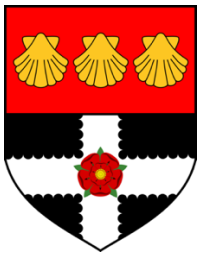


Comparing different types of EFL vocabulary instruction for Chinese senior secondary school learners of English

Pengchong Zhang

Institute of Education

UNIVERSITY OF READING

Thesis submitted for the degree of

Doctor of Philosophy

December 2017

ACKNOWLEDGEMENTS

I would first like to express my sincere and profound gratitude to my supervisor, Professor Suzanne Graham, without whom I would not have started to do a PhD or accomplished this PhD thesis. Over the past three years, Suzanne has supported me from every aspect of my study, from at the beginning guiding me to explore the theoretical background, discussing the research design and developing relevant research instruments to analysing the data, interpreting and writing up the results at the later stage. She has also inspired me with her passion of doing research and has been a most patient supervisor who helped me revising my abstract for each conference presentation and applications for funding resources. In particular, she has also been a most considerate and thoughtful supervisor who encouraged me to think positively towards every aspect of my life.

I am also deeply grateful to the Institute of Education, University of Reading, in particular to Professor Jeanine Treffers-Daller, my second supervisor, and Dr. Daisy Powell, who have provided invaluable suggestions at different stages of my PhD study. My sincere thanks go to Dr. Daguo Li, Dr. Holly Joseph and Professor Rhona Stainthorp who trusted me and assigned teaching and research assistant jobs to me. I am also thankful to Ms. Yun Wang, who helped with sample coding the stimulated recall interviews and provided insightful suggestions at the initial stage of analysing the qualitative data.

My special thanks to the students and the two senior secondary schools in Nantong, China who took part in the pilot and the main data collection phase of the study. Thanks also go to the students' parents, who gave consent for their children's participation.

Finally, I am especially indebted to my family and friends in Reading and Nantong, my parents and my grandparents, who have been supportive all the way through my PhD journey. Their warm welcome and company whenever I returned to China for holiday breaks is indeed a great motivation for my study and of course essential to the accomplishment of this thesis.

ABSTRACT

The primary aim of the study was to explore the teaching and learning of vocabulary through listening among 137 senior secondary school EFL learners in China. A quasi-experimental, mixed method design was adopted comparing incidental vocabulary learning through listening (Control Group) with different types of Lexical Focus-on-Form delivered to three treatment groups: post-listening vocabulary explanations in the L2; codeswitched explanations; and explanations providing additional cross-linguistic information (contrastive Focus-on-Form, CFoF). The second aim of the study was to investigate whether learners' listening comprehension developed alongside their vocabulary. Finally, the study explored what strategies were used by learners in response to the vocabulary instruction in each of the three experimental conditions.

The data collection procedure, involving a classroom intervention, lasted three months. Learners completed aural vocabulary tests at pre, post and delayed post-test and listening assessments at pre and post-test. The three treatment groups also completed an additional final vocabulary delayed post-test. Stimulated recall interviews were conducted finally with twelve learners from the three treatment groups. The findings first indicate that for short and long-term vocabulary acquisition, the three treatment groups significantly outperformed the Control group. Gains for the CFoF group were significantly greater than for the L2 and codeswitching groups. Additionally, the codeswitching group significantly outperformed the L2 group for short-term but not for long-term acquisition.

Regarding whether the vocabulary learning varied according to learners' general English language proficiency, findings reveal that compared with lower-level learners, higher-level learners benefited more from the L2-only and the CFoF vocabulary explanations for short-term vocabulary learning. Additionally, analysing the learning by word classes and for collocations, results on the one hand indicate that collocations and nouns tended to be better acquired than verbs and adjectives, on the other hand suggest that the learning of collocations and single words by the learners who received CFoF vocabulary explanations

was significantly better than those from the L2 and codeswitching group. Furthermore, regarding the impact of different repetitions on vocabulary retention, findings confirm that target lexical items receiving nine repetitions were significantly better retained than those receiving seven, five or three repetitions.

