The Development of Lexical Complexity in Sixth-Grade Intensive English Students

Mémoire

Mélissa Thériault

Maîtrise en linguistique de l'Université Laval
offert en extension à l'Université du Québec à Chicoutimi

Maître ès arts (M.A.)

Département des arts et lettres
Université du Québec à Chicoutimi
Chicoutimi, Canada

Faculté des lettres et des sciences humaines
Université Laval
Québec, Canada

RÉSUMÉ

L'étude avait pour but d'examiner le développement de la complexité lexicale chez des élèves de sixième année du primaire (n = 56) participant à un programme d'anglais intensif d'une durée de cinq mois. Plus précisément, grâce à une épreuve de narration réalisée au début et à la fin du programme, l'étude a permis de déterminer à quel point un programme intensif favorise le développement de la complexité lexicale. De plus, l'étude a permis de comparer les progrès réalisés par deux groupes d'apprenants, soit un groupe fort et un groupe faible, pour déterminer si le niveau de compétence initial influence le développement de la complexité lexicale d'un apprenant. Les principaux résultats ont montré un développement de la complexité lexicale chez tous les apprenants, sans toutefois établir de différence entre le progrès fait par l'un ou l'autre des groupes.

ABSTRACT

The present study investigated the development of lexical complexity (LC) in sixth-grade intensive English (IE) students (n = 56) over the course of a five-month program. More specifically, the study analysed oral narratives produced at two points in time (at the beginning and at the end of the intensive period) in order to determine to what extent the program promotes the development of LC. Moreover, the study also compared the progress made by low-level and high-level students to examine whether lexical development differs as a result of differences in initial levels of second-language (L2) proficiency. Findings revealed improvement in LC for all students over the course of the five-month program, but no clear advantage for either of the two proficiency groups.

TABLE OF CONTENTS

RÉSUMÉ iii

ABSTRACT v

TABLE OF CONTENTS vii

ACKNOWLEDGEMENTS ix

1. INTRODUCTION AND PROBLEM STATEMENT 1

2. RESEARCH QUESTIONS AND HYPOTHESES 7

2.1. Research Questions 7

2.2. Hypotheses 7

3. THEORETICAL FRAMEWORK 9

3.1 Lexical complexity 9

3.1.1 Conceptualization of lexical complexity 9

3.1.2 Measures of lexical complexity 11

3.2 Intensive English programs 14

4. LITERATURE REVIEW 17

4.1 Intensive English in Quebec 17

4.2 Development of lexical complexity 20

5. RESEARCH METHOD 23

5.1 Learning context 23

5.2 Participants 23

5.3 Task 24

5.3.1 Vocabulary test 24

5.3.2 Oral narration 24

5.3.3 Procedure for oral narration 26

5.4 Measures of lexical complexity 27

5.4.1 Lexical density 27

5.4.2 Lexical diversity 28

5.4.3 Lexical sophistication (use of less-frequent words) 28

5.4.4 Reliance on French 29

5.5 Procedures for data analyses 30

5.6 Statistical analysis 32

6. RESULTS 33
6.1 Development of lexical complexity from T1 to T2 33
6.1.1 Lexical density 34
6.1.2 Lexical diversity 34
6.1.3 Lexical sophistication 35
6.1.4 Reliance on French 36
6.1.5 Overall observations 36
6.1.6 New measures of lexical complexity 38
6.1.7 Overall development of lexical complexity 40
6.2 Differences in the development of low-proficiency and high-proficiency learners 40
6.2.1 Results at T1 and T2 41
6.2.2 Gains in lexical complexity from T1 to T2 42
6.2.3 Vocabulary test 43
6.2.4. Overall differences between groups 44
6.3. Summary of results 44
7 Discussion 47
7.1 Development of lexical complexity over the course of a five-month IE program 48
7.2 Differences in the development of low-proficiency and high-proficiency learners 50
7.3 Pedagogical implications 53
7.4 Limitations 55
8. CONCLUSION 59
8.1. Further research 60
REFERENCES 61
Appendix A 67
Appendix B 71
Appendix C 75

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all the people who have guided, supported and helped me throughout this project. First and foremost, many thanks to my thesis director, Dr. Leif French, whose knowledge and guidance played an invaluable role in this endeavour; his expert advice made it possible for me to complete this ambitious project. I would also like to thank Mr. Jean-Daniel Guay for his help in my first foray into the world of statistical analysis. Thank you to the Social Sciences and Humanities Research Council, as well as Université du Québec à Chicoutimi for the financial support. Last but not least, many thanks go out to my family, friends and colleagues who were there to encourage and support me every step of the way. Thank you to my parents who taught me to believe in myself and who encouraged me throughout this project. Special thanks also to Ms. Shannon Dacres for her expert proofreading skills.

1. INTRODUCTION AND PROBLEM STATEMENT

In the 1970s, French immersion programs were developed in Canada (Billy, 1980). The programs gave English-speaking students the opportunity to receive part of their daily instruction for all subject matters in French, thus increasing the amount of French they were exposed to. Given the success of French immersion throughout Canada, the programs were transposed to the second-language context of Quebec, where English is most often learned as a second language (Lightbown, 2012). However, since Quebec legislation prevents any form of content-based instruction (e.g., math, science, history, etc.) from being given in a language other than French, intensive English (IE) programs were adapted for the context and focused solely on the teaching of English (Spada & Lightbown, 1989). IE programs generally give students the opportunity to receive their regular, content-based instruction (in French) in a condensed form over half the school year and focus exclusively on English for the rest of the year. This particular model of IE, along with newer models that differ in terms of the distribution of intensive instruction, have proved successful over the years and are now offered regularly in most school boards throughout Quebec.

To date, numerous studies have been carried out in IE contexts in Quebec, and there is now a considerable body of evidence showing that these programs are quite beneficial to students, enabling them to make substantial progress in basic English communication skills over a short period of time (Collins, Halter, Lightbown & Spada, 1999; Lightbown & Spada, 1994; White & Turner, 2005). More specifically, research has shown that IE students perform considerably better than their peers who follow a regular track program, both in reading and listening comprehension tasks (Lightbown & Spada, 1994). In terms of speech production, IE students are found to be more fluent and confident than regular track students when expressing themselves orally (Lightbown & Spada) and are also more coherent and produce longer stretches of language when telling stories (Spada & Lightbown, 1989; White & Turner, 2005). As for vocabulary knowledge, IE students have also been shown to use a wider range of expressions and vocabulary items (Spada & Lightbown, 1989; White & Turner, 2005). Moreover, recent research (Collins & White, 2011) examining word recognition suggests that students are actually familiar with approximately three quarters of the 1000 most frequent words of the English language at the end of an intensive program. Moreover, research has shown that students who participate in an IE program and spend only five months covering the regular fifth- or

sixth-grade curriculum succeed just as well in regular subject areas, such as French and mathematics, as students in regular programs who devote the full ten months of the school year to those subjects (Lightbown & Spada, 1991; 1997; Spada & Lightbown, 1989). Thus, by these accounts, it is clear that intensive programs have a positive impact on various aspects of students' second-language (L2) learning.

Interestingly, one area of L2 learning to receive little research attention in intensive programs is the development of learners' oral lexical complexity (LC)[1]. One of the only studies to look directly at students' lexical development in an IE context was conducted by Horst and Collins (2006). They used Vocabprofile, a software program for lexical frequency profiling (LFP), to examine sixth-grade students' development of written LC at four 100-hour intervals. The main LFP findings suggested that students' LC actually decreased over time; however, four other proposed measures of LC (reliance on French words, reliance on cognates, morphological variety and number of K1 families[2]) showed significant development in LC. This prompted the authors to question the validity of Vocabprofile and its resulting LFP measure for young learners, as the frequency lists used by the software were based on adult language and did not necessarily reflect the frequent and infrequent language the students were exposed to on a daily basis. The authors suggested that the use of more age-appropriate frequency lists as well as other measures of complexity might be better suited for illustrating the nuances in students' progress.

In addition, there appear to be conflicting results when using LFP with other young learners in similar intensive/immersive contexts. For instance, while Horst and Collins's (2006) initial results suggested that using LFP was inconclusive with IE students, Lo and Murphy (2010) also used LFP through Vocabprofile with young learners and found significant development in students' LC. Vermeer (2004)

[1] Lexical complexity, as defined by Bulté and Housen (2012), refers to "the degree of elaboration, the size, the breadth, width or richness of the L2 learner's [lexical knowledge]" (p. 25). The term "lexical richness" (Daller et al., 2003; Horst and Collins, 2006; Laufer, 1994; Laufer and Nation, 1995; Vermeer, 2000, 2004) and "lexical diversity" Jarvis, 2002; McKee et al., 2000) is also used in various studies to refer to the concept of lexical complexity.

[2] Number of K1 families represents the number of word families comprises within the 1000 most frequent of the language

also showed that the Measure of Lexical Richness[3] (MLR), which is closely related to LFP, is indeed a valid measure of young beginning learners' lexical richness. The discrepancy in findings points to the importance of conducting further research to assess whether LFP can indeed be used to illustrate IE students' development of LC. It also highlights the relevance of exploring the use of other measures of LC in intensive contexts to supplement the results provided by studies using LFP as a measure.

An important factor that may influence learning outcomes over the course of an intensive program is students' initial level of proficiency. In terms of language knowledge, Collins and White (2012) suggest that learners in a wide range of levels stand to make substantial gains over the course of an intensive program. Their study followed sixth-grade students enrolled in a five-month intensive program to see whether increased amount and concentration of time could benefit students with a wide range of abilities. Although the results of the study did not demonstrate a clear relationship between the results obtained from the pretest and overall gains, teachers reported that by the end of the program, differences between the weaker and stronger students were much less apparent.

In 1995, Lapkin, Hart and Swain investigated the hypothesis according to which greater linguistic gains are made by learners with lesser linguistic abilities. They studied the development of English-speaking Canadian high-school students participating in a three-month exchange program in Quebec. Although test results from the most proficient students had to be excluded from the study because of a ceiling effect, the results of the study supported the hypothesis that the least proficient students were the ones who made the greatest gains over the course of the immersion program. This conclusion was supported both by test results and student self-assessments.

[3] The Measure of Lexical Richness (MLR) is a lexical measure similar to LFP developed by Vermeer (2004). MLR is an indicator of the complexity of language produced, and calculates the degree of difficulty of a word according to its frequency of use in the language. The frequency lists are based on spontaneous, oral speech produced by children from kindergarten to sixth grade, making MLR particularly well suited to the analysis of children's speech.

It is true that the intensive context presented in that study differs from the immersion context studied by Lapkin et al. (1995), where, rather than being placed in a native-speaking community, the students were in contact with the L2 five days a week for five months. Nevertheless, the fact that Lapkin et al. found initial levels of proficiency to be predictors of linguistic gains, coupled with the fact that Collins and White (2012) found IE programs to be beneficial for students in a wide range of linguistic abilities, points to the pertinence of investigating whether a relationship might also exist between the students' initial level of proficiency and their lexical development over the course of the intensive program.

Finally, another important element to consider is the type of task used to investigate the development of LC. Freed, So and Lazar (2003) suggest that fluency does not develop the same way in oral and written language. Although the development of vocabulary was not directly investigated in their study, vocabulary richness was one of the most important factors in determining written fluency. The idea that the vocabulary used in oral and written language may differ underlines the importance of considering the impact of the type of task used for analysis. Horst and Collins (2006) investigated the development of LC in written tasks produced by IE students and found that although it was not indicated through LFP, progress in LC was indeed made over the course of an intensive program. Thus, Freed et al.'s (2003) suggestion that vocabulary is used differently in oral and written tasks leads to speculation about whether similar development would be observable through an oral task and emphasizes the necessity of further investigating the development of LC in IE students through oral tasks.

Although much research has been conducted in IE contexts, showing that IE programs are beneficial to students, enabling them to make considerable gains in their L2 over a short period of time (Collins et al., 1999; Lightbown & Spada, 1994; Spada & Lightbown, 1989; White & Turner, 2005), little research has dealt directly with the lexical development of IE students. Consequently, investigating the development of LC in IE students has both theoretical and practical implications for the future. On a theoretical level, it will contribute to furthering our knowledge of how IE students' L2 skills develop, as well as how the large quantity of input to which they are exposed on a daily basis is transformed into productive knowledge. On a practical level, it can contribute to informing teachers and various

stakeholders in the L2 community on how students acquire lexical knowledge in an intensive context, enabling them to adapt and tailor their teaching practices and materials in order to ensure that students get the most out of the intensive experience.

2. RESEARCH QUESTIONS AND HYPOTHESES

2.1. Research Questions

Given the limited amount of research that has been conducted on oral LC in intensive contexts, it is important to carry out additional research that examines how LC develops in students and how other factors, such as students' initial level of LC, influence that development. The present study will examine the development of LC and will investigate the following research questions by analyzing the oral narratives of sixth-grade learners both at the beginning and end of a five-month intensive ESL program:

1. Does the oral lexical complexity of sixth-grade intensive English students improve over the course of a five-month program?
2. Does the development of lexical complexity differ between low- and high-level learners over the course of the five-month program?

