

## Evaluating the Oral Language Skills of English-Stream and French Immersion Students: Are the CLB/NCLC Applicable?

Diana Burchell, Kathleen Hipfner-Boucher, Janani Selvachandran, Patricia Cleave and Xi Chen

Volume 23, Number 2, Fall 2020

Special Issue: The Canadian National Frameworks for English and French Language Proficiency: Application, Implication, and Impact  
Numéro spécial : Niveaux de compétence linguistique canadiens pour la compétence langagière en français et en anglais : impact, application et implication

URI: <https://id.erudit.org/iderudit/1072972ar>

DOI: <https://doi.org/10.37213/cjal.2020.30461>

[See table of contents](#)

Publisher(s)

University of New Brunswick

ISSN

1920-1818 (digital)

[Explore this journal](#)

Cite this article

Burchell, D., Hipfner-Boucher, K., Selvachandran, J., Cleave, P. & Chen, X. (2020). Evaluating the Oral Language Skills of English-Stream and French Immersion Students: Are the CLB/NCLC Applicable? *Canadian Journal of Applied Linguistics / Revue canadienne de linguistique appliquée*, 23(2), 118–140. <https://doi.org/10.37213/cjal.2020.30461>

Article abstract

This study examined the oral language skills of grade-two anglophone children enrolled in French Immersion and English-stream programs. The study had two objectives: (a) to compare performance between the groups on measures of receptive vocabulary, narrative comprehension, and narrative production (i.e., structure and language) in English, and (b) to explore the applicability of the Canadian Language Benchmarks/Niveaux de compétences linguistiques canadiens (CLB/NCLC) to assessment of their conversational competency. All children (English-stream n = 27, French Immersion n = 33, aged 7-8 years) were tested in English. In addition, the French Immersion students were tested using equivalent measures in French. The results comparing performance in English revealed no differences between the groups on receptive vocabulary, narrative comprehension and narrative structure. However, the English-stream children outperformed their French Immersion peers in narrative language. Furthermore, CLB/NCLC listening and speaking criteria were applied to conversational samples yielding level scores in English (both groups) and French (French Immersion only). The range of benchmarks that are appropriate for this population is discussed in detail.

Copyright (c) Diana Burchell, Kathleen Hipfner-Boucher, Janani Selvachandran, Patricia Cleave, Xi Chen, 2020



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

Érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

## Evaluating the Oral Language Skills of English-Stream and French Immersion Students: Are the CLB/NCLC Applicable?

Diana Burchell  
*University of Toronto*

Kathleen Hipfner-Boucher  
*University of Toronto*

Janani Selvachandran  
*University of Toronto*

Patricia Cleave  
*Dalhousie University*

Xi Chen  
*University of Toronto*

### Abstract

This study examined the oral language skills of grade-two anglophone children enrolled in French Immersion and English-stream programs. The study had two objectives: (a) to compare performance between the groups on measures of receptive vocabulary, narrative comprehension, and narrative production (i.e., structure and language) in English, and (b) to explore the applicability of the Canadian Language Benchmarks/*Niveaux de compétences linguistiques canadiens* (CLB/NCLC) to assessment of their conversational competency. All children (English-stream  $n = 27$ , French Immersion  $n = 33$ , aged 7-8 years) were tested in English. In addition, the French Immersion students were tested using equivalent measures in French. The results comparing performance in English revealed no differences between the groups on receptive vocabulary, narrative comprehension and narrative structure. However, the English-stream children outperformed their French Immersion peers in narrative language. Furthermore, CLB/NCLC listening and speaking criteria were applied to conversational samples yielding level scores in English (both groups) and French (French Immersion only). The range of benchmarks that are appropriate for this population is discussed in detail.

### Résumé

Cette étude compare les habiletés langagières orales en anglais et en français d'élèves anglophones de deuxième année, scolarisés en anglais ou en immersion française. L'étude avait deux objectifs : (1) comparer les performances des deux groupes d'enfants sur un test de vocabulaire réceptif, de compréhension narrative et de production narrative (c.-à-d., structure et langage narratifs) en anglais (2) tester l'applicabilité des Niveaux de compétences linguistiques canadiens (CLB/NCLC) à l'évaluation de la fluidité

conversationnelle chez les enfants en anglais et en français. Vingt-sept enfants scolarisés en anglais et 33 enfants scolarisés en immersion française, tous âgés de 7 à 8 ans, ont été évalués en anglais. Les enfants en immersion française ont aussi été évalués en français au moyen de tests de vocabulaire réceptif et de compétences narratives équivalents à ceux utilisés en anglais. Les résultats comparatifs en anglais ne révélaient aucune différence entre les deux groupes quant au vocabulaire réceptif et à la compréhension et la structure narratives en anglais. Cependant, la performance des élèves scolarisés en anglais était supérieure à celle des élèves en immersion française quant au langage narratif. De plus, les critères des CLB/NCLC portant sur la compréhension et l'expression orales ont été transposés aux échantillons de langage conversationnel recueillis, donnant des scores de niveau en anglais et en français. Enfin, la gamme de niveaux pouvant être transposés à cette population est discutée en détail.

### **Evaluating the Oral Language Skills of English-Stream and French Immersion Students: Are the CLB/NCLC Applicable?**

English-French bilingualism is a highly valued societal goal in Canada. (Paradis et al., 2011). For many Canadian children growing up in English-speaking families, competency in French as an additional language is achieved in school through French Immersion. French Immersion is predicated on the concept of additive bilingualism, the development of proficiency in a second language (L2) while maintaining proficiency in the first language (L1) (Genesee, 2004). Through early French Immersion programs, non-francophone children receive integrated language and content instruction in French beginning in kindergarten or grade 1 (age 5-6).

Studies conducted over the past forty years reveal the overall effectiveness of Early French Immersion in achieving its goal of additive bilingualism (Genesee & Jared, 2008; Lazurak, 2007). However, to our knowledge, no studies have compared broad oral discourse-level skills in English or French (conversational competency, narrative competence) among students in French Immersion and English-stream programs in the elementary grades. Moreover, recent data comparing vocabulary outcomes among students in French Immersion and English-stream programs indicate the need to revisit the question of English vocabulary outcomes among early French Immersion students (Au-Yeung et al., 2015; Hipfner-Boucher et al., 2014). The current study, therefore, was guided by two objectives: 1) to compare the outcomes of French Immersion and English-stream students in grade two on measures of English vocabulary and narrative competence; and 2) to investigate whether the Canadian Language Benchmarks/Niveaux de compétence linguistique canadiens (CLB/NCLC) criteria in English and French can be adapted to monolingual English and emergent English-French bilingual students in the early elementary grades.

We situate our study within a conceptual framework proposed by Cummins (1979; 2003) that distinguishes between basic interpersonal communication skills (BICS) and cognitive academic language proficiency (CALP). Whereas BICS refers to conversational competency, CALP refers to the academic language register (Cummins, 2003). BICS is the language of face-to-face conversation. It is contextualized, meaning that speakers and listeners draw on gesture, intonation, questioning, and feedback to interactively negotiate meaning in real-time (Sulzby, 1985). BICS has been shown to be important for social

integration both within and outside of school (Brown, 2004). We based our assessment of the children's conversational competency on the criteria laid out by the CLB/NCLC. We investigated benchmark levels achieved in conversational competency in English in native-English speaking students in French Immersion and English-stream classrooms as a means of qualitatively assessing the adaptability of the CLB/NCLC to these two populations. At the same time, we investigated benchmark levels achieved in conversational French among children in French Immersion. We expected that the CLB/NCLC would provide a useful tool for evaluating BICS in students as early as grade two, the children's third year of Immersion.