Looking at the impact of vocabulary intervention on learners' listening comprehension, findings indicate that the L2, codeswitching and Control group showed significant pre to post-test improvement in listening comprehension, with most progress for the Control group. However, the CFoF group did not make significant progress and their performance was significantly worse than the Control group's at post-test. Finally, the qualitative analysis regarding the strategies used in response to the vocabulary instruction suggest that learners used L2 listening comprehension strategies to understand the listening input as well as employing vocabulary learning strategies to guess the meaning of the unfamiliar lexical items and to further remember these items. In addition, in general, higher proficiency level learners tended to use more different strategies than low proficiency level learners. Moreover, although certain patterns of strategy use were shared by both higher and lower proficiency level learners within each treatment condition, higher proficiency level learners tended to use these strategy patterns in a more active way, compared with lower proficiency level learners who employed the strategies in a passive manner. The thesis concludes by discussing these findings in relation to theories of vocabulary acquisition and listening comprehension, as well as their implications for pedagogy, the limitations of the study and areas for future research.

DECLARATION OF ORIGINAL AUTHORSHIP

I confirm that this is my own work and the use of all materials from other sources has been properly and fully acknowledged.

Signature:

Print name: Pengchong Zhang

Date: 08/12/2017

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

CHAPTER 1 INTRODUCTION

This study investigates the acquisition of L2 vocabulary through listening, an area which has received relatively little research attention. This research objective has two main aspects. First, from a teaching perspective, it investigates the impact of three types of vocabulary instruction on Chinese senior secondary school English as a foreign language (EFL) learners' vocabulary learning. Second, from a learning perspective, it explores what strategies a group of senior secondary school EFL learners used in response to different types of vocabulary instruction provided by the teacher, covering strategies used while listening, for understanding the teacher's explanations, and any strategies used to learn the vocabulary items. As a background to the study, the introductory chapter will first discuss the current trends and challenges of English language education in China. Thereafter, the research rationale of the present study will be briefly discussed. Then, research aims will be listed. Finally, a brief overview of the structure of the thesis will be presented.

1.1 English language education in China

1.1.1 Opportunities

With rapid economic growth and modernisation, English language education has received prominent attention by the Chinese government as well as by the general public (Hu, 2002). Since July 2001, when Beijing was announced to be the host city for the 2008 Summer Olympics, the number of both English teachers and learners in China has been rapidly increasing. These English learners are not only school-aged children or university-level students, for whom English language courses are compulsory, but also those from all age groups and different industrial sectors who regard good English language proficiency as a competitive personal asset when seeking jobs (Cortazzi & Jin, 2002).

Corresponding to this national call for 'learning English', the development of English proficiency has become an important goal for school-aged children in China (Wu, 2001). An educational reform initiated by the Ministry of Education of People's Republic of China (MoE), specifically targeting school English language education (primary school level: Grade 1-6; secondary school level: junior secondary school Grade 7-9 and senior secondary school Grade 10-12), took place in 2001. It enforced formal and compulsory English language courses to be implemented in primary school, starting from Grade 3 level (aged 9-10) (MoE, 2001). Before that, formal English courses were only provided by those primary schools in large cities. Most school-aged learners started learning English formally at junior secondary school Grade 7 (aged 13-14) (Hu, 2005). There were three main factors which led to this school-level English language education reform. Firstly, China was about to be accepted as a member of the World Trade Organization. Secondly, as discussed above, Beijing was about to host the 29th Summer Olympics (Nunan, 2003). Thirdly and most importantly, the early 1990s' education reform, aiming at improving the quality of secondary school English language teaching, had not shown positive effects on secondary school graduates' English language proficiency, indicating that there is a need for learning English at an earlier education stage (Hu, 2005).

In addition, as part of this education reform, a national-level curriculum for senior secondary school English language teaching was introduced by MoE (2003). This new curriculum replaced the English syllabus for full-time senior secondary schools (MoE, 2000a) which had only been implemented for less than two years. The following six features are emphasised in this new English curriculum: "one way, two assessments, two processes, four tasks, five components and nine bands" (Wang & Lam, 2009, p. 71). Details of these features are presented in Table 1.1.

Table 1.1

Features of the new curriculum (Wang & Lam, 2009, p. 71)

Features	Details
One way	Enquiry-based learning
Two assessments	Formative and summative, mainly formative
Two processes	Students developing language use ability and developing as people
Four tasks	To develop students' interest in learning English, language use ability, creativity, and cultural awareness
Five components	Linguistic skills, linguistic knowledge, learning strategies, affect, and cultural awareness
Nine bands	The above components are graded into nine bands (two bands for primary school, three bands for junior secondary school and four bands for senior secondary school).