2.2. Hypotheses

The above research questions also give rise to the following two hypotheses:

1. With respect to question 1, it is hypothesized that IE students will indeed show development in LC over the course of a five-month program. IE programs are shown to have a positive impact on several aspects of students' L2 development (Collins et al., 1999; Lightbown & Spada, 1994; White & Turner, 2005), including gains in vocabulary (Spada & Lightbown, 1989). While Horst and Collins (2006) did not find traces of development through LFP, other measures led them to conclude that there was progress in students' LC over the course of a five-month program. Research has also shown that immersion contexts favour the development of LC in young learners (Bournot-Trites, 2007; Lo & Murphy, 2010). The results therefore lead researchers to believe that IE programs could promote gains in lexical abilities, including the development of LC.

2. With respect to question 2, it is hypothesized that there will indeed be a difference in the progress made by learners of different levels, with the most significant progress being made by the least proficient students. In a study involving high-school students participating in a linguistic exchange program, Lapkin et al. (1995) reported that less proficient students appeared to make the most progress. Collins and White (2012) found that students from a wide range of proficiency levels stand to make substantial progress when participating in an IE program, and teachers reported that by the end of the program, all students were "more or less at the same level" (p. 60). The fact that the students who were less proficient at the beginning of the program seemed to catch up to those who were more proficient adds weight to the hypothesis that the weaker students stand to make greater gains than the ones who appear to be the strongest at the beginning of the intensive program.

3. THEORETICAL FRAMEWORK

3.1 Lexical complexity

3.1.1 Conceptualization of lexical complexity

In the language-learning literature, lexical complexity (LC) has been referred to using a variety of terms, such as *lexical diversity* (Jarvis, 2002; Malvern & Richards, 1997; McKee, Malvern & Richards., 2000), *lexical richness* (Daller, Van Hout & Treffers-Daller, 2003; Horst & Collins, 2006; Laufer, 1994; Laufer & Nation, 1995; Ovtcharov, Cobb & Halter, 2006; Read, 2000; Tidball & Treffers-Daller, 2008; Vermeer, 2000, 2004) and *lexical proficiency* (Harley & King, 1989). The wide variety of terminology used, coupled with the fact that, as Butlé and Housen (2012) point out, there is no commonly accepted definition for this concept, can lead to confusion.

As Bulté and Housen (2012) note, many studies do not give a clear definition of the terms, which makes ensuring uniformity in the interpretation of the concepts rather difficult. These observations emphasize the importance of clearly defining the concept of lexical complexity in order to enable a more correct and precise interpretation and comparison of results and conclusions across a number of studies. OK

Bulté and Housen (2012) present the concept of linguistic complexity as a vast system that involves many aspects of language. The subcategory of linguistic complexity that is of interest here is lexical complexity. The authors define this concept as "the degree of elaboration, the size, breadth, width, or richness of the learner's L2 system or 'repertoire', that is, to the number, range, variety or diversity of different structures and items that he knows or uses . . . " (p. 25). The definition refers to the complexity of the lexical knowledge the learner has acquired to date rather than the complexity of the language itself. The authors further divide LC into three observable and measurable constructs, which can be used to assess LC: lexical density, lexical diversity and lexical sophistication.

Lexical density can be expressed by two measures: the ratio between the number of lexical words (nouns, adjectives, verbs and adverbs) and the number of function words (pronouns, prepositions, conjunctions, articles and auxiliary verbs), and the ratio between the number of lexical words and the total number of words (Bulté & Housen, 2012). A high ratio indicates a lexically dense text; the higher the ratio, the more lexical words are contained in the text, in comparison either to function words or to the total number of words in the text.

Lexical diversity, as defined by Bulté, Housen, Pierrard and Van Daele (2008), refers to the extent of the learner's lexical knowledge, or the number of different words he or she knows and uses. It can be expressed as the number of different words or types a learner uses, or it can be expressed as a ratio, such as the Type-Token Ratio (TTR), which is calculated by dividing the number of different words by the total number of words in the text (Bulté et al.). Read (2000) also uses the term *lexical variation* to refer to this concept.

Bulté et al. (2008) define lexical sophistication as "the perception of a L2 user's lexical proficiency formed by, among other things, his use of semantically more specific and/or pragmatically more appropriate different words from among a set of related words" (Bulté et al., p. 279). Lexical sophistication can also be considered as the proportion of less-frequent words a learner uses in a text. Frequency lists, which divide words of a given text into lists according to their frequency of occurrence in the language (for example, the 1000 most-frequent word families, the 1001-2000 most-frequent word families, etc.), are used to determine the proportion of less-frequent words used by the learner. Vermeer (2004) establishes a relationship between the degree of difficulty of a word and its order of acquisition. She emphasizes the fact that adults and children learn language differently, and states that in young learners, the frequency of a word can be used to define its degree of difficulty. A greater proportion of less-frequent words therefore indicates a more lexically sophisticated text.

The three aspects of lexical complexity (density, diversity and sophistication) combine to paint a complete picture of the complexity of a learner's lexical knowledge. The development of a learner's

LC can be observed through variation, over a period of time, in the measures used to represent these three aspects.

3.1.2 Measures of lexical complexity

Many different measures have been used to assess lexical complexity. As Daller et al. (2003) point out, the consensus among researchers appears to be that there is a need for investigating lexical knowledge; however, how to investigate this lexical knowledge remains an area of dispute. Indeed, there is no commonly accepted measure of lexical complexity, and the validity of some measures is contested by other researchers. The most frequently used measure of lexical complexity appears to be the Type-Token Ratio (Vermeer, 2000). This measure is calculated by dividing the number of types (different words) by the number of tokens (total number of words) in a given text. However, the validity of this measure has been criticized by many researchers (such as Daller et al., 2003; Jarvis, 2002; Laufer & Nation, 1995; Malvern & Richards, 1997; McKee et al., 2000; Serrano Serrano, 2011; Vermeer, 2000), mainly because of its dependence on text length.

One of the main problems in assessing lexical complexity is the influence of text length. On the one hand, a longer text appears to give learners more opportunity to use unique words, making the longer text more lexically complex; on the other hand, a longer text also means greater opportunity for repetition, as the pool of words to choose from becomes more restricted as the text lengthens. TTR is a measure that is directly influenced by text length: as the text lengthens, TTR decreases, giving the text the appearance of being less lexically rich (Malvern & Richards, 2002). The comparison between texts of different lengths through TTR is therefore rendered invalid.

In order to overcome the problem of TTR's dependence on text length, mathematical transformations have been developed as an alternative. The Guiraud Index is one of those. It is calculated by dividing the number of types by the square root of the number of tokens (Daller et al., 2003). Although it appears to be a more valid measure than TTR (Daller et al., 2003; Vermeer, 2000), the Guiraud Index is similarly criticized for its dependence on text length. Vermeer (2000) claims that as the text

lengthens, the measure will first increase, then reach a maximum and begin to decrease. The fact that the measure only seems valid for texts of a certain length prompted Daller et al. (2003) to question its validity for more advanced learners.

Collentine (2004) did not use the traditional measure of TTR or one of its variations to assess LC, but used the number of unique occurrences per 1000 words of text. Similar to TTR in that it provides insight as to the number of different words a learner is able to produce, this scaled measure made it possible to compare texts of different lengths, as the comparison was made on equivalent proportions of text. The method assessed the proportion of unique occurrences according to seven lexical categories (adjectives, adverbs, conjunctions, nouns, prepositions, pronouns, and verbs) and was adopted in order to discriminate between the four primary parts of speech (nouns, verbs, adverbs and adjectives), which constitute the learner's core lexical base, and more complex parts of speech, which are used to mark discursive coherence. A high number of unique occurrences, or a high ratio in the case of TTR, indicates a lexically diverse text.

In 1995, Laufer and Nation proposed a new measure to assess lexical richness: lexical frequency profiling (LFP). This measure categorizes words of a given text into separate lists according to their frequency of occurrence in the language; thus, the proportion of less-frequent (and therefore more complex) words illustrates the lexical complexity of the text. They set out to prove the reliability and validity of LFP, and their analyses led them to conclude that it was a reliable and valid measure of lexical richness. However, in 2005, Meara put forth a critical analysis of Laufer and Nation's (1995) use of LFP and questioned the validity and sensitivity of their results. He used a set of simulations (Monte Carlo analysis) to examine the validity of LFP in estimating productive vocabulary size. He concluded that while LFP may allow researchers to distinguish between groups of learners whose vocabulary sizes are very different, it is not as sensitive with groups that have small differences in vocabulary size. He also reported that the sensitivity of LFP seemed to decline with larger vocabulary sizes. This critique was addressed by Laufer (2005), who attempted to clarify the matter by explaining that LFP is a measure of lexical use in writing, rather than a measure of vocabulary size. She maintained that the reason LFP did not distinguish between groups showing only slight differences in

vocabulary size was that those learners do not use vocabulary differently. She argued that all areas of lexical competence do not develop the same way, and that an increase of 500 words in vocabulary size did not necessarily lead to a change in vocabulary use.

Edwards and Collins (2011) postulated that the underlying mathematical model, which was based on Zipf's law and used in the simulations conducted by Meara (2005), allowed the analysis to be done directly, without recourse to simulations. The results of their analysis suggested that LFP was not particularly sensitive with respect to estimating individual productive vocabulary sizes, and confirmed Meara's (2005) suggestion that LFP's ability to distinguish between groups decreased as the size of the vocabulary increased. In 2013, Edwards and Collins delved further into the analysis by exploring a new model of vocabulary learning that was actually an adapted version of the previously proposed model. The motivation for their analysis was the fact that it was assumed, with the earlier model (the "naïve" model), that L2 learners acquired words according to their frequency of occurrence in the language, which was not the case. They attempted to develop a more valid model that took into account the presence of less common words at different points throughout the acquisition process and the fact that words may be acquired through repeated exposure. The model took into consideration the lower-frequency words that are learned before certain higher-frequency ones and therefore better reflected real-world L2 vocabulary acquisition. They concluded that this model provided a modified view of the relationship between LFP and students' productive vocabulary size, and gave lower estimates of vocabulary size than the naïve model. Although LFP has been used by certain researchers to estimate vocabulary size it will only be used in the present study for the initial purpose for which it was developed in 1995 by Laufer and Nation, i.e., to evaluate the lexical complexity of a text by categorizing words according to their frequency of occurrence.

In 2006, Horst and Collins (2006) assessed the development of IE students' LC. They initially used Vocabprofile to create LFP as the measure of LC for their study, but this analysis did not produce conclusive results. They then turned to four additional measures (reliance on French, reliance on cognates, morphological variety and number of K1 families) to supplement their initial results. The students participating in the study were IE students who lived in a French environment and had little

or no exposure to the English language outside of class. They used French to fill the lexical gaps in their oral productions, thus pointing to the pertinence of studying how their reliance on French changed over time. The researchers looked at the amount of French used by the students, as well as the nature of the French words used. Further to that analysis, they emphasized the importance of using more than one measure to assess LC.

The studies presented above show that a variety of measures are used to assess the development of lexical complexity. Although there is no commonly accepted measure of LC, an element many researchers (Bulté and Housen, 2012; Daller et al., 2003; Horst and Collins, 2006) appear to agree on is the importance of using a number of measures that combine to paint a complete picture of the complexity of a learner's lexical knowledge. The development of a learner's LC can be observed through a variation, over a period of time, in the measures used to represent LC. For the purpose of the present study, measures related to the three constructs proposed by Bulté and Housen will be used: lexical density (through the proportion of lexical words versus total words), lexical diversity (through the number of unique occurrences) and lexical sophistication (through LFP, or the proportion of less-frequent words used). Reliance on French words, a measure suggested by Horst and Collins (2006), which appears to be particularly pertinent in the context of IE programs, will be used as an additional measure to supplement the results provided by the three first measures.

3.2 Intensive English programs

In the field of second language acquisition, research demonstrates that repeated exposure and practice are favourable to learning new items or structures (Serrano, 2012). Traditional L2 programs, which offer a few hours of instruction every week, have not proven to be particularly effective in teaching a second language, and have been shown to leave students with limited L2 abilities (Collins & White, 2011). Stern (1985) suggests that this drip-feed approach is less effective, for an equivalent number of hours of instruction, than a more concentrated model in which instruction is administered in larger blocks on a daily basis. He states that a compact or intensive course, if planned correctly, "has undoubtedly considerable potential to remedy . . . weak and straggling language programs" (p. 24). More concentrated forms of L2 instruction, such as intensive programs, have therefore been

developed to offer an alternative to traditional programs (Serrano & Munoz, 2007). Research in intensive programs has shown that concentrating the hours of instruction can allow students to make substantial L2 gains in a short amount of time (Collins et al., 1999; Collins and White, 2011; Germain, Lightbown, Netten & Spada, 2004; White & Turner, 2005) without any detrimental effects to their first language (L1) (Lightbown & Spada, 1991; 1997; Spada & Lightbown, 1989).

In Quebec, the main goal of IE programs is to help students become functionally bilingual in order to be able to face typical everyday situations in their L2. This is done by presenting students with authentic learning situations that are meaningful and that relate to the students' real-life needs and interests. Age-appropriate topics are presented through authentic learning situations to prepare the students for real-life communication (SPEAQ, 2012). A communicative approach is used in all IE classrooms, and although all four skills (speaking, listening, reading and writing) are utilized, the main focus is on listening and speaking (Germain et al., 2004).

