The Acquisition of Formulaic Sequences in High-Intermediate ESL Learners

Hoa Nguyen

University of Pennsylvania, nghoa@gse.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/edissertations

Part of the Education Commons

Recommended Citation
http://repository.upenn.edu/edissertations/1385

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/1385
For more information, please contact libraryrepository@pobox.upenn.edu.
The Acquisition of Formulaic Sequences in High-Intermediate ESL Learners

Abstract
This study investigates the relative effectiveness of three types of form-focused instruction on the acquisition of English formulaic sequences (FSs), which learners of all proficiency levels seem to struggle with. 40 Mandarin-speaking graduate students were randomly assigned to 4 groups: 1 control group and 3 treatment groups. Over 2 weeks all groups received 3 reading comprehension lessons based on 3 reading passages with 10 target FSs in each. The control group received no instruction on FSs, while in the three treatment groups, after the reading comprehension activity, learners received three types of intervention: (i) Input Enhancement in combination with Explicit Instruction, (ii) Collaborative Gap-fill tasks, and (iii) Spot-the-Difference tasks. A Vocabulary Knowledge Scale test and an Awareness test were used as pre-tests, while immediate and delayed post-tests included a cued gap-fill test followed by a multiple-choice question test and the same Awareness test. Findings obtained from ANOVAs and Cohen's d effect size calculations showed that three types of form-focused instruction benefited learners in acquiring higher levels of productive and receptive knowledge of new FSs. Form-focused instruction was particularly successful in helping learners produce the newly learnt FSs in a different context. Results also revealed that effective retention of the target FSs' form was associated with higher levels of productive knowledge. Furthermore, learners' engagement in understanding the meaning of new FSs in their context had a durable positive effect on their retention of the form and productive knowledge of these FSs. Direct instruction of new FSs' meaning helped learners retain meaning most efficiently, while explicit strategy teaching tended to enhance learners' ability to notice FSs in L2 input. Correlation analyses also suggested a complex interaction of factors related to the acquisition of FSs as frequency, n-gram length and MI Score separately could not fully account for the levels of success in acquiring new FSs receptively and productively among learners.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Education

First Advisor
Yuko G. Butler

Keywords
form-focused instruction, formulaic sequences, second language acquisition, task-based language learning

Subject Categories
Education

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/1385
THE ACQUISITION OF FORMULAIC SEQUENCES
IN HIGH-INTERMEDIATE ESL LEARNERS

Hoa Thi Hong Nguyen

A DISSERTATION

in

Education

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

[Signature]

Yuko G. Butler, Associate Professor of Education

Graduate Group Chairperson

[Signature]

Stanton E. F. Wortham, Judy & Howard Berkowitz Professor of Education

Dissertation Committee

Yuko G. Butler, Associate Professor of Education

Diane Larsen-Freeman, Professor Emerita, Education and Linguistics, University of Michigan

Nancy H. Hornberger, Professor of Education

Christina Frei, Adjunct Associate Professor of Education
To Quang Anh and Nai (Ha Anh) for the love they bring to my life
ACKNOWLEDGMENTS

I would like to thank Dr. Yuko Butler, for her constant support and prompt, detailed feedback. I am also thankful to Dr. Christina Frei, who has been giving me valuable advice on designing testing instruments and refining my arguments. I have received tremendous support from Professor Diane Larsen-Freeman, whose knowledge and advice have been crucial to the completion of my study. Professor Larsen-Freeman’s knowledge inspires me, while her warm-heartedness and her confidence in my abilities touch my heart and give me strength. I am also indebted to Professor Nancy H. Hornberger in many indescribable ways. Without Professor Hornberger’s care and support, especially after Professor Teresa Pica, my initial supervisor at Penn, passed away in 2011, completing the program would have been much more challenging for me.

I would also like to thank all of my research participants, as well as my colleagues and friends, many in the Educational Linguistics division, who helped with piloting my research instruments.

I especially want to express my gratitude and thankfulness to members of my family: my husband, my daughter, my parents, my sister, my in-laws, for their unfaltering support of my academic pursuit.
ABSTRACT

THE ACQUISITION OF FORMULAIC SEQUENCES IN HIGH-INTERMEDIATE ESL LEARNERS

Hoa T. H. Nguyen
Yuko G. Butler

This study investigates the relative effectiveness of three types of form-focused instruction on the acquisition of English formulaic sequences (FSs), which learners of all proficiency levels seem to struggle with. 40 Mandarin-speaking graduate students were randomly assigned to 4 groups: 1 control group and 3 treatment groups. Over 2 weeks all groups received 3 reading comprehension lessons based on 3 reading passages with 10 target FSs in each. The control group received no instruction on FSs, while in the three treatment groups, after the reading comprehension activity, learners received three types of intervention: (i) Input Enhancement in combination with Explicit Instruction, (ii) Collaborative Gap-fill tasks, and (iii) Spot-the-Difference tasks. A Vocabulary Knowledge Scale test and an Awareness test were used as pre-tests, while immediate and delayed post-tests included a cued gap-fill test followed by a multiple-choice question test and the same Awareness test. Findings obtained from ANOVAs and Cohen’s d effect size calculations showed that three types of form-focused instruction benefited learners in acquiring higher levels of productive and receptive knowledge of new FSs. Form-focused instruction was particularly successful in helping learners produce the newly learnt FSs in a different context. Results also revealed that effective retention of the target FSs’ form was associated with higher levels of productive knowledge. Furthermore, learners’ engagement in understanding the meaning of new FSs in their context had a durable positive effect on their retention of the form and productive knowledge of these FSs. Direct instruction of new FSs’ meaning helped learners retain meaning most efficiently, while explicit strategy teaching tended to enhance learners’ ability to notice FSs in L2 input. Correlation analyses also suggested a complex interaction of factors related to the acquisition of FSs as frequency, n-gram length and MI Score separately could not fully account for the levels of success in acquiring new FSs receptively and productively among learners.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ iii

ABSTRACT .......................................................................................................................... vi

LIST OF TABLES ................................................................................................................ vii

LIST OF ILLUSTRATIONS ................................................................................................. ix

CHAPTER 1: INTRODUCTION ............................................................................................... 1
  1.1. Vocabulary acquisition in SLA research ................................................................. 2
  1.2. Vocabulary research and the “word” ................................................................. 5
  1.3. Formulaic sequences in second language acquisition ........................................ 7

CHAPTER 2: LITERATURE REVIEW .................................................................................. 12
  2.1. Beyond the word: Formulaic sequences ............................................................... 12
  2.2. The ubiquity of formulaic sequences ................................................................ 16
  2.3. The roles of formulaic language in second language acquisition ..................... 22
  2.4. L2 learners have difficulty with FSs ................................................................. 25
  2.5. The processing of FSs ...................................................................................... 30
  2.6. Intervention to facilitate the acquisition of FSs ................................................ 35
  2.7. Task-based intervention in teaching FSs ........................................................... 43
  2.8. Research questions ............................................................................................ 50

CHAPTER 3: METHODOLOGY ......................................................................................... 52
  3.1. Participants and Setting ....................................................................................... 52
  3.2. Variables: Independent and Dependent ............................................................ 52
  3.3. Materials ............................................................................................................. 53
  3.4. Pre-tests, post-tests and delayed post-tests ......................................................... 56
  3.5. Treatment ........................................................................................................... 60
  3.6. Qualitative data .................................................................................................. 64
  3.7. Data Analysis ..................................................................................................... 64

CHAPTER 4: FINDINGS ..................................................................................................... 66
  4.1. Productive Knowledge ....................................................................................... 66
    4.1.1. Overall Results ............................................................................................ 67
    4.1.2. Results from the immediate post-test ....................................................... 70
    4.1.3. Results from the Delayed Post-test ............................................................ 74
  4.2. Receptive Knowledge ......................................................................................... 79
    4.2.1. Overall Results ........................................................................................... 80
    4.2.2. Results from the Immediate Post-test ....................................................... 81
    4.2.2. Results from the Delayed Post-test ............................................................ 88
  4.3. The Awareness Test ............................................................................................ 93
    4.3.1. Overall results ............................................................................................ 93
    4.3.2. Results from the Immediate Post-test ....................................................... 95
4.3.3. Results from the Delayed Post-test ................................................................. 98
4.3.4. Words underlined in the Awareness test ..................................................... 101
4.4. Frequency, n-gram length and MI score in correlation to test scores .......... 102

CHAPTER 5: DISCUSSION ......................................................................................... 104
  5.1. Do IEEI, Gap-fill and SpotDif help learners acquire higher levels of productive
       knowledge of FSs? If yes, which type of treatment is more effective? ............ 104
  5.2. Do IEEI, Gap-fill and SpotDif help learners acquire higher levels of receptive knowledge
       of FSs? If yes, which type of treatment is more effective? .......................... 109
  5.3. Do IEEI, Gap-fill and SpotDif help learners improve their ability to notice FSs in a reading
       passage? If yes, which type of treatment is more effective? ...................... 114
  5.4. Are frequency, MI-score, and n-gram length of FSs related to learners’ acquisition of
       receptive and productive knowledge of FSs? ............................................ 117
  5.5. Other important findings ............................................................................. 122
       Control group: Did incidental learning occur? ........................................... 122
       Acquisition of FSs is incremental and complex .......................................... 124
  5.6. Limitations ..................................................................................................... 126
  5.7. Implications ................................................................................................... 128
  5.8. Concluding remarks ...................................................................................... 131

APPENDICES ........................................................................................................... 135
Appendix A: List of target FSs and Reading Passages .......................................... 135
Appendix B: Pre-test 1 – Vocabulary Knowledge Scale ........................................ 143
Appendix C: Pre-test 2 Underlining useful words and phrases in passage ............ 145
Appendix D: Post-tests 1 - Productive knowledge tests ...................................... 147
Appendix E: Post-tests 2 - Receptive knowledge tests (Multiple-choice) .......... 153
Appendix F: Post-test 3 Underlining useful words and phrases in passage .......... 165
Appendix G: Test scores by FSs ........................................................................... 167
Appendix H: Number of target FSs underlined by members of the Control group in 3 treatment
       sessions ........................................................................................................ 169
REFERENCES .......................................................................................................... 170
LIST OF TABLES

Table 1: The Vocabulary Knowledge Scale (VKS) (Paribakht and Wesche, 1997) 19
Table 2: Hypothetical levels of Need, Search and Evaluation in the three treatment types 50
Table 3: Mean scores of the four groups on the Productive Knowledge pre-tests, immediate and delayed post-tests 67
Table 4: Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Productive Knowledge pre-test, immediate and delayed post-test 68
Table 5: Multiple comparisons of groups for Productive Knowledge immediate and delayed post-tests using Games-Howell 69
Table 6: Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Productive Knowledge pre-test and immediate post-test 71
Table 7: Multiple comparisons of groups for Productive Knowledge immediate post-test using Games-Howell ($\alpha = .05$) 72
Table 8: Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Productive Knowledge pre-test and delayed post-test 74
Table 9: Multiple comparisons of groups for Productive Knowledge delayed post-test using Games-Howell ($\alpha = .05$) 75
Table 10: Mean scores of the four groups on the Receptive Knowledge pre-tests, immediate and delayed post-tests 78
Table 11: Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Receptive Knowledge pre-test, immediate and delayed post-test 79
Table 12: Multiple comparisons of groups for Receptive Knowledge immediate and delayed post-tests using Games-Howell ($\alpha = .05$) 80
Table 13: Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Receptive Knowledge pre-test and immediate post-test 82
Table 14: Multiple comparisons of groups for Receptive Knowledge immediate post-test using Games-Howell ($\alpha = .05$) 83
Table 15: Mean scores of by the four groups on the Form and Meaning sections of the Receptive Knowledge immediate and delayed post-test 85
Table 16: Multiple comparisons of groups for Receptive Knowledge immediate post-test (Form section) using Games-Howell ($\alpha = .05$) 86
Table 17: Multiple comparisons of groups for Receptive Knowledge immediate post-test (Meaning section) using Games-Howell ($\alpha = .05$) 87
Table 18: Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Receptive Knowledge pre-test and delayed post-test

Table 19: Multiple comparisons of groups for Receptive Knowledge delayed post-test using Game-Howell (α = .05)

Table 20: Multiple comparisons of groups for Receptive Knowledge delayed post-test (Form section) using Bonferroni (α = .05)

Table 21: Multiple comparisons of groups for Receptive Knowledge delayed post-test (Meaning section) using Bonferroni (α = .05)

Table 22: Mean scores of the four groups on the Awareness pre-test, immediate and delayed post-test

Table 23: Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Receptive Knowledge pre-test, immediate and delayed post-test

Table 24: Results of mixed between-within subjects ANOVA for the mean scores of four groups on the Awareness pre-test and immediate post-test

Table 25: Multiple comparisons of groups for Awareness immediate post-test using Bonferroni (α = .05)

Table 26: Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Awareness pre-test and delayed post-test

Table 27: Multiple comparisons of groups for Awareness immediate post-test using Bonferroni (α = .05)

Table 28: Mean scores of the four groups on the Awareness pre-test, immediate and delayed post-test with regards to number of words underlined

Table 29: Correlation between FREQ, n-gram length and MI Score and test scores

Table 30: Summary of all results
LIST OF ILLUSTRATIONS

Figure 1: Productive Knowledge test scores in the pre-test, immediate and delayed post-tests 70

Figure 2: Productive Knowledge test scores in the pre-test and immediate post-test 73

Figure 3: Productive Knowledge test scores in the pre-test and delayed post-test 76

Figure 4: Receptive Knowledge test scores in the pre-test, immediate and delayed post-tests 81

Figure 5: Receptive Knowledge test scores in the pre-test and immediate post-test 84

Figure 6: Receptive Knowledge test scores in the pre-test and delayed post-test 90

Figure 7: Awareness test scores on the pre-test, the immediate post-test and the delayed post-test 95

Figure 8: Awareness test scores in the pre-test and immediate post-test 98

Figure 9: Awareness test scores in the pre-test and delayed post-test 101
CHAPTER 1: INTRODUCTION

This study aims to contribute to the recent escalation of research on formulaic language in second language (L2) learners. As will be discussed in later parts of the study, scholars in the fields of L2 vocabulary acquisition, corpus linguistics and L2 classroom pedagogy have brought to light the need for learners to accumulate a sizeable pool of L2 formulae in order to comprehend and produce the target language. These formulae range from phrasal verbs (e.g. *bank on*), collocations (e.g. *pay attention to, firmly entrenched*), to idioms and proverbs (e.g. *all the rage, haste makes waste*). This learning process is daunting given the length, semantic opaqueness of many, and low frequency of most formulaic sequences (FSs). Moreover, the obstacles stem from learners’ inclination to notice and learn new words, rather than lexical units beyond words, and their commonly held belief that semantically equivalent L2 words are interchangeable in larger units of language. Abundant research evidence shows that learners, regardless of their proficiency levels struggle with learning and appropriately using formulaic language.

Facilitating the acquisition of such FSs through classroom activities entails aiding learners to notice such sequences, which otherwise tend to be clouded by individual words, and to incorporate these lexical units into their receptive and productive knowledge. As necessary as this is, however, only a limited number of research studies have put different pedagogical interventions specifically designed for promoting the acquisition of L2 formulaic language to the test. Most of these studies attest to the effectiveness of instructional techniques and activities in raising learners’ awareness of formulaic language, improving their perceived fluency, and enhancing the retention of FSs. Additionally, none of those few experimental studies use, as a
treatment condition, interactive tasks, a type of classroom focus-on-form activities empirically proven to be effective for instructed L2 learners. This study, consequently, aims at comparing the effectiveness of two such tasks, collaborative Gap-fill and Spot the Difference with that of enhancing the input in combination with explicit instruction.

This chapter will discuss first the necessity for more research in vocabulary acquisition, then the necessity to extend the object of vocabulary research from individual words to larger chunks of language. It will end with a brief discussion of the ubiquity of formulaic sequences followed by a working definition of the term, a discussion of the roles of FSs in the language of both native and non-native speakers, as well as brief synopsis of research that inspires the current study.

1.1. Vocabulary acquisition in SLA research

Much of second language acquisition (SLA) research is concerned with the acquisition of grammar, since in the history of linguistics and applied linguistics, grammar and phonology have generally been treated as the core and vocabulary as periphery. The situation of lexis is known in Bloomfield’s statement (1933): “The lexicon is really an appendix of the grammar, a list of basic irregularities” (as cited in Bogaards, 1996, p. 274). Holley (1972) notes that in the three decades before the 1970s, there had been little interest in foreign language vocabulary learning, due to the influence of behaviorist psychology and linguistic structuralism, which held vocabulary learning as a complex case of stimulus-response learning; however, research findings in the field of human memory and linguistics at that time suggested that vocabulary learning is more serious and more interesting than expected. Gleitman and Landau (1994) opine that “for many years, the topic of lexical acquisition was a stepchild in linguistic inquiry” (p. 1). First language lexical acquisition is considered a simple and particularistic associative procedure that
maps perceptual experience onto phonetic entities. Mainstream linguistic theories regarding lexical rules and constructions, levels of representations, constraints on rules, underlying principles and parameters are not comparable to those of syntax (Wunderlick, 2006). Several theories besides the mainstream were developed and in the late 1980s, there was a “strong tendency towards favouring the lexicon as a structure-giving reservoir” (Wunderlick, 2006, p. 2; for a more detailed discussion, see Wunderlick, 2006). Wunderlick explains that each lexical item bears phonological, semantic and categorical properties; the lexicon is highly structured; and the distributed information of the lexicon determines the working of the components of grammar. On the upside, many researchers continue to side with lexical aspects of language. For example, Pinker (1984) and Hoey (1994, 2005) propose that the rules of a grammar are acquired by learning lexical items that instantiate them (Pinker, 1984) and that “lexis is complexly and systematically structured and grammar is an outcome of this lexical structure” (Hoey, 2005, p.1).

Second language acquisition research has long been influenced by developments in linguistic theories; hence it is no surprise that lexical acquisition theorizing in the field has not been a central concern (Appel, 1996; Gass, 1988; Laufer, 1986; Levenston, 1979; Meara, 1980; Richards, 1976; Zimmerman, 1997). Levenston (1979) sees second language lexical acquisition as not merely neglected, but a “victim of discrimination” (p. 147) and notes that in all fields of research, “language” or “interlanguage” is synonymous with “grammar” or “interlanguage grammar”. Syntactic structures are considered to be the most complex, worth researching aspects of language, while lexis is regarded as superficial, trivial and brittle pieces of crystalline structures (Aitchison, 1987). Syntax is also considered unique to humans, while animals can learn “words”. Levenston was among the first scholars to pose important questions regarding L2
lexical acquisition, including questions of whether L1 vocabulary acquisition is parallel with processes in SLA, the stages of lexicon building and the factors that influence learners’ lexicon growth and expansion, and the relationship between active/productive and passive/receptive knowledge of vocabulary. Meara (1980) reviews research on L2 vocabulary acquisition, which he describes as a research area that has been neglected in applied linguistics, and concludes that much of research in the field has been “atheoretical and unsystematic.” (p. 221). He also notes that most scholarly work has been concerned more with vocabulary teaching but has not revealed much about the processes whereby learners’ lexicon is acquired. Vocabulary learning, as Laufer (1986) and Appel (1996) argue, is the core of second language acquisition: without lexical items, grammatical rules are useless, and lexical problems are more relevant to second language didactics than phonology and syntax (Laufer, 1986). Lexical errors in learners’ language can cause nonunderstandings or misunderstandings, and second language lexicon is crucial for successful communication in the language (Gass, 1988). The task of acquiring L2 vocabulary, which consists of tens of thousands of words, most of which are polysemous, can be more daunting than the acquisition of grammar (Appel, 1996). Schmitt (2010) notes that the constant need for L2 learners to look up new words in a dictionary illustrates the importance of lexical acquisition. Furthermore, there are always strong correlations between vocabulary and various measures of language proficiency, between vocabulary and language skills (Alderson, 2005).

Vocabulary acquisition research has proliferated especially in recent years (Haastrup & Henriksen, 2001; Juffs, 2009; Laufer, 2009; Schmitt, 2008, 2010), resulting in more lexical acquisition theories and a wider pool of empirical research. However, research on L2 vocabulary acquisition continues to be eclipsed by research in other areas of language teaching and learning, and as Schmitt (2008) observes, much of the research on vocabulary learning has been
“slow to filter into mainstream pedagogy” (p. 330). Hastrup and Henriksen (2001) acknowledge that research in the field of second language vocabulary acquisition has flourished in the last decade of the 20th century, but also point out a much less favorable truth – “vocabulary is studied as a thing apart, since mainstream SLA theory has syntax as its core” (p. 69). Juffs (2009) agrees that regardless of the importance of the lexicon, it has not always been the focus of mainstream second language acquisition research. Commenting on the importance of studying the lexicon, Jackendoff asserts that almost all theories of language can agree that the lexicon contains items that contain “a long-term memory association of phonological, syntactic and semantic features” (p. 130).

1.2. Vocabulary research and the “word”

Vocabulary is commonly equaled with a collection of “words.” In fact, words are doubtlessly the central concept of much vocabulary research. Aitchison’s (1987) entire book is devoted to answering questions centering on the concept of word: How are words stored in the mind? How do people find the words they want when they speak? Do children remember words in the same way as adults? According to Aitchison (1987), it is a common folk belief that one needs to know a lot of words, and people start to worry if they cannot recall a word they want. A dictionary, a must-have companion of language learners, is compiled as a list of words in a language, each with its meanings and perhaps illustrative examples of how the word is used in a phrase or sentence. This reinforces the assumption that a word is the basic lexical unit that carries meaning (Sinclair, 2004b). Most research in second language vocabulary acquisition begins by defining what a word is and what word knowledge entails. Nation (2001) begins his classical work on L2 vocabulary learning by investigating how many words there are in a language and the ways in which words can be counted, distinguishing between high-frequency words and low-
frequency words. In literate societies, native speakers seem to master wordhood, but what segments one word from another in their perception is mostly orthographical (Himmelmann, 2006; Jones & Sinclair, 1974; Sinclair, 2004b). Apparently the segmentation of language into words is heavily influenced by orthography in literate societies. Himmelmann goes on to note that people in illiterate societies have some perception of what seems to look like the word, but it becomes less clear where the word ends and compounds, particle constructions and formulaic sequences begin. The separation of one word from another by conventionalized physical space gives words the “appearance of discrete units” (Sinclair, 2004b, p. 24); learning how to read begins with the recognition of individual words, hence the position of the word as the basic unit of linguistic analysis is naturalized.

In the same vein, Schmitt (2010) points out that “languages like English indicate individual words in text by placing spaces around them” (p. 10). Sinclair (1991) calls it word-form and defines it as “an unbroken succession of letters” (p. 28) with a special note on the hyphen and the apostrophe, about which he did not go into detail. Sinclair also points out the spacing between words is “structurally bogus, may disappear in time, as we see in maybe, anyway, another” (p. 110). According to Bybee and Scheibman (1999), echoed by Diessel (2007), it is the frequency effect that causes words which are frequently used next to each other to be automatized and become a single storage and processing unit, e.g. don’t know and dunno, hence the disappearance of word boundaries. Gardner (2007) opines that the greatest challenge of corpus-based vocabulary research is in determining what constitutes the “word”, emphasizing that this does not only influence how such items are counted and analyzed, but also has important ramifications for pedagogical theories and practices that derive from them. In reconstructing the concept of “word”, Gardner discusses multiword items as being possibly
subsumed by ‘Word’, and hypothesizes that if the work of Sinclair and his colleagues is accurate, this will be a powerful indictment of the traditional approach of counting individual word-forms in corpus linguistics, at least with regards to alphabetical languages.

The approach in corpus linguistics to analyzing language based on words stems from the tendency to associate meaning with the shortest possible segment of language and the one that shows the least variation, i.e. the word (Sinclair, 2004a). The same approach can be also found in psychological experiments since its early days, an approach whereby language is deemed a string of discrete words which can be further analyzed into morphemes, syllables, phonemes and letters, or combined together to form phrases and sentences (Shaoul and Westbury, 2011). While Sinclair names this approach the *minimal approach* (p. 280), Shaoul and Westbury refer to it as the *reductionist approach*. According to Sinclair as well as Shaoul and Westbury, this approach is limited in its ability to illuminate vocabulary research. In language teaching, Schmitt (2010) observes that “vocabulary instruction has tended to focus on individual words because they have been considered the basic lexical unit, but also because they are easier to work with than formulaic language” (p. 8).

**1.3. Formulaic sequences in second language acquisition**

Formulaic language has now become more acknowledged by researchers as an important part of the lexicon. The most well-known research that highlights the importance of formulaic sequences includes work by scholars such as Moon (1998), Nattinger and DeCarrio (1992), Pawley and Syder (1983), Schmitt and Carter (2008), Sinclair (1991). In this study, the term *formulaic sequences* (henceforth FSs), instances of formulaic language, will be used to indicate any multiword lexical items, including phrasal verbs (e.g. *mull over, wind down*), collocations
(e.g. firmly entrenched), idioms (e.g. dropped the ball, short end of the stick) - expressions that are mostly prefabricated but also allow the insertion or deletion of variable items. There are several reasons why FSs are gaining more attention in SLA research. First, they comprise a large proportion of any type of discourse, both spoken and written. Researchers have procured different percentages, depending on their methodologies and definitions of formulaic language, but they all fall within the range of 30-50% of a given text. Second, FSs are useful in acquiring a second language. Early L2 learners often resort to FSs in order to acquire L2 rules directly or indirectly, starting with simple unanalyzed chunks, then low scope patterns and fully analyzed sequences. These sequences also play a number of functions in communication. They also provide processing benefits as they reduce the time of online language processing and increase fluency.

The formation and ubiquity of these sequences as a type of linguistic construction can be explained by theories in usage-based approach to first and second language acquisition. The sequences are, in fact, the result of a typical cognitive process, chunking, which is in turn the result of repetitive use (Bybee & Beckner, 2010; N. Ellis & Simpson-Vlach, 2009; Tomasello, 2003; Wulff, 2008). I will revisit the concept of ‘chunking’ in section 2.3.). Each use of such sequence has an effect on the entrenchment of the form-meaning link, which reduces the semantic autonomy of the individual components (N. Ellis, 2001, 2003; Wulff, 2008). Frequency and repetition not only affects but “ultimately bring about form in language” (N. Ellis & Simpson-Vlach, 2009). The plethora of formulaic sequence is a counter-argument against the Chomskyan concept of universal grammar or an innate language acquisition device since if we are to accept the validity of a universal grammar, there would be no prefabricated sequences of language and semantically equivalent words can be used interchangeably in FSs (Beckner et al.,
2009; Sinclair, 1991) as long as grammaticality is present. Proponents of the usage-based approach to language learning instead are in favor of a general cognitive principle in which language use is primarily influenced by frequency besides other factors such as salience, prototypicality of meaning, contingency, redundancy, overshadowing and attention (N. Ellis, 2006; N. Ellis & Larsen-Freeman, 2006).

FSs serve some important functions in the storage and retrieval of language in the mind of the native speakers. L2 learners, including those who have reached advanced levels of proficiency, continue to experience difficulty when processing (Gerald, 2007) or producing FSs in both the spoken and written channels (see, for example, Howarth, 1998a, 1998b; Laufer & Waldman, 2011; Nesselhauf, 2003, 2007). How second language learners process and acquire these sequences is a newly emergent research area. In the past few years there has been research about the acquisition, processing and use of formulaic sequences among ESL learners (e.g. Schmitt, 2004), but research results remain tentative and inconclusive. Further theoretical discussion on this issue is necessary in the field of second language acquisition.

There is a need for further theoretical discussions on how FSs are processed and acquired by L2 learners. Wray’s proposition that L2 learners tend to store FSs analytically, as opposed to the holistic processing of native speakers, and the known difficulty that L2 learners have when processing FSs suggest that the acquisition of FSs is likely to be analytical and incremental. The nature of acquiring FSs in L2, according to Schmitt and Carter (2004), has important implications for selecting pedagogical materials and intervention. In this research I will look at corpus-informed variables of target FSs, namely frequency, MI-score and n-gram and how they correlate with learners’ acquisition of FSs as a result of instruction in terms of their Productive and Receptive knowledge. There is reason to speculate that these variables will be correlated
with learners’ acquisition process at different levels at different times, as some have argued that SLA is a complex dynamic system (Larsen-Freeman, 2006; Larsen-Freeman & Cameron, 2010; Beckner et al., 2009), and these three variables are not the only ones playing a role in the acquisition process.

From pedagogical perspectives, research on classroom intervention that helps facilitate the acquisition of such sequences is even more needed. Empirical evidence of the effectiveness of such techniques and activities can be found only in a few research studies, such as those by Bishop (2004), Boers, Demecheleer and Eyckmans (2004), Boers, Eyckmans, Kappel, Stengers and Demecheleer (2006), Boers and Lindstromberg (2005), Jones and Haywood (2004), Rott (2009) (see also Boers and Lindstromberg (2012) for a review of these intervention studies). Intervention types, such as awareness raising, using corpora, stimulating lookups through input enhancement or explicit instruction, and semantically focused instruction to facilitate learners’ memorization of formulaic sequences (Boers & Lindstromberg, 2012), have generated some positive results and also less encouraging evidence of the effect of particular types of instruction on the learning of formulaic sequences. Moreover, since much of this research area is concerned with techniques such as input enhancement, explicit instruction that encourages learners to look up certain sequences, explaining semantic aspects of FSs, there is a pressing need for research on intervention types that fit better into a communicative classroom.

This study was inspired and informed by the work in FSs by previous scholars and aims at contributing to a much needed research focus in the field. The overarching aim of this study is to compare the effectiveness of techniques and activities that have been proved to be effective in different studies conducted by other researchers, as well as a new activity that has never been implemented in any research study of FSs. This communicative task, namely Spot-the-
Difference, or SpotDif, widely recognized in the field of SLA as promoting learners’ ability to notice otherwise low-saliency language features as well as promoting learners’ acquisition of such forms, has never been used in experimental research related to FSs. I hope to put to the test the effectiveness of this communicative task, together with Gap-fill, as used in Jones and Haywood (2004), two tasks easy to design and implement in the classroom, with the view to providing new empirical findings for language instructors and other researchers. Following in the footsteps of senior scholars, I examined the effectiveness of different techniques and activities on learners’ ability to notice FSs in a reading text, and on learners’ receptive as well as productive knowledge of certain FSs.
CHAPTER 2: LITERATURE REVIEW

In this chapter, I will first attempt to define formulaic sequences and examine the characteristics of these sequences that English L2 learners are sensitive to and need to be sensitized to. Next, I will draw on evidence in corpus-driven and corpus-based research to prove the prevalence of formulaic sequences in the English language. A discussion of the various roles that FSs play in different stages of SLA will then follow. After that, I will delve into the challenges that L2 learners face using data from both cross-sectional and longitudinal studies, revealing that learners tend to underuse certain FSs, overuse others, or avoid using FSs. An analysis of somewhat contradictory evidence as to how learners process FSs will then be conducted to explain the obstacles that learners face when acquiring FSs. Most importantly, I will review experimental studies that attempt to test the effectiveness of different instruction types in teaching FSs, then argue for the need of using task-based intervention to teach these sequences. The chapter will end with my research questions, which build on all of these theoretical issues and previous research.

2.1. Beyond the word: Formulaic sequences

In his most widely-cited work, Sinclair (1991) introduces the very influential *idiom principle*, which states that “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments’ (1991, p. 110). The idiom principle is complementary to, but no less important than, the *open choice principle*, whereby the language user has structures and can freely fill in each slot in the structures with any word, as long as the grammaticality of the
structure is retained. However, not all words are open choice items: for example, it is unnatural to say *totally functional* or *big catastrophe*; more widely accepted would be conventionalized phrases such as *fully functional* and *major catastrophe*, even though the former phrases are perfectly grammatical.