According to this new curriculum, students are now encouraged to actively enquire during the process of learning English, rather than passively receive and remember the knowledge from the teacher. In addition, there is a new interpretation of assessment. Compared with traditional summative assessment whereby a final measurement is used to assess learners' English language proficiency, formative assessment is promoted. The main purpose of formative assessment is to help learners to develop the above 'four tasks' and 'five components' through the stages of 'nine bands', involving both the development of learners' language proficiency and their own personality "beyond the confines of education" (Wang & Lam, 2009, p. 71).

1.1.2 Challenges

It can be seen from the above features of the new curriculum that, although this curriculum targets the senior secondary school EFL learners, it is also applicable to English education at basic (primary school and junior secondary school) levels. However, in reality, primary and junior secondary schools have their own curricula to follow. For example, English language education in junior secondary schools followed the English syllabus for nine-year compulsory education full-time junior secondary schools (MoE, 2000b) until this was

replaced by the English basic education curriculum (MoE, 2012) in 2012. This inconsistency of curriculum use is due to the fact that different education stages in China are overseen by different government departments in the MoE (Jin, Wu, Alderson, & Song, 2017), which then led to the specification of inconsistent learning objectives in the curricula for Chinese EFL learners at different educational levels.

One example of this inconsistency, pointed out by Jin et al. (2017), is that the requirement for the amount of vocabulary to be learnt is not coherent across the different learning stages. To be more specific, it is unclear whether "the vocabulary sizes are cumulative across educational stages or each stage is independent of the others" (p. 4). In addition, it is also unclear whether lexical items appearing in the word-lists for lower educational levels are included in those lists for higher educational levels. Moreover, since the curriculum for each educational stage was developed independently, it was difficult for the curriculum writers to identify essential vocabulary items for each learning stage. Therefore, no separate lists of core lexical items were provided, which then made it unclear for "teachers, textbook writers and assessment developers to focus on the items essential to education at a particular stage" (Jin, et al., p. 5).

Another challenge of current English education in China is that the country has a rather complex picture of methodological approaches explored in English language teaching (Silver, Hu, & Iino, 2002). As discussed above, one feature of the new curriculum is to promote the transition from the traditional 'learning from the teacher' approach to the 'enquiry-based learning' approach, which draws on principles for Communicative Language Teaching (CLT). However, the real situation in China is that, at the secondary school level, "CLT has not gained wide currency" (Silver et al., 2002, p. 33). Only teachers from secondary schools in the cities, who have relatively good proficiency in English listening and speaking, are able to adopt CLT in their classroom teaching. Besides, the successful implementation of CLT also depends on the school's teaching facilities and resources (i.e., whether the classroom is well-equipped with audio-visual equipment).

Regarding secondary school teachers in rural China, the Grammar-Translation Method is still widely used in their classrooms not only because they do not have a clear understanding of the principles of CLT, but also because these teachers doubt that there are clear pedagogical advantages of CLT over the traditional methods to which they have been used for many years. In addition, they may resist implementing CLT as this "runs counter to traditional practice" (Rao, 2013, p. 37). In a traditional Chinese classroom, the teacher always acts like a 'leader' or 'boss' and takes charge of everything, while the students are required to follow the teacher's instruction with fewer chances to express their opinions or to interact with the teacher.

In fact, a growing number of secondary school English teachers tend to adopt "an eclectic approach, drawing on different methods and trying to reconcile traditional practices with more recent innovations to meet the demand of their particular (often diverse) teaching situations" (Silver, et al., 2002, p. 33). For these teachers, there is no best teaching method. In order to meet the teaching requirements, they prefer to select and use the appropriate elements which they think are useful for their classroom practices from the available teaching methods.