IE in Quebec is generally offered in sixth grade, but some school boards choose to offer the program in fifth grade (Germain et al., 2004). While the first IE program was developed on a five-month/five-month model, other models have been created over the years. These are grouped into two categories: massed programs and distributed programs. In the former, the intensive period is condensed into five months, with students devoting half of the school year to the regular curriculum, and the other half to ESL; in the latter, the intensive program is spread out over the ten months of the school year (Collins et al., 1999). Some examples of distributed programs include alternating between 4 days/1 day and 1 day/4 days; 2.5 days/2.5 days; 1 day/1 day; half a day/half a day. Schools and school boards have the freedom to decide on the structure of the intensive program. Of all these models, the massed program appears to be the most widely used (SPEAQ, 2012).

Overall, studies on IE programs in Quebec show that intensive instruction can have a positive impact on learning a L2. In fact, reading and listening comprehension, basic communication skills, along with fluency, confidence, and mean length of runs are some of the areas that appear to benefit from a

five-month intensive program (Collins et al., 1999; Lightbown & Spada, 1994; Spada & Lightbown, 1989; White & Turner, 2005). The present study looks at the learning process of students participating in a five-month program, which provides an ideal context in which to investigate the development of LC in a relatively short amount of time.

4. LITERATURE REVIEW

4.1 Intensive English in Quebec

Much of the research involving IE programs in Quebec has focused on the development of global components of students' language, such as fluency (Spada & Lightbown, 1989), listening and reading comprehension (Collins & White, 2011; Spada & Lightbown, 1989), correctness of language (Lightbown & Spada, 1991) and oral production or performance (Collins & White, 2011; Spada & Lightbown, 1989; White & Turner, 2005). There is now a substantial body of evidence showing that IE programs allow students to make considerable gains in their L2 in a short amount of time, and that show that IE students outperform their regular-program counterparts on a variety of tasks pertaining to all four language skills (Collins & White, 2012).

In terms of vocabulary development, research has shown that word recognition steadily improves over the course of an intensive program, suggesting that students' vocabulary has grown by the end of the intensive period. Collins and White (2011) suggest that by the end of the program, IE students are familiar with approximately 75% of the 1000 most-frequent words of the language. It has also been observed that IE students tend to be more talkative (i.e., to produce more words) than students following a regular ESL curriculum (Spada & Lightbown, 1989), and that they use a wider range of expressions and a more varied vocabulary when participating in an oral task (White & Turner, 2005).

While little research has specifically dealt with lexical knowledge and development, Spada & Lightbown's (1989) finding that IE students use a more varied vocabulary than their regular-program counterparts might suggest that IE programs could promote gains in lexical abilities, including the development of LC. This is confirmed in Laufer's (1994) idea that:

> The process of acquiring an additional language, second or foreign, has often been described and discussed in terms of the learner's progress along the Interlanguage continuum, from a non-existent knowledge of L2 towards native-like competence, without necessarily reaching this ideal stage. If this is the view we take of language acquisition, then lexical acquisition research would have to account for the gradual

> increase in the learner's vocabulary size, as the most striking difference between the vocabulary of native speakers and that of language learners is in the number of words they can control, particularly in free production: speech or writing (p. 21).

One of the few studies to investigate lexical development in an IE context was conducted by Horst and Collins (2006), who studied the development of lexical richness in written productions of sixth-grade IE students. They based their research on the hypothesis that the students' writing would show improved lexical richness over the course of an intensive program. The study involved a total of 230 students, all of whom were 11- or 12-year-old French-speaking Québécois who lived in an environment that provided little exposure to English outside the classroom. The researchers used LFP to measure the lexical richness of students' texts at four times during the program; after 100, 200, 300 and 400 hours of instruction. At each of the four testing times, students were shown a picture prompt and asked to write about what they thought had happened before, during, and after the event depicted. They were given 20 minutes to complete the task, and were allowed to use French words to fill lexical gaps.

Students' texts were analysed using Vocabprofile. The software divided the words of the text into four categories: the 1000 most-frequent word families of English (K1), the 1001-2000 most-frequent word families of English (K2), an academic word list (AWL), and an off-list, which included all the words that did not fit into one of the three previous categories. The authors hypothesized that over time, the proportion of words in the K1 band would decrease and the proportion of words in the other three categories would increase, thus showing that the students were using a greater amount of less-frequent (and therefore more complex) words as time went by. However, the proportion of words from the K1 band slightly increased over time, and the proportion of words in the K2 band actually decreased, contrary to hypothesis. These initial results therefore showed that the young beginning learners of English did not use greater amounts of less-frequent words over the course of an intensive program. However, the authors were not prepared to say that there had been no lexical development. They conducted further analyses using other measures and were able to conclude that although Vocabprofile did not suggest any development in LC, the use of four alternate measures (reliance on French words, reliance on cognates, variety of word families within the 1000 level and

morphological variety of items within word families) showed substantial improvement in LC over the course of the IE program.

One of the additional measures they looked at was reliance on French. When writing their texts, students had been informed that they could use words from their L1 to fill lexical gaps. The number of French words per text decreased considerably from one testing time to the next, and the kinds of French words that the students used also changed; over time, the French words students produced were, in increasing proportion, relatively low-frequency words. This decrease in the use of French suggested that they were able to use a greater number of English words. The number of different word families and the number of types per family were two other measures that were also used to further investigate the development in lexical complexity, and it was noted that over time, students used more word families in their compositions, along with a greater number of types per family.

The change in the use of French cognates was also measured. The most important difference was observed in the proportion of words of romance language origin at all four testing times. Horst and Collins (2006) indicate that this decline in the proportion of cognates might account for part of the relatively high proportion of beyond-2000 words in the initial measure of LFP and its subsequent decrease. Certain words (such as *search* and *respond*) appeared in the compositions written by the students at the beginning of the program. Although not necessarily incorrect in usage, the words were not the ones best suited to the context. It was hypothesized that the students used them because of a transfer effect from their L1, as these words had a corresponding French cognate that would have been appropriate for the context. All of these cognates appeared beyond the first 2000 word families of the language, therefore indicating use of less-frequent words by the students. In the final composition, students used words that were more appropriate to the context (such as *look for* and *answer*), but also appeared on higher frequency lists, therefore suggesting a regression in LC.

Overall, the results of the studies presented above (Collins & White, 2011; Collins et al., 1999; Horst & Collins, 2006; Lightbown & Spada, 1991; Lightbown & Spada, 1991; White & Turner, 2005)

suggest that IE programs have a positive effect on the development of various aspects of students' L2, such as fluency, accuracy, loquacity and correctness of the language. However, only one study (Horst & Collins, 2006) has examined the effect of intensive programs on the development of LC. The results of the study suggest that over the course of an intensive program, there is noticeable development in students' lexical complexity when measured through a written task assessing reliance on French, use of cognates, number of word families and types per family.

4.2 Development of lexical complexity

As pointed out above, only one study (Horst & Collins, 2006), has analysed the development of LC in IE students, thus highlighting the pertinence of examining other similar contexts in order to obtain a clearer picture of how the development of LC has been assessed.

While much of the research on the development of LC in learners of a second language has been conducted with adults or teenagers, some researchers have also studied the phenomenon in children. This is the case of Bournot-Trites (2007), who studied French immersion students' development in various aspects of language, including lexical development. The study followed two groups of students and analysed their progress from fifth grade to seventh grade through a composition task that was completed at the beginning of fifth and the end of seventh grade. Three measures were used to assess the students' lexical development: diversity (number of different verbs divided by the total number of words), sophistication (number of less-frequent verbs divided by the total number of verbs) and the total number of words contained in the text. Results showed greater diversity in fifth grade, but greater sophistication and mean total number of words in seventh grade. These results led the author to suggest that the immersion had a positive impact on students' lexical development at an advanced stage of learning.

Another study involving young learners was conducted by Lo and Murphy (2010). In that study, the immersion context was a central element, as the researchers attempted to determine whether the language-learning context had an impact on the learners' vocabulary knowledge and growth. The two

learning contexts studied were regular second-language instruction programs and immersion programs in Hong Kong. The students participating in the study were in seventh and ninth grades, and for either grade level, some of the students were following a regular program, while the others were enrolled in an immersion program (the immersion program began in seventh grade; therefore, immersion-program students were either in their first or third year of English immersion).

A free-composition task, which was analyzed through LFP, was used to assess development of LC. The results of the analysis showed that both in seventh and ninth grade, immersion students include a significantly larger proportion of less-frequent words (which appear beyond the first 2000 words on the frequency lists). Hence, the authors concluded that the students in immersion programs used a greater proportion of low-frequency words in their compositions. They also noted that, by as soon as the end of the seventh grade, an advantage was observed on a number of measures for the immersion students, indicating that even after only one year in an immersion program, students were able to show more advanced levels of vocabulary than their regular-program counterparts. Overall, these findings suggest a clear advantage for immersion programs in terms of the development of lexical complexity when measured through written tasks.

The findings presented above have shown that, through the use of LFP, immersion contexts tend to favour the development of LC in children. This differs from research in IE contexts, where research has yet to use LFP to demonstrate significant improvement in LC. The fact that conclusive results have been found in immersion contexts introduces the idea that it may be pertinent to further investigate whether LFP could be used to measure development of LC in an IE context.

5. RESEARCH METHOD

The data for this research project comes from an ongoing large-scale research project that examines IE programs and is conducted by Dr. Leif French at the Université du Québec à Chicoutimi. The data collected for the project covers numerous students enrolled in programs based on a variety of intensive models. However, the present study will focus on a sample of those students and will only present data based on tasks directly relevant to the research questions and hypotheses outlined in the introduction.

5.1 Learning context

The students in the study were all enrolled in a sixth-grade IE program. Although various models exist in terms of time distribution for intensive ESL programs, the model applied was among the most frequently used ones (Germain et al., 2004). For the first five months of the school year, the students focused on the regular sixth-grade curriculum. The students then spent the five remaining months learning English. During the five-month English period, the students received approximately 370 hours of instruction based on a communicative approach, which emphasized the development of oral skills. No formal academic content was taught in English, and project-based pedagogy was adopted, focusing on activities that were stimulating for the students and that aimed at helping them develop functional language that would be useful to them in everyday life, were they to be placed in an English-speaking environment.

5.2 Participants

Participants for this study (n = 56)[4] were French-speaking students (11 to 12 years old) from two different elementary schools in the Saguenay region (one class of 28 and one class of 29 students). Initially, 57 students were to participate in the study, but one participant had to be excluded, as he was absent at one of the two testing times. All students had begun to learn English in fourth grade and, prior to the start of the intensive program, had received approximately 120 hours of formal

[4] For the analyses involving proficiency groups, one student was excluded (n = 55) due to an absence at the time of the vocabulary test.

classroom instruction. All had completed the regular sixth-grade academic program in the first part of the school year and were devoting the remainder of the school year to learning English.

5.3 Task

5.3.1 Vocabulary test

A two-part vocabulary test was administered to the students both at the beginning (T1) and at the end (T2) of the program (see Appendix A). In the first part, students were presented with a list of 60 English words, which they were asked to translate into French. In the second part, they were presented with a list of 60 French words, which they were asked to translate into English. All of the items on both lists were common words taken from fifth-grade ESL textbooks. A score out of 60 was computed for both parts of the test, and a composite score out of 120 was also calculated. For the purpose of the present study, the results of the vocabulary test were used to assess students' initial level of proficiency. A median split was used to divide the students into high- and low-proficiency groups.

This measure was chosen in order to group the students on a measure which was independent from the four other measures of LC, but which was still related to the students' lexical knowledge.

5.3.2 Oral narration

An oral narration task based on picture cues was completed by the students at T1 and T2. Although picture-cue tasks are generally not used for teaching purposes in communicative contexts, as they do not represent an authentic communication situation, they are used in research to study various aspects of students' language, and are deemed particularly useful in studying vocabulary development. An important reason for using picture stories in research is the fact that while they provide researchers with the opportunity to control the language elicited, they also give students the freedom to produce language that adequately reflects their speaking abilities (Rossiter, Derwing & Jones, 2008).

Rossiter et al. (2008) also emphasize the importance of choosing a picture story that is clear and void of any ambiguity so that the students focus on the language rather than the meaning of the pictures. The picture story used in this task corresponded well to the criteria suggested by Rossiter et al. More specifically, each picture frame depicted only one action, was detailed enough for readers to understand the context and action, contained no cultural content that might be difficult for students to interpret, and had a clear sequence of events with no flashbacks or flash-forwards.

Students were provided with a 20-minute planning phase, during which they were able to look at the story and take notes as to what was happening in the pictures, and how they planned to relate the story. Although they were not allowed to use their notes for the actual narration, this planning phase enabled them to gain familiarity with the task and, thus, to reduce the cognitive load on language processing during the narration. Foster and Skehan (1996) postulate that planning time gives rise to greater linguistic complexity, as it gives the students the opportunity to attempt more ambitious ideas and to provide greater clarity in the relationship between ideas. The planning phase therefore makes it likely that students were able to use vocabulary that was more contextually appropriate, and that they were able to use the extent of their vocabulary to a greater degree.

The specific picture-cue task used in this study was an adapted version of the book *Frog, Where Are You?* by Mercer Mayer (1969). The adapted version, created by Leif French at the University of Québec at Chicoutimi, was composed of a series of 16 images and designed specifically to reduce the time needed to tell the Frog Story in a classroom situation while maintaining the integrity of the original storyline (see Appendix B). Thus, the story is that of a young boy who catches a frog and puts it in a jar by his bed. During the night, the frog gets away. In the morning, the young boy is alarmed to find that the frog is no longer in the jar. He looks for it everywhere in his bedroom, but to no avail. He therefore decides to take the search outside with his dog. He goes into the woods, where he encounters a deer, and gets pushed off a low cliff into a brook. He then returns home with his dog,

but without the frog. The next day, the young boy goes back outside to look for the frog, and finds it sitting on a log by the river with its family.