CALP, on the other hand, focuses on the academic language register (Cummins, 2003). The language of CALP is decontextualized, conveyed in the absence of real-world cues (Curenton & Justice, 2004). It is marked by the use of precise and elaborate vocabulary, procedures for making information and ideas linguistically explicit, and sophisticated syntactic markers that link propositions to one another. The acquisition of academic language has been shown to relate to school success (Storch & Whitehurst, 2002). This study compares performance on measures of English academic language between English speaking students in English-stream and French Immersion classrooms to examine the impact of French Immersion on L1 skills, with a particular focus on oral discourse-level skills. CALP was assessed on the basis of vocabulary, narrative comprehension, and narrative production.

## Review of the Literature

### Vocabulary

There is evidence that French Immersion programs promote vocabulary development in English and French over time (Swain et al., 1990; Cummins, 2001). Notably, while bilingual children's vocabulary may exceed that of their monolingual peers when the words in each of their lexicons are combined, it may be inferior to that of monolingual speakers in each of their languages taken separately (Barik & Swain, 1978). Much of the earlier research comparing English vocabulary outcomes of native-English speaking students in French Immersion to those of their English-stream peers reported that French Immersion students had significantly lower scores in grades 1 and 2. The finding was attributed to students' lack of formal schooling in English (Barik & Swain, 1978; Genesee, 1978). Data indicating that French Immersion students close the initial gap in scores following the introduction of English language arts instruction in the middle and late elementary grades support that suggestion (Barik & Swain, 1978).

More recent studies, however, revealed a somewhat different pattern of results. A study by Au-Yeung and colleagues (2015) looked at English-speaking students in French Immersion programs in senior kindergarten and grade 1. This study found that French Immersion students performed comparably to monolingual students on English receptive vocabulary from grade 1 to grade 3 as evidenced by the mean standard scores. These findings suggest that the English vocabulary skills of native-English speaking students in French Immersion may not be impeded by school-based instruction delivered exclusively in French. The discrepancy in findings related to vocabulary highlights the need to revisit the question. With this in mind, we compared the English

receptive vocabulary scores of native-English speaking children in French Immersion and English-stream programs in grade 2.

### **Narrative Competence**

The ability to comprehend and produce stories is referred to as narrative competence (Pellegrini & Galda, 1993). Narrative tasks yield rich data documenting children's discourse-level oral proficiency (Gagarina, et al., 2016). Generally, one of two techniques is used in the narrative literature to assess narrative comprehension. The first is retell, in which the child repeats a story following an initial narration (Spencer et al., 2019; Squires et al., 2014). Alternatively, narrative comprehension is assessed by requiring children to respond to a series of comprehension questions related to a story narrated by the experimenter (Roch et al., 2016; Rodina, 2017). The latter technique was used in our study.

Narrative production, on the other hand, is assessed by asking a child to generate a story in response to an illustration (Lever & Sénéchal, 2011; Luchero & Uchikoshi, 2019). In our study, the children's stories were assessed in terms of two dimensions: overall story structure and story language. Measurement of story structure generally focuses on inclusion of the discrete elements of story grammar (character, setting, problem, actions and reactions, resolution) (Mandler & Johnson, 1977; Stein & Albro, 1997). Story language, however, is assessed in different ways (e.g., number of different words [Uccelli & Páez, 2007], mean length of propositions [Muñoz et al., 2003]). In our study, story language was assessed in terms of complexity, i.e., the inclusion of temporal and causal conjunctions, grammaticality, inclusion of dialogue, and completeness and elaborative quality.

Research comparing L1 and L2 narrative outcomes within subjects largely attests to the invariance of story structure across languages (Bohnacker, 2016; Gagarina et al., 2015). From a theoretical perspective, the finding of invariance across a child's two languages is supported by Cummins' (1979) linguistic interdependence hypothesis, which suggests that higher-order elements of linguistic processing and organization are subject to cross-linguistic transfer. Story language, however, has largely been shown to be language-specific, hence less subject to cross-linguistic transfer (Berman, 2001; Pearson, 2002).

To our knowledge, no previous study has examined oral narrative competence within the context of French Immersion. Therefore, we compared outcomes achieved by students in English-stream programs to students in French Immersion on measures of narrative competence in English in grade 2. To assess narrative comprehension, all children were asked to respond to a series of literal and inferential comprehension questions based on a story narrated in English. In addition, the French Immersion children were asked to respond to comprehension questions related to a different story narrated in French. To assess narrative production, narratives were elicited in response to a single illustration by all children in English. The French Immersion children were also asked to generate a story in French in response to a different illustration.

### **Canadian Language Benchmarks (CLB/NCLC)**

Previous to the creation of the CLB, there was no Canadian framework available to measure language proficiency (Peirce & Stewart, 1997). As the Canadian population

diversified, the need was felt to measure listening, speaking, reading and written language proficiency to certify Canadian immigrants for the workforce. From this gap emerged the CLB (Peirce & Stewart, 1997). Originally released in 2000, the benchmarks were updated twice in French (2002, 2006) and once in English (2012) (CCLB, 2012). Currently, the benchmarks are used to place adult learners of English or French in language classes that are appropriate to their skill level. The benchmarks were created to assess BICS, the real-world communicative competencies on which CALP is founded (Cummins, 2003, Pawlikowska-Smith, 2002). The CLB framework evaluates four areas of language (listening, speaking, writing, reading) and five competencies (linguistic, textual, sociocultural, functional and strategic (Bachman et al., 2010).

While the benchmarks have been adapted multiple times for adults (Epp & Stawychny, 2001; Watt & Lake, 2000), their applicability to school-age children has not been tested. Yet they may provide educators with a common framework for evaluating conversational language competence. Fox and Courchène (2005) found that most evaluations by teachers focus on linguistic knowledge using task-based assessments, whereas the CLB focuses on communicative abilities using real-world tasks. Furthermore, Watt and Lake found that the listening and speaking components of the CLB were predicted by length of residence in the country, suggesting that the benchmarks might be sensitive to the Canadian immigration context. There is also some evidence that the CLB can act as a supplementary curriculum planning document in adult-level ESL classrooms (Fleming, 1998, 2014), suggesting its potential applicability to French L2 learners. In the current study, we extended the CLB/NCLC to both English monolingual and French Immersion students in grade 2.

### **The Present Study**

This study examined the oral language skills of grade 2 children in English-stream and French Immersion programs. The objectives of the study were: (a) to compare the two groups on English receptive vocabulary and narrative competence, key indicators of academic language essential to school success (Cummins, 2003), and (b) to explore the applicability of the CLB/NCLC to children in English-stream and French Immersion programs. To our knowledge, our study is the first to systematically examine the discourse-level oral language proficiency of French Immersion students and the first to evaluate children's conversational competency using the benchmarks. Conversational skills provide the foundation on which academic language is built (Cummins, 2003) but methods of evaluating them are lacking.