In his later works, Sinclair (1996, 2004b) argues that the notion of *word* being the central unit of lexical meaning should be challenged. His corpus-driven work has proven that beyond the word (or what he terms *word form*), there are a large number of multiword units, those which are usually “tucked away, well off-centre” because they seem to be “anarchic, individual, unstable, one-off items that just do not fit into a tidy description” (1996, p. 78). The anarchic items include phrasal verbs, idioms, fixed phrases, variable phrases, clichés, proverbs - lexical units where the sum of meaning of constituent words cannot account for the meaning of the holistic item.

Over time work by Corrigan, Moravcisk, Ouali, and Wheatley (2009), Moon (1998), Nattinger and DeCarrio (1992), Schmitt (2010), Schmitt and Carter (2008), Sinclair (1991, 1996, 2004a, 2004b), Wray (2002, 2008), to name a few, has drawn attention to the importance of formulaic language in language learning and teaching. Observations of any instance of language use, be it oral or written, will reveal that there are a lot of *formulae* – chunks of words that seem to concur. Formulae are common in language, just as they are in other parts of life:

Linguistic formulae are not unparalleled outside language. Frequently performed routines such as playing a favorite piano piece, starting a car, brushing one’s teeth, or even walking are akin to linguistic formulae in that they, too, form unified chunks of behavior. A seemingly paradoxical feature of such behavioral chunks is that while they may be conceptualized as single wholes, under certain conditions, users can also readily analyze them into components (Corrigan et al., 2009, p. xxiii).
So far the terms *formulaic sequences*, *formulaic language* and *formulae* have been used to refer to lexical units beyond words. These are not the only terms used by researchers interested in the topic. Wray (2002) lists nearly 60 terms that have been used, though they are not identical, ranging from *chunks, collocations, fixed expressions, fossilized forms, lexical simplex, lexical(ized) phrases*, to *prefabricated routines and patterns, schematas set phrases, multiword items/units*. Another term not included in Wray’s list but relatively well-known is *lexical bundles* (Biber et al, 1999; Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004; Cortes, 2004, 2006; Tremblay, Derwing, Libben, & Westbury, 2011). Other terms that did not make Wray’s list are *multiword sequence* (Bybee, 2002; Siyanova-Chanturia, Conklin & Van Heuven, 2011; Stubbs, 2007; Wade & Mobius, 2010), *routinized building block* (De Cock, 2007), *conventional sequence* (Forsberg, 2010). *Formulaic sequence*, however, seems to stand out and is used by many researchers, most notably Bishop (2004, 2005), Boers et al. (2006), Conklin and Schmitt (2008), Jiang and Nekrasova (2007), Schmitt (2004), Simpson-Vlach and N. Ellis (2010), Wood (2005, 2006, 2009), Wray (2000, 2002, 2008). They all conceptualize FSs more or less similar to Wray and Perkins (2000) and Wray (2002), who defines formulaic sequence as:

> “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray & Perkins 2000, p. 1; Wray 2002, p. 9).

Wray (2002) points out that “The word *formulaic* carries with it some associations of ‘unity’ and of ‘custom’ and ‘habit’, while *sequence* indicates that there is more than one discernible internal unit, of whatever kind”, (p. 9) not necessarily words.
The above definition also highlights two important characteristics of FSs: (i) they are sequences of words or other elements and (ii) in terms of mental representation, they are or appear to be represented as wholes.

This research builds on the parameters of FSs that have been established in previous research and examines the correlation between these parameters and learners’ post-treatment performance. These parameters include: n-gram, frequency, Mutual Information score. FSs can be short (e.g. bigrams such as wind down, mull over) or very lengthy (e.g. You can lead a horse to water, but you can’t make him drink). n-gram is simply the number of words in a FS. Frequency of FSs also varies. One in the higher frequency band is usually rendered as possessing a higher teaching worth (N. Ellis, Simpson-Vlach & Maynard, 2008). Another key feature of FSs is Mutual Information score (MI score). MI score is “a statistical measure commonly used in the field of information science designed to assess the degree to which the words in a phrase occur together more often than would be expected by chance (N. Ellis et al., 2008, p. 380). A high MI score reflects a stronger association between the words. N. Ellis et al.’s research suggests that ESL learners are sensitive to the frequency of FSs, while native fluents are more sensitive to the MI score of FSs. Results from this study also show both native speakers and ESL learners to be sensitive to n-gram length of the formulas.

All of these features are likely to affect, to different degrees, learners’ acquisition of FSs. As discussed previously, learners tend not to be sensitized by MI score. If that holds true in this study, it means that teachers are responsible to draw learners’ attention to these high-MI FSs more than, for example, frequently used FSs. Larger n-gram are likely to pose difficulty for learners as well. Investigating these hypothetical effects will help teachers to adapt their
instructional techniques, even within one activity, in order to help learners overcome the disheartening task of acquiring different types of FSs.

2.2. The ubiquity of formulaic sequences

Corpus-based and corpus-driven evidence shows that FSs are prevalent (Biber, 2009). In corpus-based studies, FSs pre-selected from the corpus are studied to examine how they are used, whereas in corpus-driven studies, the inductive approach allows the researcher to find what linguistic structures emerged from the corpus. Both approaches have lent evidentiary support for the preponderance of formulaic sequences. Moon (1998) is one of the first major corpus-based studies on formulaic sequences. The retrieving of formulaic sequences was accomplished not by automatic retrieving but by preselecting a number of fixed expressions and idioms. Other corpus-based studies include those of Liu (2012). The work of Sinclair (1991), Biber and his colleagues (Biber & Conrad, 1999; Biber, et al., 1999: Biber et al., 2004; Biber, 2009) and Simpson-Vlach and N. Ellis (2010) exemplify the corpus-driven strand.

From a small sample of the London-Lund corpus of spoken English, Altenberg (1998) found over 201,000 recurrent word-combinations, which are more or less conventionalized expressions. He estimates that over 80 percent of the words in the corpus form part of a recurrent word combination in one way or another. Altenberg’s (1998) study also provides a tentative answer to the question as to how many types of conventional expressions are known to the average native speaker of a language: over 68,000 of varying length and frequency. Erman and Warren (2000) claim that over 50% of both spoken and written English language they analyzed was formulaic, and Foster’s (2001) findings reveal a figure of 32% using different procedures and criteria. The Google Web1T Database (Brants & Franz, 2006, as cited in Shaoul and Westburry, 2011), which
consists of approximately one trillion word tokens of text found in publicly accessible Web pages, lists approximately 78 million formulaic sequences of two words, 244 million sequences of three words, 328 million of four words, and 294 million of five words.

According to Pawley (2007), it is not even necessary to study large corpora to show that conventional expressions are ubiquitous. He cited studies which analyzed small samples of spoken and written text and highlighted that “about 35-45 percent of all sequences of a given structural type consisted of restricted collocations” (p. 20). Pawley and Syder (1983) showcase the fact that the fluency of advanced language users is not merely grammatical but also natural and idiomatic. This fluency depends on knowledge of a large number of what they term lexicalized or institutionalized sentence stems, which are largely or wholly fixed units of clause length or longer. Besides, semi-lexicalized sequences also play a role in high-level fluency, as will be shown in the next section.

While research in corpus linguistics has shown the prevalence of FSs, this abundance of these sequences lends evidence to usage-based approach to language learning, i.e. experience with languages gives rise to its cognitive organization (N. Ellis & Larsen-Freeman, 2009; Goldberg, 2006, 2013) and counters the proposition that grammar is an abstract set of rules or structures. The existence of such a great number of such patterns can only be explained by the cumulative effect of language experience as language users retain statistical information about which words co-occur with each other, and by acknowledging the effect that frequency of use has on language use (N. Ellis & Larsen-Freeman, 2009). Over time frequency reinforces entrenchment and the frequently co-occurring words as holistic, unified entities, become accepted as legitimate structures in people’s mental representation.
Before progressing to a discussion of the roles of FSs, there is a need to return to the concept of word. I will now survey how vocabulary researchers have been striving to define what it means to know a word, that is, what word knowledge entails.

Richards (1976) was the first modern scholar to attempt an answer to the question “what does it mean to know a word?” His word knowledge framework states that knowing a word means (i) knowing the degree of probability of encountering that word in speech or print (ii) knowing the limitations imposed on the use of the word according to variations of function and situation, such as temporal, geographical, social variations and variations in social role, field and mode of discourse (iii) knowing the syntactic behavior associated with that word (iv) having knowledge of the underlying form of a word and the derivations that can be made from it (v) having knowledge of the network of the associations between that word and other words in language (vi) knowing the semantic value (vii) knowing many of the different meanings associated with the word.

Drawing on Richards’ word knowledge framework, Nation (1990, 2001) identifies various aspects of word knowledge both at the receptive and productive levels. Nation’s list includes the form (spoken, written, word parts), meaning (form-meaning relationships, concepts and referents, associations) and use (grammatical functions, collocations, constraints on use). One of the characteristics that distinguish Nation’s from Richards’ word knowledge framework is that Nation specifies each feature of word knowledge at two levels – receptive and productive. In addition, Nation’s word knowledge aspects seem to follow a logical order from word form to word meaning to word use, while Richards seem to place his assumptions about word knowledge in no such order. As Schmitt (2008) suggests, some of these word knowledge aspects are more amenable to intentional learning, such as word meaning and form, while others, such
as collocation and intuitions of frequency, are likely to be acquired only through repeated exposure. While both Richards and Nation are concerned with describing different aspects of “knowing a word”, Paribakht and Wesche (1997) develop an instrument to measure vocabulary knowledge, called the Vocabulary Knowledge Scale (VKS). The VKS uses a five-point scale (see Table 1 below) and can test the learners’ familiarity with word meaning at both the receptive and productive levels. Although not all aspects of vocabulary knowledge are included in this scale, it is a valid measure to track the early development of specific word knowledge (Kim, 2008).

**Table 1**
The Vocabulary Knowledge Scale (VKS) (Paribakht and Wesche, 1997)

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seen this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means __________________________</td>
</tr>
<tr>
<td>IV</td>
<td>I know this word. It means ________________________________</td>
</tr>
<tr>
<td>V</td>
<td>I can use this word in a sentence ________________________________</td>
</tr>
</tbody>
</table>

These theoretical frameworks regarding individual words are, although not easily discernible from the surface, concerned with the co-occurrence of words. Richards puts “knowledge of the network of the associations between that word and other words in language” fifth in his framework, while Nation includes collocation in the last aspect of word knowledge: word use (besides word form and word meaning). In order to assess learners’ vocabulary knowledge at the productive level, Paribakht and Wesche’s scale requires learners to use the word in a
sentence, which is made possible by many factors, among which the activation of words that usually co-occur with the word being tested. “Most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up text” (Sinclair, 1991, p. 108), and as other researchers suggest, academic words are co-dependent as well (Durrant, 2007; Nattinger & DeCarrio, 1993).

All words take on a specific meaning depending on which words they co-occur with (Sinclair, 2004b; Stubbs, 2002). According to the lexical priming theory (Hoey, 2005), a word is not acquired on its own. Instead, language learners are bound to acquire it in its context, i.e. they are primed by the encounter. They build up concordance knowledge of the word in our mind, and when need arises, they draw from this knowledge. As discussed previously, Sinclair (2004a) attributes the wide use of the minimal approach, i.e. researching language based on the minimal unit – the word, is because the word is the shortest segment and shows least variation; however, he suggests that the alternative approach, one that is based on “units of meaning” (Sinclair, 1991, p. 6). The lexical item can be a word, two or more words that make up a unit of meaning, allowing and accounting for the variation that may occur and reducing ambiguity that usually is a problem with the word being the basic lexical item. Williams (1994) shares the same viewpoint, conceding that the lexicon consists of different forms, many of which are monomorphemic words, the lexicon also include composed, phrasal units, and speculating that there might be more lexical phrases than there are lexical words.

Similar to Sinclair who emphasizes researching “units of meaning” and lexical item instead of words, Bogaards (1996, 2001) re-introduced Cruse’s term “lexical unit” in lieu of the vague concept “word” as relevant and important to L2 learning (Cruse, 1986, as cited in Bogaards, 1996, 2001). Cruse defines a lexical unit as satisfying two criteria: it must have at least one
semantic constituent and it must be at least one word. Thus lexical units are senses of individual words (e.g. crayon) and expressions (e.g. be over the moon).

In second language acquisition, to restrict studying the lexis to word boundaries can have adverse consequences for the language learner. The widely-known Academic Word List (AWL) (Coxhead, 2000, 2002) has been proven to be highly beneficial for learners in academic contexts. The list has been most commonly used to teach academic English and is compiled from a corpus of 3.5 million words of written academic text. The AWL is a collection of 570 word families that account for approximately 10% of the total words in academic texts. Although researchers and teachers agree on the usefulness of Coxhead’s AWL, it fails to take into consideration multi-word sequences. Coxhead later contends that:

“One of the challenges of the AWL is that it was released solely as a list of individual words and their families, with no indication of the context and patterning in which these words occurred. As a result, learners and teachers often focus merely on the recognition of individual AWL words alone, without considering wider and vital aspects of knowing a word in including learning and using common collocations and phrases containing these words” (2008, p. 152).

In examining the viability of a collocation list for EAP, Durrant (2009) finds that such a list does not strongly overlap with Coxhead’s AWL, which indicates a “methodological flaw” (p. 165) in the compilation of the AWL, and alerts teachers of this deficiency. Similarly, Biber et. al (1999), Nattinger and DeCarrio (1993), Shin and Nation (2008), Simpson-Vlach and N. Ellis (2010) have attempted to create lists of essential lexical items beyond words for ESL learners as well.
2.3. The roles of formulaic language in second language acquisition

As early as the 1970s and 1980s, different types of formulaic sequences such as formulaic speech and collocations were recognized by several researchers as an important part of second language acquisition and L2 teaching pedagogy (Brown, 1974; R. Ellis, 1983; Hakuta, 1974; Murphy, 1983; Pawley & Syder, 1983; Raupach, 1984; Vihman, 1979). Hakuta (1974) drawing from the sample speech of a 5-year-old Japanese learning English finds evidence supporting the hypothesis that early L2 learner possesses a simple learning system involving prefabricated routines. Brown (1974) and Murphy (1983) point out that teaching collocations, e.g. for the time being, is necessary in teaching vocabulary to advanced learners and suggests exercises as well as techniques to integrate the teaching of meaningful words in contexts into the L2 classroom. Vihman (1979) views formulaic speech as a learning strategy among children acquiring the first language to a limited extent and to an even greater extent in their second language acquisition, attributing L2 formulaic speech errors to their L1 transfer because of the use of this language learning strategy. Weinert (1995) reviews second language acquisition literature, focusing only on the linguistic development of early L2 learners’ speech, and synthesizes that formulaic language serves three main functions: as communicative, production and learning strategy.

Schmitt (2010) discusses why FSs play an important role in second language acquisition research. First, since both written and spoken discourse contains large (but not yet fully determined) percentages of formulaic language, proficient language users know a large number of formulaic expressions. Second, formulaic language is used to realize a number of different communicative purposes in language use, including functional use, social interaction (phatic communion), discourse organization, precise information transfer. Schmitt claims that “FSs can be used for most things society requires of communication through language” (p. 10). Finally,
the knowledge of formulaic sequences help speakers process and produce language at a higher fluency level.

In line with Schmitt’s final argument, research by Wood (2005, 2006, 2009) indicates a strong correlation between formulaic sequences and L2 fluency since they seem to be cognitively stored and retrieved holistically, hence making pauses shorter and less frequent and allowing longer runs of speech between pauses. Learners used pragmatic, functional or strategic sequences in their narrative retells in order to improve their spontaneous speech fluency over a period of 6 months (Wood, 2005, 2006). Focused instruction on formulaic sequences, including collocations, idioms, phrasal verbs, in a shorter period (6 weeks) also led to more fluent speech (Wood, 2009). Fluency is accounted for by an important mechanism in human learning theory and cognitive science – chunking. The concept was developed by De Groot and Miller in the 1940s and 1950s based on their studies on problem solving and perception and memory (Gobet et al, 2001). According to Gobet et. al., the main contribution of Miller is the concept of chunk, which combines different pieces of information into a single unit. Miller (1956, as cited in Durrant and Schmitt, 2010) establishes the capacity of short term memory at about 7 items; thus by “chunking” smaller pieces of information into larger, more complex chunks, human information processing capacity is increased despite the constraint of short term memory. N. Ellis (2001, 2003) contends that this mechanism is also accountable for language processing and language learning, whereby words that frequently co-occur are fused into larger holistic units, enabling language users to activate a larger amount of information in producing fluent language in their memory. Chunking is not limited to language use, but is a universal mechanism used in acquiring other skills, such as musical skill, driving, typing, or cooking (Wray, 2002).
Chunking may be motivated by the pressure of real-time language production (Sinclair, 1991). Spontaneous speech presents similar problems as those faced by auctioneers or commentators doing play-by-play reports on rapid sports (Pawley, 2007), which explains why formulaic sequences characterize the speech of smooth talkers (N. Ellis, 2002a). Native speakers are able to make use of the processing advantage of formulaic sequences by storing and retrieving them from memory as wholes, without having to construct them from individual lexical items or referring to grammatical rules (N. Ellis, 2001, 2002a, 2003; Nation, 2001; Pawley & Syder, 1983; Sinclair, 1991; Wray, 2002, 2008).

According to N. Ellis (2002b), the use of formulaic sequences is not restricted to initial stages of second language acquisition, but also in the subsequent stages; however, there is a need for more studies on the role of formulas in second language acquisition. In this article, N. Ellis emphasizes the role of formula in sequences of acquisition - from formula, through low-scope pattern, to construction. Formulaic language is considered a promising indicator of very advanced levels of L2 acquisition (Bartning, Forsberg & Hancock, 2009). Similarly, findings from Boers et al. (2006) reveal that learners’ repertoire of formulaic sequences contributes to the improvement of perceived oral language proficiency, as judged in an interview after they have received instruction that emphasizes the noticing of L2 formulaic sequences. Oral proficiency ratings correlate well with counts of formulaic sequences. The self-paced reading task in Millar’s (2011) experiment revealed that native speakers find it more cognitively demanding to process learner-produced malformed formulaic sequences (in this experiment, bigram collocations were exclusively used as instances of such sequences). Ushigusa’s (2008) research has yielded similar findings: temporal measures of fluency correlate with the use of prefabricated sequences of
words, and both set of variables are related to examinees’ scores on the Oral English Proficiency Test.

Nattinger and DeCarrio (1992) view lexical phrases, i.e. formulaic sequences with pragmatic functions, as the center of second language acquisition, in that all learners seem to go through a stage where they make use of unanalyzed chunks of language in certain predictable social situations.

A limited role played by learners’ knowledge of lexical collocation is demonstrated by Hsu and Chiu’s (2008) research, where they only find a significant correlation between lexical collocation knowledge, not actual collocation usage, and speaking proficiency.

Formulaic sequences, accordingly, play a crucial role in second language acquisition, serving communicative, pragmatic and cognitive purposes. Not only do they facilitate earlier stages of language learning, these “chunks” remain a pivotal part of more advanced stages in language acquisition. Findings regarding learners’ fluency levels, both in spoken and written channels, have proven that there is correlation between learners’ knowledge of FSs and their proficiency/fluency levels. A mastery of FSs yields both processing and production benefits for second language learners and is thus essential to achieving higher levels of second language proficiency.

2.4. L2 learners have difficulty with FSs

Given the pervasiveness of formulaic sequences in both spoken and written language and the roles they play in second language acquisition, L2 learners should develop a mastery of formulaic sequences in addition to individual words to reach desirable levels of proficiency. However, acquiring knowledge of FSs in an L2 is a path paved with obstacles both in ESL and EFL
settings. Bogaards (1996) proposes that it is because of the complexity of the *lexical unit*, not the *word*, that accounts for lexical errors and inadequacies in many advanced learners and for making vocabulary learning a very difficult and almost endless activity.

From a pedagogical perspective, it is not uncommon to see formulaic sequences receiving minimal attention both in textbooks and by language teachers, while single word items are usually the central concern. Vocabulary instruction centering on the word is probably one reason why L2 learners, even very advanced learners, do not display near-nativelike knowledge and proficiency in the use of formulaic sequences. Schmitt and Carter (2004) note that learners’ formulaic language use usually lags behind other aspects of their second language acquisition, and the reason might be lack of input or learners’ deliberate avoidance due to the degree of L1-L2 similarity instead of intrinsic difficulty. However, research into learners’ use and knowledge of formulaic sequences so far has indicated significant difficulty that learners encounter. Many of these studies address the difficulty of EFL learners, but others also present strong evidence that even in an ESL setting, where learners are immersed in the second language input, learners continue to struggle with using FSs appropriately.

The underuse of collocations in ESL/EFL learners has been demonstrated in several studies. Erman (2009a) compares the use of collocations in the writings of native and non-native speakers of English in a university (ESL setting) and finds that collocational usage among non-native speakers lags behind that of native speakers, resulting in less fluent, less directly comprehensible, and less pragmatically appropriate discourse. The same can be said about advanced learners’ passive knowledge of semantically opaque idioms which is not equivalent with that of native speakers, despite their comparable knowledge of low-frequency single words (Arnaud & Savignon, 1997). In another study by Erman (2009b), L2 learners underuse
collocations, resulting in less native-like writing. Over time, the frequency, accuracy and variation of learners’ use of formulaic sequences can improve; however, the corpus-derived data in Qi and Ding’s (2011) study show that despite this improvement found among Chinese speaking EFL learners, they still lag behind native speakers in terms of frequency and accuracy. Qi and Ding also find that not only did learners underuse formulaic sequences compared to native speakers, errors persisted over time (from the beginning to the end of an academic year): formulaic sequences containing prepositions and articles were most challenging for these learners.

Foster (2001) also provides evidence of the underuse of FSs by ESL learners based on data collected from native and non-native speakers in a classroom task. Native speakers use a greater variety of chunks in preplanned task performance, but planning only reduces non-native speakers’ use of chunks. Evidence for underuse of verb-noun collocations by non-native speakers is also found from a larger corpus of about 300,000 words of argumentative and descriptive essays (Laufer & Waldman, 2011), compared to that of native speakers in the LOCNESS corpus. Laufer and Waldman point out that learners of all proficiency levels make less use of selected verb-noun collocations, but their use increases with proficiency levels. However, errors persist even among advanced learners, which is consistent with Bahns and Eldaw’s (1993) finding of the productive knowledge of verb-noun collocation production of advanced EFL learners whose first language is German. Based on students’ performance on a translation task, Bahns and Eldaw find that more than half of the unacceptably translated lexical words were collocates. They also find that when students failed to find the intended collocates, they tended to paraphrase them. However, not all collocations allow paraphrasing, adding to advanced learners’ difficulty. Studying 810 adjective-noun collocations in 31 essays written by Russian
learners of English, Siyanova and Schmitt (2008) discovered that around 45% of these collocations were appropriate. Yet the non-native speakers demonstrated poorer intuition about the frequency of these collocations than the native speakers, and they were slower than native speakers in processing collocations. Research on collocational knowledge of learners from different L1 backgrounds, such as Jordanian (Farghal & Obiedat, 1995) and Hong Kong learners (Fan, 2009), also show that this knowledge among learners is not comparable with that of native speakers. Howarth (1998a) drawing on native speaker data (the Lancaster/Oslo/Bergen corpus) and essays written by nonnative speakers conclude that nonnative speakers experience difficulty differentiating between free collocations (e.g. blow a trumpet, under the table) and restricted collocations (e.g. blow a fuse, under attack), a finding similar to Howarth (1998b). They also have difficulty with figurative collocations (e.g. blow your own trumpet, under the microscope). Because of this lack in collocational knowledge, learners have to resort to several strategies, such as avoidance, transferring from L1 collocation and repetition. In analyzing the use of verb-noun collocations (e.g. take a break, shake one’s head) by advanced German-speaking learners of English in free written production, Nesselhauf (2003) finds that the influence of L1 on triggering collocational errors goes far beyond what earlier (small-scale) studies have predicted. In another study on advanced German-speaking learners of English, out of more than 2,000 verb-noun collocations produced by these learners in argumentative and descriptive essays, about a quarter were judged unacceptable and a third inappropriate (Nesselhauf, 2007). Types of deviations range from the very frequent verb usage in collocations e.g. *make an experience (have), or noun e.g. *make a cut (distinction) to the less frequent structure type as in *set somebody an example (set an example for somebody). Avoidance is also found to be a strategy by intermediate Chinese learners of English in Liao and Fukuya (2004)
when it comes to the use of phrasal verbs. The study finds that intermediate learners use significantly fewer phrasal verbs than native speakers and more advanced learners in the multiple-choice test, recall test and translation test.

Other researchers have shown that language learners overuse some FSs but do not employ a wide range of FSs as native speakers do. As Granger (1998) notes, only a small number of familiar and safe sequences become learners’ repertoires or “islands of reliability” and they tend to overuse them. In her study that focuses on the use of amplifiers that modify adjectives (e.g. *bitterly cold, unbearably ugly*), she finds that non-native learners significantly overused *totally* and *completely*, suggesting that learners were ignorant of or reluctant to use other conventionalized amplifiers, such as *bitterly* in *bitterly cold*. Durrant and Schmitt (2009) criticize previous studies which claim that non-native language lacks formulaicity for failing to take account of frequency information and individual differences. In their study that accounts for these two factors, Durrant and Schmitt find that L2 learners tend to overuse high-frequency collocations but underuse low-frequency ones. In a longitudinal study by Li and Schmitt (2009), the writings of one Chinese student show that she overuses only a limited number of lexical phrases to the point where raters consider the usage non-native, although over time she has successfully learned to use more lexical phrases with better accuracy. DeCock, Granger, Leech, & McEnery (1998) point out that adult advanced EFL learners whose L1 is French do use prefabricated routines, sometimes more often than NSs. For example, they underuse most vagueness tags (e.g. *and things like that, or something, and everything*), but overuse one (*and so on*).

Some formulaic sequences even have an adverse effect on L2 learners’ reading comprehension. Gerald (2007) looks at how native and non-native speakers process formulaic sequences such as...
racked his brains, tossed and turned, but to no avail in a whole written text and examine participants’ eye fixation patterns. While native speakers showed fewer eye fixations on words in the formulaic condition than in the baseline condition, the pattern for the nonnative speakers was reversed: there were a higher percentage of words fixated in the formulaic sequences than in the baseline condition, suggesting that they have difficulty processing these expressions. Gerald speculates that the semantic opacity of the formulaic sequences and the learners’ lack of prior exposure to these formulaic sequences bring out this adverse effect among non-native readers.

2.5. The processing of FSs

The difficulty that even advanced L2 learners experience with formulaic sequences is noteworthy and thus worth investigating further. One hypothesis to account for the difficulty in acquiring these chunks of language is that the tradition of focusing on individual words draws most of the attentional resources of learners, especially in classroom environments. Evidence from research presented in the previous section has revealed that in terms of the usage of formulaic sequences, learners’ performance is not satisfactory even after they have reached higher levels of proficiency in the L2. The fundamental question, one that may illuminate the reasons why learners’ advanced levels of proficiency in many aspects of the L2 do not accompany near native-like usage of formulaic sequences, lies in examining the “processes of memory storage and production underlying performance.” (Howarth, 1998a, p. 26) Researchers hold opposing viewpoints as to how L2 learners store and process formulaic sequences. Some opine that formulaic sequences are stored and processed analytically (Schmitt, Grandage & Adolphs, 2004; Wray, 2002), i.e. word-by-word, while others have found evidence that learners
indeed do so in a holistic manner, similar to native speakers (e.g. Durrant & Schmitt, 2010; Jiang & Nekrasova, 2007).

Wray (2002) claims that adult second language learners take an essentially non-formulaic approach to language learning, relying more on creativity in their L2 production rather than on prefabricated patterns. To illustrate, she cites Yorio’s (1989) research findings, which indicate that written English of ESL learners showed many attempts at formulaic sequences, but were riddled with errors, e.g. *take advantages of*, *being taking care of*. There is no straightforward explanation for this type of error, but one possibility, according to Wray, is that learners’ interlanguage grammar interferes with either the storage or the retrieval of these formulaic sequences. Another example for the analytic approach to learning formulaic sequences among adult learners is that on encountering the collocation *major catastrophe*, they would break it down to a word meaning *big*, and the other meaning *disaster*. When need arises later, they would use any synonym of *big* to combine with any synonym of *disaster*, without knowing which combinations are predominantly used and thus falling into the trap of “deceptive compatibility” (Laufer, 2011, p. 44).

In an experiment aimed specifically at finding out whether corpus-derived recurrent clusters are stored holistically or analytically by non-native speakers, Schmitt et al. (2004) use an oral dictation task which requires participants to repeat short passages orally, one after another, under time pressure. Results showed that non-native speakers did not reproduce pre-selected, corpus-derived recurrent clusters (e.g. *as a consequence of*, *in the middle of*) with fluency and accuracy, suggesting that they did not store these clusters holistically.
Contrary to Wray’s model of formulaic sequences in a second language, other experiments have lent evidence to the holistic retention and processing of formulaic sequences. Durrant and Schmitt’s (2010) lab-based study of collocation learning shows that adult L2 learners are able to retain information about the co-occurrence of words. In the treatment session, participants were exposed to adjective-noun collocations either once or twice in the same linguistic context, or twice in different contexts. Results from recall tests immediately following treatment show that participants in both repetition training conditions outperformed those in the single presentation condition, with a slight advantage to the verbatim repetition condition group. Durrant and Schmitt conclude that retention of collocation among adult L2 learners did occur, even though it was an implicit process (learners were unaware of what and how they were going to be tested after treatment).

Another study that provides evidence against the allegedly analytical approach to formulaic sequences by adult L2 learners was conducted by Jiang and Nekrasova (2007). In their experiments, both native and high-proficiency non-native speakers took two online grammaticality judgment tests that include both non-formulaic and formulaic sequences, matched for word length, frequency, and in the second experiment, even the visual word shape of the phrases (all materials were in uppercase). In both tests, all participants, native as well as non-native, showed significantly faster reaction to formulaic sequences than they did to non-formulaic expressions. The much faster reaction time and lower error rate for formulaic sequences provide evidence in support of the holistic representation and processing of formulaic sequences in both native speakers and L2 learners. Jiang and Nekrasova believe that their instrument, the online grammaticality judgment test, is a more direct and sensitive task to measure the effect of formulaicity, compared to the oral-dictation task (Schmitt et al., 2004) or

The self-paced reading task that Conklin and Schmitt (2008) used to examine the processing advantage of formulaic sequences measures the reading times for formulaic sequences (specifically idioms) compared to matched nonformulaic expressions. Both native and non-native speakers showed faster reading rate for formulaic sequences, both in literal and idiomatic contexts. Conklin and Schmitt’s findings suggest that formulaic sequences are also processed more easily than nonformulaic ones in the mind of proficient L2 learners. The ease in cognitive efforts formulaic sequences provide could be the by-product of their holistic representation and processing.

Two studies provide mixed evidence as to how L2 learners process formulaic sequences. The eye-movement experiment conducted by Underwood, Schmitt and Galpin (2004) only provides partial support for the holistic processing of formulaic sequences among non-native speakers, although it clearly shows that native speakers process these expressions as a whole. By comparing the difference between the fixations of the eye movement on target words within and outside formulaic sequences, with the underlying assumption that holistic processing would allow fewer fixations and shorter duration of fixations while participants read a target word, Underwood et al. found fewer and shorter fixations among native speakers; among non-native speakers, results revealed fewer fixations but no significant difference in the duration of the fixations.