The same picture story was presented to the students at T1 and T2. On both occasions, the students were provided with a booklet in which they could take notes and plan each of the three parts of the story. Both the booklet and story were clearly labelled with a time frame (*yesterday and this morning, now, tomorrow*) in order to provide a basis for indirectly eliciting specific temporal elements (see Appendix B). However, students were not explicitly instructed to use the past, present and future tenses for the different parts of the story.

5.3.3 Procedure for oral narration

Instructions for the task were presented to the students in French by the researcher. They were given 20 minutes to prepare and plan the story. To do so, a planning booklet was given to them, and they were given specific instructions as to how the story had to be constructed. The booklet was divided into three distinct parts, and each one included a space for specific vocabulary the students deemed might be useful to them for the task. The instructions for the task, which were given to the students by the researcher at the beginning of the task, were also contained in the preparation booklet.

At the end of the 20-minute period, students were asked to step into another room, where a researcher was waiting to record their story. Students were asked to narrate the story in English. They were allowed to use French to compensate for lexical gaps, but were not encouraged to do so. They could take as long as they wished to tell the story and were recorded on a computer using the Audacity software program.

5.4 Measures of lexical complexity

Several authors (Bulté & Housen, 2012; Daller et al., 2003; Horst & Collins, 2006) stress the importance of using multiple measures to provide a complete picture of the students' progress over time. In the present study, four different measures were therefore used in conjunction to illustrate students' development of LC over the course of the five-month intensive program (see Table 1).

Table 1: Overview of measures of lexical complexity

Measure	Studies
Lexical density The ratio between the number of lexical words and the total number of words contained in a text.	-Ishikawa, 2007 -Michel, Kuiken and Vedder, 2007 -Ortega, 1995
Lexical diversity (unique occurrences): The size of the learner's productive lexical knowledge, or the number of different words he or she uses.	-Bulté et al., 2008 -Collentine, 2004 -Isbell, Sobol, Lindauer and Lowrance, 2004
Lexical frequency profiling (LFP): Categorization of all the words contained in a text into four categories according to the frequency with which they appear in the language: the 1000 most-frequent word families of the language, the 1001-2000 most-frequent word families of the language, an academic word list, and an off-list, which comprises all of the words that do not fit into one of the three previous categories.	-Horst and Collins, 2006 -Laufer, 1994 -Laufer, 1998 -Laufer and Nation, 1995 -Lo and Murphy, 2010 -Ovtcharov et al., 2006[5] -Vermeer, 2004[6]
Reliance on French: Proportion of French words used in the text.	-Horst and Collins, 2006

5.4.1 Lexical density

The lexical density of the students' narrations was calculated by dividing the number of lexical words (nouns, verbs, adjectives and adverbs) by the total number of words in the text. The result was expressed as a ratio, with a high ratio indicating a more lexically dense text and a higher ratio at T2 than at T1 indicating growth in students' lexical density.

[5] Used the Measure of Lexical Richness (MLR), which is closely related to LFP but uses different frequency lists
[6] Used *Profil de fréquence lexicale*, a French adaptation of LFP

5.4.2 Lexical diversity

The lexical diversity of the narrations produced by the students was gauged by the number of unique occurrences (different words) per 100 words of text produced by the students both at T1 and T2. The diversity provided an indication of whether the students' productive vocabulary seemed to have grown.

Since the texts produced by the students were of different lengths and were likely to be longer at T2 than at T1, a measure independent of text length had to be calculated. By expressing the total number of unique occurrences per 100 words, and thus matching the texts on a comparable ratio rather than simply using the number of unique occurrences in the text, narrations of various lengths could be analysed and compared.

5.4.3 Lexical sophistication (use of less-frequent words)

Students' use of less-frequent words was measured through lexical frequency profiling, with the Vocabprofile software program. Vocabprofile analyzes texts and categorizes each of the words (tokens) into one of four lists: the 1000 most-frequent word families in English (K1), the 1001 to 2000 most-frequent word families in English (K2), an academic word list (AWL), and an off-list, which is a list of all the words that did not fall into one of the three previous categories. All of the words contained in the off-list therefore appear beyond K2. Calculating the proportion of words that fall into each of the categories, both at T1 and T2, made it possible to assess whether students' vocabulary seemed to include a greater amount of less-frequent words and therefore become more complex over time.

The reason for using LFP is that, contrarily to measures that indicate that learners are using a more varied vocabulary, it demonstrates whether the learners' vocabulary is becoming more complex over time. As Horst and Collins (2006) point out, LFP offers an objective way of determining the degree of

difficulty of a word, as it is based on recognized frequency lists, and therefore renders valid comparisons across studies.

Although much of the research on lexical complexity has been conducted with adult learners of various levels, Laufer and Nation (1995) suggest that LFP could be a valid way of measuring young learners' lexical complexity. This position is further supported by Vermeer (2004), who postulates that word frequency can be used to operationalize the degree of difficulty of words and who used the MLR, which is similar to LFP, to compare the lexical richness of children learning Dutch as a first and as a second language. Her results led her to suggest that LFP could be a valid way of measuring LC with beginners, including children with low proficiency levels. Finally, Lo and Murphy (2010), also used LFP (through Vocabprofile) with young learners, and the conclusive results of their study led them to believe that LFP is indeed a valid measure of children's lexical richness.

5.4.4 Reliance on French

In their study, Horst and Collins (2006) did not succeed, using only the measures provided by LFP, to determine that IE students' lexical richness improved over a five-month program. They therefore turned to other measures, such as reliance on French words, to illustrate the progress made by the students.

In the data for the present study, it was observed that many students used French sporadically during the narrations to fill lexical gaps. It was specified, in the instructions for the task, that the students could, if needed, resort to using their L1, but they were instructed to limit their use of French as much as possible. Observing how the use of French changed over time therefore became another indicator of the development of LC.

Reliance on French was measured through the proportion of French words contained in each student's text and the number of French words divided by the total number of words in the text. This

measure was then converted to a percentage in order to be able to compare the students' texts, which were of different lengths.

5.5 Procedures for data analyses

The narrative data used for the analysis had previously been transcribed from the audio format into a written format. The resulting text contained no spelling mistakes, which helped assure proper analysis by the software. Grammatical errors, however, had been left intact.

Before the data was analyzed, all French words contained in the transcriptions were tagged. This was an important step because it made the French content easier to identify for the analysis and helped avoid any misinterpretation by the software: some words, such as *original*, exist both in English and in French. Such a word, if used in French in one of the students' texts, could have been considered an English word by the software had it not been tagged.

In order to go through with the first phase of the analysis, which included measures of lexical diversity and density, several measures had to be calculated for each learner's text. First, the total number of words[7] contained in the text needed to be calculated. In order to do so, the transcriptions had to be altered by removing all pauses and sounds that did not constitute words. The overall number of words was then calculated using a word processor's statistics function. Then, the total number of English words was calculated. This was also done using a word processor; all the words tagged as French words were removed, and then the total number of words was once more identified using the word processor's statistics function. It is important to know that the French words that were removed from the text were copied and saved in another file, for an analysis of the use of French. Finally, texts were submitted to Vocabprofile first to identify the lexical density of texts (expressed as a ratio calculated by the software by dividing the number of lexical words by the total number of words), and then to identify the number of unique occurrences, which is the number of different words of the text.

[7] A word is defined as "the smallest unit of grammar that can stand alone as a complete utterance, separated by spaces in written language and potential pauses in speech" (Crystal, 1997, p. 440).

Following this, the measure of lexical density was calculated by dividing the number of unique occurrences by the total number of English words and multiplying by 100. It should be mentioned that the text submitted to the software for this analysis was the version that had been pruned of all French words.

Next, lexical sophistication was assessed through LFP with Vocabprofile. The text submitted for this analysis was void of all pauses, sounds and French words. These items were removed so as not to influence the results of the analysis; all items that fell into the off-list were therefore words that appeared beyond the 2000 most-frequent in the language but not on the academic list. As for proper nouns, Vocabprofile has a function that considers all capitalized words as part of the 1000 most-frequent of the language. This allows the proper nouns to be considered among the easier rather than the more difficult words of the language. The fact that no punctuation, and therefore no capitalization of new sentences, was used in the transcriptions made the use of that function possible. Other words, such as onomatopoeias, were also considered to be among the first 1000 of the language. The results provided by the software included fours measures: the proportion of the 1000 (K1) and the proportion of the 2000 (K2) most-frequent words of the language, the proportion of academic words (AWL), and the proportion of off-list words. All of these measures were expressed as percentages.

To assess reliance on French, the total number of French words (tokens) for each text was calculated. This was done using the file containing the French words for each text, extracted from the data for the first analysis. The number of French tokens was then divided by the overall number of words in the text, and the measure was converted to a percentage to make comparisons between texts easier.

The results obtained through the measures calculated above were used to answer the first research question. In order to answer the second research question, the same measures were used, but they were divided according to the students' level of proficiency. Proficiency was assessed through results

on the vocabulary test at T1, and a median split was performed in order to divide the students into two proficiency groups (low and high). Results according to proficiency group were then calculated in order to compare them and determine whether differences existed between the two groups.

5.6 Statistical analysis

In order to interpret the results obtained through the analyses, the results were submitted to statistical testing using *t*-tests and ANOVA tests, along with their non-parametric equivalents, Wilcoxons and Mann-Whitneys. This made it possible to determine whether differences between results obtained at T1 and T2 were significant and therefore suggested that progress had been made during the five-month intensive period. Moreover, the statistical analyses also made it possible to determine whether the difference between the progress made by low-level and high-level learners was significant, which established whether the two groups progressed differently during the intensive program.

6. RESULTS

In order to investigate the development of LC over time, measures relating to a combination of constructs were assessed. In order to do so, the intact group, the low-proficiency group and the high-proficiency group[8] were tested on all measures to determine whether they followed a normal distribution. The results were then submitted to inferential testing in order to determine where significant differences existed. For analyses with groups presenting a normal distribution, parametric tests (independent and paired *t*-tests) were used; for groups whose distribution was not normal, the non-parametric equivalents of independent and paired *t*-tests (Wilcoxon and Mann-Whitney) were performed.[9] Given the necessity of multiple comparisons, Bonferroni corrections were applied to all analyses at T1 and T2, and the α level was set at .006 to adjust for seven variables.

6.1 Development of lexical complexity from T1 to T2

For the first research question, which aimed at assessing whether there was development in students' LC over the course of a five-month IE program, four constructs relating to lexical complexity were assessed: lexical density, lexical diversity, lexical sophistication and reliance on French. Bulté and Housen (2012) suggest that the first three constructs (density, diversity and sophistication) combine to paint a complete picture of a learner's lexical complexity. Reliance on French was also added to these constructs, as it has been shown to be particularly pertinent in the case of IE programs (Horst and Collins, 2006).

In order to paint a complete picture of the students' lexical complexity over the course of an IE program, the descriptive statistics and change over time for each of these measures are presented together in Table 2. It is important to note that an increase in most measures from T1 to T2 is indicative of progress; however, for measures of K1 and reliance on French, progress is conveyed by a decrease in the measures from T1 to T2.

[8] See section 5.3.1 for more information on how proficiency groups were formed.
[9] See Appendix C for the list of measures that follow a normal distribution and those that do not.

TABLE 2: ALL LEXICAL MEASURES: MEANS, STANDARD DEVIATIONS (TIME 1 AND TIME 2) AND CHANGE IN PERFORMANCE OVER TIME (TIME 1 TO TIME 2)

		TIME 1				TIME 2				CHANGE OVER TIME (T1 to T2)	
		Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
Density		0.43	0.07	0.30	0.58	0.42	0.05	0.28	0.55	-0.01	-0.02
Diversity		31.90	6.50	20.29	46.43	35.51	5.99	20.74	49.57	3.61	5.36
Sophistication	K1	89.19%	2.25%	77.78%	96.39%	90.10%	2.90%	84.24%	95.77%	0.91%	3.94%
	K2	3.72%	2.25%	0%	8.93%	4.13%	2.16%	0%	8.70%	0.41%	2.33%
	AWL	0.03%	0.16%	0%	0.97%	0.04%	0.17%	0%	0.81%	0.01%	0.20%
	Off-list	7.02%	3.32%	1.83%	17.17%	5.72%	2.01%	0.99%	12.12%	-1.30%	3.04%
Reliance on French		7.17%	5.19%	0%	21.82%	2.59%	2.21%	0%	8.67%	-4.59%	4.51%

Note: Density = lexical words/total words; Diversity = unique occurrences per 100 words; K1 = 1000 most-frequent word families; K2 = 1001-2000 most-frequent word families; AWL = academic word list; Off-list = all words except K1, K2 and AWL; Reliance on French = French words

6.1.1 Lexical density

Lexical density was assessed first. Expressed as a ratio, it indicates what proportion of a text consists of content versus total words. The results presented in Table 2 show that the mean lexical density of students' texts tended to decrease slightly from T1 to T2. However, a paired *t*-test revealed that this decrease was not significant; $t(55) = -1.681$, $p = .099$, suggesting that, overall, students' lexical density remained relatively stable over time.