A final challenge is that although school-aged English learners in China have a large amount of formal contact with English as a subject in school, the general language proficiency of these learners is relatively low. Since 2001 when English became a compulsory subject starting from primary school Grade 3, learners are required to have an average two to three hours of English lessons every week in primary school. At both junior and senior secondary levels, learners spend three to four hours each week learning English as a school subject (Silver, et al., 2002). Therefore, Chinese senior secondary school graduates would have had approximately 1000 hours of formal contact with English during their 10 years' learning. In reality, this amount of contact time tends to be higher in the cities, where children normally start learning English at primary Grade 1 or even earlier when they are still at nursery school.

Unfortunately, this fairly large amount of formal contact with English as a school subject does not lead to high levels of English language proficiency. According to the International English Language Testing System (IELTS) test taker performance report 2015 (see Table 1.2), the IELTS (Academic) mean overall score of the Chinese test takers was lower than that of the test takers from surrounding countries or regions. Looking more specifically at the average mean score for each of the four skills (listening, reading, writing and speaking), the performance of the Chinese test takers was also less satisfactory than those from the neighbourhood countries or regions. Normally, test takers of IELTS (Academic) are senior secondary school graduates or university undergraduates who wish to pursue further education in an English-speaking country. Therefore, these learners' IELTS results can to some extent provide a picture of the general English proficiency of the current secondary school and university English learners in China.

Table 1.2

Mean band score of IELTS (Academic) for China and its surrounding countries or regions (IELTS, 2015)

Place of origin	**Listening**	**Reading**	**Writing**	**Speaking**	**OVERALL**
China (People's Republic of)	5.9	6.1	5.3	5.4	5.7
Hong Kong	6.9	6.7	5.9	6.2	6.5
Taiwan	6.2	6.2	5.6	6.0	6.1
Japan	5.9	6.1	5.3	5.6	5.8
Korea (Republic of)	6.2	6.2	5.4	5.7	5.9
Thailand	6.2	6.0	5.5	5.9	6.0
Vietnam	6.0	6.2	5.6	5.8	6.0

To sum up, English language education is an area of focus in China as both the general public and the policy makers have realised the importance of this language and wish to support the development of the nation's English language proficiency. However, English language education in China is still facing several challenges regarding the current government policies (lack of a unified English curriculum) and the classroom teaching methods (CLT or traditional teaching methods). In addition, Chinese and English are very distinct from each other as both the written and spoken systems are different between the

two languages. Moreover, there is not much common cultural background shared by the two languages. Therefore, there is still a need for more research to investigate how English can be better acquired by Chinese learners.

1.2 Research rationale

For senior secondary school learners in China developing English proficiency is an important goal, a central aspect of which is the acquisition of a wide range of vocabulary (Bogaards & Laufer, 2004; Meara, 1980; Nation, 2013). Nation (2013) argues that knowing a word is far more than just knowing the meaning of the word. Therefore, he proposed a framework indicating all aspects of knowing a word involve both receptive and productive knowledge of the word in three dimensions: form, meaning as well as use (details for these will be presented in Section 2.2.1).

Regarding how the three aspects of word knowledge can be learnt, Ellis (1994), from a psychological point of view, argued that the form aspect strongly relies on implicit learning and the meaning aspect can be better learnt through explicit learning while the use aspect may need both implicit and explicit learning (differences between the two types of learning will be discussed in Section 2.2.1). Considering implicit and explicit learning from a slightly different angle, a view that sees vocabulary learning as involving intentional learning and/or incidental learning, has attracted particular attention (Sonbul & Schmitt, 2010). In incidental vocabulary learning, learners learn "vocabulary as the by-product of any activity not explicitly geared to vocabulary learning". Conversely, intentional vocabulary learning refers to "any activity aiming at committing lexical information to memory" (Hulstijn, 2001, p. 267).

Although there has been much research about how a new vocabulary item can be acquired from both a theoretical perspective (Kroll & Stewart, 1994; Jiang, 2000) (to be discussed in Section 2.2.4) and a pedagogical point of view (Ellis, 1994; Hulstijn, 2001; Nation, 2013), relatively little is known, however, regarding the relative benefits to vocabulary

development of different classroom practices. It has been argued that certain types of intentional vocabulary learning (Laufer, 2009) or, at the very least, incidental plus intentional learning, may lead to better learning outcomes compared with incidental learning alone (Hill & Laufer, 2003; Zhao & Macaro, 2014), as incidental learning through reading always involves intentional inferring meaning of the unknown vocabulary items (Hennebry, Rogers, Macaro, & Murphy, 2013). These types of incidental plus intentional learning methods may include the meaning-given method (Mondria, 2003), in which the meaning of unfamiliar lexical items is directly translated by the teacher so that learners can immediately start memorising, reading plus vocabulary-enhancement activities (Min, 2008) and lexical Focus-on-Form (Laufer & Girsai, 2008).