The multi-study research on adjective-noun collocations carried out by Siyanova and Schmitt (2008) shows that L2 learners’ collocational knowledge is emergent, supported by the finding
that about 45% of their adjective-noun collocation usage is appropriate. However, they do not have as subtle intuitions about the frequency of adjective-noun collocations as native speakers do, and this may account for their slower processing of collocations in general, and for their failure to process high frequency collocation more rapidly than less frequently-used collocations.

Formulaic expressions extracted from different corpora are used in N. Ellis et al.’s (2008) experiments to test which aspects of formulaicity native and non-native speakers are sensitive to. Results reveal that the accuracy and fluency of processing these formulaic expressions among native speakers is affected by MI score, while processing among advanced L2 learners of English is predominantly affected by frequency (occurrence per million). Findings from this study are somewhat in conflict with Siyanova and Schmitt’s, probably because each study looks at a different type of formulaic sequences or different proficiencies of learners. In another study by Siyanova-Chanturia et al. (2011), eye-tracking was used to examine the processing of 3-word binominal phrases e.g. bride and groom and their reversed form (groom and bride), all of which were extracted from the British National Corpus. Results indicate that similar to native speakers, non-native speakers of different proficiency levels are sensitive to the frequency of these phrases. Regardless of the frequency of the content words in the phrases, non-native speakers’ reading fluency continues to correlate with the frequency of the whole phrases. Higher proficiency learners are also sensitive to the configuration of these sequences (binominal vs. reversed).

Learners’ processing of FSs can also be influenced by their self-perceived knowledge of these larger lexical units. In Laufer’s (2011) study, learners seem to overestimate their knowledge of verb-noun collocations. When asked to supply the missing verb in such combinations, with the
option of using a dictionary (learners were provided with entries of specific words from bilingual dictionaries), they often thought that they knew the collocations, or in her words, fell in the trap of “deceptive compatibility” (p. 44). That is, the learners were unaware of the nature of these restricted collocations and thought that the nouns provided in the sentences could be combined freely with any verb that is semantically compatible. Examples of their non-targetlike production are: *do action (take action), *put an eye on (take an eye on), *notice attention (pay attention), *save costs (reduce costs), or *try an attempt (make an attempt). Learners like those in Laufer’s study are more prone to ignore the conventionalized nature of word combinations, especially with familiar words. They did not recognize the need to look for the collocationally compatible verb in the dictionary but used a deceptively compatible verb.

2.6. Intervention to facilitate the acquisition of FSs

On one hand, extensive reading, with no explicit teaching or intentional learning of vocabulary learning, has been proven to be an essential part of building learners’ lexicon. In a survey of empirical studies on the effectiveness as well as problems of incidental lexical acquisition, Huckin and Coady (1999) highlight a consensus among vocabulary acquisition researchers on the necessity of extensive reading. This type of reading expands learners’ lexicon through a secondary type of learning, i.e. learning vocabulary is the by-product of the activity while making sense of the reading texts is the primary purpose.

The incidental acquisition of L2 vocabulary (and grammar) is the main assumption of Krashen’s Input Hypothesis (1989). According to Krashen, learners acquire language by understanding the message, and the pleasure of reading provides abundant comprehensible input for language acquisition to take place without the need of explicit teaching or intentional learning. In his 2004
book, Krashen summarizes research evidence that advocates FVR, i.e. *Free Voluntary Reading* (no book reports, no questions required of learners) and claims that FVR is a powerful tool that is usually missing in foreign language teaching (Krashen, 2004).

Schmitt (2008) synthesizes from the findings of different vocabulary learning studies that the learner needs exposure to reading and listening materials, or both (reading-while-listening). Multimodal input (e.g. pictures, movies) will facilitate L2 vocabulary learning since such input resembles that available to L1 learners, whose acquisition of vocabulary is based on not just listening and reading texts.

On the other hand, research has shown that incidental vocabulary acquisition yields low pick-up rate and is not likely to lead to productive mastery of the word. According Schmitt (2008), research in vocabulary acquisition suggests that approximately 8-10 reading exposures may just give learners a chance of acquiring an initial receptive knowledge of words. Schmitt concludes that incidental vocabulary learning cannot be used as the primary method of building L2 lexicon. Krashen (2004) also points out that FVR does not single-handedly result in high levels of language proficiency, although FVR is likely to contribute to a strong lexical stock for these advanced levels.

Hulstijn (2002) questions the frequency effect when it comes to lexical knowledge, which requires explicit knowledge of form-meaning pairs. He cites studies by neuroscientists that identify the different locations in the brain that store explicit knowledge as opposed to implicit knowledge, usually knowledge of forms such as phonemes, letters, grammatical gender or inflection. Hulstijn postulates that lexical knowledge is less susceptible to input frequency effects than other types of implicit knowledge. Therefore, sometimes a single encounter suffices
while other times, many encounters do not guarantee the learning of a lexical entry (fast mapping, see e.g. Heibeck and Markman, 1987).

If incidental acquisition of single words is not always predictable, it is even more unreasonable to expect that learners would acquire larger chunks of individual words in formulaic sequences merely through exposure to L2 (Boers & Lindstromberg, 2009). Similarly to how they pick up or do not pick up single words, learners tend to focus more on comprehending the text rather than paying attention to its exact wording. With formulaic sequences, it is even more challenging since some FSs have low frequency despite a strong association among their constituent words. Native speakers tend to be very sensitive to high-MI FSs, regardless of their frequency, while non-native advanced ESL learners seem to be only attuned to frequency (N. Ellis et al., 2008).

With regards to collocation, what makes it even more challenging for learners is “deceptive compatibility” (Laufer, 2011, p. 44), i.e. the belief that semantically compatible words are combinatory. Deceptive compatibility can lead to the production of nontarget-like collocations such as *notice attention or *try an attempt.

Many researchers have conducted studies on different classroom activities that promote learners’ engagement with target individual words, from using an online database, Internet chat, to seeing the new words in a reading text and retelling the passage using those words, or temporarily isolating words from their context and processing them elaborately (Schmitt, 2008).

There are numerous proponents of the instruction of formulaic sequences in the language classrooms, e.g. in Brown (1974), Cowie (1998), N. Ellis et al. (2008), Granger and Meunier (2008), Murphy (1983), Nattinger and De Carrio (1992), Pawley and Syder (1983), Sinclair (1991, 2004). A relatively extremist advocate for the teaching of such formulaic sequences is Michael
Lewis, who introduced the *lexical approach* (Lewis, 1993, 1997, 2000). This is an instructional approach that de-emphasizes syntax to the minimal level and centers around the instruction of lexical items, consisting of words, polywords (*by the way, on the other hand*), collocations (*to raise capital, a short-term strategy*), institutionalized utterances (*I’ll get it, If I were you, I’d..., I’ll be back in a minute*) and sentence frames or heads (*secondly...and finally, we come now to a number of important reservations*). Lewis’s work is similar to other above-mentioned scholars in the sense that they do not provide any empirical evidence regarding the effectiveness of such instructional practices in the language classroom (see also Baigent (1999), Ceh (2007), Chang & Bao (2008), Harwood (2002), Lewis (1997, 2000), Moudraia (2001) for other arguments for the lexical approach with no or weak empirical evidence). As Boers et al. (2006) comment, before his research there had been no ‘hard’ empirical evidence of the effectiveness of ‘chunk-noticing’, which Lewis considers the heart of the lexical approach, although students responded positively to the activities in some action research. Boers et al. (2006) put the ‘chunk-noticing’ principle of Lewis’s lexical approach to the test by conducting a small-scale experiment among 32 proficient L2 learners and found positive results for the treatment group, who received awareness-raising instruction regarding standardized phrases such as collocations and idiomatic expressions.

Empirical evidence for or against instructional intervention in helping learners acquire formulaic sequences like the kind found in Boers et al.’s (2006) study is scarce in literature, but there has been a growing number of studies attempting to do so, although they are not to prove the validity of the approach that Lewis proposed in the 1990s. Most studies show that teaching formulaic sequences in one way or another has positive effects on learners’ knowledge and
usage of such expressions, suggesting that L2 learners systematically benefit from instruction focused on FSs.

Several research projects examine the effectiveness of general awareness-raising. Using a type of treatment close to what Lewis suggests, Boers et al. (2006) encourage learners to first identify and then learn the lexical phrases in the input. In a total of 22 hours of classroom instruction, learners in the experimental group were taught with a special attention to formulaic sequences. Texts and exercises used for both experimental and control group were identical; the only difference is that the teacher intentionally directed students’ attention to the co-occurrence of words in the experimental group, and to individual words or grammar patterns in the control group. In order to raise students’ awareness of formulaic sequences, all input was presented twice: the first time to focus on meaning and the second for ‘exploration’. It was in this ‘exploration’ stage that learners were encouraged to identify useful chunks in the material through two types of activities: gap-fill exercises targeting either words in formulaic sequences or in freer combinations, encouraging learners to highlight or underline useful and interesting formulaic sequences or individual words. The experimental group was perceived as more orally proficient than the control group. The number of formulaic sequences in their speech was found to correlate with their perceived oral proficiency.

Rott (2009) suggests from her research that since learners are often unaware of the importance of formulaic constructions, a pre-writing brainstorming task was successful in drawing their attentional resources to these special constructions and thus was able to enhance the use of such constructions in learners’ writing. The degree of this task’s effectiveness depended on the genre: if the nature of the genre, e.g. a recipe, requires the use of FSs, learners made more
effort to use them than when the use of FSs is only useful, but not essential to the genre (e.g., descriptive essay).

The second type of instructional intervention used in previous research is input enhancement in the studies by Bishop (2004) and Jones and Haywood (2004), the former using only input enhancement as treatment, while the latter using input enhancement in combination with other awareness-raising activities. The purpose of input enhancement techniques is also to raise students’ awareness of the salience, though this is achieved not by explicit instruction from the teacher.

Bishop (2004) conducted a study on 44 participants to examine the effect of typographical salience on the lookup and comprehension of FSs. The saliency of FSs was increased in the text provided for the treatment group by changing the font color (red) and underlining. This small-scale study led to a possible conclusion that typographically enhanced FSs encouraged learners to look up the glosses provided if they click on these FSs and that looking up FSs correlates with better comprehension results.

Jones and Haywood (2004) conducted a study in which several techniques were used to raise learners’ awareness of formulaic sequences in a reading-writing class. First, after learners became familiar with the meaning of the text, their attentional resources to formulaic sequences were drawn through enhanced input, that is, important phrases were highlighted in bold italics. They were then explicitly encouraged to remember and use these formulaic sequences in their writing; strategies to learn formulaic sequences, such as classifying FSs semantically and structurally, comparing FSs with less formal writing styles, were introduced. Results showed that the awareness of learners in the treatment group improved after the
intervention; however, their improvement in producing FSs in their essays was less noticeable than in their c-test results, despite the fact that before learners wrote the essay, there was a revision activity which sometimes included showing learners a list of useful FSs and their grammatical structure.

A third type of instructional intervention is awareness raising instruction which focuses on the phonological motivation of formulaic sequences (structural elaboration). Based on the findings of Boers and Lindstromberg (2005), which indicated the strong mnemonic effects of alliteration, i.e. the repetition of the word-initial consonant as in time will tell, spic and span and preliminary evidence of explicitly drawing learners’ attention to such sound patterns, Lindstromberg and Boers (2008b) conducted three experiments and found that (i) even very brief teacher-led noticing of alliteration had a positive effect on learners’ recall of these expressions and (ii) the mnemonic potential of such alliteration was not detected and used by learners without this brief awareness-raising instructional technique. Another type of phonemic repetition, assonance, was also found to have significant mnemonic effect (Lindstromberg & Boers, 2008a) on L2 learners, but it was not clear if learners can autonomously make use of assonance.

The fourth type of instructional intervention is semantic elaboration, which is in accord with cognitive semantic theory and dual coding theory. A technique called etymological elaboration of transparent and opaque idioms was used with the purpose of helping learners decode the semantic opaqueness of figurative idioms. In a study by Boers et al. (2004), the control group invented contexts to use idioms in; the experimental group, on the other hand, hypothesized the origin of the expressions from several options. Identify-the-source task proved beneficial for learners’ retention of idioms since the experimental group performed significantly better than the control group, both with semantically opaque and transparent idioms.
Wood (2009) uses tasks in the communicative classroom with one participant. The participant, who is a Japanese EFL learner, attended a special 6-week (total of 9 hours) fluency workshop. The workshop aimed at developing participants’ fluency by first providing them with the input, then activities to facilitate automatization (shadowing, dictogloss, mingle jigsaw, and chat circle). After that, the participants had opportunities to practice and produce, and finally free talk was encouraged. Results indicate a strong increase in fluency after six weeks of focused instruction, and a relationship between the instruction and the fluency and use of formulaic sequences in the learner speech samples.

Hsu (2010) conducted a study with a 30-minute instructional treatment that requires learners to discuss in groups a list of collocations. Then learners were given another list of Chinese equivalents and asked to compose a sentence using the collocation and present their sentences to the class. There was another treatment group that did the same things, only with single words instead of collocations, and a control group, which received no instruction. Results from a vocabulary recall test and a reading comprehension test showed that the collocation instruction treatment group improved their scores in the vocabulary recall test more than in the reading comprehension test. The group also outscored the other two groups in the vocabulary recall test.

In addition to the need for more research on instructional types that benefits FSs, it is also important to investigate the nature of this acquisition process over time. Schmitt and Carter (2004) comment that the acquisition of FSs, similar to that of individual words, is likely to be incremental: these sequences are learned over time, even among native speakers. Schmitt and Carter argue that determining the question of whether, and which, FSs are acquired in an “all-or-nothing” manner or in an incremental manner, is important in second language acquisition.
because the answer may inform teachers of what FSs are practical to teach. Schmitt and Carter’s hypothesis suggests that the incremental nature of the acquisition of FSs is likely to affect the pedagogical decisions that instructors make in the classrooms, how they design activities or choose which FSs to incorporate into their lesson plan, and what numbers of exposure will be necessary, and how to implement activities that can foster learners’ ability to continue noticing and learning FSs beyond the classroom.

2.7. Task-based intervention in teaching FSs

Assuming that a usage-approach to SLA is valid, i.e. L2 learners acquire the language through experience with it, and that frequency of structures is conducive to L2 development, we need to ponder the characteristics of input on SLA and how to optimize the kind of input usually available to L2 learners (N. Ellis, 2009), especially those whose primary contact with L2 is through classroom materials. Despite the evidence against the role of frequency (e.g. Hustijn, 2002), it is likely that frequency has a dominant effect on how L2 is acquired, with exceptions. Thus pedagogical intervention has to create to conditions for such exceptions to occur – in other words, L2 pedagogy needs to use certain input characteristics compensating for the poverty of L2 stimulus (Tomasello, 2003). Poverty of L2 stimulus, or impoverished input (Goldberg, 2006, 2013) is when there is no or ambiguous evidence in the language they (children) hear around them. In SLA, the input learners are exposed to is in many circumstances much more limited than L1 input. In a setting where English is a foreign language, input is even more limited than in an ESL setting. Instruction in L2 classrooms has to compensate for the insufficient breadth of L2 input with depth of engagement processes evoked, as will later be revisited in the discussion of the Involvement Load Hypothesis. As Larsen-Freeman (2002) points out, acknowledging frequency effects does not mean that abundant comprehensible input is all learners need.
As discussed in previous parts, L2 learners do not always pay attention to formulaic sequences as a language form that is difficult to master; consequently, they may overestimate their knowledge of these FSs, especially when the constituent words are semantically familiar. There have been several research studies that examined the effectiveness of various types of classroom activities: gap-fill exercises, encouraging learners to highlight FSs (Boers et al, 2006; Jones & Haywood, 2004), a pre-writing brainstorming task (Rott, 2009), discussion (Hsu, 2010), input enhancement (Bishop, 2004; Jones & Haywood, 2004), raising students awareness of phonological or FSs’ etymological features (Boers et al., 2004; Lindstromberg & Boers, 2008b), and communicative tasks (Wood, 2009). All these instruction types are to retune learners’ selective attention through form-focused instruction or consciousness raising (N. Ellis, 2008).

Among these types of instructional intervention methods, tasks seem to be most suitable for the communicative classroom. Many researchers in the field of instructed second language acquisition (mostly regarding L2 grammar) concur that the optimal classroom activities should be primarily meaning-focused and at the same time allow for learners to focus on L2 form. Long (1991) introduces the concept of focus on form as a type of instruction that “overtly draws students’ attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication” (pp. 45-46). Even though focus on form is generally discussed in second language acquisition as related to grammar, Laufer (2005) argues that ‘form’ could be lexical items as well. In this study, we use “focus on form” to refer to any kind of instruction that draws learners’ attention to FSs, whether it be in a meaning-based context or in a decontextualized vocabulary activity, following Laufer (2005)’s definition.

Tasks are usually implemented in form-focused classrooms to provide learners with the opportunities to notice low-saliency forms, modify input, interact with each other, modify
output, negotiate and give corrective feedback to each other (see for example Doughty & Pica, 1986; Long, 1981; Nunan, 1989; Pica, Kang & Sauro, 2006).

Despite the effectiveness of communicative tasks in the L2 classroom, not much research has examined the use of such tasks in teaching vocabulary. However, a few studies have lent evidentiary support to the usefulness of communicative tasks, such as R. Ellis, Tanaka and Yamazaki (1994), R. Ellis and He (1999), Gass and Torres (2005), Kim (2008), Kowal and Swain (1994), Loewen and Philp (2006). These studies have found that communicative tasks such as jigsaw, dictogloss (both individual and interactive), opinion gap, story narration, match/label tasks are useful in a number of ways. First, tasks encourage learners to modify their input and to help each other with finding out the meaning of unknown vocabulary, even when the task’s main aim is not to teach vocabulary (Kowal & Swain, 1994; Loewen & Philp, 2006). Second, tasks lead to better vocabulary acquisition, both at receptive and productive levels (R. Ellis & He, 1999; Gass & Torres, 2005). Third, tasks benefit learners even when they do not actively participate in the interaction (R. Ellis et al., 1994). An important finding by Kim (2008) is that a collaborative dictogloss is more effective than the individual dictogloss. From these studies, interaction between learners seems to benefit vocabulary learning in general.

Little research, however, has been done on the possible contribution of interactive tasks to the acquisition of L2 formulaic sequences, or on their effectiveness on such acquisition in comparison to other types of instruction. The study by Wood (2006) is the only one that looks into this issue, but with very limited data from only one participant. Moreover, the tasks used in Wood’s study were not specifically designed to facilitate the participant’s acquisition of FSs.
This niche in research leads me to believe in the need for designing and conducting a study that aims at testing the possible effectiveness of communicative tasks in relation to other types of instructional techniques and activities. As Robinson (2011) contends, experimental research in task-based language teaching and learning has been helpful in exploring “connections between pedagogical practice and the second language acquisition processes they may stimulate” (p. 5). Furthermore, tasks can be used as a research tool to investigate theoretically informed hypotheses about SLA processes (Candlin, 1987, as cited in Robinson, 2011). In this research study, I see interactive tasks such as Gap-fill and SpotDif not only as pedagogical activities, but also as constructs that have the potential of casting light on the theory of usage-based language learning. Tasks are capable of redirecting learners’ selective attention to offset the “poverty of the stimulus” (Tomasello, 2003, p. 288) of ESL and EFL input. Tasks provide a useful way of engaging students in “meaningful interaction because engaging them so provides the optimal way for learners to benefit from frequency and saliency” (Larsen-Freeman, 2002, p. 283). Tasks thus can create conditions not otherwise available to learners due to the impoverished L2 input. Tasks are usually designed to allow learners to attend to FSs, which would be obscured by other language features, such as individual words or grammatical structures. Since constructions like FSs are, in essence, form-meaning pairings, tasks need to facilitate learners’ attention to both. In the two types of tasks proposed as intervention for this experimental study, one is more geared towards the meaning (and use) of the FSs (Gap-fill), and the other directs learners’ attentional resources more to the form (SpotDif). These interactive tasks aim at optimizing the L2 input by increasing the saliency of the target FSs by activating different levels of engagement processes such as Need, Search and Evaluation.
Gap-fill, as used in Jones and Haywood’s study (2004), was chosen because it requires learners to collaborate with each other to find the appropriate FSs to fill in the blanks. Another potentially successful interactive task, Spot-the-Difference (henceforth SpotDif), a type of task as modeled in Pica et al.’s (2006) study, is chosen for this study because, similar to Gap-fill, its implementational procedures seem to guarantee learners’ attention to FSs and to allow for inter-learner negotiation of meaning. Other types of tasks that were used in previous research studies, such as dictogloss, jigsaw, match and label, while very useful in teaching individual words, are not easy to tailor to the need of the teachers who want to teach FSs.

In order to complete a Gap-fill task, learners first read the original passage. Next, they read a version of the original passage, this time with all target FSs taken out, then in pairs they choose from a group of FSs to decide which FS suits best with each blank.

SpotDif is a 5-step communicative task. In Step 1, learners read the original passage. In Step 2, learners read either version A or version B of the original passage. In Step 3, learners work with each other to choose between sentences or phrases in versions A and B and justify their choices based on what they think is more targetlike. Step 4 requires learners to recall choices in Step 3 and complete cloze version of the original passage. In Step 5, learners compare their choices in Step 4 with the original passage.

My belief in the potential of two communicative tasks used as treatment in this study, SpotDif and Gap-fill, is also in line with a vocabulary acquisition hypothesis advanced by Laufer and Hulstijn (2001), the “Involvement Load Hypothesis” (see also Laufer & Girsai, 2008). This hypothesis posits that a classroom task requiring higher involvement load will be more effective than tasks with lower involvement load. Involvement load is determined by three factors: need,
which is the motivational dimension of involvement, *search*, the attempt to find the meaning of an unknown word or trying to find a word in L2 to express a certain concept, and *evaluation*, which is the process of comparing between meanings of the same words, or different words to make the task-essential decision. Even though Laufer and Hustijn’s hypothesis is concerned with the acquisition of individual words, this hypothesis seems applicable to the acquisition of FSs.

SpotDif and Gap-fill both entail opportunities for learners to engage in a crucial cognitive process: evaluation, which helps increase their involvement load. In Gap-fill, in order to fill in the gaps, learners will likely feel a strong *need* to know the meaning of all FSs in the word bank, then *search* for their meaning and *evaluate* between different options made available to them. In this task, learners have to evaluate the meaning of the FSs in the word bank in relation to the meaning of a particular sentence (or that of several sentences) in the original passage. However, once learners figure out the meaning of the FSs in the word bank, they probably have no reason to attend more to the form of these FSs, because they do not have to evaluate between different forms of the same FS, as in the SpotDif task.

In completing the SpotDif, learners have to work with two versions of the passage, where FSs differ. For example, when the target FS is *on the books* (being part of the law), the altered version will be *in the book*. Learners will have to evaluate the appropriateness of the form of these two alternatives and choose one that fits the context. In the process of evaluating in SpotDif, learners probably will have the motivation, or *need*, to learn about the meaning of this sequence, and this will encourage them to *search* for its meaning, or to ask the instructor about it. However, the intensity of *search* in SpotDif is probably not as high as in the Gap-fill exercise because they do not have to choose from different FSs. Laufer and Hustijn believe that *need* is moderate when learners are asked by the teacher to use a word in a sentence, and *need* is
strong when the learners themselves deem it necessary to know a certain word. SpotDif involves a moderate level of *evaluation* where learners have to evaluate the correct form of the target FSs.

Comparing this task with other instructional methods and techniques that have been proven in previous research to be relatively beneficial to learners in acquiring FSs, such as input enhancement, it seems that Gap-fill and SpotDif induce a more rigorous involvement load. Input enhancement, for example, probably will result in a moderate *need, search* and very little *evaluation*. Bishop’s (2004) study shows that learners look up (*search*) the sequences more often when these sequences are typologically enhanced. To compensate for the lack of *evaluation* in the input enhancement treatment, I decided to include an explicit instruction component in which the teacher provides learners with some examples of FSs that usually go unnoticed or create the “deceptive compatibility” effect. For example, the sequence *all a rage* is not a target-like sequence, although it seems to be semantically equivalent to *all the rage*. This type of *evaluation* is likely to be low-level, because it is done by the teacher, and the learners do not receive any task that requires them to evaluate different forms of the target FSs.

The following table summarizes how the three different types of instructions measure against the Involvement Load Hypothesis.
Table 2
Hypothetical levels of *Need, Search and Evaluation* in the three treatment types

<table>
<thead>
<tr>
<th></th>
<th>Need</th>
<th>Search</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td>low (meaning)</td>
<td>low (meaning)</td>
<td>low (meaning)</td>
</tr>
<tr>
<td></td>
<td>low (form)</td>
<td>low (form)</td>
<td>low (form)</td>
</tr>
<tr>
<td>Input Enhancement + Explicit Instruction (IEEI)</td>
<td>moderate (meaning)</td>
<td>moderate (meaning)</td>
<td>low (meaning)</td>
</tr>
<tr>
<td></td>
<td>moderate (form)</td>
<td>low (form)</td>
<td>low (form)</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>moderate (meaning)</td>
<td>high (meaning)</td>
<td>moderate (meaning)</td>
</tr>
<tr>
<td></td>
<td>low (form)</td>
<td>low (form)</td>
<td>low (form)</td>
</tr>
<tr>
<td>SpotDif</td>
<td>moderate (meaning)</td>
<td>moderate (meaning)</td>
<td>moderate (meaning)</td>
</tr>
<tr>
<td></td>
<td>moderate (form)</td>
<td>high (form)</td>
<td>high (form)</td>
</tr>
</tbody>
</table>

This research study through comparing the effectiveness of three different instructional intervention types hopes to provide additional empirical evidence for the Involvement Load Hypothesis. If this hypothesis holds true, the two tasks (Gap-fill and SpotDif) will be more effective than the IEEI instruction and the IEEI group in turn will outperform the control group. More specifically, the Gap-fill group will better facilitate learners’ acquisition of the meaning of FSs, and SpotDif will probably be more effective in fostering learners’ acquisition of the FSs’ form.

2.8. Research questions

The general aim of this study is to examine the effectiveness of three types of instruction: (i) a combination of input enhancement (as in Bishop’s (2004) study) and explicit instruction (IEEI) (ii) Gap-fill task (modified from what was used in Jones & Haywood’s (2004) study), and (iii) Spot-the-Difference (SpotDif)
The research questions are:

1. Do IEEI, Gap-fill and SpotDif help learners acquire higher levels of productive knowledge of FSs? If yes, which type of treatment is more effective?

2. Do IEEI, Gap-fill and SpotDif help learners acquire higher levels of receptive knowledge of FSs? If yes, which type of treatment is more effective?

3. Do IEEI, Gap-fill and SpotDif help learners improve their ability to notice FSs in a reading passage? If yes, which type of treatment is more effective?

4. Are frequency, MI-score, and n-gram length of FSs related to learners’ acquisition of receptive and productive knowledge of FSs?
CHAPTER 3: METHODOLOGY

This study is a quasi-experimental study conducted to test the relative effectiveness of different types of form-focused instruction on the acquisition of FSs of high-intermediate ESL students. This chapter will describe how the experiment was conducted, including its participants and setting, variables, materials, pre-tests, post-tests and delayed post-tests, types of treatment, supplementary qualitative data, and data analysis methods.

3.1. Participants and Setting

Participants in the study are students enrolled in graduate programs at a large research university in the Northeast United States, who have a TOEFL iBT score ranging from 90-110 (or an equivalent test score) and have studied in the United States for 1-2 years. They all use Mandarin-Chinese with native-like proficiency. A total of 40 students were randomly assigned to 4 groups, one control group and three experimental groups. There are 10 participants in each group.

3.2. Variables: Independent and Dependent

Independent variables are

(i) Type of instruction: Input Enhancement + Explicit Instruction, Collaborative gap-fill exercise, SpotDif

(ii) Properties of the FSs: n-gram length, frequency and MI score

I opted to use the Corpus of Contemporary American English (Davies, 2008-) (CoCA) to determine the frequency and MI score of our target FSs as of May 2013. CoCA was used because
the passages chosen for our research procedures are from magazines and newspapers in the U.S. In addition, the corpus size (450 million words), its contemporariness (all texts are from the period of 1990-2012) and the inclusiveness of genres (spoken, fiction, popular magazines, newspapers, academic journals) can guarantee its representativeness. In this study I include FSs of different frequency bands and MI score levels to see if these variables affect learners’ reception and production of FS. The process of selecting FSs to be included in this study will be described in more detail in the following section.

Dependent variables are (i) the participants’ ability to notice FSs in a reading passage (not applicable to research question No. 4) (ii) receptive knowledge of FSs in multiple-choice tests (iii) productive knowledge of FSs in c-tests.

3.3. Materials

Three texts chosen for the three treatment sessions are adapted from online newspaper articles from The Atlantic, New York Times, Forbes, Medical News Today, and PC Magazine (see Appendix A). Each text is compiled by the researcher drawing on multiple sources. For example, the first passage was adapted from the textbook Concepts for Today, which originally was an article in The Los Angeles Times, and from an article in The Huffington Post. Adapting techniques include shortening the passage, inserting target FSs where appropriate, adjusting the tone of the passage to ensure stylistic consistency. Each passage contains approximately 400 words in length. All texts share a common theme: information and communication technology and its influence on people’s life. Each passage include 10 target FSs (for all 30 target FSs, see Appendix A). The selection process of these FSs is described below.
As Read and Nation (2004) note, the validity of measuring FSs is one of the most prominent challenges in this research area. The primary criterion of a FS is its holistic storage and retrieval (Wray, 2002, 2008), which is “a difficult one to operationalize” (Read & Nation, 2004, p. 35). Thus to ensure internal validity of FS research, Read and Nation argue that a triangulation of methods is necessary. This study follows Underwood et al. (2004) and Schmitt and Underwood’s (2004) methodological triangulation. Three methods are used to select FSs from the text: intuition, corpus reference, and cloze test.

First, I used my own intuition to determine possible FSs to be used in the reading passages, then have two native speakers read the passages, target FSs highlighted, to ensure that the FSs are appropriately used. Then these items are looked up in the CoCA (Davies, 2008-) for frequency and MI score. N. Ellis et al. (2008), for instance, included in their studies FSs with a minimum frequency of 10 per million tokens, and minimum MI score of approximately 3.3. However, the studies by N. Ellis et al. were corpus-derived with the aim of formulating a list of most popular and useful FSs that should be included in the English for Academic Purpose classroom, such as at the beginning of, it should be noted that. These FSs are probably too frequent and are likely already a part of our participants’ lexicon; thus I aim for a lower frequency threshold. Moreover, as N. Ellis et al. point out, advanced ESL learners were as sensitive to high-frequency FSs as native speakers were; however, they were not sensitive to FSs with high MI but low frequency. These researchers argue that these high-MI FSs usually have distinctive meanings and functions. They are likely to be stored as wholes in the mind of native speakers. In this study, I chose FSs with a MI score of at least 3, as N. Ellis et al. recommend, but with a minimal raw frequency (as determined by CoCA) of 50.
A manual checking of the hits after a search query is completed, according to Read and Nation (2004), is necessary to ensure that each hit is a valid cluster of the words in the FS. For FSs that have a potentially inflected word, i.e. the word is one word form of a lemma, e.g. *push the envelope*, one of the target FSs of the present study, the corpus search included all forms of the lemma *push*. The search query for this particular FS was *[push] the envelope*. In addition, I also allowed for collocates to be of a certain distance from each other. Therefore, in this example, the search command was *[push]* with the collocate *[the envelope]* within 3 words (following Read & Nation, 2004).