6.1.2 Lexical diversity

The number of unique occurrences per 100 words was measured to assess the change in lexical diversity. From T1 to T2, the mean number of unique occurrences per 100 words increased, as shown in Table 2. A paired *t*-test showed that this increase was indeed significant: $t(55) = 5.042$, $p = .000$, indicating that the students produced a greater proportion of different words at the end of the program than at the beginning. This suggests growth in students' vocabulary knowledge and is in line with the idea of development of lexical complexity.

6.1.3 Lexical sophistication

Lexical sophistication was assessed through four sub-measures: the proportion of K1 (1000 most frequent), K2 (1001-2000 most frequent), AWL (academic word list) and off-list words. Horst and Collins (2006) point out that in the case of IE students, the size of the vocabulary is quite small at the beginning of the program and suggests that the divide between the proportion of the K1 list and the three other lists is a reliable indicator of LC. Thus, at T1, the proportion of K1 words would be expected to be high, and the proportion of K2, AWL and off-list words would be expected to be quite low. However, at T2, the proportion of K1 would be expected to have decreased and the proportion of words from the other lists would be expected to have increased in order to show development of LC. In the present study, however, the proportion of K1 words did not decrease significantly but remained relatively stable from T1 to T2 ($t(55) = 1.737$, $p = .088$). In terms of the proportion of words from the K2 list, the variation from T1 to T2 was not significant ($t(55) = 1.303$, $p = .198$), suggesting that this measure also remained stable over time. As for the proportion of words from the AWL, it was quite small at both testing times and did not vary significantly ($Z = -0.405$, $p = .686$). Overall, this suggests that lexical sophistication, as referenced by these three measures, remained virtually unchanged during the five-month program.

The last measure of lexical sophistication computed was the proportion of words pertaining to the off-list. The off-list comprised all the words that did not fall into one of the three other categories (K1, K2 and AWL). Since all proper nouns and onomatopoeias had been re-categorized as K1, the words in the off-list were all words that appeared beyond the 2000 most-frequent words of the language, and were therefore among the least-frequent or most "difficult." It was hypothesized that, in line with development of lexical sophistication, the proportion of words in the off-list would increase from T1 to T2, indicating an increase in the use of less-frequent words. However, Table 2 shows that the proportion of words in the off-list actually decreased significantly ($t(55) = -3.186$, $p = .002$), suggesting that there was actually a setback in students' lexical sophistication over the course of the program.

Although stability in K1, K2 and AWL suggest stability in lexical sophistication, the decrease in the off-list suggests a setback in the development of lexical sophistication, which seems counterintuitive.

To examine this finding further, the texts were re-examined qualitatively. This analysis showed that the word frog, which fell in the off-list, was repeated several times in the narrations, suggesting that the frequent repetition of this item may have influenced the LFP results. Consequently, it was decided to prune the narrations of all repetitions of *frog*, and a second version of the text containing only unique occurrences was again submitted to LFP. Nevertheless, the results for the second LFP measure were quite similar to the first, indicating a significant decrease only in the off-list from T1 to T2. This made it possible to conclude that the repeated occurrence of the word *frog* had not affected the results of the LFP.

6.1.4 Reliance on French

Reliance on French was assessed through the proportion of French words used at T1 and T2. Results are displayed in Table 2 and support the hypothesis that the proportion of French words would decrease from T1 to T2 ($Z = -5.601$, $p = .000$). This suggests that the students were able to use a greater proportion of English words, and therefore had a wider range of L2 vocabulary at T2. The results are in line with the idea of development of LC.

6.1.5 Overall observations

Several researchers (Bulté and Housen, 2012; Daller et al., 2003; Horst and Collins, 2006) emphasize the need to assess multiple constructs in order to clearly illustrate the range of a learner's LC. On its own, each of the four constructs assessed in the present study gives only a partial idea of the students' lexical complexity. The results presented above show that significant variations in results were present in only three areas: lexical diversity, proportion of off-list words, and reliance on French.

The increase in lexical diversity and the decrease in the use of French over time both indicate growth in students' vocabulary. However, LC is defined not only by the size or breadth of a learner's vocabulary, but also by the degree to which it is composed of "rich" or "difficult" words, and while the results presented above suggest a wider and more varied vocabulary, the extent to which this

vocabulary is more complex is less certain. In the present study, the measure used to assess the difficulty or richness of the vocabulary is lexical sophistication, which was assessed through LFP. Yet, the results do not show any improvement over the course of the program in any of the four sub-constructs that compose LFP. Moreover, the significant decrease in the proportion of off-list words actually suggests a setback in LC.

Based on these results and results in Horst and Collins (2006), whose LFP measures through Vocabprofile did not show concrete signs of use of greater proportions of less-frequent words by the end of an IE program, one may wonder whether there was indeed any progress made in terms of lexical sophistication or, more importantly, whether Vocabprofile may simply not be the best-suited measure to assess this construct. Initially, the decision to use LFP to assess lexical sophistication was based on the fact that other researchers had used this measure successfully with young ESL learners. For example, Lo and Murphy (2010) used Vocabprofile to measure LFP in ESL students in Hong Kong, in an immersion context somewhat comparable to the present IE context. However, they reported only slight progress in lexical sophistication over a three-year period, suggesting that Vocabprofile may not be sensitive enough to capture the progress made in the present five-month study context.

Another element to consider is the fact that the frequency lists used by Vocabprofile are based on adult learner language, and adults and children, as mentioned by Vermeer (2004), use vocabulary differently. As Horst and Collins (2006) point out, a measure that makes use of frequency lists more appropriate for young learners may therefore be better suited to measure the progress of IE students. In fact, Vermeer (2004) used LFP to measure the lexical richness of young students learning Dutch, but did not use Vocabprofile. She developed her own measure (MLR), which was based on nine frequency classes. In this study, MLR was deemed valid with learners at a beginning level, but the researcher admitted that it was unclear whether this measure would be able to distinguish between learners at higher levels. This reinforces the idea that the measure of LFP used should be based on lists that are appropriate to the age and level of the learners.

Given these potential problems with the use of LFP through Vocabprofile, the lack of development in terms of lexical sophistication in the present study must be interpreted with caution, as other measures may indeed be more sensitive to lexical development than those provided by Vocabprofile. Consequently, it is important to go beyond lexical sophistication as a measure of LC and investigate whether other lexical measures might be better suited to illustrating progress in IE students' lexical complexity.

6.1.6 New measures of lexical complexity

Interestingly, qualitative observations made during the analysis of the data showed that while the present tense was the main verb tense used in nearly all of the narratives at T1, other verb tenses, such as the simple past, the future, the present progressive and even the past progressive were used in many of the narratives at T2. Correct use of third-person verbs in the present tense (*he look*s, vs. *he look*), along with correct use of irregular verbs in the past (*he saw*, vs. *he see*), were also noted at T2. While these examples do not consist in the use of more complex vocabulary, they do indicate a more complex use of the same word families that may have been used at T1. In order to measure whether there was actually a significant improvement in the use of these more complex forms of words, morphological variety within word families was assessed.

Initial results of LFP showed that the proportion of K1 words used at T1 and T2 was relatively stable. However, the results gave no indication as to the number of different words the students used within this category. As Horst and Collins (2006) point out, IE students are beginning learners who have a very limited range of vocabulary at the beginning of the intensive program. While the proportion of words belonging to the 1000 most-frequent may not have changed over time, the number of different words they used within this category may have grown. Such a variation would indicate progress in terms of students' lexical growth. In order to measure whether students used an increasing proportion of different words from the K1 list, Vocabprofile was used to calculate a measure of variety within K1, which was computed both at T1 and at T2.

Two additional measures of lexical complexity were therefore computed: morphological variety and variety of K1 families used. Since lexical sophistication was removed from the analysis, five measures would now be used to assess LC: lexical density, lexical diversity, reliance on French, morphological variety and number of K1 families. The results for all of these measures are presented in Table 3.

TABLE 3: NEW LEXICAL MEASURES: MEANS, STANDARD DEVIATIONS (TIME 1 AND TIME 2) AND GAINS OVER TIME (TIME 1 TO TIME 2)

	TIME 1				TIME 2				GAINS (T1-T2)	
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
Density	0.43	0.07	0.30	0.58	0.42	0.05	0.28	0.55	-0.01	-0.02
Diversity	31.90	6.50	20.29	46.43	35.51	5.99	20.74	49.57	3.61	5.36
Reliance on French	7.17%	5.19%	0%	21.82%	2.59%	2.21%	0%	8.67%	-4.59%	4.51%
Morphological variety	1.04	0.03	1.00	1.15	1.06	0.04	1.00	1.18	0.02	0.05
Variety of K1 families	31.95	9.10	14	54	46.30	9.12	31	72	14.36	9.34

Note: Density = lexical words/total words; Diversity = unique occurrences per 100 words; Reliance on French = French words; Morphological variety = types/families; Variety of K1 families = number of words from K1

6.1.6.1 Morphological variety

Morphological variety was calculated according to the number of types per family. Results are displayed in Table 3 and show an increase in the mean number of types per family from T1 to T2. Statistical analysis revealed that this increase was significant ($Z = -3.116$, $p = .002$),[10] indicating that at T2, students were able to produce more grammatical forms belonging to a same word family, and therefore, a greater number of complex forms of a word.

[10] Bonferroni corrections were applied to all analyses in order to adjust for multiple comparisons. Given the measures that were eliminated (K1, K2, AWL and off-list) and those that were added (morphological variety and variety of K1 families) the α level was set at .005 to adjust for five variables.

6.1.6.2 Variety within K1

Variety within K1 was assessed through the number of different families belonging to the K1 list, and the results presented in Table 3 indicate an increase in the number of different K1-family words used over time. Statistical analysis through showed that this increase was significant ($t(55) = 11.500$, $p = .000$), indicating growth in students' vocabulary within the 1000 most-frequent words of the language.

6.1.7 Overall development of lexical complexity

Overall results on the five measures of lexical complexity used suggest development in IE students' lexical complexity. The increase in lexical diversity, along with the decrease in reliance on French, showed that students' vocabulary grew over the course of the five-month program. Variety within K1 showed that the students used an increased amount of word families within the 1000 most frequent of the language, and changes in morphological variety showed that students used an increased proportion of more complex forms of words by the end of the IE program. In all, these results appear to indicate that students' vocabulary not only grows and becomes wider over the course of an IE program, but also incorporates more "difficult" or complex words, which, combined, are a clear indicator of growth in LC.

6.2 Differences in the development of low-proficiency and high-proficiency learners

The second research question aimed at determining whether learners of different proficiency levels stood to make different gains in terms of lexical complexity over the course of a five-month IE program. In order to answer this question, participants were divided into two proficiency groups—a low proficiency group (LG) and a high proficiency group (HG). This proficiency measure was based on results of a two-part vocabulary test (see Table 4) that the students completed at the beginning of the IE program. Students' results on both sections of the test were compiled to create a composite score, and a median split was used to divide the students into two groups. As a result of the median split at T1, the groups were significantly different on the vocabulary measure ($t(53) = -9.084$, $p = .000$).

TABLE 4: VOCABULARY TEST: MEANS, STANDARD DEVIATIONS (TIME 1 AND TIME 2) AND CHANGE IN PERFORMANCE OVER TIME (TIME 1 TO TIME 2) ACCORDING TO PROFICIENCY GROUP

			TIME 1		TIME 2		CHANGE OVER TIME	
		N	Mean	SD	Mean	SD	Mean	SD
VOCABULARY TEST	LG	27	52.85	6.39	72.81	8.74	19.96	6.67
	HG	28	70.96	8.24	90.79	9.01	19.82	8.73

LC for the proficiency groups was assessed using the same five measures that were used with the intact group: lexical density, lexical diversity, reliance on French, morphological variety and variety of K1 families. Results at both testing times, along with differences between T1 and T2 for both groups, were calculated for each of the measures. Groups were compared on all measures at both testing times and statistical analyses were then conducted to determine whether differences between groups were significant.[11]

6.2.1 Results at T1 and T2

Five measures of lexical complexity were assessed for both LG and HG at the two testing times, and results are displayed in Table 5. Means for both groups show that at T1, HG had higher scores than LG, except for morphological variety, where both groups had the same mean, and lexical density, where LG's average was actually higher than HG's. At T2, the means show that HG had a higher score than LG on all measures except lexical density, where scores were the same for both groups. The trend that emerges from the data is that generally, HG appears to perform better than LG on most of the measures.

The trends presented above, which appear to indicate that HG tends to perform better than LG on most measures, were further explored through statistical analysis to determine whether they were indeed a true reflection of reality. Differences between HG and LG were calculated for all measures

[11] Both LG and HG were tested on all measures at both testing times to determine whether they followed a normal distribution or not. In the case of normal distributions, parametric tests were used, and in the case of groups that did not present a normal distribution, non-parametric tests were used. See Appendix C for a complete list of the groups that followed a normal distribution and those that did not.

of lexical complexity at the two testing times, and then submitted to inferential analysis in order to determine whether the differences between HG and LG were significant on any of these measures. The results of these analyses showed that none of the differences were significant (all *p*s > . 005).[12]

The lack of significant differences between groups on all measures at both testing times suggests that both groups had similar lexical complexity both at the beginning and at the end of the IE program.