Lexical Focus-on-Form is defined by Laufer (2009) as an attempt to promote intentional vocabulary learning through communicative language teaching. However, lexical Focus-on-Form taking the form of teacher-learner oral interactions is an area which only a few studies have explored (Tian & Macaro, 2012). In addition, it is still questionable which type of vocabulary instruction, second language (L2) explanations or first language (L1) equivalents, should be delivered by the teacher in order to more efficiently enhance vocabulary learning (Hennebry et al., 2013). While current educational policy in China encourages maximum second language (L2) use by senior secondary school teachers in the classroom, it is unclear whether this is in fact the optimum teaching approach (Silver, et al., 2002). Moreover, there is no clear evidence indicating how much or what kind of explanation given by teachers is required for learners to acquire a certain lexical item, and particularly whether it would be valuable to add other explanations, such as how to use a word, in addition to providing the word meaning.

The promotion of L2 exclusivity became dominant in L2 classrooms when the new CLT method was introduced in the area of second language acquisition (Cook, 2010; Widdowson, 2003). Since the late 1970s, three theoretical hypotheses have been raised to support L2-only explanations for vocabulary teaching (Zhao & Macaro, 2014). First, the L1 = L2 Hypothesis (Ellis, 1985) proposed that the process of L2 acquisition is same as L1

acquisition. According to the proponents of this theory, extensive exposure to the L2 is sufficient for second language acquisition (Dulay & Burt, 1974; Ellis, 1995). The two other hypotheses, the Comprehensible Input Hypothesis (Krashen, 1982) and the Interaction Hypothesis (Long, 1983), indicated that L2 knowledge would be efficiently acquired if learners were exposed to sufficient comprehensible L2 input or were able to interact with the teacher in L2 to a large extent. In more depth, Nation (1996) suggested four important strands of a language course: learning through meaning-focused input, learning through deliberate attention to language items and language features, learning through meaning-focused output as well as "developing fluent use of known language items and features over the four skills of listening, speaking, reading and writing" (p. 7). These four strands emphasise the importance of extensive target language input both in listening and reading for direct vocabulary learning.

By contrast, it has been argued by Cook (2001) and Macaro (2005) that limiting the teacher's use of L1 is likely to eliminate learners' opportunity to use teachers as bilingual dictionaries, considered to be useful tools in dealing with reading comprehension and writing tasks. In addition, in the context of rural China, where many EFL teachers still resist using CLT in their classrooms, it seems impossible and unnecessary for them to be deprived of the use of L1 and, in fact, the case of L2-only in the classroom is still quite rare. Moreover, even if L1 is allowed in the classroom context, the basic principles of CLT can still be maintained. For example, Tian and Macaro (2012) made a distinction between teacher codeswitching (CS) and teacher use of L1 by arguing that 1) teacher CS follows such frameworks where a hook to "hang principled rather than ad hoc L1 use" is offered (p. 369); and 2) teacher CS which occurs in classroom discourse can be authentic and it may not conflict with naturalistic CS because it is employed in the classroom context with the purpose of assisting communication and teaching/learning target language.

A tentatively positive role for teacher CS in vocabulary learning has been confirmed as learners' vocabulary learning outcomes can be improved through providing L1 meaning by teacher CS in both L2 listening (Hennebry et al., 2013; Tian & Macaro, 2012) and reading

(Lee & Macaro, 2013; Zhao & Macaro, 2014) modalities. However, two problems have been raised. Firstly, in an examination of learners' long-term vocabulary learning outcomes by Tian and Macaro (2012), the benefits of teacher CS are unsustainable compared with L2-only. Secondly, some potential drawbacks of teacher CS should be taken into account, as with the case when learners may be relying too much on L1 explanations and thus neglecting to develop the strategy of inferencing for vocabulary learning (see Section 2.3.1 for more details of this study). It is possible to argue that teacher CS can be effective to some extent. Nevertheless, the potential costs that it may bring are still worth investigating.