To assess conversational competency, children were engaged in one-on-one conversations with an experimenter on a topic of their choice. Their conversational samples were then scored using CLB/NCLC criteria that were modified to be developmentally appropriate. Assessment criteria in the domain of grammatical knowledge needed little modification. With respect to textual knowledge, we assessed cohesion within topics of conversation and not across the conversation as a whole. Functional knowledge was evaluated in terms of imagination and ideation only since manipulation and heuristics were deemed to be developmentally inappropriate (Schneider, 2008). Likewise, strategic knowledge was deemed developmentally inappropriate (Carr et al., 1989), and was not

evaluated. The children were therefore evaluated on grammatical, functional, sociolinguistic and textual knowledge.

## **Methods**

### ***Participants***

A total of 60 children in Grade 2 were recruited from 10 public schools in a linguistically and culturally diverse metropolitan area of Canada as part of a larger project. The sample consisted of 27 children in English-stream programs ( $N = 94.59$  months/7.9 years,  $SD = 7$  months) and 33 children in French Immersion programs ( $N = 92.36$  months/7.8 years,  $SD = 6$  months). The children in English-stream programs received school instruction exclusively in English. The French Immersion children had been instructed exclusively in French since entering Immersion in senior kindergarten. All students were native English speakers, and in the families where another language was spoken, it was less than 10% of the time. The average level of parental education was post-secondary.

### ***Measures and Procedure***

A battery of language and cognitive measures was administered to all children. Conversation samples were also collected to measure conversational competency. Children in the English-stream program were tested in English only; children in the French Immersion program were tested in both English and French. The students were tested at their schools by trained research assistants who were fluent in English and/or French. Children completed testing individually with a research assistant in a quiet environment outside the classroom.

#### ***ALEQ: Demographic Questionnaire***

All parents filled out a demographic questionnaire adapted from the Alberta Language Environment Questionnaire (ALEQ, Paradis et al., 2010). They reported on the following topics: the home language environment, extra-curricular activities, parental language, socioeconomic status (as indicated by parental education levels), sibling interaction and immigration history.

#### ***Nonverbal Reasoning***

The Reasoning by Analogy subtest of the Matrix Analogies Test (Naglieri, 1985) was chosen to assess nonverbal reasoning ability. This subtest contained 16 items of increasing difficulty. The child was asked to choose one of six patterns that best completes a given matrix. Testing was discontinued after four consecutive errors.

### ***Receptive Vocabulary***

The Peabody Picture Vocabulary Task, Fourth Edition (PPVT-IV) (Dunn & Dunn, 2007) was administered to all students to assess receptive vocabulary in English. The test consisted of 228 items of increasing difficulty. The child was asked to select one of four illustrations that best depicted a stimulus word presented orally by the experimenter. Testing was discontinued when the child made 8 errors in a set of 12 items. A parallel French measure, the *Échelle de vocabulaire en images Peabody* (EVIP) (Dunn et al., 1993) was administered to all French Immersion students. Since this test was normed on a francophone population, we ignored the basal rule and started all students at Item 1. The test included 170 items of increasing difficulty. Testing was discontinued when the child made 6 errors in a set of 8 items. Scores represent the number of correct responses.

### ***Narrative Comprehension and Production***

A shortened version of the Test of Narrative Language (TNL, Gillam & Pearson, 2004) was used to assess narrative skills in English. The English version was adapted from the Spanish version of the TNL (Gillam et al., 2006). For our purposes, it was translated into French by the research team. The TNL includes two subtests: narrative comprehension and narrative production. In the story comprehension test, students were shown a picture as the experimenter narrated a story related to it. They then answered 6 literal and 7 inferential questions worth 1 or 2 points each. In the story production task, students were shown a picture and asked to make up a story about it. The students were scored on 18 story structure items worth 1 point each, and 6 story language items worth 2 points each (for a maximum of 30 points). The same testing protocol was followed in French. The TNL has two alternate versions (Form A and Form B), which were counterbalanced in both the English and French Immersion children. The stories were audio-recorded for later transcription and scoring.

### ***Conversational Samples and Canadian Language Benchmarks***

Conversational samples were collected through one-on-one conversations between experimenter and child. Conversations were 11-12 minutes in duration. Research assistants were trained to steer conversations away from narratives or expository topics (e.g., book plots or how to play soccer). Instead, conversation topics focused on family, school, friends, hobbies and pets. More specifically, we chose to focus on Task Type C from both the beginner and intermediate stages of the listening/speaking skills. Task Type C is defined as “Takes part in short informal conversation about personal experience” for the beginner stage and “Discusses concrete information on a familiar topic” for the intermediate stage (Peirce & Stewart, 1997, p. 22-23). We chose to include prompts from both stages in response to concerns that not all students would progress through the levels in a linear fashion (Pierce & Stewart, 1997). For this reason, as well as the variation in proficiency in our sample, we gave prompts which would allow us to place students between level 1-8. For context, the benchmarks are broken down into three categories: beginner (levels 1-4), intermediate (levels 5-8) and advanced (levels 9-12) (Gauthier, 2019). All conversations were recorded, transcribed, and coded according to CLB criteria.



All transcripts were coded by at least two research assistants. Inter-rater reliability was 95% in French and 93% in English.

## Results

Table 1 shows the descriptive statistics for the French Immersion and English-stream children. Raw scores are presented for all measures. Three subscores are reported for the TNL: story comprehension, story structure and story language. Two items in the original scoring scheme for the production task measured story coherence and completeness. Since these items are better aligned with story structure (Botvin & Sutton-Smith, 1977), they were included as part of the story structure subscore in our analysis, bringing the maximum scores for story language to 8 and for story structure to 22. Standard scores were calculated for English and French receptive vocabulary. No multivariate outliers were found using the Mahalanobis distance (McLachlan & McLachlan, 1999). No univariate outliers were found through a boxplot examination. All variables were normally distributed based on skewness and kurtosis statistics. The CLB levels for the two groups are reported in Table 2.

Initial analyses were carried out to compare the French Immersion children and English-stream children on parent education and non-verbal reasoning. No significant differences were found ( $t(58) = -.699$ ,  $p = .487$  for parent education, and  $t(58) = .764$ ,  $p = .448$  for nonverbal reasoning), suggesting the two groups were well matched. As displayed in Table 1, children in both groups performed well on English receptive vocabulary. Standard scores reveal that the French Immersion students scored more than one standard deviation above the population mean. The English-stream students also scored above the population mean. Means scores for English narrative comprehension were 8 and 8.5 for the English-stream and French Immersion children, respectively. With respect to narrative production, the mean score for story structure in both groups was 9, whereas, for story language, the mean scores were 5 and 4 for the English-stream and French Immersion students, respectively. Thus, the raw scores show the two groups with very similar means in English.

With respect to French receptive vocabulary, the French Immersion students scored approximately two standard deviations below the mean. This is not surprising given that this test was normed on francophone students. Mean scores for French narrative comprehension, narrative structure, and narrative language were 4, 5, and 2 respectively. Thus, the French Immersion children scored higher in English (their L1) than they did in French on all narrative measures.