TABLE 5: ALL LEXICAL MEASURES: MEANS, STANDARD DEVIATIONS AND RANGE (TIME 1 AND TIME 2) ACCORDING TO PROFICIENCY GROUP

		TIME 1					TIME 2				
		Mean	**SD**	**Min**	**Max**	**Range**	**Mean**	**SD**	**Min**	**Max**	**Range**
Density	**LG**	0.44	0.06	0.34	0.56	0.22	0.42	0.04	0.36	0.55	0.19
	HG	0.42	0.07	0.30	0.54	0.24	0.42	0.05	0.28	0.51	0.23
Diversity	**LG**	30.69	6.05	20.29	46.15	25.86	33.81	5.36	26.32	48.39	22.97
	HG	32.62	6.56	22.06	46.43	24.37	37.11	6.31	20.74	49.57	28.83
Reliance on French	**LG**	8.71%	5.57%	0%	21.82%	21.82%	3.25%	2.58%	0%	8.67%	8.67%
	HG	5.90%	4.43%	0%	18.75%	18.75%	1.88%	1.58%	0%	5.73%	5.73%
Morphological variety	**LG**	1.04	0.03	1.00	1.10	0.10	1.06	0.04	1.00	1.18	0.18
	HG	1.04	0.04	1.00	1.15	0.15	1.07	0.04	1.00	1.18	0.18
Variety of K1 families	**LG**	29.89	9.49	14.00	54.00	40.00	43.11	8.32	31.00	63.00	32.00
	HG	33.89	8.58	19.00	51.00	32.00	49.29	9.12	32.00	72.00	40.00

6.2.2 Gains in lexical complexity from T1 to T2

6.2.2.1. Individual gains

Results of the differences between T1 and T2 on all five measures of LC are presented in Table 6, and show that over the course of the IE program, both LG and HG made gains in LC. Statistical analysis showed that for the LG, significant gains were made in terms of lexical diversity ($t(26) = -5.235$, $p = .000$), reliance on French ($t(26) = 7.677$, $p = .000$), and number of K1 families

[12] Bonferroni corrections were applied to all analyses, and the α level was set at .005 to adjust for multiple comparisons (five variables).

(Z = -4.520, *p* = .000). For the HG, significant gains were made in terms of reliance on French (Z = -3.388, *p* = .001) and number of K1 families (*t*(27)=-6.046, *p*=.000). These results suggest that students from both proficiency groups made gains in the same areas during the IE program. Overall, seen in terms of these measures, both proficiency groups made significant progress in LC over the five-month period.

TABLE 6: DIFFERENCES IN ALL LEXICAL MEASURES: MEANS, STANDARD DEVIATIONS, MINIMUM AND MAXIMUM (TIME 1 AND TIME 2) ACCORDING TO PROFICIENCY GROUP

Measure	Group	Mean	SD	Min	Max
Density	LG	-0.01	0.06	-0.12	0.07
	HG	-0.01	0.04	-0.08	0.06
Diversity	LG	3.12	4.98	-5.24	13.39
	HG	4.49	5.40	-6.15	14.99
Reliance on French	LG	-5.46%	4.43%	-15.82%	3.73%
	HG	-4.02%	4.38%	-15.20%	3.98%
Morphological variety	LG	0.02	0.05	-0.05	0.13
	HG	0.03	0.05	-0.10	0.18
Variety of K1 families	LG	13.22	9.39	-13.00	28.00
	HG	15.39	9.50	-2.00	35.00

6.2.2.2. Differences between gains

The gains made by LG and HG (see Table 6) were compared in order to determine whether they were significantly different. Statistical analysis through a paired *t*-test showed that the gains for both groups were not significantly different on any of the measures (all *p*s > .005), indicating that the two groups made gains that were comparable in size.

6.2.3 Vocabulary test

At both testing times, students also completed a two-part vocabulary test, which required them to translate a series of 60 items from French to English, and another series of 60 items from English to French. Overall results on the test at T1 were used to divide the students into two proficiency groups and statistical analysis through a paired *t*-test showed that the groups were significantly different on this vocabulary measure (*t*(53) = -9.084, *p* = .000). At T2, scores on the vocabulary test for all

students were once again compiled, and the two groups were compared. Once again, statistical analysis showed that the groups formed at T1 were still significantly different on the vocabulary measure at T2 ($t(53) = -7.505$, $p = .000$). Gains from T1 to T2 were then calculated, and results showed that both LG and HG made significant gains between the two testing times ($ps = .000$), and that there was no significant difference between the size of the gains made by the two groups ($t(26) = 0.037$, $p = .505$)

6.2.4. Overall differences between groups

Overall, the results presented above suggest that there are no significant differences in terms of the gains in LC made by low-proficiency and high-proficiency students over the course of a five-month IE program. Although analysis of the median split showed that the two groups of students were significantly different with respect to vocabulary knowledge, there was nonetheless some overlap in the vocabulary scores. As such, in order to make highly distinct proficiency groups, the entire cohort was divided into three groups: low, mid and high. All of the analyses in section 6.2 were then performed a second time, but using only the low (n = 18) and high groups (n = 18) (i.e., the bottom 33% and the top 33%) in order to ensure completely distinct proficiency groups. For this second analysis, results were virtually similar to the first, suggesting that the lack of significant inter-group differences reported above was not due to potential problems resulting from the use of the median split to create proficiency groups.

6.3. Summary of results

Results for the intact group suggest that there is indeed development of LC over the course of IE programs, as there appears to be development both in terms of the size and the level of difficulty of the vocabulary, as shown by progress in measures of diversity, reliance on French, morphological variety and number of K1 families.

In terms of differences between proficiency groups, initial results lead to believe that students with different proficiency levels do not differ in terms of the progress that they make. Results showed that

although students appeared to be significantly different in terms of their vocabulary knowledge, they were quite similar in terms of their ability to use the language to communicate. This may be due to the fact that prior to the IE program, the students were exposed to different types and quantities of input but had little opportunity to practise using the language (see discussion for further details).

7 Discussion

The present study aimed at investigating IE students' development of lexical complexity through oral production over the course of a five-month program. While much research has assessed the development of different linguistic constructs in IE students, few studies have dealt with LC. In 2006, Horst and Collins conducted a study that investigated the development of LC through IE students' written productions. However, Freed et al. (2003) postulate that language develops differently in oral and written forms, and research has shown that although there is no single, unique characteristic that distinguishes spoken from written language, certain differences exist between these two types of language. In terms of the lexical complexity, spoken language appears to be somewhat less complex than written language by its simpler, less varied vocabulary, lower lexical diversity and lower type-token ratio (Chafe and Tannen, 1987). However, the fact that there could potentially be less observable progress in oral than in written language did necessarily predict a lack of development, and accentuated the pertinence of conducting further research focusing on oral development of LC. More specifically, the present study aimed at answering the two following research questions:

1. Does the oral lexical complexity of sixth-grade intensive English students improve over the course of a five-month program?
2. Does the development of lexical complexity differ between low- and high-level learners over the course of the five-month program?

In order to address these research questions, narrations produced by 56 students at the beginning (T1) and at the end (T2) of a five-month IE program were analyzed. The hypotheses put forth were directional in nature and predicted that there would be development in IE students' LC, and that the low proficiency learners would make greater lexical gains than the high proficiency learners.

In this study, lexical complexity was defined as "the degree of elaboration, the size, breadth, width, or richness of the learner's L2 system or 'repertoire', that is, to the number, range, variety or diversity of

different structures and items that he knows or uses . . . " (Bulté and Housen, 2012, p. 25). The concept of lexical complexity therefore refers to the complexity of the language the learner has acquired to date rather than the complexity of the language itself. Bulté and Housen further subdivide LC into three specific constructs: lexical density, lexical diversity and lexical sophistication, which combine to paint a complete picture of the learner's lexical complexity. These three constructs, along with reliance on French (a pertinent construct in an IE context) were initially measured for this study to assess the students' lexical complexity. Lexical sophistication, however, was later removed from the measures as it posed certain problems in terms of internal validity. Furthermore, morphological variety and the variety of K1 families were also measured, as qualitative observations showed that these measures might be particularly pertinent in illustrating the lexical development of IE students.

7.1 Development of lexical complexity over the course of a five-month IE program

Overall, learners' lexical density did not change but remained stable from T1 to T2. While an increase in lexical density is usually needed to suggest improvement in terms of LC, the observed stability in the measure does not necessarily entail that there was no development at all in the present context. For example, density is measured through the number of content words divided by the total (content and function) number of words, and an increase in density would accordingly suggest greater growth in content than function words. Since the measure of lexical diversity shows that students generated a greater number of different words at T2, the stability in the measure of lexical density implies that both categories (content words and function words) actually grew over the five-month period, in turn suggesting a change in LC development when taking both categories into consideration.

More specifically, for students at a more advanced level, who already have considerable mastery of a certain number of function words, growth in lexical density would go hand in hand with development of LC. However, the present context provides data from students who are at the very beginning of their L2 learning and who have little vocabulary knowledge at the start of the IE program. The fact that beginners such as these had relatively little knowledge of function words at T1 makes the increase in the use of function words an interesting element to consider when assessing LC. As

Collentine (2004) explains, function words are used by a learner to provide discursive coherence to a text. Therefore, the stability in lexical density, combined with an increase in lexical diversity (which indicates use of a greater number of different words), suggests that students in the present study used a greater number of both function and content words, which can be clearly argued as a change in LC.

Bulté and Housen (2012) also argue that lexical complexity not only consists of a combination of a broader vocabulary but also of a vocabulary that uses more difficult or "richer" words. In the present study, the notion of lexical richness was based on four measures of lexical sophistication provided by the LFP software program Vocabprofile: the proportion of K1 (1000 most frequent), K2 (1001-2000 most frequent), AWL (academic word list) off-list words. However, results in the first three categories showed no significant differences in development from T1 to T2, and results from the off-list actually suggested a setback in learners' LC. Interestingly, Horst and Collins (2006) reported similar LFP findings using Vocabprofile. One plausible explanation is that Vocabprofile may not be sensitive enough to capture changes in lexical sophistication in an IE context, since it uses frequency lists based on adult language and, as Vermeer (2004) points out, children and adults do not learn languages in the same way. Another potential explanation that should be taken into account is that the categories used by Vocabprofile were quite broad (there was evidence to support the idea that students were making progress within the K1 band, but this category was too broad to pick up on this progress) and consequently not adapted to beginner learners such as IE students.

Although Vocabprofile did not show any clear change in LC, certain qualitative findings hinted at the fact that other measures did indeed provide evidence of development of LC. First, it was observed that at T2, students appeared to use a wider range of grammatical forms and tenses than at T1. While the present tense was the main verb form used at T1, the simple past, the future, the present progressive and even the past progressive were used at T2 by a fair number of students. Certain grammatical forms, such as the use of the third person singular in the present, were also used inaccurately at T1 (*he go*) but accurately at T2 (*he goes*) by several students. Moreover, based on these findings, the types-per-family index revealed that morphological variety (i.e., number of different

grammatical forms relating to a same word family) improved significantly over time. This strongly suggests that students' LC did change, as it was clear they were able to produce more complex forms of a same word by the end of the IE program.

Another finding suggesting improvement in LC was the use of frequent words. In the present study, students appeared to use a greater variety of words at T2, but this increase was restricted to the K1 band because LFP failed to show any increase in the use of less-frequent words (i.e., at the K2 level and beyond). Other research (e.g., Johnson & Swain, 1994; Meara, Lightbown and Halter, 1997) has also shown that learners possess a very limited vocabulary at the onset of the learning programs in contexts similar to that of the present study. This would suggest that vocabulary development seems confined mainly to K1 boundaries over the course of the intensive experience. However, in this study, there was an increase in the use of words within the K1 list at the end of the program, which would suggest that there was lexical development and that this development simply took place within the K1 band. Given this finding, it is clear that a more sensitive LFP measure using smaller frequency bands is needed in order to assess IE students' lexical development more reliably.

Overall, the main findings from this study support the hypothesis that IE students' LC develops over the course of a five-month program, as shown through progress in a combination of lexical measures. These findings are in line with previous research conducted by Horst and Collins (2006), who found that while LFP did not suffice to demonstrate development of LC, additional measures could be used to show that there is indeed important lexical development made over the course of a five-month IE program.

7.2 Differences in the development of low-proficiency and high-proficiency learners

In order to investigate whether the initial level of proficiency had an impact on students' lexical development, participants were split into two groups: low proficiency (LG) and high proficiency (HG) using a written vocabulary test completed at T1. Statistical analyses of performance on this test, as a

result of the planned group split, showed that proficiency groups were significantly different. However, the results at both testing times also showed that inter-group differences (LG vs. HG) on five measures of lexical complexity were not significant. These findings therefore seemed counter-intuitive to the idea of low proficiency and high proficiency groups, as they indicated that the groups were not significantly different in terms of LC at T1.

In terms of the gains made in LC over the course of the IE program, both proficiency groups made gains. For example, LG made significant gains in three areas: lexical diversity, reliance on French, and variety of K1 families, while HG made significant gains in reliance on French and variety of K1 families. However, in terms of the actual size of the gains made, there were, again, no significant differences between the two groups.