Quite different from teacher CS and L2-only, Laufer and Girsai (2008) suggested the contrastive Focus-on-Form (CFoF) method for lexical explanations, in which cross-linguistic information about the target vocabulary is provided by the teacher through a communicative task. Cross-linguistic information may not be simply referred to as bilingual definitions of target words, "but to the kind of instruction which leads to learners' understanding of the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system" (p. 696). Empirical evidence (Laufer & Girsai, 2008) shows that L2 language learners who received CFoF vocabulary explanations performed better than those who received L2-only explanations not only for learning single words and but also for learning collocations, groups of words "that belong together, either because they commonly occur together…or because the meaning of the group is not obvious from the meaning of the parts" (Nation, 2001, p. 317) (see Section 2.3.2 for more details of this study).

Another issue related to CFoF is its possible value for teaching near synonymy, pairs of lexical items sharing very similar cognitive meanings, yet having different collocational behaviour. Xiao and McEnery (2006) examined collocational behaviour and semantic prosody for near synonyms from a cross-linguistic perspective (L1: Chinese; L2: English). Semantic prosody refers to "a form of meaning which is established through the proximity of a consistent series of collocates" (Louw, 2000, p. 57). They found that the possibility of misuse of L2 words by learners could be largely reduced if teachers make distinctions

between “the collocational behaviour and semantic prosody/preference of near synonyms in L1 and their close translation equivalents in L2” (p. 126).

It should be noted that Laufer and Girsai’s study is set mainly within the context of incidental vocabulary learning where general non-communicative tasks (reading comprehension, gap-fill exercise and translation tasks) are adopted, while Xiao and McEnery’s research based on data collected from several corpora is not in a classroom context. However, in senior secondary school Chinese EFL classrooms, where vocabulary is taught through spoken texts by the CLT method, it is still unclear whether these benefits for CFoF exist.

Additionally, looking into the pedagogical effects of different types of instruction for L2 vocabulary explanation also raises four other issues: 1) Whether the impact of each type of instruction varies according to learners’ English language proficiency level; 2) whether the different types of vocabulary instruction lead to different learning outcomes for different word classes (nouns, verbs, adjectives), and for collocations, especially in the context where vocabulary knowledge is delivered through listening (Tian & Macaro, 2012); 3) whether repeated teaching of the target lexical items can make a difference in learning, i.e., whether the items receiving more repetitions can be retained better; and 4) which strategies are employed by L2 learners in response to the teacher’s vocabulary instruction (Macaro, 2014).

Moreover, existing research has tended to explore vocabulary learning through reading, while there has been a much less focus on investigating how vocabulary knowledge can be developed through listening (Vidal, 2003; 2011; van Zeeland & Schmitt, 2013a). It is indeed necessary to call for more research looking at vocabulary acquisition through oral input primarily because compared with written input, oral input arguably has a more central position within CLT (Richards & Rodgers, 2014). In addition, in the context of Chinese secondary school EFL classroom, a transition has been taking place, whereby the traditional Grammar Translation approach in EFL classrooms has been gradually replaced

by a more communicative and oral-based approach (Lee, 2010), which emphasises the importance of vocabulary development through listening.

Finally, a question raised from vocabulary acquisition through listening is whether any increasing vocabulary knowledge would have a potential impact on learners' general listening comprehension, a question which seemingly has not yet been explored. Nevertheless, there has been evidence from the L1 reading context (Wright & Cervetti, 2016) indicating that vocabulary interventions through simply teaching the meaning of the words, although having positive impacts on reading comprehension of written texts including these target words being taught, does not support the improvement of general reading comprehension. However, in the context of L2 listening, this still remains uninvestigated.

1.3 Significance and aim of the study

While several previous studies explored the role of teacher use of L1 in L2 language classrooms, few studies have identified the role of L1 for learners' L2 learning, especially for L2 vocabulary learning. In addition, previous empirical studies mainly focused on learning lexical items through written texts, yet little research has touched upon the area of vocabulary learning through oral input. Moreover, even fewer studies have investigated what types of vocabulary explanation given by the teacher are most beneficial for learners to acquire a certain lexical item.