**Table 1***Descriptive Statistics of the Measures organized by Language Program (n = 60)*

Measures	English-stream			French Immersion		
	Min-Max	<i>M</i>	<i>SD</i>	Min-Max	<i>M</i>	<i>SD</i>
1 PAR_ED	1-4	2.6	.79	0.5 – 4	2.7	.92
2 MAT	0 - 10	3.6	3.2	0 - 13	3.6	3.2
3 PPVT	112 - 175	142.1	20.4	112 - 181	146.4	17.7
4 PPVT_SS	85 - 141	110.5	16.7	87 – 146	117.2	14.8
6 TNLE_C	4– 12	8.48	1.91	3 – 13	8.03	2.40
7 TNLE_S	3 – 20	8.96	3.24	3 – 18	9.48	4.25
8 TNLE_SL	2 – 7	4.74	1.43	1 – 6	3.52	1.35
11 EVIP	-	-	-	12 - 96	48.3	21.2
12 EVIP_SS	-	-	-	40 - 111	70.6	18.1
14 TNLF_C	-	-	-	0 – 9	4.03	2.60
15 TNLF_S	-	-	-	1 – 8	4.78	2.17
16 TNLF_SL	-	-	-	0 – 5	1.78	1.72

*Note.* MAT= Non-verbal reasoning; PAR\_ED= Parental education (socioeconomic status); PPVT = English receptive vocabulary raw scores; PPVT\_SS = English receptive vocabulary standard scores; TNLE\_C= English TNL Comprehension; TNLE\_S= English TNL Story Structure; TNLE\_SL= English TNL Story Language; EVIP = French receptive vocabulary raw scores; EVIP\_SS = French receptive vocabulary standard scores; TNLF\_C= French TNL Comprehension; TNLF\_S= French TNL Story Structure; TNLF\_SL= French TNL Story Language.

**Table 2***Descriptive Statistics of the Canadian Language Benchmarks (CLB/NCLC) by Language Group (CLB n = 45; NCLC n = 18)*

Measures	English-stream			French Immersion		
	Min-Max	<i>M</i>	<i>SD</i>	Min-Max	<i>M</i>	<i>SD</i>
9 CLB_C	3 – 6	4.5	.82	3 – 8	6	1.39
10 CLB_P	3 - 6	4	.88	3 - 8	6	1.51
17 NCLC_C	-	-	-	1 – 6	4	1.56
18 NCLC P	-	-	-	1 - 6	4	1.53

*Note.* CLB\_C = English Canadian Language Benchmarks Comprehension; CLB\_P= English Canadian Language Benchmarks Production; NCLC\_C = French Canadian Language Benchmarks Comprehension; NCLC\_P = French Canadian Language Benchmarks Production.

### **Quantitative Analysis: English Oral Language**

An ANCOVA was conducted on English receptive vocabulary with group as a between-subject factor (French Immersion vs. English-stream) and parental education and non-verbal reasoning as covariates. No significant difference was found between the two groups,  $F(1, 58) = 2.1, p = .110$ . A MANCOVA was run to determine the effect of group as a between-subject factor (French Immersion vs. English-stream) on English narrative comprehension and narrative production (story structure, story language). Parental

education and nonverbal reasoning were again included as covariates. There was a significant difference on English story language in favour of the English-stream group,  $F(3, 58) = 3.458, p < .05$ . In contrast, no significant differences were found on story structure,  $F(3, 58) = .640, p = .593$ , or narrative comprehension,  $F(3, 58) = .307, p = .820$ .

### ***Qualitative Analysis: Applying the Canadian Language Benchmarks (CLB)***

Conversation samples were collected for 20 students in French Immersion programs and 25 students in English-stream programs. The listening and speaking criteria of the CLB/NCLC were applied to the conversation data. Given that the applicability of the CLB/NCLC to this population has not previously been empirically studied, this study is exploratory in its approach. Research assistants were interviewed after coding the conversational data in order to evaluate the strengths and challenges encountered in applying this assessment to children (see Table 3). Furthermore, since neither floor nor ceiling effects were observed, there is preliminary evidence suggesting that the CLB may be a useful indicator of second-grade children's language proficiency. Inter-rater reliabilities were 89% and 92% for listening and speaking respectively.

**Table 3**

#### *Description of Applicability by Coders*

Component	Strengths	Challenges
Range of Levels	Since there were few levels and they were clearly defined, it was not overly difficult to place students. This allowed for high inter-rater reliability.	As students are developing their language together at school, they fall into a small section of the benchmarks (especially the English students).
Evaluating Speaking Elicited Conversation	Given that the conversation was student-focused and lasted for 10-12 minutes, there was ample material to evaluate speaking skills.	Our elicitation task allowed the students to choose the conversational topics. Depending on the topic they chose, children may not have the opportunity to demonstrate competency in certain domains of knowledge.
Evaluating Listening through Elicited Conversation	Research assistants reported that it was typically easy to tell from the children's responses whether they understood the questions.	Some children may provide few verbal cues to indicate confusion, which can lead to bias. For example, shy students may not ask for clarification. Listening scores were more difficult to determine.
English CLB	Students clearly varied in their level of receptive and expressive skills. Scoring in English was easier than in French.	While students at this age fit into the criteria easily, they may soon outgrow the benchmarks given that they are native speakers. RA's thought older students may hit ceiling.
French NCLC	Our L2 learners clearly fit the criteria.	Language anxiety could be a confounding factor here.
Training	The initial training was not too overwhelming and could be achieved in a one-day PD session.	Research assistants mentioned the importance of a second opinion to help determine some children's benchmark level.

Tables 4 and 5 describe the benchmark levels observed in English and French, respectively. Please note that, in the interest of concision, the listening and speaking criteria are presented together. The column on the left, labelled “description,” is a summary of common descriptors found at each level for this population. The column on the right, labelled “examples,” lists short excerpts from conversations that best illustrate the description. These levels were determined relative to L1 adult proficiency. Among French Immersion students, listening and speaking levels ranged from 3-8 (beginner-intermediate) in English, with a mean of 6, meaning that the average French Immersion student had an intermediate-level mastery of English, described as “moderately complex.” Performance among the English-stream students ranged from levels 3-6 (beginner-intermediate) with a mean of 5 in listening and 4 in speaking. This means that the average English-stream student straddled the line between beginner and intermediate, between “simple” and “moderately complex.” The levels ranged from 1-6 (beginner-intermediate) in French, with a mean of 4. This means that the average French Immersion student was still in the upper beginner stage.

**Table 4**

*Description of English CLB Benchmarks applied to Grade 2 French Immersion and English-stream Children*

Level	Description: English	Examples: English
3	Participant sometimes has to ask for clarification from the examiner and responds with simple phrases. Participant makes grammatical errors and does not always stay on topic. Participant makes little or no inferences and does not engage in imaginative scenarios.	E do you have any other favourite subjects in school besides gym? C I like swimming. (ID228)
4	Participant makes grammatical errors but generally stays on topic. Participant makes little to no inferences and may try to engage in imaginative scenarios. Participant makes some sociocultural references.	E how about your mom? C (um) she likes cooking. E what else does she like? C (um) she likes cooking food. (ID226)
5	Participant understands moderate questions and can respond appropriately. Participant may mention a few ideas but does not go into genuine imagination. Participant makes one or two sociocultural references and breaks cohesion a few times. Participant may make some grammatical errors.	E two hundred? that is really really high. C I do that every day when I fall asleep. E does that help? C I (d*) only do it sometimes when I am very bored and do not have anything to do. (ID234)
6	Participant is engaged with the examiner and can elaborate on questions asked. When responding, participant starts to venture into ideational and imaginal topics. Participant makes some sociocultural references and is mainly cohesive while speaking on a particular topic. Participant may make a few, small grammatical errors.	E yes, tell me about your brother. C my brother, he likes sharks. and (um) he plays this game called hungry sharks. and he keeps talking to me about what his favourite shark is. and it keeps changing every single day. and I sometimes get annoyed from him because he keeps

