performance. *Child Development*, *86*(6), 1922-1934. https://doi.org/10.1111/cdev.12412

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & cognition*, *29*(6), 850-859. https://doi.org/10.3758/BF03196414

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of educational psychology*, *96*(1), 31-42. https://doi.org/10.1037/0022-0663.96.1.31

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children. *Language, Speech, and Hearing Services in Schools*, *32*, 38-50. https://doi.org/10.1044/0161-1461(2001/004)

Chall, J. S. (1983). *Learning to read: The great debate*. McGraw-Hill.

Currie, N. K., Francey, G., Davies, R., Gray, S., Bridges, M. S., Restrepo, M. A., ... & Cain, K. (2021). The process and product of coherence monitoring in young readers: Effects of reader and text characteristics. *Scientific Studies of Reading*, *25*(2), 141-158. https://doi.org/10.1080/10888438.2020.1831503

D'Angelo, N., Krenca, K., & Chen, X. (2020). The Overlap of Poor Reading Comprehension in English and French. *Frontiers in Psychology*, *11*,120. https://doi.org/10.3389/fpsyg.2020.00120

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific studies of Reading*, *6*(1), 51-77. https://doi.org/10.1207/S1532799XSSR0601_03

Dunn, L. M., and Dunn, D. M. (2007). *Peabody Picture Vocabulary Test*, 4th Edn. NCS, Pearson.

Dunn, L. M., Theriault-Whalen, C. M., and Dunn, L. M. (1993). *Échelle de Vocabulaire en Images Peabody (EVIP)*. Psycan.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906. https://doi.org/10.1037/0003-066X.34.10.906

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the

relationship in English. *Review of Educational Research*, *84*(1), 74-111. https://doi.org/10.3102/0034654313499616

Genesee, F., & Jared, D. (2008). Literacy development in early French immersion programs. *Canadian Psychology*, *49*(2), 140-147. https://doi.org/10.1037/0708-5591.49.2.140

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of educational research*, *71*(2), 279-320. https://doi.org/10.3102/00346543071002279

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial Special Education*, *7*, 6-10. https://doi.org/10.1177/074193258600700104

Hogan, T., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children*, *44*(3), 1-20. https://digitalcommons.unl.edu/specedfacpub/79

Jeong, J., McCoy, D. C., & Fink, G. (2017). Pathways between paternal and maternal education, caregivers' support for learning, and early child development in 44 low-and middle-income countries. *Early Childhood Research Quarterly*, *41*, 136-148. https://doi.org/10.1016/j.ecresq.2017.07.001

Johnson, R. K., & Swain, M. (1997). *Immersion education: International perspectives*, Cambridge University Press. https://doi.org/10.1017/CBO9781139524667

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of learning disabilities*, *47*(2), 125-135. https://doi.org/10.1177/0022219412439326

Kenny, D. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, *82*, 887–903. https://doi.org/10.1037/0033-2909.82.6.887

Kim, Y. S. G. (2015). Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading research quarterly*, *50*(4), 459-481. https://doi.org/10.1002/rrq.107

Kim, Y. S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of experimental child psychology*, *141*, 101-120. https://doi.org/10.1016/j.jecp.2015.08.003

Kim, Y. S. G., & Pilcher, H. (2016). What is listening comprehension and what does it take to improve listening comprehension?. In *Interventions in Learning Disabilities* (pp. 159-173). Springer, Cham.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, *95*(2), 163. https://doi.org/10.1037/0033-295X.95.2.163

Language and Reading Research Consortium & Yeomans-Maldonado, G. (2017). Development of comprehension monitoring in beginner readers. *Reading and Writing*, *30*(9), 2039-2067. https://doi.org/10.1007/s11145-017-9765-x

Language and Reading Research Consortium, Jiang, H., & Logan, J. (2019). Improving reading comprehension in the primary grades: Mediated effects of a language-focused classroom intervention. *Journal of Speech, Language, and Hearing Research*, *62*(8), 2812-2828. https://doi.org/10.1044/2019_JSLHR-L-19-0015

Lee, K., & Chen, X. (2019). An emergent interaction between reading fluency and vocabulary in

the prediction of reading comprehension among French immersion elementary students. *Reading and Writing*, *32*(7), 1657-1679. https://doi.org/10.1007/s11145-018-9920-z

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological bulletin*, *144*(4), 394. https://doi.org/10.1037/bul0000142

Lesaux, N. K., Crosson, A. C., Kieffer, M. J., & Pierce, M. (2010). Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. *Journal of applied developmental psychology*, *31*(6), 475-483. https://doi.org/10.1016/j.appdev.2010.09.004

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, *83*(404), 1198-1202. https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478722

MacGinitie, W. H., & MacGinitie, R. K. (1992). *Gates-MacGinitie Reading Tests*, 2nd Edn. Nelson Canada.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates- MacGinitie Reading Tests*, 4th Edn. Riverside Publishing.

Morrison, L. (2004). Comprehension monitoring in first and second language reading. *Canadian modern language review*, *61*(1), 77-107. https://doi.org/10.3138/cmlr.61.1.77

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*(6697), 71-73. https://doi.org/10.1038/25731

Naglieri, J. A. (1985). *Matrix Analogies Test – Short Form*. The Psychological Corporation.

National Reading Panel (US), National Institute of Child Health, Human Development (US), National Reading Excellence Initiative, National Institute for Literacy (US), & United States Department of Health. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*, 657-686. https://doi.org/10.1007/s11145-005-3355-z

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific studies of reading*, *16*(2), 91-121. https://doi.org/10.1080/10888438.2010.529219

Oster, L. (2001). Using the think-aloud for reading instruction. *The Reading Teacher*, *55*(1), 64-69. https://www.jstor.org/stable/20205012

Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders, 43,* 474–497. https://doi.org/10.1016/j.jcomdis.2010.01.002

Pyykkönen, P., & Järvikivi, J. (2012). Children and situation models of multiple events. *Developmental Psychology*, *48*(2), 521.

Skarakis-Doyle, E. (2002). Young children's detection of violations in familiar stories and emerging comprehension monitoring. *Discourse Processes*, *33*(2), 175-197. https://doi.org/10.1207/S15326950DP3302_04

Strasser, K., & Río, F. D. (2014). The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Reading Research Quarterly*, *49*(2), 169-187. https://doi.org/10.1002/rrq.68

Swain, M., & Lapkin, S. (2005). The evolving sociopolitical context of immersion education in Canada: Some implications for program development. *International Journal of Applied Linguistics*, *15*(2), 169-186. https://doi.org/10.1111/j.1473-4192.2005.00086.x

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6th ed.). Pearson Education.

van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, *25*(7), 1665-1690. https://doi.org/10.1007/s11145-011-9337-4

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *CTOPP-2: Comprehensive test of phonological processing* (2nd ed.). PRO-ED.

Wassenburg, S. I., Beker, K., van den Broek, P., & van der Schoot, M. (2015). Children's comprehension monitoring of multiple situational dimensions of a narrative. *Reading and writing*, *28*(8), 1203-1232. https://doi.org/10.1007/s11145-015-9568-x

Wassenburg, S. I., Bos, L. T., de Koning, B. B., & van der Schoot, M. (2015). Effects of an inconsistency-detection training aimed at improving comprehension monitoring in primary school children. *Discourse Processes*, *52*(5-6), 463-488. https://doi.org/10.1080/0163853X.2015.1025203

Wechsler, D. (2005). *Wechsler Individual Achievement Test* (2nd ed., Canadian). Pearson Canada Assessment.

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside Publishing.