Thus, the present study attempts to fill this gap and make a contribution to our understanding of classroom language teaching and lexical learning for senior-secondary school EFL learners in China and explore the role of L1 for vocabulary teaching and learning. It has four primary aims:

1. to investigate the extent to which L2 learners learn words from listening;

2. to determine whether CFoF instruction, L2-only explanations or teacher CS is more beneficial for vocabulary learning through oral input by evaluating learners' vocabulary learning outcomes in the Lexical Focus-on-Form context;

3. to investigate whether the type of vocabulary intervention that L2 learners receive may have an impact on their general listening comprehension;

4. to gain insights into the strategies used by learners within different types of vocabulary instruction received in a Chinese EFL context.

1.4 Thesis overview

This Introductory Chapter has already set out the scene of the present study by initially giving the general background of English language education in China, listing its current opportunities and challenges. Then, by narrowing down the scope, the chapter highlighted vocabulary acquisition as a fundamental goal of improving English proficiency for Chinese senior secondary school learning. Thereafter, the research rationale was discussed with a general review of theories and empirical evidence. Finally, the significance and research aims of the present study were identified.

The following Chapter Two, the literature review chapter, will further explore some fundamental issues regarding vocabulary learning including what is a word, what does it mean to know a word. This will be followed by a brief consideration of some aspects relating to general second language acquisition. Then, theories relating to L2 vocabulary acquisition will be systematically reviewed. Subsequently, existing empirical studies will be discussed with a specific focus on vocabulary learning through listening. Afterwards, issues regarding vocabulary learning strategies and L2 listening comprehension will be reviewed. Finally, research gaps will be identified and research questions will be stated.

Chapter Three, the methodology chapter will begin with a brief discussion of the research design. Thereafter, a justification of the selection of the research methods will be given.

This will be followed by detailed discussion of sampling, data collection procedures as well as the selection and development of relevant research instruments. Subsequently, the details of the pilot phase will be given with a discussion of its implications for the main study. Eventually, data analysis procedures and ethical issues will be considered.

In Chapter Four, the findings chapter, reliability of the test instruments adopted will be given firstly. Secondly, the procedure of allocating participants into different proficiency levels will be discussed. Thirdly, findings for each research question will be presented. Quantitative data collected will be used to address the first five research questions focusing on vocabulary acquisition through listening, while the final research question exploring what strategies do learners use in response to the vocabulary instruction offered by the teacher in each of the three experimental conditions will be addressed by the qualitative data.

In Chapter Five, the discussion chapter, findings for each research question will be discussed with reference to the theories and empirical background discussed in the literature review chapter. In the instances where the findings of the present study differ from previous empirical evidence, possible reasons will be identified.

Chapter Six, the conclusion chapter, will initially summarise the findings of the present study. Then, its methodological contribution and limitations will be discussed. Recommendations for future research will be made afterwards. This will be followed by a discussion of the pedagogical implication of the study. Finally, concluding remarks will be given, again highlighting the significance of the present study.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

In this chapter, literature relating to the present study will be reviewed. As the primary aim of the present study is to explore how vocabulary knowledge can be acquired and taught through listening, the review will include four major issues: 1. the theoretical background of how vocabulary knowledge can be developed; 2. empirical evidence for how vocabulary knowledge can be acquired within the classroom context; 3. the impact of different types of explicit lexical Focus-on-Form on vocabulary learning with a specific focus on the role of learners' first language; 4. the relationship between vocabulary knowledge and L2 listening comprehension.

In the first section, initially the definition of 'word' will be given and what is involved in knowing a word will be discussed. Then, a consideration of some aspects of second language acquisition, especially regarding implicit and explicit knowledge will be presented, followed by a specific focus on theories about L2 vocabulary development, in order to identify how learners are believed to process L2 vocabulary items cognitively and how this differs from processing L1 lexical items. In the second section, subsequently, from a pedagogical perspective, possible ways of how vocabulary knowledge can be developed in and outside the second language classroom will be discussed with support of evidence from previous empirical studies.