		talking about that game. (ID 1016)
7	Participant can clearly understand and elaborate on complex questions such as future aspirations. Participant makes little or no grammatical errors and shows moderate imagination and ideation. Participant makes many sociocultural references and shows cohesion in their chosen topics.	E so do you want to be something like your parents when you grow up? like an artist or something like that? what do you want to be? C I want to be an actor. I have been talking to my dad about going to theatre camp. (ID607)
8	Participant understands and is clearly engaged in a number of complex topics. Can give complex and clear directions. Participant shows complex ideas and imagination and makes little or no grammatical errors. Participants show cohesion throughout the conversation and refer to sociocultural items consistently.	E do you like Christmas? E Halloween? C (: ) Christmas a little. because I get some time to fake sleep. and I am really good at it. So, I can make my eyes close. and if Santa Claus touches me. or tries to tickles me with a feather or anything. I will not laugh. because (um) I keep my (um : ) head close.

*Note.* C=Child, E=Examiner. Please note that all samples are broken into utterances, not sentences.

**Table 5**

*Description of French NCLC and English CLB Benchmarks applied to Grade 2 French Immersion Children*

Level	Description: French	Examples: French
1	Le participant utilise un mot pour répondre à la question ou ne répond pas à la question. Il ne comprend que les phrases les plus simples et il utilise souvent des mots en anglais. Il fait peu de références socioculturelles ou imaginatives.	E on doit avoir une conversation en français alors de quoi aimerais-tu parler? C (um). E qu'est-ce que tu aimes à propos de l'école? C l'école (: ) what?
2	Le participant donne des réponses courtes à des questions. Il demande des éclaircissements à l'examineur et il doit parfois faire ces éclaircissements en anglais. Le participant fait souvent des erreurs grammaticales. Il fait peu de références socioculturelles ou imaginatives.	E qu'est-ce que tu as ? C un chat. E un chat ? il s'appelle comment? C Sam.

3	Le participant doit parfois demander des éclaircissements à l'examineur. Pour répondre, le participant utilise des phrases simples. Le participant fait des erreurs grammaticales et ne reste pas toujours sur le sujet. Le participant fait peu ou pas d'inférences et ne s'engage pas dans des scénarios imaginatifs. Le participant fait quelques références socioculturelles.	E As-tu des animaux à la maison ? C (um:) chien. E Il s'appelle comment ? C Pepper. E Est-ce que c'est un petit chien ou un grand chien ? C un petit chien.
4	Le participant fait des erreurs grammaticales mais reste généralement sur le sujet. Le participant fait peu ou pas d'inférence et peut essayer de s'engager dans des scénarios imaginatifs. Les participants font quelques références socioculturelles.	E Dormir ? Alors les samedis tout ce que tu fais, tu dors? C non j'[je ai les activités j'ai les>. E Quelles activités ? C les[ew] natation et le ballet.
5	Le participant comprend les questions complexes et peut répondre de manière appropriée. Le participant mentionne quelques idées, mais montre une imagination limitée. Le participant fait une ou deux références socioculturelles et rompt quelquefois la cohésion. Le participant peut faire des erreurs grammaticales.	E qu'est-ce que tu as fait pendant le weekend ? C alors ma cousine a viens de Wales et c'est son première fois de venir au Canada. C et on est allé au chalet de le ami de mon papa.
6	Le participant est engagé et peut développer les réponses complexes. En répondant, le participant commence à s'aventurer sur des sujets idéaux et imaginaires. Le participant fait quelques références socioculturelles et il est principalement cohérent en parlant sur un sujet particulier. Le participant peut faire quelques erreurs grammaticales.	C parce que tu ne joues pas dans la classe. seulement (c'est comme) sur le recreé. et (tu) tu es dans tes poches. mais tu joues pas dans les classes. tu peux jouer dans la recreé avec. (mais pas) mais pas dans les classes. (cause pas* parce) parce que beaucoup de personnes a fait ça dans mon classe.

*Note.* C=Child, E=Examiner. Please note that all samples are broken into utterances, not sentences.

### ***Examples: Further Analysis***

Three examples are presented below to illustrate the application of the CLB/NCLC to students in French Immersion and English-stream programs. We evaluated their grammatical, textual, functional and sociolinguistic competencies in English and French. Each student was chosen because they obtained the average score for their respective group in English or French. As our intent was to illustrate average benchmark performance, we discuss only performance in the language in which the child achieved the mean score. In all

three examples, we expand upon their performance in other areas in order to give a more complete picture of what a typical student would look like.

***Example 1: Paulina (English: English-Stream)***

Paulina (pseudonym), an average English-stream student who spoke only English at home, scored a 5 on listening and a 4 on speaking on the English benchmarks. Paulina made only a few grammatical errors that did not interfere with sentence comprehension. For example, she omitted a conjunction in the sentence, “because like he stays home for one day, he leaves.” Paulina also had few problems with cohesiveness (i.e., textual knowledge). For example, when asked about her family, Paulina started by saying “my sister,” but did not finish her thought without prompting from the examiner. With respect to functional knowledge, Paulina described highly imaginative scenarios twice (i.e. saying that goggles, which fell down a deep well months previously, might still be falling). Paulina engages in few cultural references as most of her conversation is based on personal experience which limited use of sociolinguistic knowledge. Paulina was fairly close to the average on all the other measures with the exception of receptive vocabulary, on which she scored well above average (127 versus 110). However, as previously mentioned, she engaged in highly speculative scenarios which was a particular strength of hers.

***Example 2: Zahra (English: French Immersion Stream)***

Zahra (pseudonym), a French Immersion student who spoke only English at home, scored a 6 (the overall average score for French Immersion students) on both English benchmarks. In English, Zahra demonstrated grammatical knowledge, and any errors made were small enough that they did not impede the evaluator’s understanding. For example, when talking about Spiderman, she said, “he got bit.” With respect to textual knowledge, Zahra shows a general ability to cohesively tie utterances together. There are some occasions where cohesion is lacking (jumping from talking about spells to her being on “the train” with no clear connection between the two), but they are minor in that they do not impede comprehension. In terms of functional knowledge, Zahra displays some imagination in her description of her favourite characters or TV shows, but her conversation is mainly descriptive. Conversely, Zahra’s cultural references can be found throughout this conversation and therefore placed her high in sociolinguistic knowledge. Overall, Zahra did very well in grammatical and sociolinguistic knowledge. She scored slightly lower on functional and textual knowledge. Zahra was fairly close to the average on all measures with the exception of English receptive vocabulary, on which Zahra scored 129 (M = 117). The sociolinguistic references were Zahra’s biggest strength throughout the conversation.