With such a protocol, frequency and MI scores obtained from CoCA are those of the lemma, not of a particular form. For example, the frequency and MI score of *[push] the envelope* include the parameters of all possible inflections of the word *push*, including *pushes the envelope, pushing the envelope, pushed the envelope, (has/have/had) pushed the envelope*, in addition to *push the envelope*.

Lastly, to prevent the corpus from searching sequences where, for example, *beef* functions as a noun, I specified the part of speech of the key words. The search query therefore becomes *[beef].[v*] in collocation with *up*.

Finally, all the FSs were included in a cloze test, which provides initial, sometimes middle, letters of content words in the FSs. FSs were put in short contexts, for example:

\[
\begin{align*}
\text{City councils and city governments are going to have to } & \mathbf{s} & \mathbf{t} & \mathbf{u} & \mathbf{p} & \mathbf{l} & \mathbf{e} \\
\text{and make some concessions to keep economic development on track ( } & \text{take responsibility) }
\end{align*}
\]

Answer: *step up to the plate*
The cloze tests were administered to native speakers, whose feedback prompted me to revise the test to include more cue letters. This test was later used as the Productive Knowledge post-test for learners.

3.4. Pre-tests, post-tests and delayed post-tests

*Pre-tests*

Two kinds of pre-tests were administered, the first to examine the knowledge levels of key words in the FSs, and the second to learn about learners’ ability to notice FSs in a reading text.

The first pre-test is to ensure that participants have not had knowledge of the target FSs (see Appendix B). This pre-test is modified based on the 5-point Vocabulary Knowledge Scale proposed by Paribakht and Wesche (1997). Instead of using the full scale with 5 different levels of knowledge, this study uses one with 3 levels (1) neither receptive knowledge nor productive knowledge of the item (2) receptive knowledge only and (3) both receptive and productive knowledge. For each word or phrase in the test, the learners were instructed to choose one of the 3 following levels:

(1) I don’t know this word/phrase

(2) I know the meaning of this word/phrase, but I never use it in my writing/speaking

(3) I use this word/phrase quite often in my writing/speaking

30 distracters, individual words and phrases, were used alongside with 30 target FSs. The reason why individual words are also included is that it helps disguise the purpose of this research study. Following Paribakht and Wesche’s model, learners were asked to write down the meaning of the word/phrase in question if they choose level 2 or 3. They can explain the word
using as many words as they wish, or using a sentence that contains the word. If they chose level (1), no point was given. 1 point was given each time level (2) was chosen, and 3 points for each answer at level (3) with the correct meaning of the FSs provided. In later parts of the dissertation, “Productive Knowledge pre-tests” will be used to refer to the scores learners received based on the number of target FSs with a level (3) of perceived knowledge multiplied by 3, and Receptive Knowledge pre-tests indicate the scores learners received based on the number of target FSs that they indicated a level (2) of knowledge. For example, if learners chose level (3) for 2 target FSs in the Vocabulary Knowledge Scale Test, their Productive Knowledge pre-test score would be 3 X 2 = 6 points. If learners chose level (2) of knowledge for 3 target FSs in the VKS Test, their Receptive Knowledge pre-test score would be 3. It should be noted that the type of productive knowledge indicated by these scores is perceived productive knowledge.

The second pre-test is a 400-word reading passage adapted from www.helium.com (see Appendix C). The topic of this passage is consistent with the theme of all passages used in 3 treatment sessions: problems of the Information Age. This passage includes FSs different from target FSs. This test is used to gauge learners’ ability to notice FSs, on which they were not trained, when they encounter reading materials in L2. Learners were instructed to underline or highlight words and phrases in this passage that they deem useful for improving their language skills (in the study by Jones and Haywood (2004), participants were instructed to highlight words and phrases they would advise other students to learn). According to Jones and Haywood (2004), the ratio between the number of individual words highlighted and that of FSs is indicative of the learners’ awareness of the importance of attending to FSs. In this study, learners’ answers are scored based on the number of FSs they were able to underline as compared to a list of pre-determined FSs. This list consists of FSs with a minimal frequency of 50
and a minimal MI score of 3.0. Two raters, one being the researcher, rated the participants’ underlining test. 1 point was given to each qualified FS underlined by the learners. To further examine learners’ rationale for their underlining, they were asked at the end of this test to specify the reasons why they chose to underline those words/phrases. In later parts of the dissertation, this underlining test will be referred to as the Awareness test.

**Immediate post-tests**

After each session, learners in all groups were asked to complete the Productive Knowledge test (Appendix D) and the Receptive Knowledge test (Appendix E). One exception is that after the third lesson, in addition to these two tests, learners also completed the Awareness test.

The Productive Knowledge test, which is used to measure learners’ productive knowledge of the target FSs, is the c-test, also following the format in Schmitt et al.’s (2004) study. In this test, the initial letter(s), at times middle letters, of the target FSs are provided in the same sentences as in the Receptive Knowledge test, which will be described in the later section. Schmitt et al. argue that in order to ensure the validity of this test, learners should not spend time on guessing the meaning of the FSs being tested. Therefore, in addition to the context, the learners are also provided with a definition of the target FSs at the end of the context. For example,

Some sellers on eBay complain that the multinational internet corporation is giving them the

\[ s\hat{h}______ e______ _____ _____ st______ \]

as it only protects the rights of the buyers. (*disadvantages, the smaller or less desirable part*)

A scale Jones and Haywood (2004) employed in their study was used to assess learners’ performance on the c-test. Data analysis procedures assume that this scale is an equal-interval scale.
Learners’ performance on the c-tests was evaluated by two raters. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among raters. The interrater reliability for the raters was found to be $\text{Kappa} = 0.84$ ($p < .001$), which reaches the almost perfect agreement level as determined by Landis and Koch (1977).

The Receptive Knowledge test is to tap into learner’s receptive knowledge of the target FSs, both in terms of form and meaning, after receiving the treatment. This study adopted the two test models by Schmitt et al. (2004), with several modifications. In order to measure learners’ receptive knowledge of FSs, Schmitt et al. used a multiple choice test, in which participants were asked to choose from four options in a short context. The distractors were semantically similar to the correct option and as similar in form and length as possible. In the multiple-choice test of our study, I retained the content word(s) of all FSs in all distractors. At least one distractor includes the content word(s) but with a different part of speech or ungrammatically inflected. The other distractors have the content word(s) intact but other components of the FSs will be altered (e.g. prepositions, articles). Choice E is always “I don’t know” so that the participants do not have to guess when they do not know an answer (Schmitt et al., 2004). Schmitt et al.’s test seems effective in evaluating learners’ receptive knowledge regarding the form of target FSs. To
tap into learners’ retention of their meaning, I added a new, supplementary component to each test item, which requires learners to choose the FS’s correct meaning. For each FS, five options are provided. One choice is the correct meaning; two are near synonyms; one conveys an unrelated meaning; and the final choice is “I don’t know”. The length of all 5 choices is controlled to ensure that no choice stands out to the test-takers. Cronbach’s alpha for all 60 items in the Receptive Knowledge test was .904, suggesting a high level of reliability of the test.

Both the multiple choice-test and c-tests were piloted with 20 native speakers.

At the end of session 3, learners read the same passage that they read before session 1 and were instructed to underline useful words and/or phrases they deem useful for their English language learning (the Awareness test). The assumption is that, whether their ability to notice FSs in a reading text changed or not, it would be reflected in the number of FSs underlined in the passage.

Delayed post-tests
2 weeks after the last treatment session, participants completed the same multiple choice test and c-test. They also completed the Awareness test.

Sequencing of multiple-choice tests and c-tests
Each time these two tests are administered, whether as pilot tests for native speakers, or immediate and delayed post-tests for learners, participants completed the c-tests first so that the results from this test are not contaminated by the multiple-choice test.

Measures were taken to ensure that learners’ completion of the c-test would not influence their performance on the multiple-choice test.

For example, in the following item:
In the past, farmers lived ______ ______ ______ ______ the weather. (completely dependent on)

I did not provide the initial letters of the prepositions and article lest learners recall the initial letters in the c-test in order to eliminate the choices in the multiple choice test. For example, if the learners memorize that the first letter of this sequence is a_____, they will be able to eliminate choice A on the mercy of. If the learner memorize that the second word in the article starts with a t, they will be able to eliminate choice B at a mercy of.

In summary, two pretests include the Vocabulary Knowledge Test and the Awareness Test. Post-tests, both immediate and delayed, include the Productive Knowledge Test, the Receptive Knowledge Test, and the Awareness Test.

3.5. Treatment

Since this study aims at comparing the effectiveness of different instructional techniques in the teaching of FSs and in enhancing learners’ ability to notice these sequences in a reading text, there were one control group and three treatment groups, each consisting of 10 learners.

I taught all four groups to minimize the effects of different teaching styles. Materials for instruction were identical across all four groups. All groups met for 3 sessions in 3 weeks. They completed the pre-tests before session 1, the immediate post-tests (c-tests then multiple choice tests) at the end of each session, and the underlining immediate post-test at the end of session 3. Lessons for all groups started with a warm-up activity to introduce the topic of the reading passage. Then learners read the passage for the first time in order to answer a few comprehension questions (main idea and specific details). In order to answer these
comprehension questions, learners may or may not use FSs – none of the questions require that FSs be used. They then read the passage for the second time, this time with different activities.

The control group spent the whole lesson reading and answering several comprehension questions. After reading the passage for the first time, they collaborated with a partner to answer these questions. Learners may refer to the passage for necessary information. After they had finished answering all comprehension questions, the instructor led an all-class discussion to provide suggested answers to them. If learners asked vocabulary questions, the teacher would provide them with an English explanation of the lexical items with no special focus on formulaic sequences.

The first treatment group, Input Enhancement + Explicit Instruction (IEEI), began their lessons by individually reading the passage and then in pairs answered several comprehension questions. After that, they were given an explicit instruction about the meaning of target FSs and the importance of FSs in general. The reading texts for this group has all target FSs bolded, highlighted and formatted in larger fonts. For example:

Sex offenders often use matchmaking sites to find victims. Jeffrey Marsalisz, a Pennsylvania man, is a smooth talker who would meet women on the popular dating website Match.com, telling them he was an astronaut, doctor or a spy and then slip something into their drinks to incapacitate them. Lawmakers in several states are mulling over legislation to help make online daters more aware of the potential pitfalls of the process. Connecticut's bill, mirroring a law in New York, requires Internet dating services to provide a safety awareness notice during registration that offers advice such as never including one’s last name, email address, place of work, phone numbers or identifying information in an Internet profile. Similar laws are already on the books in Florida and New Jersey. Other states are expected to follow suit.
The second treatment group, the Gap-fill group, participated in collaborative Gap-fill tasks following regular reading comprehension activities. After reading for the first time and answering comprehension questions, they were, in pairs, given a gap-filling exercise based on the reading passage. The gap-fill exercise is similar to that used in Jones and Haywood’s (2004) study. Students read the original passage again with target FSs deleted. They were given a word bank from which to choose and were asked to complete the exercise collaboratively. Only one word bank was given to each pair/group of students.

The third treatment group, SpotDif, received instruction that contained a SpotDif task in every session. In order to maintain similar time on task across control and treatment groups, a shortened version of SpotDif was used, which excluded the cloze activity. After reading the passage for the first time and answering comprehension questions, learners worked in pairs to complete the SpotDif task. Two students were given a version of the original passage, with the target FSs modified in one version. Each version of the passage contained both target FSs and modified FSs. Since the reading passages are relatively long, the sentences that include the target FSs were numbered so that learners know where to look for the differences in their two versions of the passage. The target FSs were not typologically modified in this task. Upon completing SpotDif, the instructor helped learners find out what the FSs in the original passage were. Below is an example of a part of two different versions of a paragraph (part of the whole passage given to learners in session 3). The target FSs are *mull over* and *on the books*. 
<table>
<thead>
<tr>
<th>Student A</th>
<th>Student B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex offenders often use matchmaking sites to find victims. Jeffrey Marsalisz, a Pennsylvania man, is a smooth talker who would meet women on the popular dating website Match.com, telling them he was an astronaut, doctor or a spy and then slip something into their drinks to incapacitate them. (9) Lawmakers in several states are mulling through legislation to help make online daters more aware of the potential pitfalls of the process. Connecticut's bill, mirroring a law in New York, requires Internet dating services to provide a safety awareness notice during registration that offers advice such as never including one’s last name, email address, place of work, phone numbers or identifying information in an Internet profile. (10) Similar laws are already on the books in Florida and New Jersey. Other states are expected to follow suit.</td>
<td>Sex offenders often use matchmaking sites to find victims. Jeffrey Marsalisz, a Pennsylvania man, is a smooth talker who would meet women on the popular dating website Match.com, telling them he was an astronaut, doctor or a spy and then slip something into their drinks to incapacitate them. (9) Lawmakers in several states are mulling over legislation to help make online daters more aware of the potential pitfalls of the process. Connecticut's bill, mirroring a law in New York, requires Internet dating services to provide a safety awareness notice during registration that offers advice such as never including one’s last name, email address, place of work, phone numbers or identifying information in an Internet profile. (10) Similar laws are already in the book in Florida and New Jersey. Other states are expected to follow suit.</td>
</tr>
</tbody>
</table>

3.6. Qualitative data

Qualitative data were collected to supplement quantitative data. All classes were audiotaped. In addition, in the SpotDif and Gap-fill groups, learners’ pair interactions were recorded on digital voice recorders. In addition, learners were asked at the end of the underlining task why they underlined such words and phrases.

3.7. Data Analysis

Since this is a between-within design study, it is important to establish the nature and arrangement of the predictor variables. In terms of arrangement, type of instruction is considered the main between-subjects predictor variable. Other variables, such as MI score,
frequency, \( n \)-gram of FSs are secondary. In terms of nature, these secondary variables are continuous (MI score, frequency, \( n \)-gram), while type of instruction is a categorical variable. I did not examine the interaction effects of instruction type and different features of FSs, since such design would require a large number of treatment groups that combine these two factors, with one factor having 2 levels (instruction type), and one factor having 5 levels (features of FSs). Such a research study would exceed logistical practicality to implement. Instead, I explored the effects of instruction type and the effects of different features separately.

In order to answer the first three research questions, mixed between-within subjects ANOVAs were conducted to determine whether learners’ performance showed statistically significant improvement between pre-test, immediate post-test and delayed post-test. Timing of pre-tests, post-tests and delayed post-tests is the within-subject (repeated measures) factor. Treatment type is the between-subject factor. Post hoc analyses were conducted to determine if there are statistically significant differences between test score means across control and treatment conditions. Besides statistical significance testing, effect sizes using Cohen’s \( d \) were also calculated in order to serve as indicators of educational significance. Finally, in order to answer the last research question, correlations between frequency, \( n \)-gram length and MI Score were calculated.
CHAPTER 4: FINDINGS

This chapter summarizes all results obtained from the statistical tests conducted on students in the pre-tests, immediate post-tests and delayed post-tests. Mixed between-within subjects ANOVA were conducted on students’ test scores. Effect size values using Cohen’s $d$ were also calculated. First, I will report learners’ performance on the Productive Knowledge tests, then the Receptive Knowledge tests and Awareness tests. Last, I will report how the features of FSs (frequency, $n$-gram length, MI Score) affected the learners’ performance on the Productive and Receptive Knowledge tests.

4.1. Productive Knowledge

*In response to research question:* Do IEEI, Gap-fill, and SpotDif help learners acquire higher levels of productive knowledge of FSs? If yes, which type of treatment is more effective?

Mixed between-within subjects ANOVAs conducted on the Productive Knowledge pre-tests, immediate, and delayed post-tests revealed statistically significant improvement among learners of all groups from pre-tests to post-tests. Results also showed that the three treatment groups outperformed the Control group, with the Gap-fill group achieving the highest scores, followed by the SpotDif group and the IEEI group. In the immediate post-test, all treatment groups outperformed the Control group at a significant level. In the delayed post-test, statistically significant difference was found only between the Gap-fill group and Control group. Effect sizes using Cohen’s $d$ were large for all treatment groups in both post-tests, except for the moderate effect size of the IEEI group as compared to the Control group in the delayed post-test.
### 4.1.1. Overall Results

Table 3 summarizes the raw scores obtained by learners in the Productive Knowledge pre-tests (see chapter 2, section 3.4. for information about how Productive Knowledge pre-test scores were calculated) immediate, and delayed post-tests. The maximal score learners could receive for the Productive Knowledge test is 90 as there are 30 target FSs and the highest score for each FS in the Productive Knowledge test is 3.

**Table 3**
Mean scores of the four groups on the Productive Knowledge pre-tests, immediate and delayed post-tests

<table>
<thead>
<tr>
<th>Timing</th>
<th>Pre-test</th>
<th>Immediate Post-test</th>
<th>Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control group</td>
<td>IEEI</td>
<td>Gap-fill</td>
</tr>
<tr>
<td>Mean</td>
<td>0.30</td>
<td>0.00</td>
<td>0.30</td>
</tr>
<tr>
<td>SD</td>
<td>.95</td>
<td>0.00</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Normality, independence and sphericity assumptions were met for a mixed between-within subjects ANOVA to be conducted. Mauchly’s test of sphericity yielded $p > .05$. The homogeneity of variance assumption was violated, with Levene’s test of homogeneity of variance showing significant unequal variances in the pre-test (but not the immediate and delayed post-tests). Thus ANOVA was conducted with Games-Howell post hoc analyses ($\alpha = .05$) which do not assume equal variance. Results revealed significant differences between the Control group and all treatment groups, and no statistically significant difference across experimental groups. In sections 4.1.2 and 4.1.3 similar procedures will apply.

A mixed between-within subjects ANOVA was conducted to compare the mean scores of the four groups in order to examine whether the scores improved significantly from pre-test to immediate and delayed post-test and if there were significant differences across groups. Independent variables are Group and Time; dependent variable is the test scores on the
Vocabulary Knowledge Scale Test and the immediate as well delayed Productive Knowledge test.

There were a significant effect for Time across the pre-test, the immediate and delayed post-test, $F(2,72) = 574, p < .001, \eta^2_p = .94$, a significant interaction effect between Group and Time $F(6,72) = 17, p < .001, \eta^2_p = .58$, and a significant effect for Group $F(3,36) = 16, p < .001, \eta^2_p = .58$.

Table 4 summarizes the results of this mixed ANOVA.

### Table 4
Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Productive Knowledge pre-test, immediate and delayed post-test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-Subjects Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>81286</td>
<td>2.0</td>
<td>40643</td>
<td>574</td>
<td>.000</td>
<td>.94</td>
</tr>
<tr>
<td>Time * Group</td>
<td>6985</td>
<td>6.0</td>
<td>1164</td>
<td>16</td>
<td>.000</td>
<td>.58</td>
</tr>
<tr>
<td>Error</td>
<td>5097</td>
<td>72</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between-Subjects Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>8140</td>
<td>3</td>
<td>2713</td>
<td>16</td>
<td>.000</td>
<td>.57</td>
</tr>
<tr>
<td>Error</td>
<td>6088</td>
<td>36</td>
<td>169</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All four groups displayed significant improvement from the pre-test to the immediate post-test; however, the significant interaction for group and time suggested that the gain in scores differed across groups. Games-Howell post hoc analyses with $\alpha = .05$, as shown in Table 5, revealed two important findings: (1) The three experimental groups outperformed the Control group in the immediate and delayed post-test at a statistically significant level; and (2) There were no statistically significant differences among the three different types of treatment.
Table 5
Multiple comparisons of groups for Productive Knowledge immediate and delayed post-tests using Games-Howell

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-14.57 *</td>
<td>3.571</td>
<td>.005</td>
<td>-24.82</td>
<td>-4.31</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-22.47 *</td>
<td>3.678</td>
<td>.000</td>
<td>-32.97</td>
<td>-11.96</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-16.27 *</td>
<td>3.763</td>
<td>.002</td>
<td>-26.97</td>
<td>-5.56</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>14.57 *</td>
<td>3.571</td>
<td>.005</td>
<td>4.31</td>
<td>24.82</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-7.90</td>
<td>2.896</td>
<td>.061</td>
<td>-16.09</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-1.70</td>
<td>3.003</td>
<td>.941</td>
<td>-10.21</td>
<td>6.81</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>22.47 *</td>
<td>3.678</td>
<td>.000</td>
<td>11.96</td>
<td>32.97</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>7.90</td>
<td>2.896</td>
<td>.061</td>
<td>.29</td>
<td>16.09</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>6.20</td>
<td>3.130</td>
<td>.232</td>
<td>-2.65</td>
<td>15.05</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>16.27 *</td>
<td>3.763</td>
<td>.002</td>
<td>5.56</td>
<td>26.97</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>1.70</td>
<td>3.003</td>
<td>.941</td>
<td>-6.81</td>
<td>10.21</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-6.20</td>
<td>3.130</td>
<td>.232</td>
<td>-15.05</td>
<td>2.65</td>
</tr>
</tbody>
</table>

Based on observed means.
The error term is Mean Square(Error) = 56.370.

* The mean difference is significant at the .05 level

Figure 1 illustrates the changes in the four groups’ mean scores in the Productive Knowledge pretest, immediate and delayed post-tests. Note that the differences in the groups’ pre-test scores are so small (0.3; 0.0; 0.3; 0.6) and the improvement was so significant (the highest mean score is 80 out of 90 in the immediate post-test) that in this figure it appears like all groups received the same score of 0.0.
4.1.2. Results from the immediate post-test

In the pre-test, students’ self-reported productive knowledge of the target formulaic sequences was extremely limited, as shown in the results of Vocabulary Knowledge Scale Test (pre-test). Means (with standard deviations in parentheses) for the control, IEEI, Gap-fill and SpotDif in order were 0.30 (SD = 0.95), 0.00 (SD = 0.00), 0.30 (SD = 0.95), and 0.60 (SD = 1.27). The one-way ANOVA conducted on the groups’ means showed no statistically significant differences, $F(3,39) = .71$, $p > .05$, $\eta_p^2 = .56$. Post hoc analyses using the Games-Howell post hoc criterion for significance indicated homogeneity of the four groups’ performance on the pre-test.

In the immediate post-test, mean scores (with standard deviations in parentheses) for the Control group, IEEI group, Gap-fill group and SpotDif group were 34 (SD = 18), 70 (SD = 11), 80 (SD = 8.1), and 72 (SD = 11) respectively, with the highest score possible being 90. Levene’s test for equality of variances found $p > .05$ in the immediate post-test, indicating that the differences
in standard deviations in the pre-test scores did not endure to the immediate post-test. The effect of unequal variances is mitigated by equal sample sizes, although it should be noted that the $p$ value might be not as significant as it suggested. However, most of our $p$ values are less than .001, hence the possibility of exaggerating the significance level is minimal.

A mixed between-within subjects ANOVA was conducted to compare the mean scores of the four groups in order to examine whether the scores improved significantly from pre-test to immediate post-test and if there were significant differences across groups. Independent variables are Group and Time; dependent variable is the test scores on the Vocabulary Knowledge Scale Test and the immediate Productive Knowledge test.

Results revealed a significant effect for Time across the pre-test and the immediate post-test, $F(1,36) = 1076, p < .001, \eta^2_p = .97$, a significant interaction effect between Group and Time $F(3,36) = 28.4, p < .001, \eta^2_p = .70$, and a significant effect for Group $F(3,36) = 28, p < .001, \eta^2_p = .70$. Table 6 summarizes the results of this mixed ANOVA.

**Table 6**
Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Productive Knowledge pre-test and immediate post-test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>81154</td>
<td>1</td>
<td>81154</td>
<td>1077</td>
<td>.000</td>
<td>.97</td>
</tr>
<tr>
<td>Time * Group</td>
<td>6432</td>
<td>3</td>
<td>2144</td>
<td>28</td>
<td>.000</td>
<td>.70</td>
</tr>
<tr>
<td>Error</td>
<td>2714</td>
<td>36</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>6444</td>
<td>3</td>
<td>2148</td>
<td>28</td>
<td>.000</td>
<td>.70</td>
</tr>
<tr>
<td>Error</td>
<td>2787</td>
<td>36</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All four groups displayed significant improvement from the pre-test to the immediate post-test; however, the significant interaction for group and time suggested that the gain in scores differed across groups. Games-Howell post hoc analyses with $\alpha = .05$, as shown in Table 7,
revealed two important findings: (1) The three experimental groups outperformed the Control group in the immediate post-test at a statistically significant level; and (2) There were no statistically significant differences among the three different types of treatment.

Table 7
Multiple comparisons of groups for Productive Knowledge immediate post-test using Games-Howell (α = .05)

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>p</th>
<th>95% Confidence Interval Lower Bound</th>
<th>95% Confidence Interval Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-18.15*</td>
<td>3.257</td>
<td>.000</td>
<td>-27.53</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-23.25*</td>
<td>3.058</td>
<td>.000</td>
<td>-32.24</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-19.40*</td>
<td>3.288</td>
<td>.000</td>
<td>-28.85</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>18.15*</td>
<td>3.257</td>
<td>.000</td>
<td>8.77</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-5.10</td>
<td>2.160</td>
<td>.124</td>
<td>-11.25</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-1.25</td>
<td>2.475</td>
<td>.957</td>
<td>-8.25</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>23.25*</td>
<td>3.058</td>
<td>.000</td>
<td>14.26</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>5.10</td>
<td>2.160</td>
<td>.124</td>
<td>-1.05</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>3.85</td>
<td>2.208</td>
<td>.334</td>
<td>-2.44</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>19.40*</td>
<td>3.288</td>
<td>.000</td>
<td>9.95</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>1.25</td>
<td>2.475</td>
<td>.957</td>
<td>-5.75</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-3.85</td>
<td>2.208</td>
<td>.334</td>
<td>-10.14</td>
</tr>
</tbody>
</table>

Based on observed means.
The error term is Mean Square(Error) = 38.7
* The mean difference is significant at the .05 level
The effect sizes using Cohen’s $d^2$ for this analysis, respectively for the IEEI group ($d = 2.49$), Gap-fill group ($d = 3.38$), and SpotDif group ($d = 2.64$), as compared to the Control group, were found to considerably exceed Cohen’s convention for a large effect ($d = .80$), suggesting a large magnitude of instructional effect. $d$ was largest for the Gap-fill group, followed by the SpotDif group and IEEI.

Figure 2 illustrates the differences among four groups’ mean scores in the Productive Knowledge immediate post-test despite the commonalities displayed in the pre-test. Note that there were small differences in the pre-test scores across groups that were not reflected in this figure.

Figure 2
Productive Knowledge test scores in the pre-test and immediate post-test

---

1 Effect sizes are calculated using Campbell Collaboration formula
http://www.campbellcollaboration.org/resources/effect_size_input.php
As can be seen from Figure 2, even the Control group seemingly displays a significant increase in mean scores from the pre-test to the immediate post-test. A repeated measures ANOVA for the Control group’s scores on these two tests showed a significant effect of Time on their performance in the tests, $F(1,9) = 35, p < .001, \eta_p^2 = .79$.

4.1.3. Results from the Delayed Post-test

Results from Table 3 indicated that the Gap-fill group obtained the highest scores ($M = 46, SD = 15$), followed by the SpotDif group ($M = 35, SD = 12$), IEEI group ($M = 32, SD = 10$) and the Control group ($M = 25, SD = 12$). A mixed between-within subjects ANOVA conducted for the pre-test and delayed post-test scores revealed that all groups performed significantly better on the delayed post-test than on the pre-test since a significant effect for Time was found, $F(1,36) = 296, p < .001, \eta_p^2 = .89$ (see Table 8), as was a significant effect for Group, $F(3,36) = 4.7, p < .05, \eta_p^2 = .28$ and a significant interaction effect between Group and Time $F(3,36) = 4.7, p < .05, \eta_p^2 = .28$.

Table 8

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-Subjects Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>23,222</td>
<td>1</td>
<td>23,222</td>
<td>296</td>
<td>.000</td>
<td>.89</td>
</tr>
<tr>
<td>Time * Group</td>
<td>1,117</td>
<td>3</td>
<td>372</td>
<td>4.7</td>
<td>.007</td>
<td>.28</td>
</tr>
<tr>
<td>Error</td>
<td>2,825</td>
<td>36</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between-Subjects Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>1,133</td>
<td>3</td>
<td>378</td>
<td>4.7</td>
<td>.007</td>
<td>.28</td>
</tr>
<tr>
<td>Error</td>
<td>2,890</td>
<td>36</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The significant interaction effect between Group and Time suggests that the gain in Productive Knowledge test scores vary across groups. Post hoc analyses using Games-Howell’s criterion of
significance ($\alpha = .05$) revealed that there was a significant difference between the Control group and the Gap-fill group. There was no statistically significant difference in the mean scores of the Control group, IEEI and SpotDif groups; neither was there a statistically significant difference in the mean scores of the IEEI, SpotDif and Gap-fill groups (see Table 9).

**Table 9**
Multiple comparisons of groups for Productive Knowledge delayed post-test using Games-Howell ($\alpha = .05$)

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>$p$</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-3.55</td>
<td>2.492</td>
<td>.501</td>
<td>-10.61</td>
<td>3.51</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-10.45*</td>
<td>3.042</td>
<td>.015</td>
<td>-19.09</td>
<td>-1.81</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-5.15</td>
<td>2.767</td>
<td>.279</td>
<td>-12.97</td>
<td>2.67</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>3.55</td>
<td>2.492</td>
<td>.501</td>
<td>-3.51</td>
<td>10.61</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-6.90</td>
<td>2.899</td>
<td>.122</td>
<td>-15.20</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-1.60</td>
<td>2.609</td>
<td>.926</td>
<td>-9.01</td>
<td>5.81</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>10.45*</td>
<td>3.042</td>
<td>.015</td>
<td>1.81</td>
<td>19.09</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>6.90</td>
<td>2.899</td>
<td>.122</td>
<td>-1.40</td>
<td>15.20</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>5.30</td>
<td>3.138</td>
<td>.358</td>
<td>-3.59</td>
<td>14.19</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>5.15</td>
<td>2.767</td>
<td>.279</td>
<td>-2.67</td>
<td>12.97</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>1.60</td>
<td>2.609</td>
<td>.926</td>
<td>-5.81</td>
<td>9.01</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-5.30</td>
<td>3.138</td>
<td>.358</td>
<td>-14.19</td>
<td>3.59</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

The only significant mean difference is between the Control group and Gap-fill group, indicating that the Gap-fill group was the only treatment group that significantly outperformed the Control group on the Productive Knowledge delayed post-test.

However, using Cohen’s convention for the magnitude of effect sizes, the effect size for the IEEI group ($d = 0.66$) as compared to the Control group was moderate (Cohen, 1977), but the
magnitude of instructional effect on the Productive delayed test scores of both the Gap-fill group \((d = 1.53)\) and SpotDif group \((d = 0.87)\) as compared to the Control group exceeded Cohen’s convention for a large effect \((d = 0.8)\), suggesting a large magnitude of instructional effect for both treatment conditions. As in the immediate post-test, \(d\) was largest for the Gap-fill group, followed by the SpotDif group and IEEI. It should be noted that even though the effect size for the IEEI group as compared to the Control group \((d = 0.66)\) was no longer at the large effect level according to Cohen’s standard, it exceeded the clinically significant level \((d = 0.5\) i.e. something really changed) as determined by Wolf (1986) guideline.

Figure 3 illustrates the increase in test scores from pre-test to delayed post-tests of all four groups. Note that there were small differences in the pre-test scores across groups not reflected in this figure.