The pattern of findings presented above show that at the start of the IE program, students of different proficiency levels were distinct in terms of their vocabulary knowledge, but not in terms of their ability to use this vocabulary to communicate. This might be attributable to the drip-feed approach used in elementary schools in Quebec, where students in all grades from first to fifth receive very limited ESL instruction (one to two hours weekly) and have virtually no opportunity to practise English either inside or outside the classroom. As a result, while students are able to build some receptive vocabulary knowledge throughout the elementary school years, there is insufficient time and practice to proceduralize the knowledge into active communication skill. Thus, in the present context, while intensive learners did differ at T1 in terms of the size of their vocabulary, perhaps as a result of individual differences or the quality of previous instruction, they were nevertheless strikingly similar with regard to their ability to use the vocabulary to communicate.

Studies involving students placed in study-abroad contexts, where they are immersed daily in the L2, have contradicting results. While some researchers (Collentine and Freed, 2004; Freed, 1995; Hernandez, 2010; Lapkin et al., 1995; Regan, 1995) have found a relationship between language exposure and the development of linguistic abilities, others (Magnan and Back, 2007; Martinsen,

2008; Segalowitz and Freed, 2004) have found no significant relationship between the two variables. Some who have observed a positive effect of study-abroad contexts on linguistic development have also highlighted the importance of students' participation in the development of oral skills (Hernandez, 2010; Lapkin et al., 1995; Ranta and Meckelborg, 2013). Their results show that the development of oral skills depends on more than simple exposure: students' involvement and the use they make of opportunities to practice influence the progress they make, underlining the important role that practice plays in the development of oral skills. Thus, the limited opportunities that the students in this study have had prior to the start of the IE program may account for the fact that, regardless of their proficiency level, they all appeared to have similar abilities to use their vocabulary knowledge in communication situations.

In terms of the progress made over the five-month period, all students, regardless of their proficiency group, made similar gains both in terms of vocabulary knowledge and LC. The fact that all the students were highly selected, and therefore, were top academic performers, might also explain why, when exposed to the same sources of input and when given the same opportunities to practise using the language, they made similar progress. Since all students were at a very basic level in terms of using their vocabulary knowledge to communicate, it would be natural for them to have comparable ability in this area by the end of the program. In terms of vocabulary knowledge, even though both groups made similar gains, it is logical for higher proficiency students to have greater vocabulary knowledge at T2, since they initially (at T1) had a larger lexcial base to build upon.

The findings of the present study are in line with those of Collins and White (2012) who, in a study involving sixth-grade IE students, did not find pretest performance to be a valid predictor of gains to be made over the course of an intensive program. One interesting finding in their study was the fact that some teachers reported that by the end of the program, there did not appear to be any really weak or strong students; they all seemed to be at a similar level. Thus, seen in light of the present context, if students' ability to use their vocabulary knowledge in communication situations at T1 in Collins and White's study were limited or nonexistent, the differences in their vocabulary knowledge may have been more apparent at the beginning of the program, creating a clearer distinction between

the weak and strong students. If all the students had developed their ability to use the vocabulary in communication situations to a comparable degree by T2, the differences that still existed (in terms of vocabulary knowledge, for example) might have been less apparent, especially if they had also developed strategies to compensate for lexical gaps by developing their communication skills.

Overall, the findings presented above do not support the hypothesis that progress in LC differs as a result of different proficiency levels. Regardless of their proficiency levels, all students made comparable progress in measures of LC. This may be explained by the fact that at T1, although students differed in terms of their vocabulary knowledge, they did not differ on any of the LC measures, and appeared to have similar, basic ability when it came to using their vocabulary knowledge in communication situations. The fact that they were all at a same starting point in terms of ability to use the vocabulary to communicate and that they all had relatively similar exposure and opportunities to use the language over the five-month period may explain why there were no significant differences in how their LC developed.

7.3 Pedagogical implications

The findings presented above indicate that there is indeed development of LC during a five-month intensive program. However, the findings also show that the basic nature of the vocabulary acquired by students before the start of the program limits the amount of progress they are able to make. The results have shown that students are learning new words within the 1000 most frequent word families of the language. Although these students also do know some words from less-frequent lists, Meara et al. (1997) point out that there are significant gaps in the students' knowledge of the 2500 most frequent word families of English. Despite the fact that vocabulary size for IE students in Quebec has yet to be assessed, it is reasonable to believe that there is a gap between the number of word families these students know and the threshold level, or the critical number of word families a learner should be familiar with in order to be able to communicate in the language. Thornbury (2002) reports that 2000 and 3000 word families are figures which are often quoted when referring to the threshold level for ESL learners, and Nation (2006) reports that in order to ensure proper comprehension of spoken texts, a vocabulary size of 6000 to 7000 word families is needed. This reinforces the

importance of helping students fill the gaps that exist in their knowledge of the 2500 most frequent word families quickly in order to provide them with the necessary vocabulary to be able to communicate and benefit from the IE experience.

For the most part, intensive English programs adopt, a communicative approach to language teaching in which, as Meara et al. (1997) point out, many teachers rely on implicit learning of vocabulary through comprehensible input, and shy away from direct, explicit vocabulary instruction. While the authors tend to suggest that much vocabulary can be acquired through comprehensible input, it is interesting to ponder whether more focused vocabulary teaching could benefit the development of LC, especially at the start of IE programs, when students have weaker vocabulary skills. While the validity of implicit or incidental learning through exposure is not contested, there is also argument for not focusing solely on this type of instruction, and for combining it with more explicit or intentional instruction (Sökmen, 1997). Sökmen raises various potential problems that may stem from focusing solely on implicit vocabulary teaching, such as the fact that it is a slow, error-prone process and that it does not always lead to word retention. Moreover, focus on implicit instruction could foster low levels of comprehension when vocabulary knowledge is insufficient, hindering not only the vocabulary learning process, but also comprehension during communication situations and the completion of the task at hand. It may therefore be appropriate to consider whether the addition of explicit vocabulary teaching in the IE classroom could lead students to master a greater proportion of the most-frequent words of the language faster.

Lee (2003) also reports that explicit vocabulary instruction helps students convert recently learned or recognized vocabulary into productive knowledge. Although explicit instruction alone may not be sufficient in order to ensure long-term memorization, Lee argues that, combined with use of vocabulary in context, explicit instruction could also favour vocabulary retention. In teaching vocabulary explicitly, Sökmen mentions the importance of building a large sight vocabulary, either by learning lists of frequent or less-frequent words. Although there are arguments in favour of both approaches, in the present ESL context, learning to quickly master the more frequent words of the language would benefit the students, as it would equip them with the basic knowledge needed to make the most of the intensive experience.

Research on the lexical environments provided by ESL classrooms in Quebec (Meara et al., 1997) has shown that the great majority of the vocabulary to which learners are exposed on a daily basis is limited to the 2500 most-frequent word families of English. Although their initial assumption was that the IE classroom did not provide a lexically-rich environment for the students, the authors revised their position, partly due to the fact that when students' vocabulary knowledge was taken into consideration, an environment composed of the 2500 most frequent word families provided ample opportunity for the students to learn new words. Explicit vocabulary instruction aimed at equipping the students with basic vocabulary lists composed of the 2000 most frequent word families of the language would therefore enable teachers to progress to using a richer, more complex vocabulary in the classroom, thereby exposing students to more varied sources of input that could lead to the students making greater gains in terms of LC over the course of an IE program.

7.4 Limitations

The study's main goal was to assess IE students' development of LC, and although this overall goal was achieved, certain limitations should be taken into account when interpreting the findings and should be considered for future research on the topic.

Firstly, as Bulté and Housen (2012) point out, there is no commonly accepted definition of LC. This also means that there is no single measure or set of measures used to assess this concept in research. From one study to another, the definition and operationalization of LC may vary, rendering it difficult to make valid comparisons of results across studies. The only element that appears to be consistent across studies was the importance of using multiple measures to assess LC, which is why the present study used a combination of five measures in order to assess different sub-constructs of LC.

Secondly, problems were found with the way lexical sophistication was assessed, creating issues as to the validity of the results of the LFP presented in the study. It was believed that the lack of progress shown by the results may not have accurately reflected the changes in students' lexical sophistication; consequently, the measure had to be removed from the analysis. While Vocabprofile's general validity is not being questioned in the present study, a measure that would have made it possible to create a more accurate profile of students' lexical sophistication through frequency lists that were adapted to the learners' lexical ability would have been better suited to assessing the development of lexical sophistication.

Thirdly, in terms of the second research question, which aimed at determining whether students with different initial levels of proficiency stood to make different gains in LC, it is important to consider the influence that the measure used to group the students may have had. Indeed, students were divided into two groups (high and low proficiency) according to their score on a two-part vocabulary test. The choice of this measure was based on the fact that there were a number of different measures of LC, and splitting the groups on one of these measures may have influenced the results. By splitting the groups on a different vocabulary measure, one that was independent from the measures of LC, none of the LC measures was favoured. While the two groups were significantly different on the results of this vocabulary test (both at T1 and T2), they were not significantly different on any of the LC measures. Although the lack of significant difference was attributed to the fact that students may all have had a similar level of LC at T1, it is also important to consider the possibility that splitting the groups on a different measure (lexical diversity at T1, for example) may have yielded different results, and may have produced significant differences between groups in terms of the development of LC.

Finally, the oral task used to elicit vocabulary may also have constituted a limitation to the study. Picture stories are used in research because they are believed to give the researcher the ability to control the language, while also providing the participant with the opportunity to truly reflect on his or her linguistic abilities (Rossiter et al., 2008). In the present study, however, it should be noted that the adaptation of the Frog Story was quite simple, and did not necessarily offer the students the opportunity to use a much wider range of vocabulary at T2. Thus, the task itself may have limited the

vocabulary produced by the students, and it is possible that the use of a different picture story or a different type of task may have given the students the opportunity to better showcase their lexical abilities.

8. CONCLUSION

The main goal of the present study was to investigate the development of LC in sixth-grade students participating in a five-month IE program. The study followed 56 students, who were asked to participate in an oral narration task both at the beginning and at the end of the IE program. Transcripts of these narrations were used to assess LC through a series of measures, as several researchers (Bulté and Housen, 2012; Daller et al., 2003; Horst and Collins, 2006) stress the importance of the use of a combination of measures to paint a complete picture of a learner's LC.

Results showed that although clear progress cannot be established for all measures of LC, significant gains in a number of areas indicate that there is indeed development in students' overall LC. The decrease in the use of French, the increase in lexical diversity and the increase in the number of words from the K1 list demonstrate that by the end of the program, students have a larger L2 vocabulary. The increase in morphological variety suggests that students are also using a greater number of more complex forms of words. Although no significant development was found in terms of lexical density, the results presented above illustrate progress that is in line with Bulté and Housen's (2012) definition of LC, according to which LC refers to the size but also the richness of a learner's vocabulary. In terms of development according to proficiency groups, the study was unable to show that students at a lower proficiency level stand to make greater gains over the course of an IE program. All students, regardless of their proficiency group, appeared to make comparable progress in terms of LC. This may be due to the fact that, at the beginning of the program, the students did not differ on any of the LC measures, and all appeared to have basic skills in using the vocabulary in communication situations. Finally, when exposed to the same language and given similar opportunities to practise and use the language in the intensive program, they made comparable gains in LC.

The findings presented above contribute to furthering our knowledge of how IE students' lexical abilities develop. Previous research by Meara et al. (1997) has shown that most of the input available to the students in the classroom is composed of the 2500 most-frequent words of the language. Such

findings, combined with the idea that students' vocabulary does not appear to grow beyond K1, could lead one to question what kind of impact classroom input has on students' development of LC, and whether a more lexically rich environment (containing a greater proportion of words beyond the 2500 most frequent), coupled with explicit vocabulary teaching, could contribute to students' use of an increased proportion of less-frequent words.

8.1. Further research

The findings of the present study give rise to potential questions for further research. First, the inability to show any progress in terms lexical sophistication through LFP (as measured through Vocabprofile) underscores the importance of further research conducted with measures that are better suited to young IE learners. Although LFP measured through Vocabprofile appears to have been successful in demonstrating progress made by ESL learners in a study conducted by Lo and Murphy (2010), it did not succeed in showing progress in lexical sophistication in the present study. It was hypothesized that this was due to the fact that students' progress within the frequency categories presented by Vocabprofile was too small for the software to pick up on. Moreover, the progress made by the students appears to be, for the most part, contained within the K1 band. Thus, measures based on frequency lists of young ESL learner language, and that also use smaller frequency bands might be better suited to illustrate how IE students' lexical knowledge develops in the early stages of L2 learning.

Finally, although the present research has shown development of LC over the course of a five-month program, it has not investigated the long-term effects of the IE program on students' LC. IE programs in Quebec are offered to students for one year only, in fifth or sixth grade, and although enriched ESL programs are offered in some high schools, these programs are not comparable to IE programs in terms of the amount of L2 exposure and practice they provide (Lightbown and Spada, 1991). It would therefore be important to conduct further research to investigate whether students' LC remains stable, develops further or even perhaps regresses after they have left the IE program and returned to regular ESL instruction.

REFERENCES

Billy, L. (1980). Expérimentation d'une nouvelle approche en immersion. *The Canadian Modern Language Review, 36*, 422-433.

Bournot-Trites, M. (2007). Quality of writing on the intermediate level in French immersion: Effect of an intensive program and hypothesis of a plateau effect. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée, 10*(1), 7-24.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamin Publishing.

Bulté, B., Housen, A., Pierrard, M., & Van Daele, S. (2008). Investigating lexical proficiency development over time – The case of Dutch-speaking learners of French in Brussels. *Journal of French Language Studies, 18*(3), 277-298.