***Example 3: Brandon (French: French Immersion Stream)***

Brandon’s (pseudonym) score of 4 on the French benchmarks was our average. His first language was English, and he spoke no other languages at home. Brandon has been enrolled in a French Immersion program since senior kindergarten. Brandon made grammatical mistakes throughout the conversation but was still able to communicate

meaning. For example, when talking about a trip to Cuba, he said “on a allé” instead of “on est allé.” Confusion over which auxiliary verb to use is common among young French Immersion students. Generally speaking, grammatical knowledge is weak among French Immersion students given their low level of exposure to the language. For textual knowledge, Brandon’s utterances were always cohesive. For example, when speaking about his trips to Cuba and Florida, he consistently stayed on subject. He did not, however, elaborate without further prompting. In terms of functional knowledge, Brandon’s answers were mainly descriptive and did not delve into underlying meanings. For example, when talking about his vacation, he could describe an activity but did not engage in any imaginative scenarios. Brandon also did not make many cultural references but was able to discuss topics such as his favourite subject at school or why his family went to Florida in the winter. Overall, Brandon’s strengths were textual and sociolinguistic knowledge. He performed less well on grammatical and functional knowledge in this task. Finally, he occasionally used English words to supplement his descriptions, especially when talking about the attributes of his family members. This is likely due to limited French vocabulary, as shown on his receptive score: 66 ( $M = 70$ ). For the other variables, Brandon was a bit below average. This code-switching is not typically addressed by the CLB/NCLC and will be discussed below.

### Discussion

The first objective of this study was to evaluate the oral language proficiency of grade 2 children enrolled in English-stream and French Immersion programs. We compared the English performance of the two groups on receptive vocabulary and narrative comprehension and production. No significant differences were found on English receptive vocabulary or English narrative comprehension. With respect to narrative production, no difference was found on story structure, but the English-stream students outperformed their French Immersion peers on story language. On French receptive vocabulary, the French Immersion students scored about two standard deviations below the mean on standard scores. This, however, was expected because the test was normed on native French-speaking children. The second objective was to explore whether the CLB can be used to evaluate conversational skills of Grade 2 English-speaking children in both the English-stream and French Immersion. Among English-stream children, CLB scores fell into the beginner-intermediate stage, ranging from levels 3-6 with a mean of 5 in speaking and 4 in listening. Among French Immersion children, scores ranged between 3-8 in English with a mean of 6 (beginner-intermediate) and 1-6 with a mean of 4 (beginner-intermediate) in French. Furthermore, the examples provided show that four out of the five competencies (grammatical, textual, sociocultural and functional) may be applicable to young children.

Our finding that the children in French Immersion programs performed similarly to those in English-stream programs on English receptive vocabulary deviates from earlier research reporting an initial gap in English vocabulary (Barik & Swain, 1978; Genesee, 1978). However, it aligns with more recent studies revealing similar performance between the two groups (Hipfner-Boucher et al., 2014). The French Immersion children’s strong performance on English vocabulary, despite their reduced exposure to the language, may be attributed to a number of factors. First, French Immersion programs are typically attended by children from families of high socioeconomic status (SES, Allen, 2004;



Makropoulos, 1998). Although the two groups of children in our study did not differ on parent education level, an indicator of SES, enrolling in French Immersion programs may still involve a self-selective process. Children are more likely to be enrolled in an Early Immersion program by their parents if they are perceived to have strong language skills (Turnbull et al., 2003). Relatedly, the attrition rate of French Immersion programs remains high (Sinay et al., 2018). Children with language and/or reading difficulties tend to be transferred to English-stream programs, leaving stronger children in French Immersion programs over time (Turnbull et al., 2003). Notably, French Immersion programs have become much more diverse in recent years, both in family SES status and the language spoken at home (Sinay et al., 2018). Future studies need to examine the performance of students from diverse backgrounds in the French Immersion program.

Aside from self-selection factors, we argue that cross-language transfer also plays a key role in French Immersion children's strong performance on English measures. The interdependence hypothesis (Cummins, 1979, 1991) postulates that for bilingual children, proficiency developed in one language can transfer to support learning in the other language. While extensive research has demonstrated transfer of language and literacy skills from L1 to L2, recent studies also provide evidence for transfer in the other direction (Chung et al., 2019). English and French are closely related in many linguistic features, including vocabulary, morphology, and syntax. Thus, for French Immersion children, linguistic knowledge and metalinguistic awareness acquired in French can be transferred to support vocabulary learning in English (Burchell, 2019; Sohail et al., 2019). This cross-language transfer, combined with daily exposure to English at home and in the broader community, may explain why French Immersion children acquire English vocabulary at the same rate as their peers in English-stream programs. Finally, the societal context of French Immersion programs plays an important role in the student's language competency. French Immersion programs are only attended by non-Francophone students and are usually located in anglophone communities. This means that, outside of the classroom, students are completely immersed in English. Other studies have found that trilingual students perform well in English despite being in French Immersion, partially due to their in-depth exposure to English in the community (Dagenais & Day, 1999). The availability of English stimulus outside of the classroom allows for a more successful additive bilingual model.

Our study examined discourse-level oral language competence, in addition to word-level knowledge. Receptive skill was assessed on the basis of a narrative comprehension task. Overall, the French Immersion and English-stream students responded comparably to a series of literal and inferential questions gauging story comprehension following a narration in English. Our finding suggests that receptive discourse-level skill in native-English speaking children is not impeded as a result of participation in a French Immersion program. We speculate that story comprehension skills initially developed in English within the context of the home are maintained by the French Immersion children through interaction with English speakers outside of school. Indeed, stories are a pervasive feature of children's lives and much of their interaction with peers and family is story-based (Roch et al., 2016). In particular, narrative skills have been shown to be practiced and honed in the home through parent-to-child storybook reading (Curenton et al., 2008; Levy et al., 2006). Performance on our measure of narrative comprehension suggests that French Immersion students may benefit from ongoing exposure to stories told in English in the

home. Furthermore, the supra-lexical receptive skills assessed by our narrative comprehension task are likely to be founded in part on English word knowledge.

Our findings were somewhat more nuanced with respect to narrative expressive skill. A story production task was used to elicit child narratives that were assessed in terms of story structure and story language. We found no differences between native-English speaking children in French Immersion and their English-stream peers on measures of story structure. Overall, the stories narrated by both groups of students made equal mention of characters and setting, a precipitating event that motivated characters to act, as well as subsequent actions and reactions and their consequences. Again, we speculate that the French Immersion children maintained the expressive narrative skills they initially developed in English through interactions with peers and family outside of school. At the same time, we know from past research that the structural organization of stories is relevantly invariant across languages (Gagarina et al., 2015), leaving it subject to cross-linguistic transfer. We know, too, that stories are a prevalent feature of the classroom (Cummins, 2008). Therefore, it is possible that the stories told in French in school enhanced the French Immersion children's awareness of macro-level story structure, further supporting their storytelling ability in English through the mechanism of cross-linguistic transfer.

Whereas the French Immersion children demonstrated performance comparable to their English-stream classmates on our measure of story structure, a significant difference was found between the groups favouring the English-stream students on story language. Story language was assessed in terms of linguistic complexity (on the inclusion of temporal and causal conjunctions, on grammaticality, and on inclusion of dialogue). It may be the case that acquisition of the semantic and morpho-syntactic elements that comprise story language are highly dependent on language exposure. French Immersion students in the early primary years may still be in the process of mastering the precise language required to construct a cohesive narrative in English without the benefit of English language instruction – and storytelling – in school. Moreover, previous research suggests that story language is language-specific, hence less subject to cross-linguistic transfer than story structure (Pearson, 2002). Thus, the development of story language in native-English speaking children may be impacted by participation in a French Immersion program. Future research is needed to track the development of skill on this dimension of narrative competence among French Immersion students to determine if and when they fill the performance gap. At the same time, we must keep in mind that story language was assessed on the basis of a small number of items in the current study. Therefore, our findings need to be replicated by future studies using more comprehensive measures.