**Figure 3**
Productive Knowledge test scores in the pre-test and delayed post-test

As can be seen from Figure 3, even the Control group apparently continued to perform significantly better in the delayed post-test. A repeated measures ANOVA for the Control
group’s scores on these two tests showed a significant effect of Time on their performance in the tests, $F(1,9) = 41, p < .001, \eta^2_p = .82$.

4.2. Receptive Knowledge

In response to research question: Do IEEI, Gap-fill, and SpotDif help learners acquire higher levels of receptive knowledge of FSs? If yes, which type of treatment is more effective?

Mixed between-within subjects ANOVAs conducted on the Productive Knowledge pre-tests (see chapter 2, section 3.4. for information about how Receptive Knowledge pre-test scores were calculated), immediate and delayed post-tests revealed statistically significant improvement among learners of all groups from pre-tests to post-tests. Overall results revealed significant differences between the Control group and two treatment groups, IEEI and SpotDif, and no statistically significant difference across experimental groups. In the immediate post-test, all treatment groups outperformed the Control group at a significant level. Effect sizes for all treatment groups as compared to the Control group far exceed Cohen’s convention for a large effect. In the delayed post-test, no statistically significant difference was found across groups, but effect sizes using Cohen’s $d$ remain above the large level for all treatment groups as compared to the Control group. When examining more closely learners’ performance in the Form and Meaning sections of the Receptive Knowledge post-tests, we find that

(1) in the Form section of the immediate post-test, all treatment groups obtained better scores than the Control group at a statistically significant level, and effect sizes for all treatment groups as compared to the Control group exceed Cohen’s convention for large effect sizes.
(2) In the Meaning section of the immediate post-test, statistically significant difference exists between the Control group and two treatment groups (Gap-fill and IEEI), yet effect sizes remain large for all groups, and

(3) In the Form section of the delayed post-test, only the Gap-fill group outperformed the Control group at a statistically significant level, but effect sizes continue to exceed the large level using Cohen’s convention. In the Meaning section of the delayed post-test, there were no statistically significant differences across groups, and effect size for only the IEEI group remained large.

4.2.1. Overall Results

Table 10 summarizes the raw scores obtained by learners in 4 groups in the Receptive Knowledge pre-test, immediate and delayed post-test. The highest possible score for the Receptive Knowledge test is 60 since there are 30 target FSs, each with a maximum score of 2 if learners gave the correct answers in both the Form and Meaning section of the question.

<table>
<thead>
<tr>
<th>Timing</th>
<th>Pre-test</th>
<th>Immediate Post-test</th>
<th>Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control group</td>
<td>IEEI</td>
<td>Gap-fill</td>
</tr>
<tr>
<td>Mean</td>
<td>3</td>
<td>1</td>
<td>.80</td>
</tr>
<tr>
<td>SD</td>
<td>2.2</td>
<td>1.4</td>
<td>1</td>
</tr>
</tbody>
</table>

Normality, independence and sphericity assumptions were met for a mixed between-within subjects ANOVA to be conducted. Mauchly’s test of sphericity yielded $p > .05$. The homogeneity of variance assumption was violated, with Levene’s test of homogeneity of variance showing significant unequal variances in the pre-test and immediate post-tests, but not the delayed
post-tests. Thus ANOVA was conducted with Games-Howell, analyses ($\alpha = .05$) which do not assume equal variance. In sections 4.2.2 and 4.2.3 similar procedures will apply.

A mixed between-within subjects ANOVA was conducted to compare the mean scores of the four groups in order to examine whether the scores improved significantly from pre-test to immediate and delayed post-test and if there were significant differences across groups. Independent variables are Group and Time; dependent variable is the test scores on the Vocabulary Knowledge Scale Test and the immediate as well delayed Receptive Knowledge test. There were a significant effect for Time across the pre-test, the immediate and delayed post-test, $F(2,72) = 1513, p < .001$, $\eta_p^2 = .98$, a significant interaction effect between Group and Time $F(6,72) = 126, p < .001$, $\eta_p^2 = .34$, and a significant effect for Group $F(3,36) = 5.2, p < .005$, $\eta_p^2 = .30$. Table 11 summarizes the results of this mixed ANOVA.

**Table 11**
Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Receptive Knowledge pre-test, immediate and delayed post-test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects</td>
<td>Time</td>
<td>61562</td>
<td>2</td>
<td>30781</td>
<td>1513</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Time * Group</td>
<td>753</td>
<td>6</td>
<td>126</td>
<td>6.2</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>1464</td>
<td>72</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between-Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects</td>
<td>Group</td>
<td>831</td>
<td>3</td>
<td>277</td>
<td>5.2</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>1914</td>
<td>36</td>
<td>53</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All four groups displayed significant improvement from the pre-test to the immediate post-test; however, the significant interaction for group and time suggested that the gain in scores differed across groups. Games-Howell post hoc analyses ($\alpha = .05$) were also conducted (see Table 12). Results revealed significant differences between the Control group and IEEI and SpotDif groups. There was no statistically significant difference across experimental groups.
Table 12
Multiple comparisons of groups for Receptive Knowledge immediate and delayed post-tests using Games-Howell ($\alpha = .05$)

<table>
<thead>
<tr>
<th>(l) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>Sig.</th>
<th>95% Confidence Interval Lower Bound</th>
<th>95% Confidence Interval Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-5.87*</td>
<td>1.979</td>
<td>.049</td>
<td>-11.70</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>IEEI</td>
<td>-5.70</td>
<td>2.235</td>
<td>.088</td>
<td>-12.07</td>
</tr>
<tr>
<td>SpotDif</td>
<td>IEEI</td>
<td>-6.53*</td>
<td>2.163</td>
<td>.037</td>
<td>-12.74</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>5.87*</td>
<td>1.979</td>
<td>.049</td>
<td>.03</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>IEEI</td>
<td>.17</td>
<td>1.553</td>
<td>1.000</td>
<td>-4.31</td>
</tr>
<tr>
<td>SpotDif</td>
<td>IEEI</td>
<td>-.67</td>
<td>1.448</td>
<td>.967</td>
<td>-4.81</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>SpotDif</td>
<td>5.70</td>
<td>2.235</td>
<td>.088</td>
<td>-.67</td>
</tr>
<tr>
<td>SpotDif</td>
<td>SpotDif</td>
<td>6.53*</td>
<td>2.163</td>
<td>.037</td>
<td>.33</td>
</tr>
<tr>
<td>SpotDif</td>
<td>IEEI</td>
<td>.67</td>
<td>1.448</td>
<td>.967</td>
<td>-3.48</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Gap-fill</td>
<td>.83</td>
<td>1.781</td>
<td>.965</td>
<td>-4.21</td>
</tr>
</tbody>
</table>

* Based on observed means.
The error term is Mean Square(Error) = 17.723.

* The mean difference is significant at the .05 level.

Figure 4 illustrates the changes in the four groups’ mean scores in the Receptive Knowledge pretest, immediate and delayed post-tests.
4.2.2. Results from the Immediate Post-test

As shown in the results of the VKS test (see Table 10), students’ self-reported receptive knowledge of the target formulaic sequences was very low, as shown in the results of Vocabulary Knowledge Scale Test (pre-test). All four groups started at relatively the same level. The average scores of the four groups are as follows: the Control group \( M = 3 \) (SD = 2.2), the IEEI group \( M = 1 \) (SD = 1.4), the Gap-fill group \( M = .80 \) (SD = 1), and SpotDif \( M = 2.3 \) (SD = 2.7). The one-way ANOVA conducted on the groups’ means showed statistically significant differences across group, \( F(3, 39) = 2.99, p < .05 \). However, post hoc analyses using the Games-Howell post hoc criterion for significance indicated the homogeneity of the four groups’ performance in the pre-test.

In the immediate post-test, mean scores (with standard deviations in parentheses) for the Control group, IEEI group, Gap-fill group and SpotDif group were 44 (SD = 9.4), 56 (SD = 2.7), 56
(SD = 3.4), and 55 (SD = 2.6) respectively, with the highest score possible being 60. Standard deviation of the Control group’s test scores was greater than standard deviations of the experimental groups, suggesting that the treatment conditions reduced variance in performance of members within the same group.

A mixed between-within subjects ANOVA was conducted to compare the mean scores of the four groups in order to examine whether the scores improved significantly from pre-test to immediate post-test and if there were significant differences across groups. Independent variables are Group and Time; dependent variable is the test scores on the Vocabulary Knowledge Scale Test and the immediate Receptive Knowledge post-test.

Results revealed a significant effect for Time across the pre-test and the immediate post-test, $F(1,36) = 3242, p < .001, \eta_p^2 = .99$, a significant interaction effect between Group and Time $F(3,36) = 15, p < .001, \eta_p^2 = .56$, and a significant effect for Group $F(3,36) = 8.4, p < .001, \eta_p^2 = .41$.

Table 13 summarizes the results of this mixed ANOVA.

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>51461</td>
<td>1</td>
<td>51461</td>
<td>3242</td>
<td>.000</td>
<td>.99</td>
</tr>
<tr>
<td>Time * Group</td>
<td>713</td>
<td>3</td>
<td>238</td>
<td>15</td>
<td>.000</td>
<td>.56</td>
</tr>
<tr>
<td>Error</td>
<td>571</td>
<td>36</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>406</td>
<td>3</td>
<td>135</td>
<td>8.4</td>
<td>.000</td>
<td>.41</td>
</tr>
<tr>
<td>Error</td>
<td>579</td>
<td>36</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All four groups displayed significant improvement from the pre-test to the immediate post-test; however, the significant interaction for group and time suggested that the gain in scores differed across groups. Post hoc analyses using Games-Howell ($\alpha = .05$), as shown in Table 14,
revealed two important findings: (1) All treatment groups outperformed the Control group in the immediate post-test at a statistically significant level (2) There were no statistically significant differences among the three different types of treatment.

Table 14
Multiple comparisons of groups for Receptive Knowledge immediate post-test using Games-Howell ($\alpha = .05$)

<table>
<thead>
<tr>
<th>(l) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>$p$</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-5.25*</td>
<td>1.550</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-4.90*</td>
<td>1.571</td>
<td>.038</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-5.40*</td>
<td>1.580</td>
<td>.022</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>5.25*</td>
<td>1.550</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>.35</td>
<td>.850</td>
<td>.976</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-.15</td>
<td>.866</td>
<td>.998</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>4.90*</td>
<td>1.571</td>
<td>.038</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-.35</td>
<td>.850</td>
<td>.976</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-.50</td>
<td>.904</td>
<td>.944</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>5.40*</td>
<td>1.580</td>
<td>.022</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>.15</td>
<td>.866</td>
<td>.998</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>.50</td>
<td>.904</td>
<td>.944</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

The effect sizes for this analysis, respectively for the IEEI group ($d = 1.82$), Gap-fill group ($d = 1.71$), and SpotDif group ($d = 1.68$), as compared to the Control group, were found to considerably exceed Cohen’s convention for a large effect ($d = .80$), suggesting a large magnitude of instructional effect. $d$ was largest for the IEEI group, followed by the Gap-fill group and SpotDif group.

Figure 5 illustrates the differences among four groups’ mean scores in the Receptive Knowledge immediate post-test despite their relatively equal performance on the pre-test.
As can be seen from Figure 5, even the Control group seemingly displays a significant increase in mean scores from the pre-test to the immediate post-test. A repeated measures ANOVA for the Control group’s scores on these two tests showed a significant effect of Time on their performance in the tests, $F(1,19) = 432, p < .001, \eta_p^2 = .96$.

Learners’ scores on the Receptive Knowledge immediate post-test were then broken down into two sub sections: meaning and form. One-way ANOVAs were conducted to detect possible differences across groups.

**Form section**

With regards to learners’ performance in form-related questions of the immediate post-test, the IEEI group’s mean score was highest ($M = 27, SD = 2.2$), followed by the SpotDif group ($M = 27, SD = 1.7$), the Gap-fill group ($M = 27, SD = 2.3$) and the Control group ($M = 19, SD = 5.9$) (See table 15).
Table 15
Mean scores of by the four groups on the Form and Meaning sections of the Receptive Knowledge immediate and delayed post-test

<table>
<thead>
<tr>
<th></th>
<th>Immediate Post-test</th>
<th>Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Gap- fill</td>
</tr>
<tr>
<td><strong>Form</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>19</td>
<td>27</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>5.9</td>
<td>2.2</td>
</tr>
<tr>
<td><strong>Meaning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>25</td>
<td>29</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>3.9</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Levene’s test for homogeneity of variance yielded \( p < .05 \); thus in this one-way ANOVA, Welch’s procedure was used as well as a Games-Howell post hoc test. Results from this test suggested a significant effect for Group \( F (3,19) = 5.6, p < .01 \). As seen in Table 16, post hoc analyses using Games-Howell (\( \alpha = .05 \)) indicated that (1) the Control group’s test scores were significantly lower than those of the 3 treatment groups and (2) there was no statistically significant difference among the three experimental groups.
Table 16
Multiple comparisons of groups for Receptive Knowledge immediate post-test (Form section) using Games-Howell ($\alpha = .05$)

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>$p$</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-8.300*</td>
<td>1.996</td>
<td>.007</td>
<td>-14.27</td>
<td>-2.33</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-7.800*</td>
<td>2.003</td>
<td>.011</td>
<td>-13.78</td>
<td>-1.82</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-8.000*</td>
<td>1.949</td>
<td>.009</td>
<td>-13.91</td>
<td>-2.09</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>8.300*</td>
<td>1.996</td>
<td>.007</td>
<td>2.33</td>
<td>14.27</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>.500</td>
<td>.998</td>
<td>.958</td>
<td>-2.32</td>
<td>3.32</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>.300</td>
<td>.885</td>
<td>.986</td>
<td>-2.22</td>
<td>2.82</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>7.800*</td>
<td>2.003</td>
<td>.011</td>
<td>1.82</td>
<td>13.78</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-.500</td>
<td>.998</td>
<td>.958</td>
<td>-3.32</td>
<td>2.32</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-.200</td>
<td>.901</td>
<td>.996</td>
<td>-2.76</td>
<td>2.36</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>8.000*</td>
<td>1.949</td>
<td>.009</td>
<td>2.09</td>
<td>13.91</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-.300</td>
<td>.885</td>
<td>.986</td>
<td>-2.82</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>.200</td>
<td>.901</td>
<td>.996</td>
<td>-2.36</td>
<td>2.76</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

The effect sizes for this analysis, respectively for the IEEI group ($d = 1.86$), Gap-fill group ($d = 1.74$), and SpotDif group ($d = 1.83$), as compared to the Control group, were found to considerably exceed Cohen’s convention for a large effect ($d = .80$), suggesting a large magnitude of instructional effect. $d$ was largest for the IEEI group, followed by the SpotDif group and Gap-fill group.

**Meaning section**
With regards to learners’ performance in meaning-related questions of the immediate post-test, the average score of the IEEI group was highest ($M = 29$, $SD = 1.5$), followed by the Gap-fill group ($M = 29$, $SD = 1.4$), SpotDif group ($M = 28$, $SD = 1.9$) and the Control group ($M = 25$, $SD = 3.9$) (see
Table 15, p. 82). Levene’s test result indicates unequal variances across groups, thus Welch’s procedure was implemented and Games-Howell post hoc test was used.

Results from a one-way ANOVA test revealed a significant effect for Group $F (3,19) = 3.6, p < .05$. As seen in Table 17, post-hoc analyses using Games-Howell ($\alpha = .05$) showed a statistically significant difference between the Control group and the Gap-fill as well as the IEEI groups. No statistically significant difference was found between the Control and SpotDif conditions, as between the IEEI and Gap-fill groups.

Table 17
Multiple comparisons of groups for Receptive Knowledge immediate post-test (Meaning section) using Games-Howell ($\alpha = .05$)

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>$p$</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-4.400*</td>
<td>1.322</td>
<td>.028</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>-4.100*</td>
<td>1.320</td>
<td>.041</td>
<td>-8.05</td>
</tr>
<tr>
<td>SpotDif</td>
<td>-3.600</td>
<td>1.372</td>
<td>.087</td>
<td>-7.63</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>4.400*</td>
<td>1.322</td>
<td>.028</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>.300</td>
<td>.644</td>
<td>.966</td>
<td>-1.52</td>
</tr>
<tr>
<td>SpotDif</td>
<td>.800</td>
<td>.744</td>
<td>.709</td>
<td>-1.31</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>4.100*</td>
<td>1.320</td>
<td>.041</td>
</tr>
<tr>
<td>IEEI</td>
<td>-.300</td>
<td>.644</td>
<td>.966</td>
<td>-2.12</td>
</tr>
<tr>
<td>SpotDif</td>
<td>.500</td>
<td>.740</td>
<td>.905</td>
<td>-1.60</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>3.600</td>
<td>1.372</td>
<td>.087</td>
</tr>
<tr>
<td>IEEI</td>
<td>-.800</td>
<td>.744</td>
<td>.709</td>
<td>-2.91</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>-.500</td>
<td>.740</td>
<td>.905</td>
<td>-2.60</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

The effect sizes for this analysis, respectively for the IEEI group ($d = 1.45$), Gap-fill group ($d = 1.39$), and SpotDif group ($d = 1.17$), as compared to the Control group, were found to exceed
Cohen’s convention for a large effect \((d = .80)\), suggesting a large magnitude of instructional effect. \(d\) was largest for the IEEI group, followed by the Gap-fill group and SpotDif group.

4.2.2. Results from the Delayed Post-test

Results from Table 10 indicated that the SpotDif group obtained the highest scores \((M = 50, SD = 7.7)\), followed by the Gap-fill group \((M= 48, SD = 9.2)\), IEEI group \((M = 48, SD = 5)\) and the Control group \((M = 41, SD = 8.8)\). A mixed between-within subjects ANOVA (Table 18) conducted for the pre-test and delayed post-test scores revealed that all groups performed significantly better in the delayed post-test than in the pre-test as a significant effect for Time was found, \(F (1,36) = 1442, p < .001, \eta_p^2 = .98\) (see Table 16), as was a significant interaction effect between Group and Time \(F (3,36) = 4.7, p < .05, \eta_p^2 = .28\). However, no significant effect for Group was found, \(F (3,36) = 1.5, p > .05, \eta_p^2 = .11\). Levene’s test of homogeneity of variances in the delayed post-test scores yielded \(p > .05\), suggesting equal variances of test scores across groups.

Table 18
Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Receptive Knowledge pre-test and delayed post-test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>(\eta_p^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>40186</td>
<td>1</td>
<td>40186</td>
<td>1442</td>
<td>.000</td>
<td>.98</td>
</tr>
<tr>
<td>Time * Group</td>
<td>330</td>
<td>3</td>
<td>110</td>
<td>3.9</td>
<td>.016</td>
<td>.25</td>
</tr>
<tr>
<td>Error</td>
<td>2825</td>
<td>36</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>170</td>
<td>3</td>
<td>57</td>
<td>1.5</td>
<td>.23</td>
<td>.11</td>
</tr>
<tr>
<td>Error</td>
<td>2,890</td>
<td>36</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The significant interaction effect between Group and Time suggests that the gain in Productive Knowledge test scores vary across groups. Post hoc analyses using Games-Howell’s criterion of
significance ($\alpha = .05$) revealed that there was no statistically significant differences in the mean scores of the Control group in comparison with all treatment groups (see Table 19).

**Table 19**
Multiple comparisons of groups for Receptive Knowledge delayed post-test using Game-Howell ($\alpha = .05$)

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>$p$</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-2.55</td>
<td>1.721</td>
<td>.471</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-2.55</td>
<td>2.075</td>
<td>.617</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-4.05</td>
<td>2.110</td>
<td>.255</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>2.55</td>
<td>1.721</td>
<td>.471</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>0.00</td>
<td>1.759</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-1.50</td>
<td>1.800</td>
<td>.838</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>2.55</td>
<td>2.075</td>
<td>.617</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>0.00</td>
<td>1.759</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-1.50</td>
<td>2.141</td>
<td>.895</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>4.05</td>
<td>2.110</td>
<td>.255</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>1.50</td>
<td>1.800</td>
<td>.838</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>1.50</td>
<td>2.141</td>
<td>.895</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

However, using Cohen’s convention for the magnitude of effect sizes, the effect size for the IEEI group ($d = 0.99$), the Gap-fill group ($d = 0.81$) and SpotDif group ($d = 1.06$) as compared to the Control group was large (Cohen, 1977), suggesting a large magnitude of instructional effect for all treatment conditions. $d$ was largest for the SpotDif group, followed by the IEEI group and the Gap-fill group.

Figure 6 illustrates the increase in test scores from pre-test to delayed post-tests of all four groups.
As can be seen from Figure 6, even the Control group seemingly continued to perform significantly better in the delayed post-test. A repeated measures ANOVA for the Control group’s scores on these two tests showed a significant effect of Time on their performance in the tests, $F(1,19) = 41, p < .001, \eta^2_p = .97$.

**Form section**

According to the results summarized in table 15 (p. 82), the Gap-fill group scored the highest among the four groups ($M = 22, SD = 5.2$), followed by the IEEI group ($M = 20, SD = 4.3$), SpotDif ($M = 19, SD = 2.2$) and the Control group ($M = 16, SD = 5.1$). All assumptions for a one-way ANOVA to be conducted were met, including the homogeneity of variance assumption (Levene’s test yielded $p < .05$). Results from a one-way ANOVA test revealed a significant effect for Group $F(3,39) = 3.3, p < .05$, analyses using Bonferroni ($\alpha = .05, \text{adjusted } \alpha = .083$) showed that there was only statistically significant difference between the Gap-fill group and the Control group (Table 20).
Table 20
Multiple comparisons of groups for Receptive Knowledge delayed post-test (Form section) using Bonferroni (α = .05)

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>p</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-4.200</td>
<td>1.960</td>
<td>.234</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-6.000*</td>
<td>1.960</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-3.600</td>
<td>1.960</td>
<td>.447</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>4.200</td>
<td>1.960</td>
<td>.234</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-1.800</td>
<td>1.960</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>.600</td>
<td>1.960</td>
<td>1.000</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>6.000*</td>
<td>1.960</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>1.800</td>
<td>1.960</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>2.400</td>
<td>1.960</td>
<td>1.000</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>3.600</td>
<td>1.960</td>
<td>.447</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-.600</td>
<td>1.960</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-2.400</td>
<td>1.960</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

The effect sizes for this analysis, respectively for the IEEI group ($d = 0.88$), Gap-fill group ($d = 1.14$), and SpotDif group ($d = 0.91$), as compared to the Control group, were found to exceed Cohen’s convention for a large effect ($d = .80$), suggesting a large magnitude of instructional effect. $d$ was largest for Gap-fill the group, followed by the SpotDif group and IEEI group.

**Meaning section**
In the meaning section of the Receptive Knowledge delayed post-test, the IEEI group’s scores ($M = 28$, $SD = 1.8$) were better than both the Gap-fill group ($M = 26.2$, $SD = 4.7$), the SpotDif group ($M = 26.3$, $SD = 2.5$) and the Control group ($M = 26.3$, $SD = 2.5$). All assumptions for a one-way ANOVA to be conducted are met, including the homogeneity of variance assumption (Levene’s
test yielded \( p < .05 \). A one-way ANOVA analyses indicated the absence of a significant effect for Group, \( F (3,39) = 1.3, p > .05 \). As seen in Table 21, analyses using Bonferroni (\( \alpha = .05 \), adjusted \( \alpha = .083 \)) showed no statistically significant difference across the four groups.

**Table 21**
Multiple comparisons of groups for Receptive Knowledge delayed post-test (Meaning section) using Bonferroni (\( \alpha = .05 \))

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>(J) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>( p )</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-3.200</td>
<td>1.605</td>
<td>.323</td>
<td>-7.68 - 1.28</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-1.300</td>
<td>1.605</td>
<td>1.000</td>
<td>-5.78 - 3.18</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>-1.300</td>
<td>1.605</td>
<td>1.000</td>
<td>-5.78 - 3.18</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>3.200</td>
<td>1.605</td>
<td>.323</td>
<td>-1.28 - 7.68</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>1.900</td>
<td>1.605</td>
<td>1.000</td>
<td>-2.58 - 6.38</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>1.900</td>
<td>1.605</td>
<td>1.000</td>
<td>-2.58 - 6.38</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>1.300</td>
<td>1.605</td>
<td>1.000</td>
<td>-3.18 - 5.78</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-1.900</td>
<td>1.605</td>
<td>1.000</td>
<td>-6.38 - 2.58</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>0.000</td>
<td>1.605</td>
<td>1.000</td>
<td>-4.48 - 4.48</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>1.300</td>
<td>1.605</td>
<td>1.000</td>
<td>-3.18 - 5.78</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-1.900</td>
<td>1.605</td>
<td>1.000</td>
<td>-6.38 - 2.58</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>0.000</td>
<td>1.605</td>
<td>1.000</td>
<td>-4.48 - 4.48</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

The effect sizes for this analysis, respectively for the IEEI group (\( d = 0.94 \)), Gap-fill group (\( d = 0.28 \)), and SpotDif group (\( d = 0.36 \)), as compared to the Control group, were found to be large for the IEEI group, and small for the Gap-fill and SpotDif groups. It should be noted that the effect sizes for both the Gap-fill and SpotDif groups exceeded the educationally significant level (\( d = 0.25 \)) (i.e. something was learned) as determined by Wolf (1986).
4.3. The Awareness Test

In response to research question: Do IEEI, Gap-fill, and SpotDif help learners improve their ability to notice FSs in a reading text? If yes, which type of treatment is more effective?

Mixed between-within subjects ANOVAs conducted on the Awareness pre-tests, immediate and delayed post-tests revealed no statistically significant change in test scores over time in all four groups. The IEEI and Gap-fill groups improved over time while the SpotDif and Control groups’ scores decreased from pre-test to post-tests. On both the immediate and delayed post-test, there was no statistically significant difference across groups. In the immediate post-test, the effect sizes were small for the IEEI group, null for the Gap-fill group, and negative for the SpotDif group as compared to the Control group. In the delayed post-test, effect sizes for IEEI group increased to large, small for Gap-fill, and remained negative for the SpotDif group as compared to the Control group.

4.3.1. Overall results

Table 22 summarizes the raw scores obtained by learners in 4 groups on the Awareness pre-test, immediate and delayed post-test. The highest possible score for the Awareness test is 84 since I can identify 84 qualified FSs, 1 point for each, in the reading passage (see chapter 3 for the criteria of choosing FSs).

**Table 22**

Mean scores of the four groups on the Awareness pre-test, immediate and delayed post-test

<table>
<thead>
<tr>
<th>Timing</th>
<th>Pre-test</th>
<th>Immediate Post-test</th>
<th>Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control group</td>
<td>IEEI</td>
<td>Gap-fill</td>
</tr>
<tr>
<td>Mean</td>
<td>5.7</td>
<td>5.8</td>
<td>3.8</td>
</tr>
<tr>
<td>SD</td>
<td>2.5</td>
<td>2.4</td>
<td>3.1</td>
</tr>
</tbody>
</table>
Levene’s test results yielded $p > .05$ for pre-test, immediate and delayed post-test, indicating that assumption of homogeneity of variances was met. Mauchly’s test of sphericity yielded $p < .05$, thus the degree of freedom reported below will be adjusted using the Greenhouse-Geisser correction.

A mixed between-subjects ANOVA conducted on the learners’ Awareness pre-test, immediate and delayed post-tests revealed no significant effect for Time, $F(1.8,66) = .75, p > .05, \eta^2 = .02$, no significant effect for Group, $F(3,36) = 1.89, p > .05, \eta^2 = .14$, and a significant interaction effect between Time and Group, $F(5.5,66) = 4.3, p < .01, \eta^2 = .26$. Table 23 summarizes the results of this mixed ANOVA.

**Table 23**  
Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Receptive Knowledge pre-test, immediate and delayed post-test

<table>
<thead>
<tr>
<th>Within-Subjects Effects</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>6.2</td>
<td>1.82</td>
<td>3.4</td>
<td>.75</td>
<td>.47</td>
<td>.02</td>
</tr>
<tr>
<td>Time * Group</td>
<td>107</td>
<td>5.47</td>
<td>20</td>
<td>4.3</td>
<td>.001</td>
<td>.26</td>
</tr>
<tr>
<td>Error</td>
<td>298</td>
<td>66</td>
<td>4.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>127</td>
<td>3</td>
<td>42</td>
<td>1.9</td>
<td>.15</td>
<td>.14</td>
</tr>
<tr>
<td>Error</td>
<td>807</td>
<td>36</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7 illustrates the mean scores of four groups on the pre-test, the immediate post-test and the delayed post-test.
4.3.2. Results from the Immediate Post-test

Results in Table 22 show the groups’ performance on the Awareness test. In the pretest, the IEEI and SpotDif group obtained the highest scores (M = 5.8, SD = 2.4 and SD = 2.1 respectively), followed closely by the Control group (M= 5.7, SD = 2.5), and the Gap-fill group (M = 3.8, SD = 3.1). The one-way ANOVA conducted on the groups’ means on the pre-test showed no significant differences among them, $F(3, 39) = 1.3, p = .29$. Post hoc analyses using the Tukey’s HSD criterion for significance indicated the homogeneity of the four groups’ performance on the pre-test. On the immediate post-test, mean scores (with standard deviations in parentheses) for the Control group, the scores of the IEEI group was the highest (M = 6.1, SD = 2.9), followed by the Gap-fill group (M = 5.1, SD = 3.9), the Control group (M = 5.1, SD = 3.4) and the SpotDif group (M = 2.6, SD = 1.8).
A mixed between-within subjects ANOVA was conducted to compare the mean scores of the four groups in order to examine whether the scores improved significantly from pre-test to immediate post-test and if there were significant differences across groups. Independent variables are Group and Time; dependent variable is the test scores on the Awareness pre-test and immediate post-test.

Results revealed no significant effect for Time across the pre-test and the immediate post-test, $F(1,36) = 1.9$, $p > .05$, $\eta^2_p = .05$, a significant interaction effect between Group and Time $F(3,36) = 5.9$, $p < .005$, $\eta^2_p = .33$, and no significant effect for Group $F(3,36) = 1.04$, $p > .05$, $\eta^2_p = .08$. Table 24 summarizes the results of this mixed ANOVA.

### Table 24
Results of mixed between-within subjects ANOVA for the mean scores of four groups on the Awareness pre-test and immediate post-test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-Subjects Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>6.1</td>
<td>1</td>
<td>6.1</td>
<td>1.91</td>
<td>.176</td>
<td>.05</td>
</tr>
<tr>
<td>Time * Group</td>
<td>56</td>
<td>3</td>
<td>19</td>
<td>5.9</td>
<td>.002</td>
<td>.33</td>
</tr>
<tr>
<td>Error</td>
<td>114</td>
<td>36</td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between-Subjects Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>40</td>
<td>3</td>
<td>13</td>
<td>1.04</td>
<td>.39</td>
<td>.08</td>
</tr>
<tr>
<td>Error</td>
<td>464</td>
<td>36</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All four groups displayed non-significant differences in their performance from the pre-test to the immediate post-test. The significant interaction for Group and Time suggested that the change in scores differed across groups. However, post hoc analyses using Bonferroni ($\alpha = .05$, adjusted $\alpha = .083$), as shown in Table 25, revealed no statistically significant differences in the Awareness test scores among the four groups.
Table 25
Multiple comparisons of groups for Awareness immediate post-test using Bonferroni (α = .05)

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>p</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-.55</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>.95</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>1.20</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>.55</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>1.50</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>1.75</td>
<td>1.135</td>
<td>.791</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>-.95</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-1.50</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>.25</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>-1.20</td>
<td>1.135</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-1.75</td>
<td>1.135</td>
<td>.791</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-.25</td>
<td>1.135</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

The effect size was largest for the IEEI group (d = 0.32) as compared to the Control group, which is above the small effect threshold by Cohen. The effect size of Gap-fill group is null (d = 0), and negative for the SpotDif group (d = -0.92).