Chafe, W., & Tannen, D. (1987). The relation between written and spoken language. *Annual Review of Anthropology, 16*, 383-407.

Cobb, T. (2014). Web Vocabprofile. (Adaptation of Heatley, Nation & Coxhead's (2002) *Range*) [software]. Available from: http://www.lextutor.ca/vp/

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*(2), 227-248.

Collentine, J., & Freed, B. F. (2004). Learning context and its effects on second language acquisition. *Studies in Second Language Acquisition, 26*(2), 153-356.

Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in l2 instruction. *TESOL Quarterly, 33*(4), 655-680.

Collins, L., & White, J. (2011). An intensive look at intensity and language learning. *TESOL Quarterly, 45*(1), 106-133.

Collins, L., & White, J. (2012). Closing the gap: intensity and proficiency. In C. Munoz (Ed.), *Intensive exposure experiences in second language learning* (pp. 45-65). Great Britain: Multilingual Matters.

Crystal, D. (1997). *The Cambridge encyclopedia of language* (2nd ed.). New York: Cambridge University Press.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics, 24*(2), 197-222.

Edwards, R., & Collins, L. (2011). Lexical frequency profiles and Zipf's law. *Language Learning, 61*(1), 1-30.

Edwards, R., & Collins, L. (2013). Modelling second language vocabulary learning. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 157-184). Amsterdam: John Benjamins.

Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299-323.

Freed, B., So, S., & Lazar, N. A. (2003). Language learning abroad: How do gains in written fluency compare with gains in oral fluency in French as a second language? *ADFL Bulletin, 34*(3), 34-40.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In V. Regan (Ed.), *Contemporary approaches to second language acquisition in social context: Crosslinguistic perspectives* (pp. 123-148): John Benjamins Publishing Company.

Germain, C., Lightbown, P. M., Netten, J., & Spada, N. (2004). Intensive French and intensive English: Similarities and differences. *The Canadian Modern Language Review, 69*(3), 409-430.

Harley, B., & King, M. L. (1989). Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition, 11*, 415-439.

Heatley, A., Nation, I.S.P. & Coxhead, A. (2002). RANGE and FREQUENCY [software]. Available at http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Hernandez, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *The Modern Language Journal, 94*(4), 600-617.

Horst, M., & Collins, L. (2006). From faible to strong: How does their vocabulary grow? *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 63*(1), 83-106.

Isbell, R., Sobol, J., Lindauer, L., & Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Childhood Education Journal, 32*(3), 157-163.

Ishikawa, T. (2007). The effect of manipulating task complexity along the [+ / - here-and-now] dimension on L2 written narrative discourse. In M. d. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136-156). Clevedon: Multilingual Matters.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57-84.

Johnson, R. K., & Swain, M. (1994). From core to content: Bridging the L2 proficiency gap in late immersion. *Language and Education, 8*(4), 211-229.

Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Québec. *Contemporary Approaches to Second Language Acquisition in Social Context: Crosslinguistic Perspectives* (pp. 67-94). Amsterdam: John Benjamins Publishing Company.

Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal, 25*(2), 21-33.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*(2), 255-271.

Laufer, B. (2005). Lexical frequency profiles: From monte carlo to the real world. A response to Meara (2005). *Applied Linguistics, 26*(4), 582-588.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307-322.

Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System, 31*(4), 537-561.

Lightbown, P. M. (2012). Intensive L2 instruction in Canada: Why not immersion? In C. Munoz (Ed.), *Intensive exposure experiences in second language learning*. North York: Multilingual Matters.

Lightbown, P. M., & Spada, N. (1991). Étude des effets à long terme de l'apprentissage intensif de l'anglais, langue seconde, au primaire. *Canadian modern language review, 48*(1), 90-117.

Lightbown, P. M., & Spada, N. (1994). An innovative program for primary ESL students in Quebec. *TESOL Quarterly, 28*(3), 563-579.

Lightbown, P. M., & Spada, N. (1997). Learning English as a second language in a special school in Quebec. *The Canadian Modern Language Review, 53*(2), 315-355.

Lo, Y. Y., & Murphy, V. A. (2010). Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education, 24*(3), 215-238.

Magnan, S. S., & Back, M. (2007). Social interaction and linguistic gain during study abroad. *Foreign Language Annals, 40*(1), 43-61.

Malvern, D., & Richards, B. (1997). A new measure of lexical diversity. In British association for applied linguistics (Ed.), *Evolving Models of Language* (pp. 58-71). Great Britain: Multilingual Matters.

Martinsen, R. A. (2008). Short-term study abroad: Predicting changes in oral skills. *Foreign Language Annals, 43*(3), 504-530.

Mayer, M. (1969). *Frog, where are you?* New York: Dial Books for Young Readers.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing, 15*(3), 323-337.

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26*(1), 32-47.

Meara, P., Lightbown, P. M., & Halter, R. H. (1997). Classrooms as lexical environments. *Language Teaching Research, 1*(1), 28-47.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL - International Review of Applied Linguistics in Language Teaching, 45*(3), 241-259.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-82.

Ortega, L. (1995). The effect of planning in oral narratives by adult learners of Spanish (Research note No. 15). Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center.

Ovtcharov, V., Cobb, T., & Halter, R. (2006). The lexical richness of oral productions: A reliable measure of level of language competence. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 63*(1), 107-125.

Ranta, L., & Meckelborg, A. (2013). How much exposure to English do international graduate students really get?: Measuring language use in a naturalistic setting. *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 69*(1), 1-33.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Regan, V. (1995). The acquisition of sociolinguistic active speech norms: effects of a year abroad in L2 learners of French. In B. Freed (Ed.), *The linguistic impact of study abroad* (pp. 245-268). Amsterdam / Philadephia: John Benjamins Publishing Co.

Rossiter, M. J., Derwing, T. M., & Jones, V. M. L. O. (2008). Is a picture worth a thousand words? *TESOL Quarterly, 42*(2), 325-329.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*, 173-199.

Serrano, R. (2012). Is intensive learning effective? Reflecting on the results from cognitive psychology and the second language acquisition literature. In C. Munoz (Ed.), *Intensive exposure experiences in second language learning* (pp. 3-22). North York: Multilingual Matters.

Serrano, R., & Munoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System, 35*(3), 305-321.

Serrano Serrano, R. (2011). The effect of program type and proficiency level on learners' written production. *Revista Espanola de Linguistica Aplicada, 24*, 211-226.

Sökmen, A. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. M. (Eds.), *Vocabulary: Description, acquisition and pedagogy* (p. 150-161). Cambridge: Cambridge University Press.

Spada, N. (1986). The interaction between type of contact and type of instruction: Some effects on the L2 proficiency of adult learners. *Studies in Second Language Acquisition, 8*, 181-200.

Spada, N., & Lightbown, P. M. (1989). Intensive ESL programmes in Quebec primary schools. *TESL Canada Journal/Revue TESL du Canada, 7*(1), 11-32.

SPEAQ. (2012). Guide d'implantation de l'enseignement intensif de l'anglais, langue seconde, au Québec. Montréal: SPEAQ.

Stern, H. H. (1985). The time factor and compact course development. *TESL Canada Journal, 3*, 13-27.

Tidball, F., & Treffers-Daller, M. (2008). Analysing lexical richness in French learner language: what frequency lists and teacher judgments can tell us about basic and advanced words. *Journal of French Language Studies, 18*(3), 299-313.

Thornbury, S. (2002). *How to teach vocabulary*. Essex: Pearson Education.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*(1), 65-83.

Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 173-189). Amsterdam: Benjamins.

White, J. L., & Turner, C. E. (2005). Comparing children's oral ability in two ESL programs. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 61*(4), 491-517.

Appendix A

Vocabulary Test (Administered at T1 and T2)

Name:

List A: English to French, Mixed Rating Vocabulary

1. to hurt:
2. new:
3. neighbor:
4. low:
5. to hate:
6. teeth:
7. mean:
8. to taste:
9. talkative:
10. stick:
11. cold:
12. to hear:
13. leaves:
14. shy:
15. to smell:
16. towel:
17. thirsty:
18. to cry:
19. water:
20. expensive:
21. to laugh:
22. star:
23. boring:
24. to try:
25. eraser:
26. proud:
27. to hold:
28. picture:
29. dirty:
30. to scratch:
31. movie:
32. sleepy:
33. to hunt:
34. kitchen:
35. quiet:
36. to travel:
37. birthday:
38. angry:
39. to like:
40. house:
41. hungry:

42. to jump:

43. apple:

44. slow:

45. to relax:

46. farm:

47. good:

48. to run:

49. turtle:

50. bad:

51. to fish:

52. snow:

53. big:

54. to read:

55. napkin:

56. happy:

57. to open:

58. ice cream:

59. short:

60. afternoon:

Name:

List B: French to English, Mixed Rating Vocabulary

1. plier:
2. lumière:
3. gentil:
4. brocher:
5. sable:
6. pesant:
7. porter:
8. pluie:
9. sombre:
10. laisser tomber:
11. boue:
12. sucré:
13. croire:
14. fourchette:
15. couloir:
16. mouiller:
17. apporter:
18. lait:
19. piscine:
20. fort:
21. garder:
22. avion:
23. dur:
24. oublier:
25. cadeau:
26. doux:
27. penser:
28. bonbon:
29. facile:
30. appartenir:
31. verre:
32. vieux:
33. nager:
34. chaise:
35. violet:
36. jouer:
37. soleil:
38. jaune:
39. voir:
40. foulard:
41. drôle:

42. donner:

43. oiseau:

44. petit:

45. raconter:

46. ami:

47. rose:

48. se souvenir:

49. jeu:

50. salé:

51. boire:

52. sec:

53. frère:

54. grand:

55. bateau:

56. chaud:

57. fenêtre:

58. courageux:

59. occupé:

60. trouver:

Appendix B

The Frog Story (Adapted from Mayer, 1969)

A. Voici ce qui s'est passé hier et ce matin:

1.)

2.)

3.)

B. Voici ce qui se passe <u>maintenant</u> :

5.)

6.)

7.)

8.)

9.)

10.)

Voici ce qui se passe <u>maintenant</u> (suite) :

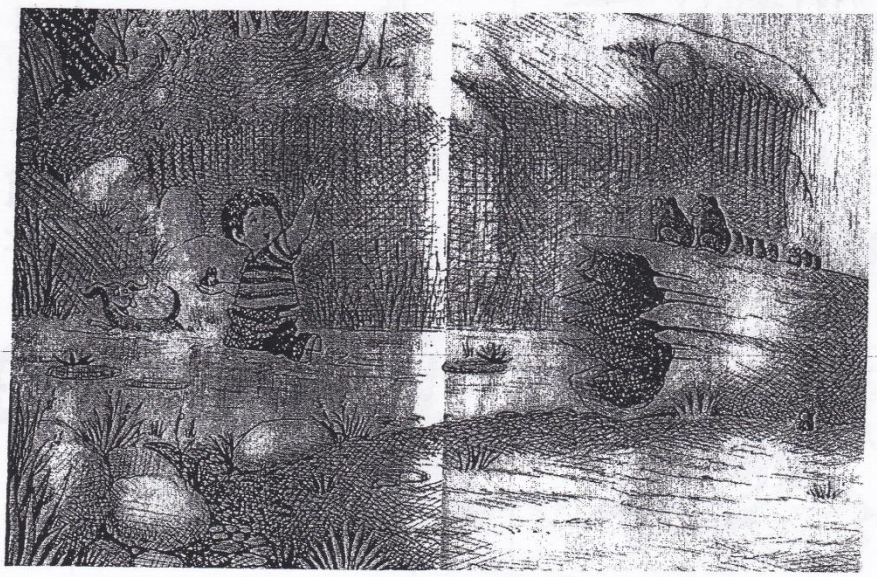
11.)

12.)

13.)

Le garçon retourne à la maison...

C. Voici ce qui se passera <u>demain</u> :



4

Appendix C

GROUP	MEASURE	TIME 1	TIME 2
Intact group	Density	Normal distribution	Normal distribution
	Diversity	Normal distribution	Normal distribution
	K1	Normal distribution	Normal distribution
	K2	Normal distribution	Normal distribution
	AWL	No normal distribution	No normal distribution
	Off-list	Normal distribution	Normal distribution
	Reliance on French	No normal distribution	No normal distribution
	Morphological variety	No normal distribution	No normal distribution
	Number of K1 families	Normal distribution	Normal distribution
Low-proficiency group	Density	Normal distribution	No normal distribution
	Diversity	Normal distribution	Normal distribution
	K1	No normal distribution	Normal distribution
	K2	Normal distribution	Normal distribution
	AWL	No normal distribution	No normal distribution
	Off-list	No normal distribution	Normal distribution
	Reliance on French	Normal distribution	Normal distribution
	Morphological variety	Normal distribution	No normal distribution
	Number of K1 families	Normal distribution	Normal distribution
High-proficiency group	Density	Normal distribution	Normal distribution
	Diversity	Normal distribution	Normal distribution
	K1	Normal distribution	Normal distribution
	K2	Normal distribution	No normal distribution
	AWL	No normal distribution	No normal distribution
	Off-list	Normal distribution	Normal distribution
	Reliance on French	No normal distribution	Normal distribution
	Morphological variety	No normal distribution	No normal distribution
	Number of K1 families	Normal distribution	Normal distribution