Clear findings emerged in this study related to the French skills of French Immersion students. On the French test of receptive vocabulary, students performed two standard deviations below the norm. Low standard scores were expected given that the task was normed on Francophone students. French Immersion students also achieved lower scores on the French narrative measure than they did in English. Notable features of the French narratives were numerous grammatical errors and code-switching. That said, it was clear that students were able to communicate basic ideas after just two years of French-language instruction. The purpose of French Immersion programs is to promote additive bilingualism (i.e., proficiency in L2 at little or no cost to L1). This study shows that overall, students had gained beginner-level proficiency in their L2 while maintaining L1

performance levels that were largely comparable to their monolingual peers in the domains of vocabulary and narrative competence.

Finally, we investigated whether the CLB/NCLC criteria could be applied to French Immersion and English-stream students in the early primary years. Conversational samples were elicited, using Task Type C from the benchmark protocols, as a means to evaluate basic interpersonal communication skills (Cummins, 2003). In this study, we took a first step toward establishing a range of performance levels for grade 2 students in English and French. As indicated in the results, our ability to place all students within the framework levels without hitting floor or ceiling levels was a first indicator of applicability.

In addition to our quantitative analysis, we provided qualitative data to describe the conversational abilities of children who achieved average levels of performance on the CLB/NCLC. We discussed the English performance of an English-stream student (Paulina), the English performance of a French Immersion student (Zahra), and the French performance of a French student (Brandon). We found that grammatical knowledge was strong in English but was a notable challenge in French for French Immersion students. Textual knowledge was judged on the basis of local cohesiveness as the researchers felt it would be unfair to rate children on the cohesiveness of an entire conversation. Functional knowledge (ideas and imagination) was largely case-dependent as the student chose the topic of discussion as well as the way in which that topic was discussed. However, the coders did find that the students with higher language skills tended to demonstrate greater functional knowledge (e.g., two of our examples discussed vacations, but one was descriptive, and the other was imaginative). Similar to functional knowledge, evidence of sociocultural knowledge was largely dependent on the student's chosen topic as well as their hobbies. That said, those with better language skills were often able to better demonstrate this knowledge in their conversations. Both functional and sociocultural knowledge seem to be developmentally appropriate for children of this age.

### **Limitations and Future Directions**

This study provides preliminary evidence supporting the application of the CLB/NCLC to young students enrolled in English-stream and French Immersion programs. It used a conversational elicitation measure which presented particular limitations. Chief amongst these is the fact that children chose conversational topics that may have restricted opportunity to demonstrate competency in certain domains of knowledge. Future studies should employ tasks that allow children to demonstrate competency across the range of knowledge domains. Furthermore, the same conversational elicitation task was used to evaluate listening and speaking. Future studies should not only assess these skills separately but also include assessments of reading and writing based on CLB/NCLC benchmark criteria.

Discussions with the coders responsible for scoring the conversational samples indicated that application of the benchmarks to child language was not without its challenges. The most substantial challenge in using these benchmarks with children related to the age-appropriateness of the CLB benchmark criteria. Because children in the early primary years are in the process of developing the full range of cognitive abilities that support language competency (Zelazo et al., 2008), adjustments to current CLB scoring criteria that reflect age-appropriate expectations are recommended before this framework

could be fully implemented with children. It would also be useful to explore whether the current range of benchmarks is appropriate, or whether sub-levels would be needed when working with children. This is due to the natural delays that may occur in children's language development that does not happen in adults.

One of the particular challenges in applying this framework to French Immersion studies relates to code-switching. In its current form, the CLB/NCLC does not take code-switching into consideration in its assessment of conversational competency. However, research has shown that code-switching is a notable feature of French Immersion students' speech (Lin, 2013; Turnbull, et al., 2011). Moving forward, it would be important to discuss what role, if any, translanguaging might play in this framework as it applies to children in French Immersion programs.

For the present study, only students in Grade 2 were included. Future studies are needed to establish CLB levels for students across the elementary grades. Finally, this study included only native English speakers although we know that English-mainstream and French Immersion classrooms in Canada are becoming increasingly diverse (Swain & Lapkin, 2005). Future research assessing conversational competency among English language learners using the CLB/NCLC would be useful for teachers, as it would provide them with a framework for assessing language abilities. For example, the benchmarks may be useful in comparing English Language Learners to their L1 English peers in both English-stream and French Immersion programs to track development of conversational competency.

Correspondence should be addressed to Diana Burchell.

Email: [diana.burchell@mail.utoronto.ca](mailto:diana.burchell@mail.utoronto.ca)

## References

- Allen, M. (2004). Reading achievement of students in French immersion programs. *Educational Quarterly Review*, 9(4), 25-30.
- Au-Yeung, K., Hipfner-Boucher, K., Chen, X., Pasquarella, A., D'Angelo, N., & Hélène Deacon, S. (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.
- Bachman, L. F., Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Barik, H. C., & Swain, M. (1978). Evaluation of a French immersion program: The Ottawa study through Grade five. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 10(3), 192.
- Berman, R. (2001). Crosslinguistic perspectives on narrative development. Epilogue. In L. Verhoeven & S. Strömquist (Eds.), *Narrative development in a multilingual context* (pp. 419-428) John Benjamins.
- Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37, 19-48.
- Botvin, G. J., & Sutton-Smith, B. (1977). The development of structural complexity in children's fantasy narratives. *Developmental Psychology*, 13(4), 377.

- Brown, C. L. (2004). Content based ESL curriculum and academic language proficiency. *The Internet TESL Journal*, 10(2), 1-6.
- Burchell, D. M. (2019). *The Direct and Indirect Effects of Syntactic Awareness on French Reading Comprehension in French Immersion Students* [Master's thesis, University of Toronto]. <https://tspace.library.utoronto.ca/handle/1807/97062>
- Carr, M., Kurtz, B. E., Schneider, W., Turner, L. A., & Borkowski, J. G. (1989). Strategy acquisition and transfer among American and German children: Environmental influences on metacognitive development. *Developmental Psychology*, 25(5), 765.
- Centre for the Canadian Language Bookmarks (2012). *CLB: ESL for Adults*. <https://www.language.ca/product/pdf-e-01-clb-esl-for-adults-pdf/>
- Chung, S. C., Chen, X., & Geva, E. (2019). Deconstructing and reconstructing cross-language transfer. *Journal of Neurolinguistics*, 50, 149-161.
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 198-205.
- Cummins, J. (1991). Interdependence of first-and second-language proficiency in bilingual children. *Language processing in bilingual children*, 70-89.
- Cummins, J. (2001). The entry and exit fallacy in bilingual education. In C. Baker & N.H. Hornberger (Eds.). *An introductory reader to the writings of Jim Cummins* (pp. 110–138). Multilingual Matters.
- Cummins, J. (2003). Basic interpersonal communicative skills and cognitive academic language proficiency. *BICS and CALP*.
- Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education*, 2<sup>nd</sup> Edition, Volume 2: Literacy. (pp. 71-83). Springer Science + Business Media LLC.
- Curenton, S.M., Craig, M.J., & Flanagan, N. (2008). Use of decontextualized talk across story contexts: How oral storytelling and emergent reading can scaffold children's development. *Early Education & Development*, 19, 161-187.
- Curenton, S. & Justice, L. (2004). African American and Caucasian preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech & Hearing Services in Schools*, 35, 24-253.
- Dagenais, D., & Day, E. (1999). Home language practices of trilingual children in French immersion. *Canadian Modern Language Review*, 56(1), 99-123.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.
- Dunn, L. M., Dunn, L. M., & Theriault-Whalen, C. (1993). *EVIP: échelle de vocabulaire en images Peabody*. Pearson Canada Assessment.
- Epp, L., & Stawychny, M. (2001). Using the Canadian Language Benchmarks (CLB) to Benchmark College Programs/Courses and Language Proficiency Tests. *TESL Canada Journal*, 18(2), 32-47.
- Fleming, D. (1998). Autonomy and agency in curriculum decision-making: A study of instructors in a Canadian adult settlement ESL program. *TESL Canada Journal*, 19 -35.
- Fleming, D. (2014). A “complicated conversation” with the Canadian language benchmarks. *Citizenship Education Research Journal (CERJ)*, 4(1), 41-55.
- Fox, J., & Courchêne, R. (2005). The Canadian language benchmarks (CLB): A critical