Figure 8 illustrates the changes in performance among four groups’ mean scores in the Awareness immediate post-test. As far as the raw scores were concerned, the IEEI group and the Gap-fill group improved their scores from pre-test to immediate post-test while the Control group and SpotDif group displayed a downward trend in test performance.
4.3.3. Results from the Delayed Post-test

Results in Table 22 also show the groups’ performance on the Awareness test from the pre-test to the delayed post-test. In the delayed post-test, mean scores (with standard deviations in parentheses) for the Control group, the scores of the IEEI group were the highest (M = 7.9, SD = 4.8), followed by the Gap-fill group (M = 5.3, SD = 4.2), the Control group (M = 4.3, SD = 3.6) and the SpotDif group (M = 2.8, SD = 2.3).

A mixed between-within subjects ANOVA was conducted to compare the mean scores of the four groups in order to examine whether the scores improved significantly from pre-test to delayed post-test and if there were significant differences across groups. Independent variables are Group and Time; dependent variable is the test scores on the Awareness pre-test and delayed post-test.
Results revealed no significant effect for Time across the pre-test and the delayed post-test, $F(1,36) = 0.15, p > .05, \eta^2_p = .004$, a significant interaction effect between Group and Time $F(3,36) = 5.4, p < .05, \eta^2_p = .31$, and no significant effect for Group $F(3,36) = 1.7, p > .05, \eta^2_p = .12$. Table 26 summarizes the results of this mixed ANOVA.

Table 26
Results of mixed between-within subjects ANOVA for the mean scores of four groups in the Awareness pre-test and delayed post-test

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
</table>
| Within-Subjects  
Effects              |                |    |             |     |    |            |
| Time                | 0.8            | 1.0| 0.8         | 0.1 | 0.7| .004       |
| Time * Group        | 87             | 3.0| 29          | 5.4 | 0.0| .31        |
| Error               | 193            | 36 | 5.4         |     |    |            |
| Between-Subjects  
Effects              |                |    |             |     |    |            |
| Group               | 80             | 3  | 27          | 1.7 | .19| .12        |
| Error               | 569            | 36 | 16          |     |    |            |

All four groups displayed non-significant differences in their performance from the pre-test to the delayed post-test. The significant interaction for group and time suggested that the change in scores differed across groups. However, post hoc analyses using Bonferroni ($\alpha = .05$, adjusted $\alpha = .083$) as shown in Table 27, revealed no statistically significant differences among the four groups.
Table 27
Multiple comparisons of groups for Awareness immediate post-test using Bonferroni (α = .05)

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>Mean Difference (I-J)</th>
<th>SE</th>
<th>p</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>IEEI</td>
<td>-.185</td>
<td>1.257</td>
<td>.898</td>
<td>-5.36</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>.45</td>
<td>1.257</td>
<td>1.000</td>
<td>-3.06</td>
<td>3.96</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>.70</td>
<td>1.257</td>
<td>1.000</td>
<td>-2.81</td>
<td>4.21</td>
</tr>
<tr>
<td>IEEI</td>
<td>Control</td>
<td>1.85</td>
<td>1.257</td>
<td>.898</td>
<td>-1.66</td>
<td>5.36</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>2.30</td>
<td>1.257</td>
<td>.453</td>
<td>-1.21</td>
<td>5.81</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>2.55</td>
<td>1.257</td>
<td>.300</td>
<td>-.96</td>
<td>6.06</td>
</tr>
<tr>
<td>Gap-fill</td>
<td>Control</td>
<td>-.45</td>
<td>1.257</td>
<td>1.000</td>
<td>-3.96</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-2.30</td>
<td>1.257</td>
<td>.453</td>
<td>-5.81</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>SpotDif</td>
<td>.25</td>
<td>1.257</td>
<td>1.000</td>
<td>-3.26</td>
<td>3.76</td>
</tr>
<tr>
<td>SpotDif</td>
<td>Control</td>
<td>-.70</td>
<td>1.257</td>
<td>1.000</td>
<td>-4.21</td>
<td>2.81</td>
</tr>
<tr>
<td></td>
<td>IEEI</td>
<td>-2.55</td>
<td>1.257</td>
<td>.300</td>
<td>-6.06</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>Gap-fill</td>
<td>-.25</td>
<td>1.257</td>
<td>1.000</td>
<td>-3.76</td>
<td>3.26</td>
</tr>
</tbody>
</table>

Based on estimated marginal means.
* The mean difference is significant at the .05 level

The effect size was large for the IEEI group ($d = 0.84$) as compared to the Control group. The effect size for the Gap-fill group is small ($d = 0.25$), and negative for the SpotDif group ($d = -0.50$).

Figure 9 illustrates the changes in performance among four groups’ mean scores in the Awareness immediate post-test. The IEEI group and the Gap-fill group improved their scores from pre-test to immediate post-test while the Control group and SpotDif group displayed a downward trend in test performance.
4.3.4. Words underlined in the Awareness test

In the pre-tests as well as both post-tests, learners were also asked to underline words they thought were useful for improving language learning. This was intentionally used to strengthen the power of the ANOVAs and to confirm that the significant effect for Group, or lack thereof, and also account for the amount of attention learners directed towards individual words. Table 28 summarizes the descriptive statistics regarding the number of words underlined. One-way ANOVAs were conducted for all tests, indicating there was no statistically significant difference across groups in terms of the number of the words underlined. Specifically, for the pre-test, $F(3, 39) = .12, p = .95$; for the immediate post-test, $F(3, 39) = .48, p = .70$; and for the delayed post-test, $F(3, 39) = .48, p = .70$. Post hoc analyses using Tukey’s HSD for all tests confirmed that scores across groups did not differ significantly.
Table 28
Mean scores of the four groups on the Awareness pre-test, immediate and delayed post-test with regards to number of words underlined

<table>
<thead>
<tr>
<th>Timing</th>
<th>Pre-test</th>
<th>Immediate Post-test</th>
<th>Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control group</td>
<td>IEEI</td>
<td>Gap-fill</td>
</tr>
<tr>
<td>Mean</td>
<td>4.8</td>
<td>4.5</td>
<td>4.8</td>
</tr>
<tr>
<td>SD</td>
<td>4.0</td>
<td>4.0</td>
<td>3.6</td>
</tr>
</tbody>
</table>

4.4. Frequency, n-gram length and MI Score in correlation to test scores

In response to research question: Are frequency, n-gram length and MI-score of FSs related to learners’ acquisition of receptive and productive knowledge of FSs?

In order to examine the possible influence of frequency levels (FREQ), n-gram length and MI Score and the acquisition of corresponding FSs, correlations were calculated. FREQ and participants’ test scores on the immediate Productive Knowledge post-test were correlated at a significant level, \( r(28) = .37, p < .05 \). FREQ and learners’ performance on the delayed Productive Knowledge post-tests were somewhat correlated at a level close to significant, \( r(28) = .34, p = .069 \). As can be seen from Table 29, there were few correlations between FREQ and all Receptive Productive tests.

n-gram length was not correlated at a significant level with any test scores, even though the correlation between n-gram length and the Productive Knowledge immediate post-tests was close to significant, \( r(28) = .36, p = .054 \).

MI Score, on the other hand, was inversely correlated with participants’ performance on the Meaning section of the immediate Receptive Knowledge test. This correlation did not endure to the delayed post-test. No correlation was found between MI Score and all Productive Knowledge post-tests, nor between MI Score and the Form section of the Receptive Knowledge post-tests.
Table 29  
Correlation between FREQ, n-gram length and MI Score and test scores

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FREQ</td>
<td>Pearson Correlation: <strong>0.368</strong></td>
<td><strong>0.336</strong></td>
<td><strong>0.105</strong></td>
<td><strong>0.185</strong></td>
<td><strong>-0.019</strong></td>
<td><strong>0.059</strong></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td><strong>0.046</strong></td>
<td><strong>0.069</strong></td>
<td><strong>0.581</strong></td>
<td><strong>0.329</strong></td>
<td><strong>0.922</strong></td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>n-gram length</td>
<td>Pearson Correlation: <strong>-0.355</strong></td>
<td><strong>-0.297</strong></td>
<td><strong>-0.236</strong></td>
<td><strong>-0.229</strong></td>
<td><strong>-0.072</strong></td>
<td><strong>0.129</strong></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td><strong>0.054</strong></td>
<td><strong>0.111</strong></td>
<td><strong>0.210</strong></td>
<td><strong>0.224</strong></td>
<td><strong>0.706</strong></td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>MI Score</td>
<td>Pearson Correlation: <strong>0.026</strong></td>
<td><strong>0.053</strong></td>
<td><strong>0.131</strong></td>
<td><strong>-0.401</strong></td>
<td><strong>0.110</strong></td>
<td><strong>-0.151</strong></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td><strong>0.892</strong></td>
<td><strong>0.779</strong></td>
<td><strong>0.489</strong></td>
<td><strong>0.028</strong></td>
<td><strong>0.563</strong></td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Statistical inferences based from results presented in this chapter will be discussed further in Chapter 5.
CHAPTER 5: DISCUSSION

This chapter discusses the implications drawn from results of statistical tests and Cohen’s $d$ effect size calculations conducted on learners’ scores on the pre-tests, immediate post-tests and delayed post-tests. I will begin by addressing each of the four research questions this study initially set out to answer. Quantitative results will be expounded using theories in Second Language Acquisition. Then I will discuss several peripheral findings crucial to understanding how second language learners acquire FSs in the classroom. Finally, I will discuss the some limitations of the study and examine pedagogical implications drawn from it.

5.1. Do IEEl, Gap-fill and SpotDif help learners acquire higher levels of productive knowledge of FSs? If yes, which type of treatment is more effective?

This study substantiates the effectiveness of instructed second language acquisition, and more specifically, corroborates previous empirical findings on the efficacy of instruction on the acquisition of FSs (Bishop, 2004; Boers et al., 2006; Boers & Lindstromberg, 2005; Boers et al., 2004; Hsu, 2010; Jones & Haywood, 2004; Lindstromberg & Boers, 2008a; Rott, 2009; Wood, 2009). From the findings discussed in section 4.1, there was strong evidence that the three types of instruction successfully help learners acquire levels of productive knowledge of FSs. Overall the IEEl, Gap-fill and SpotDif group outperformed the Control group at a statistically significant level. In the immediate post-test, all treatment groups’ mean scores were significantly higher than that of the Control group. In the delayed post-test, however, only the mean score of the Gap-fill group remained significantly higher. Statistical significance testing through mixed between-within subjects ANOVA shows that with regards to the effectiveness of different treatment types, it seems that Gap-fill is most beneficial to the acquisition of the productive
knowledge of these FSs. The Gap-fill group consistently scored the highest on both post-tests, and only the difference between its members’ mean score and that of the Control group remained statistically significant in the delayed post-test.

Despite the lack of statistical significance for the SpotDif and IEEI groups in the Productive Knowledge delayed post-test, the fact that Cohen’s $d$ effect sizes are moderate and large for all treatment groups as compared to the Control group is evidence for educational significance, especially for experimental studies with a small-sized sample. Neill (2008) claims that statistical significance can be misleading when examining effects using small sample size because such a procedure is subject to Type II errors. According to Neill, “statistical significance is not a direct indicator of size of effect, but rather it is a function of sample size, effect size and $p$ level. In these situations, effect sizes and confidence intervals are more informative than significance testing” (Significance testing, para. 8). This means interventions on a large sample with very small effect sizes can have statistical significance, while large effect sizes of treatment on smaller samples are statistically non-significant. Thus in this study it is useful to look at Cohen’s $d$ effect sizes in addition to statistical significance test results in order to determine whether instructional effectiveness is present and to evaluate the magnitude of instruction effect expected in a real setting. As Norris and Ortega (2000) recommend in their meta-analysis, researchers who want to investigate instructional effectiveness in second language teaching should calculate effect sizes because statistical significance tests will not provide answers about the presence, the size and the importance of an effect.

As far as effect sizes are concerned, the two communicative tasks (Gap-fill and SpotDif) were more successful than the IEEI group in helping learners acquire a higher level of productive knowledge of the target FSs, though not at a statistically significant level. Effect sizes using
Cohen’s $d$ in both the immediate and delayed post-tests show that $d$ was largest for the Gap-fill group, followed by SpotDif and then IEEI. In both immediate and delayed post-tests, the effect sizes for the Gap-fill and SpotDif groups as compared to the Control group exceeded Cohen’s convention for a large effect. The effect size for the IEEI group as compared to the Control group was large in the immediate post-test but decreased to the moderate level in the delayed post-test. This finding attests to the effectiveness of interactive tasks in the L2 classroom, and is in line with findings from previous studies (R. Ellis & He, 1999; R. Ellis et al., 1994; Gass & Torres, 2005; Kim, 2008; Kowal & Swain, 1994; Loewen & Philp, 2006), although these studies focus on individual words rather than FSs. This study proves that tasks lead to better productive knowledge acquisition of L2 vocabulary, similar to what R. Ellis and He (1999) and Gass and Torres (2005) found in their studies.

As discussed previously in chapter 2, the Involvement Hypothesis (Laufer & Hustijn, 2001) postulates that the effectiveness of a classroom task is contingent upon the levels of involvement factors, i.e. need, search, and evaluation. Higher levels of need, search and evaluation are conducive to better acquisition of vocabulary. In accordance with the anticipation made based on the Involvement Load Hypothesis prior to the experiment, the two task-based treatment types were more effective than the IEEI group in helping learners acquire higher levels of productive knowledge. The IEEI group did not interact with each other to solve problems created by the Gap-fill or SpotDif group. Instead of actively working on the input, they received instructor-led explanation of the meaning of readily typologically-enhanced FSs in the input text, and a brief introduction of the common mistakes they would likely make with these FSs, as well some strategies to better notice the FSs in any given input. This kind of instruction is only capable of activating low to moderate levels of need, search and evaluate, while the other
two types of treatment activated higher levels of these cognitive processes due to the implementational procedures of the tasks. (see Table 3, Chapter 3) It should be noted that as for receptive knowledge, tasks do not seem to have such an advantage over IEEI. However, the effectiveness of tasks in facilitating learners’ acquisition of Productive Knowledge, which is always much harder to achieve (Milton, 2009; Nation, 2001; Schmitt, 2008), has important implications for the classroom. I will return to this point in a section presented later in this chapter.

The higher levels of productive knowledge acquired by learners in the Gap-fill group compared to the SpotDif group could be explained by the way the two tasks were actually carried out by the learners. Although I did not have a systematic analysis of qualitative data obtained from classroom recordings, for the SpotDif group, observations in the classroom suggest that the levels of need, search and evaluate were compromised by the fact that some learners seemed to rely on their short-term memory to determine which version of the FSs should be chosen. The kind of interaction occurring among learners was relatively cursory, largely limited to inquiring of each other if they remembered what the FSs in the original passage were. Most learners thus completed the task very quickly without evaluating the meaning of the FSs or understanding the surrounding text. The mechanical focus on form during the SpotDif task with apparently little sign of attempting to understand the immediate context of the FSs, nor their meaning, might have caused some learners to acquire the FSs less effectively due to lower involvement load. An explanation of this unintended effect of SpotDif could be the nature of FSs, which are multiword lexical units. This structural composition of FSs could have induced learners’ need to mechanically recall the form of the whole unit, rather than searching for the meaning and understanding the context. In addition, in previous studies where SpotDif was used, it was of a
longer format, with a gap-fill component after learners finished comparing the two versions of the original passage. The shortened format of SpotDif in my study, intended to maintain the same time on task across treatment groups, could be another factor leading to less impressive effects compared to previous research. In comparison, the Gap-fill group seems to carry out the assigned task in a way exactly as planned: they attempted to know the meaning of the FSs, many times through asking the instructor, as well as decoded the meaning of the surrounding text in order to complete the task collaboratively.

Learners’ attention to the context of new FSs has been proven to play an important role in the learning of FSs at the productive level. Pickering (1982) conducted an experiment on the effectiveness of contextualized versus de-contextualized (i.e. translation into L1) presentation of new words and found that context leads to better learning. Furthermore, Cohen and Aphek (1980), cited in Carter (2002), conclude from their studies that learners of higher proficiency levels are more likely to benefit from learning words in context. In the treatment conditions of the current study, target FSs were introduced in context to learners in the reading text, it seems that Gap-fill, the follow-up task that requires learners to attend more closely to the context around new FSs for the second time, was most successful in helping learners to recall productively these FSs in the immediate and delayed post-tests. It is likely that as learners’ attentional resources are drawn to the task of understanding the context, this intensifies the processes of need, search and evaluation in vocabulary acquisition.
5.2. Do IEEI, Gap-fill and SpotDif help learners acquire higher levels of receptive knowledge of FSs? If yes, which type of treatment is more effective?

Overall there seems to be some evidence for the effectiveness of form-focused instruction on the acquisition of Receptive Knowledge of target FSs, although the evidence is not as strong as evidence for the effectiveness of instruction on the acquisition of Productive Knowledge. As can be seen in section 4.2, in the Receptive Knowledge test, overall only the IEEI and SpotDif groups’ mean scores were higher than that of the Control group at a statistically significant level. It should be noted that there is no statistically significant difference across treatment groups. This finding can be potentially misleading if one does not further examine the results in the immediate and delayed post-tests separately.

In terms of instructional effect durability, initially the type of direct instruction as in the IEEI group has immediate positive effects but in the long run, communicative tasks such as SpotDif and Gap-fill are more effective. In the immediate post-test, all treatment groups outperformed the Control group at a statistically significant level, with IEEI obtaining the highest scores, followed by Gap-fill, then SpotDif. In the delayed post-test, the order was slightly changed: SpotDif was the group with the highest scores, followed by Gap-fill, then IEEI. No statistically significant difference was found across groups in the delayed post-test. Even though differences in groups’ mean scores failed to reach statistically significant levels, Cohen’s $d$ effect sizes of all treatment groups as compared to the Control group exceeded Cohen’s large effect convention, even in the delayed post-test, which means there was practical difference between treatment groups and the Control group.
When learners’ performance in the Form and Meaning sections of the Receptive Knowledge tests was examined separately, results show that the effect of instruction is not the same for the acquisition of target FSs’ form and meaning.

With regards to form, statistical significance testing using one-way ANOVA shows an advantage for all 3 treatment groups in the immediate post-test, but only for the Gap-fill group in the delayed post-test. Effect size calculations yielded large effect sizes (according to Cohen’s convention) for all treatment groups at both the immediate and delayed post-tests as compared to the Control group. These findings suggest that while all instructional types were generally successful, the instruction that the Gap-fill group received seems to have the most positive long-term effect on learners’ acquiring the form of the target FSs.

Contrary to expectation based on the Involvement Load Hypothesis, even though the level of need, search and evaluation of the target FSs’ form for the Gap-fill group was predicted to be lower than the SpotDif group, in the Form section of the Receptive Knowledge test, the Gap-fill group obtained higher scores. Prior to this experiment, I hypothesized that the SpotDif activity would yield better retention of form than the Gapfill; however, findings show that the effectiveness of the Gapfill group rivaled that of the SpotDif in the immediate post-test, and was better than the SpotDif group in the delayed post-test (see Table 15, chapter 4). A possible reason that could help explain this counter-evidence for Laufer’s hypothesis is the presence of non-target forms in the SpotDif task. In order to complete the SpotDif component of the class, learners compare two versions of the target FSs, e.g. on the books vs. in the book only one of which is acceptable. Although it had been expected that the exposure to inaccurate form of the target FSs would encourage learners to evaluate the different possible forms, which leads to higher scores of the SpotDif learners in comparison with the IEEI group and the Control group,
this might have increased cognitive demands for the learners in the SpotDif group, causing them to score lower than the Gap-fill group, although not to a statistically significant degree.

Observations in the classroom show that sometimes learners chose these non-targetlike forms when they could not recall the original FSs. Unlike members of the Gap-fill group, as some participants shared with me orally after the class was over, they did not ask questions to the instructors, fearing that the answers would directly reveal which version of the FSs they should choose.

Another possible explanation for the success of the Gap-fill group was that learners’ engagement with the context of FSs might be more conducive to the successful acquisition of form in the long run than more direct explanation of meaning (IEEI) and comparing targetlike vs. non-nontarget like forms of one FSs (SpotDif). As discussed in the previous section, during the Gap-fill activity, learners once again had to read the passage with blanks, attempted to understand more closely the context around those blanks, and chose a FS in the FS bank that would fit best semantically with the gaps. This process is primarily meaning-focused, and it probably helped learners make sense of the passage as a whole, and learn about the meaning the specific FSs that they could use to fill in the blanks. In this meaning-making process, focus on form, i.e. the form of FSs, seemed to happen only incidentally yet effectively. This is in line with the theory of focus-on-form instruction (Long, 1991) and with previous findings in task-based vocabulary teaching (R. Ellis & He, 1999; R. Ellis et al., 1994; Gass & Torres, 2005; Kim, 2008; Kowal & Swain, 1994; Loewen & Philp, 2006), which prove that incidental focus on form was successful when the primary focus of the activity was on meaning.

Just as frequency should not be viewed merely as frequency of form but as frequency of form-meaning-usage composites (Larsen-Freeman, 2002), how FSs are acquired is possibly informed
not only by how form was attended to during classroom activities but also how engaged learners were with the meaning of the FSs and of the context immediately surrounding them, i.e. how FSs are used in such context. Form-focused instruction types, especially tasks, were largely successful in increasing saliency and promoting higher levels of engagement, both to offset the low frequency of FSs occurrence in input available to learners. Similar to how frequency should be viewed as frequency of form-meaning-use, salience is best promoted during a classroom activity when it is concerned with not just the form (such as in the SpotDif group), but also the saliency of all dimensions of FSs: form, meaning and use (consider again how Gap-fill demands that learners attend to all of these 3 dimensions) (Larsen-Freeman, 2002).

The success of the Gap-fill group attests to the claim that processing meaning and usage at higher cognitive and motivational levels were beneficial to retaining the form (evidenced by this group’s higher scores in the Receptive knowledge) and to producing the form when future contexts require it (higher scores in the Productive Knowledge).

This result also confirms Nation’s proposition: the classroom activity more conducive to acquiring new FSs’ form will be also helpful in facilitating learners’ acquisition of Productive Knowledge (2011). As Nation explains, one of the reasons why productive knowledge of lexical items is more difficult to acquire, in comparison with receptive knowledge, is because the form of lexical items in L2 usually differs substantially from that in L1. Productive learning is thus more challenging because it requires a more precise knowledge of the form. In this experimental study, the Gap-fill group was more successful in helping learners retaining the form of the taught FSs and was also the group with the highest scores in the Productive Knowledge test.
In the Meaning section of the Receptive Knowledge post-tests, results from the immediate post-test reveal that the Gap-fill and IEEI groups had the same mean scores, and they both outperformed the Control group at a statistically significant level. The SpotDif group’s mean score was not significantly better than the Control group nor significantly lower than the Gap-fill and IEEI groups. As with the Gap-fill and IEEI group, the effect size for SpotDif exceeded the large level according to Cohen’s convention. In the delayed post-test, however, no statistical significance was found between treatment groups and the Control group. In terms of effect size, only the effect size for the IEEI group as compared to the Control group remained large. These results suggest that the IEEI treatment was most effective among the three types of treatment in helping learners retain the meaning of target FSs at a receptive level, although IEEI was not the most effective type of instruction when it comes to assisting learners acquire receptively the form of the new FSs. Schmitt (2008) claims that creating an early form-meaning link of new L2 lexical items is essential in vocabulary instruction as it frees more cognitive resources. In this experimental study, even though all treatment types facilitated learners’ establishment of that linkage, the IEEI instruction did so with less delay, in comparison with the Gap-fill and the SpotDif. In the IEEI group, upon finishing the comprehension questions, learners were provided with an explicit explanation of the FSs’ meaning. In both the Gap-fill and SpotDif group, learners were not aware of the meaning of these FSs until later, as they were involved in the task. However, since the establishment of this form-meaning connection was not motivated by learners’ need to complete the task, it did not lead to a higher ability to produce the FSs or a better retention of the form of the FSs.

Another inference from the findings pertaining to learners’ Receptive Knowledge test scores is that instruction did provide learners with an advantage, but the instructional effect was less
impressive than analysis of learners’ performance in the Productive Knowledge shows, both in terms of statistical significance and effect sizes (with only one exception: effect sizes of the IEEI group’s scores on the delayed post-tests – see Table 25). It is possible that more advanced learners have developed highly efficient strategies to acquire new lexical items receptively. As Nation (2001), Schmitt (2008) and Milton (2009) concur, when it comes to acquiring new lexical items, it is less difficult for learners to succeed at receptive learning. Learners in this study have probably had abundant practice opportunities with learning new lexical items and thus have become experts in doing so. As a consequence, there is less difference between the treatment and Control groups’ scores in the Receptive Knowledge post-tests which is reflected in consistently lower effect sizes of all treatment groups when compared with the Control group. Learners in the Control group have learned new FSs more successfully at the receptive level than they do at the productive level, probably thanks to their experience in doing so and the relative ease of acquiring lexical knowledge receptively.

5.3. Do IEEI, Gap-fill and SpotDif help learners improve their ability to notice FSs in a reading passage? If yes, which type of treatment is more effective?

Evidence for instructional effect on the ability to notice FSs was relatively weak. The type of instruction that seems most effective is IEEI, despite lack of statistical significance. One interesting finding is that from immediate post-test to delayed post-test, the effect size for the IEEI group as compared to the Control group did not decrease; on the contrary, this effect size increased from small to large, according to Cohen’s convention. The Gap-fill group seems to have only a very slight advantage over the Control group, with effect size increasing from null to small from immediate to delayed post-test. The SpotDif group’s performance in both post-tests was less satisfactory than the Control group.
The advantage of the IEEI group suggests the necessity of explicitly teaching learners strategy in order to help them notice more FSs than they would normally do. In other words, explicit teaching of FS-noticing strategies might help learners sharpen their priming device (Gass & Selinker, 2008), which tells learners to which features to attend to in processing L2 input.

Learners’ selective attention, or noticing, of these features is one of the three psychological processes essential to vocabulary acquisition (Nation, 2001; Laufer, 2008). In the treatment sessions of the IEEI group, after receiving direct explanation of the meaning of the FSs, learners were also drawn to common FS-related errors encountered by language learners, they were cautioned of the low saliency of FSs, and were instructed to underline FSs worth learning in texts they might be exposed to in the future. Input enhancement in the three reading passages might have caused the learners to underline more FSs, as they were used to a text with highlighted chunks.

This finding is somewhat similar to what Fan (2003) found from an experimental study. Results from Fan’s study suggest that students are more likely to use learning strategies if teachers convince students of their usefulness. It is interesting to learn that learners underlined fewer words in the post-tests than in the pre-test – this suggest that the increase in the number of FSs in the IEEI and Gap-fill groups is the result of instruction. For other groups, the decrease in the number of FSs underlined is probably indicative more of the decreased interest in the passage – as they read the same passage from the pre-test, immediate to delayed post-tests - than of the ability to notice useful FSs, as one participant in the study revealed to me. It is unclear, however, why the SpotDif group’s performance was less satisfactory than that of the Control group. One possible reason is that even though the SpotDif was relatively useful in facilitating the acquisition of target FSs in particular, as can be seen from the results discussed in previous
sections, it was not successful in developing learners’ favorable attitudes towards this type of lexical unit in general and thus decreased the level of learners’ attention as they read the passage in the underlining test. This is also reflected in the decrease in number of words underlined by the SpotDif group in the pre-test, immediate then the delayed post-test. It is useful to bear in mind that the way SpotDif was actually implemented might have been boring to the learners: they were not engaged in meaning processing or meaningful interaction but directed their entire efforts to decide which given form resembled that in the original passage. I speculate that the lack of meaningful engagement with new FSs could have caused learners to lose their overall interest in FSs, and thus when they were given the awareness test, were not invested in underlining the FSs.

It should be noted that the IEEI group did not perform better than the Gap-fill and SpotDif group in many other measures, include the acquisition of FSs’ Productive Knowledge and of the form at a Receptive Knowledge level. This indicates that the students’ tendency to notice more FSs after FS learning strategy instruction does not necessarily translate into better retrieval and production of these target FSs.

Although I did not have a systematic analysis of the qualitative data obtained from the Awareness tests, there seems to be no significant difference in the responses participants provided for the open question at the end of the test explaining why they underlined the words/phrases. In the delayed post-tests, for example, responses by learners in all 4 groups indicated that the phrases they underlined would be helpful in their writing either because they are new phrases, or familiar ones they often failed to use properly in their writing or speaking. Learners in all groups also commented that the phrases helped the article sound more “advanced”, “sophisticated”, “professional”, and “native(like)".
5.4. Are frequency, MI-score, and n-gram length of FSs related to learners’ acquisition of receptive and productive knowledge of FSs?

Findings suggest that frequency is positively correlated with learners’ acquisition of productive knowledge of target FSs, that the correlation between n-gram length and the Productive Knowledge immediate post-test was close to significant, and that MI-score is inversely correlated with learners’ performance only on the Meaning section of the immediate Receptive post-test.

Most noticeable in the findings is the significant correlation between frequency and learners’ performance in the Productive Knowledge immediate post-test, and the relatively close to significant correlation between frequency and that measure in the delayed post-test. It is interesting to find that frequency was to a certain extent correlated with the Productive Knowledge levels acquired by the learners at either significant or close to significant level, but not on the Receptive Knowledge levels. However, it is inconclusive from this study as to what extent frequency affects the acquisition of different knowledge types of FSs. When compared with results from other studies, it can be inferred that frequency affects the processing and learning of FSs in complex ways, possibly contingent on which types of FSs (idioms, collocations, phrasal verbs) at what frequency bands are being examined.

This finding is in partial concordance with N. Ellis et al. (2008). Results obtained from the study by N. Ellis et al. suggest that the accuracy and fluency of processing formulaic expressions by advanced L2 learners of English is predominantly affected by frequency. The difference between the two studies is that, while learners in the 2008 study received no treatment and may have already acquired many FSs in the experiment (e.g. *the length of the, the content of, the extent to*
which, it can be seen that, in other words), learners in this study had very little knowledge of the target FSs and received some form of treatment before the completed the Receptive Knowledge test. N. Ellis et al.’s study suggests that learners’ receptive processing of already known FSs is heavily influenced by the FSs. Adding to this finding, my study shows that frequency of FSs positively correlates with the acquisition of their productive knowledge, even though these FSs have not been known before. With target FSs of higher frequency (e.g. the FREQ levels of wind down and at odds with are 1254 and 1700 tokens respectively), it is possible that learners had encountered them elsewhere, but did not know their meaning, probably because of the lack of motivation to search for meaning. Hence after being taught these familiar-lookng FSs, they are more likely to produce them when prompted.

As discussed previously, advanced learners probably have extensive experience with vocabulary acquisition and are experts at acquiring new lexical items receptively. Therefore, it is possible that in this study, learners were capable of acquiring the receptive knowledge of target FSs independent of their frequency. Another reason why learners in this study did not show sensitivity to FSs’ frequency in the Receptive Knowledge test is because of the nature of the test, in comparison with the tests administered by N. Ellis et al. (2008). N. Ellis et al.’s data collection instruments include reaction time, voice onset, articulation time, as well as the correctness of their responses, whereas in my study, learners’ response time to each question was not measured. The accuracy of learners’ answers in this study, independent of the FSs, is probably suggestive of the fact that learners’ receptive knowledge of instructed FSs might be more complex than the mere ability to select the correct answer from a selection of given options. This means that even though learners, especially those of higher proficiency levels, are able to complete items in the Receptive Knowledge test accurately and independently of FSs’
frequency, hypothetically their reaction time for each item on the test might be contingent on the frequency of individual FSs.