- appraisal. *Contact*, 31(2), 7-28.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balciuniene, I. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N.Meir (Eds.), *Language impairment testing in multilingual settings* (pp. 241–274). Benjamins.
- Gagarina, N., Klop, D., Tsimpli, I. M., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics*, 37(1), 11-17.
- Gauthier, M.-E. (2019). "Overview of CLB and NCLC Competency Levels." *Centre for Canadian Language Benchmarks*. [www.language.ca/overview-of-clb-and-nclc-competency-levels/](http://www.language.ca/overview-of-clb-and-nclc-competency-levels/).
- Genesee, F. (1978). A longitudinal evaluation of an early immersion school program. *Canadian Journal of Education/Revue canadienne de l'éducation*, 31-50.
- Genesee, F. (2004). What do we know about bilingual education for majority language students? In Tej K. Bhatia & William C. Ritchie (Eds.), *Handbook of bilingualism and multiculturalism*, 547–576. Blackwell.
- Genesee, F., & Jared, D. (2008). Literacy development in early French immersion programs. *Canadian Psychology/Psychologie Canadienne*, 49(2), 140-147.
- Gillam, R. B., & Pearson, N. A. (2004). *TNL: Test of Narrative Language*. Pro-ed.
- Gillam, R. B., Peña, E. D., Bedore, L. M., & Pearson, N. (2006). Test of Narrative Language, Spanish Adaptation Experimental Version (TNL-S). *Unpublished test*.
- Hipfner-Boucher, K., Lam, K., & Chen, X. (2014). The effects of bilingual education on the English language and literacy outcomes of Chinese-speaking children. *Written Language & Literacy*, 17(1), 116-138.
- Lazaruk, W. (2007). Linguistic, academic, and cognitive benefits of French immersion. *Canadian Modern Language Review*, 63(5), 605-627.
- Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology*, 108, 1–24.
- Levy, B. A., Gong, Z., Hessels, S., Evans, M. A., & Jared, D. (2006). Understanding print: Early reading developments and the contributions of home literacy experiences. *Journal of Experimental Child Psychology*, 93, 63-93.
- Lin, A. (2013). Classroom code-switching: Three decades of research. *Applied Linguistics Review*, 4(1), 195-218.
- Luchero, A., & Uchikoshi, Y. (2019). Narrative assessments with first grade Spanish-English emergent bilinguals: Spontaneous versus retell conditions. *Narrative Inquiry*, 29, 137-156.
- McLachlan, G. J., & McLachlan, G. J. (1999). Mahalanobis distance. *Resonance*, 4(06).
- Makropoulos, J. (1998). *Sociopolitical analysis of French immersion developments in Canada* (Doctoral dissertation, National Library of Canada/Bibliothèque nationale du Canada).
- Mandler, J.M., & Johnson, N.S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111-151.
- Naglieri, J. A. (1985). *Matrix Analogies Test: Short Form*. Psychological Corporation.
- 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(6), 474-497.

- Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning* (2<sup>nd</sup> Edition). Brookes.
- Pawlikowska-Smith, G. (2002). *Canadian Language Benchmarks 2000: Theoretical Framework*.
- Pearson, B.Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D.K. Oller & R.E. Eilers (Eds.), *Language and Literacy in Bilingual Children*. (pp. 135-174). Multilingual Matters.
- Peirce, B. N., & Stewart, G. (1997). The Development of Canadian Language Benchmarks Assessment. *TESL Canada Journal*, 17-31.
- Pellegrini, A.D. & Galda, L. (1993). Ten years after: A reexamination of symbolic play and literacy research. *Reading Research Quarterly*, 28, pp. 162-175.
- Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian-English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37, 49-67.
- Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *International Journal of Bilingualism*, 21, 617-635.
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, 2(3), 114-121.
- Sinay, E., Presley, A., Armson, S., Tam, G., Ryan, T.G., Burchell, D., & Barron, C. (2018). *Toronto District School Board French as a second language program review: Developmental evaluation*. (Research Report No. 18/19-03). Toronto, Ontario, Canada: Toronto District School Board.
- Sohail, J., Duncan, T., Koh, P., -W., Deacon, H., & Chen, X. (2019). How syntactic awareness might influence reading comprehension in English-French bilingual children. *Submitted to the Journal of Research in Reading*.
- Spencer, T., Petersen, D., Restrepo, M., Thompson, M., Arvizu, M. (2019). The effect of Spanish and English narrative intervention on the language skills of young dual language learners. *Topics in Early Childhood Special Education*, 38, 204-219.
- Squires, K., Lugo-Neris, M., Peña, E., Bedore, L., Bohman, T., Gillam, R. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Communication Disorders*, 49, 60-74.
- Stein, N., & Albro, E. (1997). Building complexity and coherence: Children's use of goal-structured knowledge in telling stories. In M. Bamberg (Ed.), *Narrative development: Six approaches* (pp. 5 – 44). Mahwah, NJ: Lawrence Erlbaum Associates.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental psychology*, 38(6), 934.
- Sulzby, E. (1985). Children's emergent reading of favourite storybooks: A developmental study. *Reading Research Quarterly*, 20, 458-481.
- 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.
- Swain, M., Lapkin, S., Rowen, N., & Hart, D. (1990). The role of mother tongue literacy in third language learning. *Language, Culture and Curriculum*, 3(1), 65-81.
- Turnbull, M., Cormier, M., & Bourque, J. (2011). The first language in science class: A

- quasi-experimental study in late French immersion. *The Modern Language Journal*, 95, 182-198.
- Turnbull, M., Hart, D., & Lapkin, S. (2003). Grade 6 French immersion students' performance on large-scale reading, writing, and mathematics tests: Building explanations. *Alberta Journal of Educational Research*, 49(1).
- Uccelli, P. & Páez, M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech and Hearing Services in Schools*, 38, 225- 236.
- Watt, D. L., & Lake, D. M. (2000). *Canadian Language Benchmarks-TOEFL Research Project: A Comparison Study of the Canadian Language Benchmarks Assessment and the Test of English as a Foreign Language*.
- Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). *The development of executive function in childhood*. (pp. 553-574). MIT Press.