As Waring (1997) shows from a study of Japanese learners of English, for a lexical item of lower frequency bands (such as a FS, in comparison with individual words), it is much more likely learners can only know them only receptively. Learners are more likely to acquire the productive knowledge of a lexical item of higher frequency. As findings revealed, learners in the present study struggled with acquiring the productive knowledge of the taught FSs. The challenge of learning new FSs productively led them to perform less satisfactorily on the Productive Knowledge test, and suggests that they are, in fact, sensitive to the frequency of FSs, as suggested by N. Ellis et al. (2008) findings.

This finding differs from what Siyanova-Chanturia et al. (2011) found in their experiment involving eye-tracking. According to Siyanova-Chanturia et al., non-native speakers and native-speakers alike are sensitive to the frequency of 3-word binominal phrases (e.g. bride and groom) versus their reversed form (groom and bride). However, Siyanova-Chanturia et al. only examined binominal phrases, which is different from the type of FSs experimented in this study (idioms, phrasal verbs, collocations). The difference suggests how frequency (and other variables affect the acquisition of specific FSs) might depend on what type of FSs are being investigated (e.g. idioms, collocations, phrasal verbs).

My findings also revealed that the correlation between n-gram length and the Productive Knowledge immediate post-test was negative and close to significant. This was not surprising since with higher n-gram length FSs, learners have to produce more words in the c-test. However, n-gram length was only somewhat influential in the immediate post-test when
learners’ knowledge of the target FSs was newly established. 2 weeks later, when the delayed post-test was administered, \( n \)-gram length no longer negatively affected learners’ performance on the Productive Knowledge test. In general my study has shown that the length of FSs does not generally affect their acquisition.

MI-score is inversely correlated with learners’ performance only on the Meaning section of the immediate Receptive post-test. As N. Ellis et al.’s study (2008) indicates that nonnative speakers are not sensitive to the MI-score of already known FSs, it is not surprising to learn that MI-score does not affect the acquisition of new FSs in general. It is surprising, however, that MI-score has a negative influence on learners’ performance in the Meaning section of the immediate post-test, especially since they had received explanation about the meaning of new FSs immediately prior to the test. This suggest that non-native speakers of English are partially sensitive to FSs’ MI score, and that further research is needed to determine the mechanism of such sensitivity.

It is not straightforward from the findings of this study how specific FSs were acquired with different levels of success, even though there was a tendency for learners to acquire receptively meaning first, then form, and finally productive knowledge. There are possibly factors other than frequency, MI Score and \( n \)-gram length that influence the acquisition process, such as the types of FSs, semantic transparency, fixedness, and the similarity between the figurative symbols used in the expression between L1 and L2. In addition, the interaction of these factors could cause complicated patterns that are not easy to formulate.

These findings reinforce the claim frequency is important but not sufficient (Larsen-Freeman, 2011). On one hand, the relatively weak, though statistically significant, correlation between frequency and the immediate Productive Knowledge test scores, can be viewed as empirical
support for the usage-based theories. As I have expounded before, it is likely that learners have encountered these high-frequency FSs elsewhere. Aligned with usage-based explanation of language and language learning, this seems to suggest that L2 learners did retain statistical information about the occurrence of these FSs, even though they were probably not engaged in need, search and evaluation of form and meaning for different reasons. This statistical information worked to the learners’ advantage, not on the recognition level (Receptive Knowledge tests), but on the recall level (Productive Knowledge test).

On the other hand, the failure to find a single causal variable from a limited number of variables gives more credibility to the complexity theory’s perspective to second language acquisition (Larsen-Freeman, 2006, 2011, 2012; Larsen-Freeman & Cameron, 2008; N. Ellis & Larsen-Freeman, 2006). According to Larsen-Freeman and Cameron (2008), since language is best construed as a complex adaptive system, single variables should be viewed in a nested way, in their interaction with other variables over time, and different factors would have different levels of impact at different points in time. These variables are likely to interact in a non-linear fashion, sometimes contradictory, other times complementing each other (N. Ellis & Larsen-Freeman, 2006) and investigating such a multiplicity of variables as well as their complex interactions is necessary albeit challenging. Frequency, although important, cannot be the sole factor in SLA (Larsen-Freeman, 2002).
5.5. Other important findings

Control group: Did incidental learning occur?

It is noteworthy that the Control group was able to improve their scores on both the Productive Knowledge and Receptive Knowledge tests at a statistically significant level, despite the fact that their test scores were lower than those of the treatment groups.

The advanced proficiency levels of participants in this experimental study might have caused their Receptive Knowledge test scores to be at a comparable level with one of the treatment groups. Because Receptive Knowledge is easier to acquire, these advanced learners could have obtained a higher level of knowledge than Productive Knowledge, despite the lack of form-focused instruction on the target FSs. Observation in the classroom and examination of handouts during these reading lessons showed that some learners intentionally attended to the target FSs. They underlined certain FSs in the passages as they read. Some learners were proficient enough to underline almost all target FSs in a passage. In the picture below, a learner in the Control group successfully underlined almost all target FSs in this section of the passage in the third lesson, muddle through, in droves, at odds with, over a barrel, bode well, jump the gun. She failed to recognize a pretty penny, firmly entrenched and jump the gun. This was not uncommon among Control group members: in the second and third sessions, 8 out of 10 learners exhibited this behavior they received, though there was interindividual variability in the number of target FSs successfully identified (See Appendix H).
Online dating

Online dating took off in 1995, with Match.com celebrating its 150th wedding two years later. By 2002, this style of dating had become firmly entrenched in the cultural mainstream. As the dating pool becomes increasingly difficult to muddle through, tech-rich, time-poor young people see no stigma in online dating and are turning to it in droves. Online dating removes barriers to meeting by offering access to potential partners whom people would be unlikely to meet through other avenues, and this access yields new romantic possibilities without costing a pretty penny.

What if, however, online dating makes it too easy to meet someone new? Gian Gonzaga, eHarmony’s relationship psychologist, acknowledges that commitment is often at odds with technology. Online dating allows people to begin relationships, learn things, and ultimately make a better selection. However, online dating could have people over a barrel, causing them to leave relationships the moment they are not working—an overall weakening of commitment. Having an array of options may diminish the attractiveness of what people actually choose because they cannot stop thinking about the attractions of the unchosen options.

Though there is no statistical evidence that the break-up rate among online daters is any different from the national average, some divorce lawyers point to anecdotal evidence. Eric Spevak, a New Jersey divorce lawyer, says that as many as one in five of his clients now comes from marriages that started on the Internet. "It's usually a relationship based on fantasy or desperation, which doesn't bode well," says Spevak. Moreover, experts say online daters tend to be more interested in marriage and therefore more likely to jump the gun.

The learners’ ability to underline target FSs in the reading texts could probably be attributed to the fact that these adult learners might have guessed what post-tests that they would be given, based on the pre-tests. As seen in Appendix H, in the first session, 9 out of 10 learners did not identify any target FS. At the end of the first session, they completed the first immediate post-tests and this had an immediate effect. In the second session, 6 learners were able to identify from 2 to 9 target FSs. In the third session, 7 learners were able to underline from 3 to 9 FSs.

Overall there were only 2 learners who did not exhibit any noticing on paper in all 3 treatment sessions. In addition, many were able to memorize the FSs in the pre-test, despite the presence of individual words as distractors. As they were reading the passages in class, they recognized the form of some target FSs in the reading texts. They probably conjectured that these FSs
would be tested after the reading lessons. Furthermore, after the first immediate post-tests were given, these advanced learners must have been able to predict the format of the second and third immediate post-tests.

Therefore, it should probably be cautioned that learners of lower proficiency levels would not reach such a high level of Receptive Knowledge of FSs encountered in the language input. Learners in natural classroom settings, without the pre-tests and post-tests to alert them to FSs, would be less likely to focus on FSs. In short, it might be too hasty to draw a conclusion that learners will be able to pick up FSs in L2 input even when there is no form-focused instruction and no assessment that focuses on FSs. In other words, incidental learning apparently present in this study for the Control group might not be “incidental” at all – perhaps many learners deliberately attended to FSs to do well on the tests.

Some might argue that the treatment groups were also subject to the effects of pre-tests and post-tests, which is one of the major concerns in experimental classroom research. Such effects obviously exist; however, the treatment groups received focused instruction on target FSs anyway, and thus in reality would attend to these otherwise non-salient lexical items regardless of the anticipation of post-tests. It would also be reasonable to expect that a group of learners who receive no FS-focused instruction and no tests would highly likely pay little attention to FSs.

**Acquisition of FSs is incremental and complex**

As discussed in Chapter 3, the acquisition of FSs, similar to that of individual words, is likely to be incremental: these sequences are learned over time among L2 learners (Schmitt & Carter, 2004). Groups’ mean scores suggest that advanced learners have least difficulty acquiring the meaning of new FSs. Even in the delayed post-test, learners’ mean scores range from 83.3% (Control
group) to 93.3% (IEEI group) of the maximum score. Recognizing the form of newly learnt FSs is likely to cause learners more challenge than recognizing their meaning. In the delayed post-test, groups’ mean scores range from 53.3% (Control group) to 73.3% (Gap-fill) of the maximum score. Acquiring the productive knowledge of new FSs proves to be most challenging part, with groups’ mean scores in the delayed post-tests ranging only from 27.8% (Control group) to 51.1% (Gap-fill) of the maximum score.

When looking at the scores learners in all groups obtained in each type of test (Productive Knowledge and Receptive Knowledge) for specific FSs, as well as two different sections of the Receptive Knowledge test (Form and Meaning) (Appendix G), it can be seen that the Meaning test scores tend to be the highest, reaching near the maximum scores, even in the delayed post-test. However, since these scores reflect learners’ ability to chose the correct meaning from a number of options, it remains to be seen whether learners can recall the meaning of FSs without the help of readily available options. In comparison, the scores in the Form section tends to lag behind. For example, for the FS *mull over*, 34 out of 40 participants choose the correct meaning of this FS in the delayed post-test, but only 17 out of 40 learners were able to choose the correct form. Learners’ ability to produce the FSs with the cues given in the context was very far from ideal. Learners’ productive knowledge scores for FSs in the delayed post-test usually range between 20 and 90 (out of 120), with some exceptions in the lower end. For instance, the mean score for *mull over* was only 27 (out of 120), less than one-forth of the maximum score. For *touch off* and *over a barrel*, learners’ mean scores were only 2 and 8 (out of 120) respectively. It can be inferred that for the same FS, learners might already know the meaning, but have some difficulty with the acquiring the form, and even more difficulty producing them when cued.
The acquisition of FSs among these learners is also characterized by complexity, with no single factor (Frequency, n-gram length, MI Score) capable of single-handedly explaining the entire process. The same factor can influence the acquisition on one particular aspect of the knowledge of target FSs but not another, e.g. frequency and n-gram length impacted learners’ performance on productive knowledge but not receptive knowledge; MI-score influenced learners’ receptive learning of FSs’ meaning but not form. In terms of the durability of these factors’ effects, it seems that overtime the magnitude of effects changed. For example, frequency and n-gram ceased to correlate with learners’ Productive Knowledge on the delayed post-tests, although significant or close to significant correlations were observed in the results of the immediate post-test. There was also a possibility of the interaction of these variables and other unknown factors not examined in this study due to its scope, and the interaction between FS-internal factors and instruction types was likely to influence the acquisition of these FSs in complex ways.

All in all, discussions here have added more understanding about complexity theory (Larsen-Freeman & Cameron, 2008) and have questioned the feasibility and credibility of determining causality through examining pre-determined etic variables. These results also suggest the necessity of investigating existing data at individual learner levels in addition to examining the average data (i.e. mean scores). Unlike previous research findings (e.g. N. Ellis et al., 2008, Siyanova-Chanturia et al., 2011), the findings obtained from my research project seem to reveal indications of a surprisingly complex acquisition process of FSs.
5.6. Limitations

This cross-sectional study was limited by its small sample size (40 participants). Results could have been more robust and conclusive if each group had more than 10 participants. In addition, the lack of longitudinal data typical of cross-sectional research makes it challenging to determine to what extent instructional effects are durable, which is a central concern of instructed second language acquisition (R. Ellis, 2008). In this study, instructional effects experienced a downward trend from the immediate post-tests to the delayed post-tests, which were administered 2 weeks after the end of the treatment period. It remains to be seen whether the positive impact different types of form-focused instruction has on learners’ acquisition of FSs would endure over a longer period of time, and how repeated exposures to the newly learnt FSs or lack thereof would affect learners’ retention of knowledge. It might also be interesting to see what means would be effective in reinforcing already learnt lexical knowledge (Carter, 1998).

It should also be cautioned that the type of Productive Knowledge measurement used in this study only requires the research participants to produce the target FSs in a controlled context, with initial letters and sometimes middle letters provided. It is unclear how participants would perform if they are to use these FSs in a freer context, without any clues provided, and whether they would be able to use the FSs in a stylistically appropriate ways.

Finally, because of the difficulty in recruiting research participants, it was impossible in this experimental study to investigate the interactions between different independent variables. It would have been interesting to examine how instructional intervention and FSs’ characteristics,
namely frequency, n-gram length and MI-score interact with each other in shaping the way learners acquire certain FSs.

5.7. Implications

The most important implication for an ESL classroom that can be drawn from this study is that form-focused instruction is conducive to the learning of FSs, especially the acquisition of receptive and productive knowledge of new FSs. Depending on what goals the teachers envision for the learners, they can choose which type(s) of activity to implement in the classroom.

Overall communicative tasks such as Gap-fill and SpotDif seem to benefit learners’ acquisition of productive knowledge, while the more direct, non-interactive instruction type IEEI was useful in helping learners retain the receptive knowledge of new FSs. More specifically, activities which require learners to process the meaning of not only the FSs but also the linguistic context of such FSs, such as in Gap-fill, might be the most effective in facilitating learners’ productive retention of the FSs’ form. Activities that are potentially useful in helping learners to acquire receptively the form of new FSs are probably also conducive to the acquisition of FSs and productive knowledge. Teachers should also be aware that interactive tasks such as SpotDif, while theoretically promoting a strong focus on form, can cause learners to neglect processing the meaning of the target FSs and their linguistic context. Such neglect will likely result in reduced overall effectiveness of the activity. If the goal is for learners to acquire learners’ receptive knowledge of new FSs in terms of meaning only, establishing an early form-meaning link is probably the most important pedagogical step. It is likely that direct, focused teaching that isolates new FSs from the context and explanation of their meaning would be most
effective in facilitating learners’ receptive learning of meaning, but it will not transfer to a higher level of receptive learning of form nor of productive learning.

Teachers can also use the Involvement Load Hypothesis to evaluate the effectiveness of pedagogical activities pre- and post-implementation. The three core processes that teachers need to consider are need, search and evaluation. Teachers can further distinguish between processes involved in learning the meaning and learning the form of FSs to be more accurate in determining whether a specific activity will be useful.

In addition to using activities that promote learners acquiring high levels of productive and receptive knowledge, teachers can complement them with more overt strategy instruction. Explicitly teaching noticing strategies might be the most effective way of increasing their tendency to single out FSs in new materials. This will help learners develop strategies to learn FSs beyond the classroom on their own, increasing learners’ agency and autonomy in the FS acquisition process.

Teachers can use the Involvement Load Hypothesis to evaluate the possible involvement load that these activities will induce in the classroom. However, anticipation of classroom activity effectiveness can be inaccurate, due to the way that learners carry out the activity in an ecological context, similar to how the actual implementation of SpotDif deviates from prior expectation. A teacher who is interested in bringing more FSs to the forefront of learners’ language acquisition process will need to observe carefully how the planned activities proceed in the classroom to make necessary adjustments.

A teacher might also use a combination of activities to develop learners’ comprehensive knowledge of new FSs. Combining, for example, explicit teaching of FS learning strategies at the
end of a communicative task could be a way of ensuring the learners reap the benefits of different instructional techniques and acquire different knowledge aspects of FSs.

In terms of selecting FSs to focus on in the classroom and anticipating the possible difficulty they might cause to the learners, teachers should probably be aware of FSs with low frequency, high MI score and high n-gram length, even though there is only some limited evidence from this study that these factors will impede learners’ acquisition of productive and receptive knowledge. Frequency and n-gram length most probably influence learners’ productive knowledge of newly learnt FSs in a negative way, and MI-score might influence learners’ retention of FSs’ meaning. Learners might need more repeated exposure of FSs with low frequency, n-gram in order to produce them, or they need to learn them in a more production-oriented to be able to retrieve them later when needs arise (Milton, 2009).

Another issue for teachers to keep in mind when is the optimal time to start teaching FSs. Carter (1998) wisely opines “...consideration of lexis in discourse raises a central question of when in second-language teaching the learning of fixed expressions is best encouraged, and at what point some clearly holistic tendencies in language are best developed (p. 224). Advanced learners in this study seem to benefit from instruction, but it is uncertain whether or to what extent form-focused instruction would benefit learners of lower proficiency levels, and how much instruction effectiveness there would be when the target FSs are those of high frequency bands (i.e. more popular) and lower MI-score for those learners.

Since the acquisition of FSs is likely to be incremental as previously discussed, it is important for teachers to explore ways of facilitating the entire acquisition process, beyond providing one-time form-focused instruction. Teachers might want to create opportunities for repeated
exposures to the target FSs, identify possible obstacles learners face and help them overcome those challenges. Given the limited classroom contact hours and numerous goals teachers aim to achieve, it might be also crucial for teachers to raise learners’ general awareness about the prevalence and usefulness of FSs, teach them basic strategies in order for learners to notice and learn FSs beyond the classroom.

There is a potential for teachers to use corpus data in making pedagogical decisions as to what FSs to focus on, and to predict the level of difficulty certain FSs could cause their learners. They should probably be beware that FSs of lower frequency bands might pose more challenge to learners when it comes to productive acquisition, not those with greater n-gram length or MI-score, if the goal is to teach learners to use them actively. Teachers, however, should be cautious about other potential factors that might influence the difficulty level of acquiring a specific FS, such as the type of FS, semantic transparency, fixedness, and the similarity between L1 and L2’s figurative meaning.

5.8. Concluding remarks

This experimental study found that three types of form-focused instruction, IEEI, Gap-fill and SpotDif, benefit learners in acquiring higher levels of productive and receptive knowledge of new FSs. This finding is in line with previous literature in the topic of teaching FSs to L2 learners (Bishop, 2004; Boers et al., 2006; Boers & Lindstromberg, 2005; Boers et al., 2004; Hsu, 2010; Jones & Haywood, 2004; Lindstromberg & Boers, 2008a; Rott, 2009; Wood, 2009). Larger magnitude of instructional effects between treatment groups and Control group were observed in the Productive Knowledge test scores, in comparison with Receptive Knowledge, proving that instruction was particularly successful in helping learners produce the newly learnt FSs in a
specific context. Results also revealed that the type of instruction effective in helping learners retain the form of FSs will also help learners acquire higher levels of productive knowledge. Furthermore, learners’ engagement in understanding the meaning of new FSs in their context contributed to their retention of the form and establishment of productive knowledge in the long run. Direct instruction of new FSs’ meaning seems to help learners retain the meaning most efficiently, while explicit strategy teaching tends to enhance learners’ ability to notice FSs in L2 input. Findings from this study also suggests a complex interaction of contributing factors to the acquisition of FSs, as frequency, n-gram length and MI Score separately cannot fully account for the levels of success in acquiring new FSs both receptively and productively among learners, even though frequency seems to positively correlate with learners’ acquisition of productive knowledge in the initial stage of learning. Table 30 summarizes all findings pertaining to the first three research questions, with a primary focus on reporting meaningful differences as measured by effect sizes, and the magnitude of effect sizes are determined as large, moderate or small based on Cohen’s convention. In this table, statistical differences are indicated by asterisks (*).

More research is needed to examine the effectiveness, both temporary and durable, of form-focused instruction on learners’ acquisition of FSs. Longitudinal data, data from a larger number of research participants, data from learners who are of different proficiency levels, and a longer instructional intervention duration are needed to probe the complexities of the acquisition process. Different types of intervention besides IEEI, collaborative Gap-fill and SpotDiff should also be put to the test.
The study proves that L2 pedagogy can provide affordances to foster SLA processes through simulating conditions for learners to engage in form-meaning-use mappings. These conditions are necessary to compensate for the lack of extensive L2 input, with the underlying assumption that frequency plays a crucial role in language acquisition - a central argument of usage-based approach to language acquisition theorists. Evidence points to a higher effectiveness level of

### Table 30

Summary of all results

<table>
<thead>
<tr>
<th>Test/Effect size level</th>
<th>IEEI group As compared to Control group</th>
<th>Gap-fill As compared to Control group</th>
<th>SpotDif As compared to Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productive Knowledge Immediate Post-test</td>
<td>Large (2.49)*</td>
<td>Large (3.38)*</td>
<td>Large (2.64)*</td>
</tr>
<tr>
<td>Productive Knowledge Delayed Post-test</td>
<td>Moderate (0.66)</td>
<td>Large (1.53)*</td>
<td>Large (0.87)</td>
</tr>
<tr>
<td>Receptive Knowledge – Immediate Post-test</td>
<td>Large (1.82) *</td>
<td>Large (1.71)*</td>
<td>Large (1.68)*</td>
</tr>
<tr>
<td>Receptive Knowledge – Delayed Post-test</td>
<td>Large (0.99)</td>
<td>Large (0.81)</td>
<td>Large (1.06)</td>
</tr>
<tr>
<td>Receptive Knowledge – Immediate Post-test – Form section</td>
<td>Large (1.86) *</td>
<td>Large (1.74)*</td>
<td>Large (1.83)*</td>
</tr>
<tr>
<td>Receptive Knowledge – Delayed Post-test – Form section</td>
<td>Large (0.88)</td>
<td>Large (1.14)*</td>
<td>Large (0.91)</td>
</tr>
<tr>
<td>Receptive Knowledge – Immediate Post-test – Meaning section</td>
<td>Large (1.45)*</td>
<td>Large (1.39)*</td>
<td>Large (1.17)</td>
</tr>
<tr>
<td>Receptive Knowledge – Delayed Post-test – Meaning section</td>
<td>Large (0.94)</td>
<td>Small (0.28)</td>
<td>Small (0.36)</td>
</tr>
<tr>
<td>Awareness immediate post-test</td>
<td>Small (0.32)</td>
<td>Null (0)</td>
<td>Negative (-0.92)</td>
</tr>
<tr>
<td>Awareness delayed post-test</td>
<td>Large (0.84)</td>
<td>Small (0.25)</td>
<td>Negative (-0.50)</td>
</tr>
</tbody>
</table>
tasks that engage learners in not only form but also two other important yet commonly neglected dimensions of language: meaning and use. Different levels of correlations observed between frequency, MI score and n-gram length over time lent support to a complexity theory perspective to language and language learning.
APPENDICES

Appendix A: List of target FSs and Reading Passages

<table>
<thead>
<tr>
<th></th>
<th>Search query: Word</th>
<th>Search query: Collocate</th>
<th>Raw FREQ</th>
<th>MI score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>at the mercy of</td>
<td>at the mercy</td>
<td>674</td>
<td>3.68</td>
</tr>
<tr>
<td>2</td>
<td>go haywire</td>
<td>haywire</td>
<td>226</td>
<td>8.04</td>
</tr>
<tr>
<td>3</td>
<td>at the beck and call of</td>
<td>beck and call at (6,0)</td>
<td>113</td>
<td>5.05</td>
</tr>
<tr>
<td>4</td>
<td>saddled with</td>
<td>with saddled (1,0)</td>
<td>528</td>
<td>3.84</td>
</tr>
<tr>
<td>5</td>
<td>bank on</td>
<td>[bank].[v*] on (0,1)</td>
<td>514</td>
<td>5.17</td>
</tr>
<tr>
<td>6</td>
<td>wind down</td>
<td>[wind].[v*] down (0, 1)</td>
<td>1254</td>
<td>6.57</td>
</tr>
<tr>
<td>7</td>
<td>short end of the stick</td>
<td>short end of the stick (0,4)</td>
<td>69</td>
<td>10.64</td>
</tr>
<tr>
<td>8</td>
<td>beef up</td>
<td>[beef].[v*] up (0,3)</td>
<td>540</td>
<td>7.14</td>
</tr>
<tr>
<td>9</td>
<td>gloom and doom</td>
<td>gloom and doom (0, 2)</td>
<td>85</td>
<td>15.02</td>
</tr>
<tr>
<td>10</td>
<td>have a lock on</td>
<td>a lock on [have].[v*]</td>
<td>127</td>
<td>3.66</td>
</tr>
<tr>
<td>11</td>
<td>all the rage</td>
<td>all the rage (0,1)</td>
<td>384</td>
<td>5.54</td>
</tr>
<tr>
<td>12</td>
<td>short shrift</td>
<td>short shrift (0,1)</td>
<td>251</td>
<td>12.55</td>
</tr>
<tr>
<td>13</td>
<td>dig in someone's heels</td>
<td>[dig] in heels (0, 2)</td>
<td>166</td>
<td>9.95</td>
</tr>
<tr>
<td>14</td>
<td>goof off</td>
<td>off [goof].[v*]</td>
<td>152</td>
<td>8.19</td>
</tr>
<tr>
<td>15</td>
<td>touch off</td>
<td>off [touch].[v*]</td>
<td>568</td>
<td>3.68</td>
</tr>
<tr>
<td>16</td>
<td>push the envelope</td>
<td>[push] the envelope (0,1)</td>
<td>461</td>
<td>7.75</td>
</tr>
<tr>
<td></td>
<td>Expression</td>
<td>Word 1</td>
<td>Word 2</td>
<td>Value</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------</td>
<td>--------</td>
<td>----------------</td>
<td>-------</td>
</tr>
<tr>
<td>17</td>
<td><strong>old hat</strong></td>
<td>old</td>
<td>hat (0,1)</td>
<td>178</td>
</tr>
<tr>
<td>18</td>
<td><strong>step up to the plate</strong></td>
<td>up to the plate</td>
<td>step (5,0)</td>
<td>383</td>
</tr>
<tr>
<td>19</td>
<td><strong>dawn on</strong></td>
<td>[dawn].[v*]</td>
<td>on (0,1)</td>
<td>1074</td>
</tr>
<tr>
<td>20</td>
<td><strong>drop the ball</strong></td>
<td>[drop] the</td>
<td>ball (0,1)</td>
<td>310</td>
</tr>
<tr>
<td>21</td>
<td><strong>firmly entrenched</strong></td>
<td>firmly</td>
<td>entrenched (0,1)</td>
<td>139</td>
</tr>
<tr>
<td>22</td>
<td><strong>muddle through</strong></td>
<td>through</td>
<td>[muddle].[v*]</td>
<td>179</td>
</tr>
<tr>
<td>23</td>
<td><strong>in droves</strong></td>
<td>in</td>
<td>droves (0,1)</td>
<td>592</td>
</tr>
<tr>
<td>24</td>
<td><strong>pretty penny</strong></td>
<td>pretty</td>
<td>penny (0,1)</td>
<td>83</td>
</tr>
<tr>
<td>25</td>
<td><strong>at odds with</strong></td>
<td>at odds</td>
<td>with (0,2)</td>
<td>1700</td>
</tr>
<tr>
<td>26</td>
<td><strong>over a barrel</strong></td>
<td>over a</td>
<td>barrel (0,2)</td>
<td>81</td>
</tr>
<tr>
<td>27</td>
<td><strong>bode well</strong></td>
<td>[bode]</td>
<td>well (0,1)</td>
<td>556</td>
</tr>
<tr>
<td>28</td>
<td><strong>jump the gun</strong></td>
<td>[jump] the</td>
<td>gun (0,1)</td>
<td>150</td>
</tr>
<tr>
<td>29</td>
<td><strong>mull over</strong></td>
<td>over</td>
<td>[mull].[v*]</td>
<td>415</td>
</tr>
<tr>
<td>30</td>
<td><strong>on the books</strong></td>
<td>on the</td>
<td>books (0,1)</td>
<td>943</td>
</tr>
</tbody>
</table>
Passage 1: Close to home

(417 words)

Some people can never get away from work these days. They are \textit{at the constant mercy of} their smart phones, tablets or other electronic devices. They feel like if they are not connected to the Internet or to one of their digital devices, everything will \textit{go haywire}. For a businesswoman, the business line still rings at home, which can be good or bad. However, it is no longer just business owners or professionals such as doctors who are \textit{at their job's beck and call}, even the average employee finds it difficult to get away from work sometimes. Owning smart phones, tablets and other devices can be a mixed blessing. These digital devices can provide access and efficiency, yet at the same time, inconvenience and interruption of family life and personal time.

This phenomenon has been explored extensively by Maggie Jackson, a columnist and author of \textit{What's Happening at Home: Balancing Work, Life and Refuge in the Information Age}. She said that technology is redefining home life and blurring the line between work and home.

While she views technological development as a gift, Jackson warns that it may require sacrifice as well. For people armed with and \textit{saddled with} digital devices, mobile technology is turning homes into workplaces. Jackson reports that 70 percent of all employers \textit{bank on} mobile technology to ensure all employees are on call 24 hours a day, seven days a week. While the majority of employers don't mind when workers start the workdays later, they in turn expect flexibility from their employees to work outside of normal business hours, even as they \textit{wind down} for the night. This fluid approach to working hours means that many employers are now
comfortable calling employees after-hours, with 80 percent saying they think it is acceptable to
call staff in the evening.

Fearing employees are getting the short end of the stick in the digital age, Jackson argues that
employers should respect employees’ personal life and beef up measures to protect that privacy.
She commends companies like Ernst & Young for having stringent regulations to limit the on-call
time of employees.

Jackson said that when the Industrial Age gave way to the computer age, there was a sense of
gloom and doom as many people worried that the new devices would have a lock on their life.
One wise observer summed their fears up when he warned: “Be sure that we control these
gadgets and not the other way around.” That is not bad advice for today, either.
Passage 2: Social Media and Schools

(406 words)

While social media platforms are all the rage nowadays, in school they are getting short shrift. Because social media are considered a distraction to the teaching and learning process, many educators dig in their heels and ban them from the classroom. These educators argue that if students are allowed to use social media in school, they will goof off or exhibit inappropriate behavior.

According to a new study at Columbia University, teens that use social media are more likely to abuse drugs and alcohol. Studies like this touched off controversy about whether social media should be completely banned from schools.

However, others believe school must push the envelope and truly experience what these free resources can do for their students if they do not want to be old hat. When structured in a pedagogically sound fashion, learning activities incorporating social media, such as Twitter, YouTube, Facebook, allow students to apply what they have learned through creation. This fosters higher order thinking skills and caters to a wide range of learning styles.

Several school administrators stepped up to the plate. Schools can use social media as a powerful public relations tool in lieu of traditional newsletters and e-mail blasts. They can also use social media to communicate information such as reports, schedule changes, and to listen to
the public’s opinions. Joseph Donzelli, the communications director for the Fort Myers, Florida regional school district – one of the largest school districts in the nation – argues that Twitter has enhanced their communication efforts. The reality today is that public schools should be using social media to communicate with their audience and to talk about their strengths. Today’s public demands it.

It is also starting to dawn on some educators that incorporating social media into education provides a golden opportunity to teach digital citizenship to learners, an area where many schools are dropping the ball. Digital citizenship is the norms of appropriate and responsible technology use. One job of educators in the 21st century is to teach students to practice safe, legal, and responsible use of information and technology, and to exhibit a positive attitude toward using technology that supports collaboration, learning, and productivity. When doing so, schools must ensure that policies are in line with this change, teachers are supported through professional development on how to effectively use social media in the classroom, and parents are educated on its value.
Passage 3: Online dating

(392 words)

Online dating took off in 1995, with Match.com celebrating its 150th wedding two years later. By 2002, this style of dating had become firmly entrenched in the cultural mainstream. As the dating pool becomes increasingly difficult to muddle through, tech-rich, time-poor young people see no stigma in online dating and are turning to it in droves. Online dating removes barriers to meeting by offering access to potential partners whom people would be unlikely to meet through other avenues, and this access yields new romantic possibilities without costing a pretty penny.

What if, however, online dating makes it too easy to meet someone new? Gian Gonzaga, eHarmony's relationship psychologist, acknowledges that commitment is often at odds with technology. Online dating allows people to begin relationships, learn things, and ultimately make a better selection. However, online dating also have people over a barrel, causing them to leave relationships the moment they are not working—an overall weakening of commitment. Having an array of options may diminish the attractiveness of what people actually choose because they cannot stop thinking about the attractions of the unchosen options.

Though there is no statistical evidence that the break-up rate among online daters is any different from the national average, some divorce lawyers point to anecdotal evidence. Eric Spevak, a New Jersey divorce lawyer, says that as many as one in five of his clients now comes from marriages that started on the Internet. "It's usually a relationship based on fantasy or
desperation, which doesn’t **bode well,**” says Spevak. Moreover, experts say online daters tend to be more interested in marriage and therefore more likely to **jump the gun.**

Sex offenders often use matchmaking sites to find victims. Jeffrey Marsalisz, a Pennsylvania man, is a smooth talker who would meet women on the popular dating website Match.com, telling them he was an astronaut, doctor or a spy and then slip something into their drinks to incapacitate them. Lawmakers in several states are **mulling over** legislation to help make online daters more aware of the potential pitfalls of the process. Connecticut’s bill, mirroring a law in New York, requires Internet dating services to provide a safety awareness notice during registration that offers advice such as never including one’s last name, email address, place of work, phone numbers or identifying information in an Internet profile. Similar laws are already **on the books** in Florida and New Jersey. Other states are expected to follow suit.
## Appendix B: Pre-test 1 – Vocabulary Knowledge Scale

Please indicate the level of your knowledge of these words or phrases by putting a check (✓) in the appropriate cell.

<table>
<thead>
<tr>
<th></th>
<th>(1) I don’t know what this word/phrase means</th>
<th>(2) I know the meaning of this word/phrase, but I never use it in my writing/speaking</th>
<th>(3) I use this word/phrase quite often in my writing/speaking</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Example: omnipresent</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>all the rage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>epoch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>benevolent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>vivacious</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>quirky</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>stringent regulations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>wind down</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>all ears</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>at the mercy of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>follow suit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>old hat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>bluster</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>fashionista</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>gloom and doom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>dig in someone’s heels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>mall over</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>over a barrel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>bank on</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>euphonious</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>eager beaver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>equivocate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>drop the ball</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>bode well</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>dawn on</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>exotic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>firmly entrenched</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>under the weather</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>in droves</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>hissy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>saddled with</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) I don’t know what this word/phrase means</td>
<td>(2) I know the meaning of this word/phrase, but I never use it in my writing/speaking</td>
<td>(3) I use this word/phrase quite often in my writing/speaking</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>31</td>
<td>jump the gun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>hit the sack</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Mazel Tov</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>eclectic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>pretty penny</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>on the books</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>go haywire</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>touch off</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>at the beck and call of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>poetic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>muddle through</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>karma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>squalid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>piece of cake</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>sentimental</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>at odds with</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>beef up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>short end of the stick</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>major catastrophe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>goof off</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>holistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>serendipity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>intelligentsia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>have a lock on</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>overhaul</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>rebuttal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>short shrift</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>epiphany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>push the envelope</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>simmering</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C: Pre-test 2 Underlining useful words and phrases in passage

Passage given to learners does not contain typological enhancement. The FSs in bold are some that might be underlined.

INSTRUCTION

Underline words or phrases that you think are useful for improving your language skills, including speaking and writing.

After finishing, please explain why you choose to underline those words or phrases.

Problems of the Information Age

The Information Age is believed to have begun roughly in the 1990’s when widespread use of computers and information sharing technologies like the Internet gave rise to anytime anywhere accessibility to information. Some characteristics of the information age include: use of information to achieve high productivity, global use of information, emphasis on innovation rather than economies of scale and high social, economic and political awareness. As opposed to the industrial age, when laborious tasks were believed to yield returns, the information age relies on the production and sharing of knowledge and information to achieve profits. As the world progresses through the information age, there are some global challenges that might be faced.

1. Information security:

Everyone, from individuals and business firms to governments, security agencies and defense forces, is under the threat of information theft and misuse. It ceases to be a mere technological issue when it transcends national boundaries and enters into the realms of terrorism and other international criminal activities.

2. Digital and Internet laws

Legal systems around the world are at varying degrees of readiness when it comes to issues concerned with the digital world. This includes copyright issues, intellectual property rights and
so on. The Internet makes such issues international in nature, so a global uniform digital law will become a necessity in future.

3. Social challenges

As businesses go global there will be a pressing need to bridge the digital divide among nations and also within national borders. Weaving in information technology into the educational systems across the world poses a formidable challenge, especially so when illiteracy still plagues populous countries like India and China. A different but relevant social challenge in the Information age will be the right deployment of human capital. As the need for knowledge workers grows, a vast part of the global workforce may still be untrained in using information systems. A critical task, here, will be to train and involve them in the information revolution.

4. Technological challenges

Another daunting task will be to find and deploy cost effective, fast, accurate and smart technologies to store, secure, verify and share information. Technologists and entrepreneurs will have a tough time satisfying the on the go information needs of Generation Y. Tiny storage devices with huge capacities and fast retrieval techniques, are what this generation of the Information Age would like to see in their pockets!

Why do you think these words and/or phrases are useful?
Appendix D: Post-tests 1 - Productive knowledge tests

Instruction: Fill in the blanks in the following sentences, using the letter(s) provided at the beginning of the key words. Other words are either prepositions (e.g. in, on, etc.) or articles (a, an, the).

In each sentence, all the words filled in will constitute a complete phrase that we have learned.

The meaning of the phrase is provided in italics at the end of each sentence.

For example

I was under the weather last night and couldn’t work on my assignments until this morning. (sick)

Passage 1

1. Area 25 – deep in the very center of our brains – is connected to other areas that control sleep, appetite and drive, all the things that g_______ h_______ when someone’s depressed. Mayberg’s theory is if you cool off area 25, you treat the disease. (become out of control)

2. Honda and Toyota used to h_______ l_______ the market for small family cars. (have control of)

3. In the past, farmers lived m_______ c_______ the weather. (completely dependent on)

4. In the evenings, he likes to w__________ with a good book. (relax)
5. It was a fine hotel. There were dozens of maids and servants ready to serve her. They were ______ her ______ ______ ______ at all times. (ready to serve, ready to comply with any wish or command)

6. Many students in the United States are ______ ______ ______ ______ student loan debt. In a little more than 10 years, the total amount of student loan debt in this country has doubled to more than $1 trillion. (burdened with)

7. Most stocks have been down all week, and unfortunately analysts forecast ______ ______ a__________ d___________. (a feeling of pessimism)

8. Some sellers on eBay complain that the multinational internet corporation is giving them the sh________ e________ ________ ________ st________ as it only protects the rights of the buyers. (disadvantages, the smaller or less desirable part)

9. We need to ______ ______ security in airports. (improve)

10. You can ba________ _______ Molly’s caterer to do a good job. (have confidence in, rely on)
Instruction: Fill in the blanks in the following sentences, using the letter(s) provided at the beginning the key words. Other words are either prepositions (e.g. in, on, etc.) or articles (a, an, the). In Question No. 1 only, the initial letter of the preposition in the phrase is provided.

In each sentence, all the words filled in will constitute a complete phrase that we have learned. The meaning of the phrase is provided in italics at the end of each sentence.

For example

I was under the weather last night and couldn’t work on my assignments until this morning (sick)

Passage 2

1. City councils and city governments are going to have to st________ u________ 
   __________ __________ pl______________ and make some concessions to keep 
   economic development on track (take responsibility)

2. I hear psychics are a____ ________ r_______ now. All the stars have their favorite 
   fortune-tellers (trendy, popular)

3. Ms. Solanas, an unemployed online journalist, was part of the core group of protesters who 
   in May occupied the Puerta del Sol, a public square in Madrid, the capital, 
   to______________ __________ a nationwide protest (triggering)

4. On the way home, it suddenly d______________ __________ him that he had never 
   returned his boss’s call, so when he got home he called his boss immediately. (became clear 
   to him)
5. Sociology helps solve many pressing problems facing the society; yet it is always in the list of sciences that get sh___________ sh_________. (little attention)

6. The engineers wanted to completely redesign the product, but couldn't p___________ _______ en___________ because of a very restricted budget. (to innovate, or go beyond commonly accepted boundaries.)

7. Though o___________ h___________ in many European and Asian countries, commuter bikes, which run generally between $400 and $900 at independent bike stores, are foreign to many Americans. (old-fashioned, familiar)

8. Though you shouldn't be disrupting your learning by go___________ __________ during class, you can make the most of your school experience by having fun with your friends when you can: at the cafeteria, at your locker, while you are walking to your next class. (avoiding work)

9. When feeling out of control, people usually di___________ _____ __________ he___________ and resist change. (to refuse to alter one's course of action or opinions; to be obstinate or determined.)

10. You can't trust John to do the job right. He's always dr___________ _______ ba_________. (fail)
**Instruction:** Fill in the blanks in the following sentences, using the letter(s) provided at the beginning the key words. Other words are either prepositions (e.g. *in, on,* etc.) or articles (*a, an, the*). In Question No. 3 only, the initial letter of the preposition in the phrase is provided.

In each sentence, all the words filled in will constitute a complete phrase that we have learned.

The meaning of the phrase is provided in *italics* at the end of each sentence.

**For example**

I was under the **weather** last night and couldn’t work on my assignments until this morning *(sick)*

---

**Passage 3**

1. A research study found that the reaction on Twitter to major political events and policy decisions is often _______ **od**_________ _____________ public opinion as measured by surveys. The Twitter conversation is usually more liberal, but can be more conservative. *(disagree with someone)*

2. After the war broke out, people left the country _______ **dr**____________. *(in large numbers)*

3. Because Boeing cannot afford a strike in September, the union is gearing up for negotiations next August, and it believes it has Boeing o___________ ________ b___________. *(out of control)*

4. Election results revealed that white racism remains f___________ **entr** _____________ in the country. *(deeply-rooted)*
5. Installation instructions were far from straightforward, but he eventually managed to **m________d________**. (*get through something despite difficulty*)

6. Max has a new car. He must be earning a **p_____________ p_____________**. (*a lot of money*)

7. Since he took office, the governor has put nearly 400 laws **bo___________**. (*part of the law*)

8. That’s an interesting idea, but I’ll have to **mu_____________ it ___________.** (*think carefully about it*)

9. The old days when the government controlled the media and licensed only a few newspapers are gone. In developing countries, the rapid expansion of new technology **bo___________ w___________** for journalistic freedom. Now more online publications and media outlets allow for more transparent news dispersal. (*to be a sign of something good to happen*)

10. When it comes to choosing a life partner, you should never **j_____________ __________ g.________**. (*to do something too soon, especially without thinking carefully about it*)
Appendix E: Post-tests 2 - Receptive knowledge tests (Multiple-choice)

Read the following sentences and circle the best answer. If you do not know the answer, please do NOT guess – please choose (e) I don’t know.

Passage 1
1. In the past, farmers lived __________________________ the weather.
   a. on the mercy of
   b. off the mercy of
   c. at the mercy of
   d. at the merciful of
   e. I don’t know

   This phrase means:
   a. completely regardless of
   b. completely dependent on
   c. partially dependent on
   d. in relation to
   e. I don’t know

2. Area 25 – deep in the very center of our brains – is connected to other areas that control sleep, appetite and drive, all the things that __________________________ when someone’s depressed. Mayberg’s theory is if you cool off area 25, you treat the disease.
   a. get haywire
   b. go haywire
   c. get haywires
   d. go haywires
   e. I don’t know

   This phrase means:
   a. go out of control
   b. get worse
   c. get better
   d. go away from
   e. I don’t know

3. It was a fine hotel. There were dozens of maids and servants ready to serve her. They were __________________________ at all times.
   a. at her back and call
   b. in her beck and call
   c. at her beck or call
   d. at her beck and call
   e. I don’t know

   This phrase means:
   a. angry
   b. lovely
   c. ready to serve
d. absent
e. I don’t know

4. Many students in the United States are ______________________ student loan debt. In a little more than 10 years, the total amount of student loan debt in this country has doubled to more than $1 trillion.
   a. saddled with
   b. saddlen with
   c. saddled by
   d. saddlen by
   e. I don’t know
   This phrase means:
   a. burdened with
   b. disappointed with
   c. saddened by
   d. frustrated by
   e. I don’t know

5. You can _____________________ Molly’s caterer to do a good job.
   a. bank in
   b. bank off
   c. bank of
   d. bank on
   e. I don’t know
   This phrase means:
   a. rely on
   b. respect
   c. call on
   d. hire
   e. I don’t know

6. In the evenings, he likes to ______________________ with a good book.
   a. wind down
   b. wind off
   c. wind away
   d. wind in
   e. I don’t know
   This phrase means:
   a. put down
   b. relax
   c. move
   d. read carefully
   e. I don’t know
7. Some sellers on eBay complain that the multinational internet corporation is giving them \underline{the _________________} as it only protects the rights of the buyers.
   a. short end of two sticks
   b. short end of these sticks
   c. short end of the sticks
   d. short end of the stick
   e. I don’t know

   *This phrase means:*
   a. irritation
   b. difficulty
   c. debt
   d. disadvantages
   e. I don’t know

8. We need to \underline{security in airports}
   a. beef off
   b. beef up
   c. beef out
   d. beef in
   e. I don’t know

   *This phrase means:*
   a. reduce the size of
   b. pay more money for
   c. improve the quality of
   d. provide food for
   e. I don’t know

9. Most stocks have been down all week, and unfortunately analysts forecast
   \underline{__________________________}
   a. gloom and doom
   b. glooms and dooms
   c. gloomy and doomy
   d. glooming and dooming
   e. I don’t know

   *This phrase means:*
   a. a sense of sadness
   b. a sense of disappointment
   c. a feeling of helplessness
   d. a feeling of pessimism
   e. I don’t know

10. Honda and Toyota used to \underline{the market for small family cars}
    a. have all locks in
    b. have all locks on
    c. have a lock in
d. have a lock on  
e. I don’t know  

This phrase means  

a. keep private  
b. keep secure  
c. have control of  
d. have trust in  
e. I don’t know
Passage 2

Read the following sentences and circle the best answer. If you do not know the answer, please do NOT guess – please choose (e) I don’t know.

1. I hear psychics are __________________ now. All the stars have their favorite fortune-tellers.
   a. all a rage
   b. all the rage
   c. all up rage
   d. all out rage
   e. I don’t know
   This phrase means:
   a. popular
   b. angry
   c. less expensive
   d. more famous
   e. I don’t know

2. Sociology helps solve many pressing problems facing the society; yet it is always in the list of sciences that get __________________
   a. short shrift
   b. short shrifts
   c. shortage shrift
   d. shortage shrifts
   e. I don’t know
   This phrase means:
   a. little application
   b. not much prospect
   c. little attention
   d. not much investment
   e. I don’t know

3. When feeling out of control, people usually __________________ and resist change.
   a. dig in a heel
   b. dig in their heels
   c. dig on a heel
   d. dig on their heels
   e. I don’t know
   This phrase means:
   a. refuse to stay the same
   b. refuse to stay calm
   c. refuse to make decisions
   d. refuse to make a change
   e. I don’t know
4. Though you shouldn’t be disrupting your learning by __________________________ during class, you can make the most of your school experience by having fun with your friends when you can: at the cafeteria, at your locker, while you are walking to your next class.
   a. goofing away
   b. goofing out
   c. goofing off
   d. goofing up
   e. I don’t know
This phrase means
   a. avoiding teachers
   b. avoiding work
   c. sleeping
   d. playing
   e. I don’t know

5. Ms. Solanas, an unemployed online journalist, was part of the core group of protesters who in May occupied the Puerta del Sol, a public square in Madrid, the capital, __________________________ a nationwide protest
   a. touching off
   b. touching to
   c. touching in
   d. touching upon
   e. I don’t know
This phrase means
   a. triggering
   b. expanding
   c. being a part of
   d. contributing to
   e. I don’t know

6. The engineers wanted to completely redesign the product, but couldn’t __________________________ because of a very restricted budget.
   a. push the envelope
   b. push an envelope
   c. push the envelopes
   d. push into envelopes
   e. I don’t know
This phrase means:
   a. do things in exactly the same old way
   b. do things that are not completely new
   c. go beyond earning a lot of money
   d. go beyond what has usually been done
   e. I don’t know
7. Though ___________________________ in many European and Asian countries, commuter bikes, which run generally between $400 and $900 at independent bike stores, are foreign to many Americans.
   a. old hats
   b. old huts
   c. old hut
   d. old hat
   e. I don’t know
   *This phrase means:*
   a. popular
   b. favorable
   c. not new
   d. not expensive
   e. I don’t know

8. City councils and city governments are going to have to ___________________________ and make some concessions to keep economic development on track
   a. step up till the plates
   b. step up till the plate
   c. step up to the plates
   d. step up to the plate
   e. I don’t know
   *This phrase means:*
   a. take responsibility
   b. increase their attention
   c. get more funding
   d. hire more people
   e. I don’t know

9. On the way home, ___________________________ that he had never returned his boss’s call, so when he got home he called his boss immediately.
   a. it dawned on him
   b. it dawned of him
   c. it dropped of him
   d. it dropped on him
   e. I don’t know
   *This phrase means:*
   a. he was suddenly scared
   b. he suddenly realized
   c. he completely forgot
   d. he was very surprised
   e. I don’t know
10. You can’t trust John to do the job right. He’s always ________________________.
   a. dropping a ball
   b. dreading a ball
   c. dropping the ball
   d. dreading the ball
   e. I don’t know
   
   This phrase means:
   a. sympathizing
   b. negative
   c. failing
   d. dreaming
   e. I don’t know
Passage 3

Read the following sentences and circle the best answer. If you do not know the answer, please do NOT guess – please choose (e) I don’t know.

1. Election results revealed that white racism remains ___________________________ in the country.
   a. fondly entrenched
   b. freely entrenched
   c. firmly entrenched
   d. furiously entrenched
   e. I don’t know

   This phrase means:
   a. popular
   b. deeply rooted
   c. crucial
   d. commonly seen
   e. I don’t know

2. Installation instructions were far from straightforward, but he eventually managed to ___________________________.
   a. meddle to
   b. meddle through
   c. muddle to
   d. muddle through
   e. I don’t know

   This phrase means
   a. get through something with money
   b. get through something despite difficulty
   c. succeed with pride
   d. succeed without any difficulty
   e. I don’t know

3. After the war broke out, people left the country ___________________________.
   a. on droves
   b. in droves
   c. on drove
   d. in drove
   e. I don’t know

   This phrase means:
   a. in small groups
   b. in complete panic
   c. in large numbers
   d. in several pairs
   e. I don’t know
4. Max has a new car. He must be earning _______________________.
   a. a peppy penny
   b. a pity penny
   c. a pretty penny
   d. a pending penny
   e. I don’t know
   *This phrase means*
   a. a lot of money
   b. an average income
   c. more than before
   d. more than other people
   e. I don’t know

5. A research study found that the reaction on Twitter to major political events and policy decisions is often ______________________ public opinion as measured by surveys. The Twitter conversation is usually more liberal, but can be more conservative.
   a. at odds for
   b. at odds into
   c. at odd with
   d. at odds with
   e. I don’t know
   *This phrase means:*
   a. similar to
   b. concerned with
   c. listening to
   d. disagreeing with
   e. I don’t know

6. Because Boeing cannot afford a strike in September, the union is gearing up for negotiations next August, and it believes it has Boeing _______________________.
   a. over a barrel
   b. over the barrel
   c. over all barrels
   d. over the barrels
   e. I don’t know
   *This phrase means*
   a. out of mind
   b. out of control
   c. out of sight
   d. out of work
   e. I don’t know
7. The old days when the government controlled the media and licensed only a few newspapers are gone. In developing countries, the rapid expansion of new technology for journalistic freedom. Now more online publications and media outlets allow for more transparent news dispersal.
   a. bodes willingly
   b. bodes with
   c. bodes well
   d. bodes within
   e. I don’t know
   This phrase means:
   a. is a sign that something is happening
   b. is a sign for something good to happen
   c. is a sign that something is doing well
   d. is a sign that something is illegal
   e. I don’t know

8. When it comes to choosing a life partner, you should never ________________________.
   a. jump a gun
   b. jump the gun
   c. jump the gate
   d. jump a gate
   e. I don’t know
   This phrase means:
   a. get married for money
   b. make a wrong decision
   c. make a hasty decision
   d. get married too late
   e. I don’t know

9. That’s an interesting idea, but I’ll have to ________________________.
   a. mull it out
   b. mull it over
   c. mull it up
   d. mull it through
   e. I don’t know
   This phrase means
   a. think about it carefully
   b. ask for other people’s opinions
   c. consider other options
   d. go a different direction
   e. I don’t know
10. *Since he took office, the governor has put nearly 400 laws __________________________.*
   a. in the books
   b. in the book
   c. on the books
   d. on the book
   e. I don’t know

   *This phrase means:*
   a. available for the public
   b. part of new theories
   c. part of the law
   d. available for the government
   e. I don’t know
Appendix F: Post-test 3 Underlining useful words and phrases in passage

Instruction: **Underline** words or phrases which you think are useful for improving your language skills, including speaking and writing.

**Problems of the Information Age**

The Information Age is believed to have begun roughly in the 1990's when widespread use of computers and information sharing technologies like the Internet gave rise to anytime anywhere accessibility to information. Some characteristics of the information age include: use of information to achieve high productivity, global use of information, emphasis on innovation rather than economies of scale and high social, economic and political awareness. As opposed to the industrial age, when laborious tasks were believed to yield returns, the information age relies on the production and sharing of knowledge and information to achieve profits. As the world progresses through the information age, there are some global challenges that might be faced.

1. **Information security:**

   Everyone, from individuals and business firms to governments, security agencies and defense forces, is under the threat of information theft and misuse. It ceases to be a mere technological issue when it transcends national boundaries and enters into the realms of terrorism and other international criminal activities.

2. **Digital and Internet laws**

   Legal systems around the world are at varying degrees of readiness when it comes to issues concerned with the digital world. This includes copyright issues, intellectual property rights and
so on. The Internet makes such issues international in nature, so a global uniform digital law will become a necessity in future.

3. Social challenges

As businesses go global there will be a pressing need to bridge the digital divide among nations and also within national borders. Weaving in information technology into the educational systems across the world poses a formidable challenge, especially so when illiteracy still plagues populous countries like India and China. A different but relevant social challenge in the Information age will be the right deployment of human capital. As the need for knowledge workers grows, a vast part of the global workforce may still be untrained in using information systems. A critical task, here, will be to train and involve them in the information revolution.

4. Technological challenges

Another daunting task will be to find and deploy cost effective, fast, accurate and smart technologies to store, secure, verify and share information. Technologists and entrepreneurs will have a tough time satisfying the on the go information needs of Generation Y. Tiny storage devices with huge capacities and fast retrieval techniques, are what this generation of the Information Age would like to see in their pockets!
Appendix G: Test scores by FSs

<table>
<thead>
<tr>
<th>FSs</th>
<th>Frequency</th>
<th>n-gram length</th>
<th>MI Score</th>
<th>Productive Knowledge - Immediate Post-test (120)</th>
<th>Productive Knowledge - Delayed Post-test (120)</th>
<th>Receptive Knowledge - Form section - Immediate Post-test (40)</th>
<th>Receptive Knowledge - Meaning section - Immediate Post-test (40)</th>
<th>Receptive Knowledge - Form section - Delayed Post-test (40)</th>
<th>Receptive Knowledge - Meaning section - Delayed Post-test (40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>go_haywire</td>
<td>226</td>
<td>2</td>
<td>8.04</td>
<td>93</td>
<td>33</td>
<td>33</td>
<td>35</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>have_a_lock_on</td>
<td>127</td>
<td>4</td>
<td>3.66</td>
<td>58</td>
<td>21</td>
<td>33</td>
<td>34</td>
<td>22</td>
<td>36</td>
</tr>
<tr>
<td>at_the_mercy_of</td>
<td>674</td>
<td>4</td>
<td>3.68</td>
<td>94</td>
<td>30</td>
<td>38</td>
<td>38</td>
<td>36</td>
<td>39</td>
</tr>
<tr>
<td>wind_down</td>
<td>1254</td>
<td>2</td>
<td>6.57</td>
<td>99</td>
<td>68</td>
<td>37</td>
<td>38</td>
<td>30</td>
<td>39</td>
</tr>
<tr>
<td>at_her_beck_and_call</td>
<td>113</td>
<td>5</td>
<td>5.05</td>
<td>65</td>
<td>28</td>
<td>25</td>
<td>40</td>
<td>24</td>
<td>38</td>
</tr>
<tr>
<td>saddled_with</td>
<td>528</td>
<td>2</td>
<td>3.84</td>
<td>105</td>
<td>61</td>
<td>38</td>
<td>38</td>
<td>36</td>
<td>39</td>
</tr>
<tr>
<td>gloom_and_doom</td>
<td>85</td>
<td>3</td>
<td>15.02</td>
<td>73</td>
<td>63</td>
<td>33</td>
<td>33</td>
<td>26</td>
<td>34</td>
</tr>
<tr>
<td>short_end_of_the_stick</td>
<td>69</td>
<td>5</td>
<td>10.64</td>
<td>66</td>
<td>28</td>
<td>34</td>
<td>32</td>
<td>31</td>
<td>35</td>
</tr>
<tr>
<td>beef_up</td>
<td>540</td>
<td>2</td>
<td>7.14</td>
<td>82</td>
<td>44</td>
<td>34</td>
<td>38</td>
<td>31</td>
<td>37</td>
</tr>
<tr>
<td>bank_on</td>
<td>514</td>
<td>2</td>
<td>5.17</td>
<td>73</td>
<td>53</td>
<td>34</td>
<td>37</td>
<td>29</td>
<td>39</td>
</tr>
<tr>
<td>step_up_to_the_plate</td>
<td>383</td>
<td>5</td>
<td>9.32</td>
<td>81</td>
<td>16</td>
<td>32</td>
<td>34</td>
<td>19</td>
<td>33</td>
</tr>
<tr>
<td>all_the_rage</td>
<td>384</td>
<td>3</td>
<td>5.54</td>
<td>85</td>
<td>23</td>
<td>37</td>
<td>40</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>touching_off</td>
<td>568</td>
<td>2</td>
<td>3.68</td>
<td>59</td>
<td>2</td>
<td>35</td>
<td>36</td>
<td>14</td>
<td>33</td>
</tr>
<tr>
<td>dawned_on</td>
<td>1074</td>
<td>2</td>
<td>6.17</td>
<td>96</td>
<td>64</td>
<td>36</td>
<td>38</td>
<td>29</td>
<td>34</td>
</tr>
<tr>
<td>short_shift</td>
<td>251</td>
<td>2</td>
<td>12.55</td>
<td>87</td>
<td>21</td>
<td>37</td>
<td>37</td>
<td>22</td>
<td>33</td>
</tr>
<tr>
<td>push_the_envelope</td>
<td>461</td>
<td>3</td>
<td>7.75</td>
<td>89</td>
<td>20</td>
<td>36</td>
<td>37</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>old_hat</td>
<td>178</td>
<td>2</td>
<td>4.23</td>
<td>104</td>
<td>63</td>
<td>34</td>
<td>38</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>FSs</td>
<td>Frequency</td>
<td>n-gram length</td>
<td>MI Score</td>
<td>Productive Knowledge - Immediate Post-test (120)</td>
<td>Productive Knowledge - Delayed Post-test (120)</td>
<td>Receptive Knowledge - Form section - Immediate Post-test (40)</td>
<td>Receptive Knowledge - Form section - Delayed Post-test (40)</td>
<td>Receptive Knowledge - Meaning section - Immediate Post-test (40)</td>
<td>Receptive Knowledge - Meaning section - Delayed Post-test (40)</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------</td>
<td>---------------</td>
<td>----------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>goof_off</td>
<td>152</td>
<td>2</td>
<td>8.19</td>
<td>85</td>
<td>20</td>
<td>35</td>
<td>33</td>
<td>21</td>
<td>31</td>
</tr>
<tr>
<td>dig_in_their_hells</td>
<td>166</td>
<td>4</td>
<td>9.95</td>
<td>80</td>
<td>38</td>
<td>35</td>
<td>36</td>
<td>27</td>
<td>36</td>
</tr>
<tr>
<td>drop_the_ball</td>
<td>310</td>
<td>3</td>
<td>7.75</td>
<td>98</td>
<td>81</td>
<td>30</td>
<td>38</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td>at_odds_with</td>
<td>1700</td>
<td>3</td>
<td>5.84</td>
<td>110</td>
<td>91</td>
<td>35</td>
<td>39</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>in_droves</td>
<td>592</td>
<td>2</td>
<td>5.32</td>
<td>95</td>
<td>44</td>
<td>33</td>
<td>39</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>over_a_barrel</td>
<td>81</td>
<td>3</td>
<td>5.35</td>
<td>53</td>
<td>8</td>
<td>17</td>
<td>39</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td>firmly_entrenched</td>
<td>139</td>
<td>2</td>
<td>12.93</td>
<td>97</td>
<td>58</td>
<td>39</td>
<td>38</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>muddle_through</td>
<td>179</td>
<td>2</td>
<td>8.08</td>
<td>97</td>
<td>69</td>
<td>39</td>
<td>38</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>pretty_penny</td>
<td>83</td>
<td>2</td>
<td>6.19</td>
<td>96</td>
<td>89</td>
<td>35</td>
<td>39</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td>on_the_books</td>
<td>943</td>
<td>3</td>
<td>6.21</td>
<td>83</td>
<td>67</td>
<td>23</td>
<td>34</td>
<td>11</td>
<td>38</td>
</tr>
<tr>
<td>mull_over</td>
<td>415</td>
<td>2</td>
<td>8.19</td>
<td>72</td>
<td>27</td>
<td>27</td>
<td>37</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>bode_well</td>
<td>556</td>
<td>2</td>
<td>8.65</td>
<td>78</td>
<td>27</td>
<td>28</td>
<td>32</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>jump_the_gun</td>
<td>150</td>
<td>3</td>
<td>9.76</td>
<td>107</td>
<td>70</td>
<td>35</td>
<td>38</td>
<td>33</td>
<td>40</td>
</tr>
</tbody>
</table>
Appendix H: Number of target FSs underlined by members of the Control group in 3 treatment sessions

<table>
<thead>
<tr>
<th>Learner/Number of target FSs underlined in each passage</th>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner 1</td>
<td>0</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Learner 2</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Learner 3</td>
<td>0</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Learner 4</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Learner 5</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Learner 6</td>
<td>0</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Learner 7</td>
<td>3</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Learner 8</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Learner 9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Learner 10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
REFERENCES


171


Haastrup, K., & Henriksen, B. (2001). The interrelationship between vocabulary acquisition theory and general SLA research. EUROSLA Yearbook, 1, 69-78.