In the third section, two forms of the teacher's use of L1, in the second language classroom will be reviewed, namely teacher codeswitching (CS) and contrastive Focus-on-Form (CFoF). In addition to defining these two forms of L1 use, their possible roles for L2 vocabulary instruction will be analysed in relation to the theoretical frameworks previously presented for vocabulary learning. Issues relating to what strategies do learners use in response to the teacher's vocabulary explanations will also be reviewed. In the fourth part, the possible relationship between L2 vocabulary knowledge and L2 listening

comprehension will be discussed with reference to models of L2 listening and empirical evidence. Finally, research gaps will be identified based on the literature review, leading to the research questions for the present study.

2.2 Vocabulary learning

2.2.1 What is a word and what does it mean to know a word

Before moving onto discussing how a word can be learnt, it is very important to define what a word is. Regarding what count as words, Nation (2013, p. 10-11) suggested that words can be counted by tokens, types, lemmas as well as by word families. Counting by tokens is to simply count every single word form within a written or oral sentence. Repeated occurrence of the same word form is also counted. Differently, counting by types is to count the words of different forms within spoken or written texts, i.e., the word does not count as a new word when it appears for the second time. To be more specific, words can be counted by lemmas, whereby "closely related words could be counted as members of the same word or lemma" (e.g., *dog, dogs*). Finally, counting words by word families means that a head word and "its inflected forms and its closely related derived forms" should be counted as a word family (e.g., *develop, developed, development, undeveloped*).

Quite similarly, Harley (2006) suggested that there are "two different notions of word: a listeme — a sound-meaning correspondence — and a phonological word, a sound unit on which the spacing conventions of written English are based" (p. 6). Harley (2006) then further explained that words in the following two instances also be regarded as phonological words. First, words which are organised by two sound-meaning correspondences, such as in contractions (*I'm, don't*) and affixes (*dogs, unhappy*). Second, where words have no specific meanings (e.g., *caboodle* in '*kit and caboodle*'; *fro* in '*to and fro*').

Regarding what is involved in knowing either a L1 or a L2 word, Hulstijn (2001) indicated that "a lexical entry in the lexicon of the average adult, literate, native speaker contains semantic, pragmatic, stylistic, collocational, syntactic, categorical, morphological, phonological, articulatory and orthographic features" (p. 259). More simply, Nation (2013) suggested that knowing a word means having both receptive and productive knowledge of a word in three knowledge dimensions including form, meaning as well as use. He then proposed a model (see Figure 2.1) of what is involved in knowing a word.

Knowledge including being able to recognise and produce the spoken form, written form and word parts (e.g., the word *underdeveloped* is made up of the parts *under-*, *-develop-* and *–ed*) of a word are listed within the form dimension. Within the meaning dimension, knowing a word requires the ability to link the word form to its meaning and use the word form to express the meaning, linking the word meaning to its concept and producing the word according to its concept, associating the word with its synonyms and producing synonyms for the word. Regarding the use dimension, knowing a word requires being able to recognise and produce grammatical functions and collocations relating to this word as well as being able to recognise the frequency level of the word and to make a decision whether or not the use of the word is able to "suit the degree of formality of the situation" (p. 50).

Form	**spoken**	R What does the word sound like? P How is the word pronounced?
	written	R What does the word look like? P How is the word written and spelled?
	word parts	R What parts are recognisable in this word? P What word parts are needed to express the meaning?
Meaning	**form and meaning**	R What meaning does this word form signal? P What word form can be used to express this meaning?
	concept and referents	R What is included in the concept? P What items can the concept refer to?
	associations	R What other words does this make us think of? P What other words could we use instead of this one?
Use	**grammatical functions**	R In what patterns does the word occur? P In what patterns must we use this word?
	collocations	R What words or types of words occur with this one? P What words or types of words must we use with this one?
	constraints on use (register, frequency …)	R Where, when, and how often would we expect to meet this word? P Where, when, and how often can we use this word?

Figure 2.1 What is involved in knowing a word (Nation, 2013, p. 49)
Note. In column 3, R = receptive knowledge, P = productive knowledge.

Then, based on the distinction between the form-learning aspects and the meaning-learning aspects made by Ellis (1994), Nation (2013) proposed an overview (see Figure 2.2) of the correspondent kinds of learning and activities which could facilitate the learning of the three knowledge dimensions most effectively.

Kinds of knowledge		Kinds of learning	Activities
Form		implicit learning involving noticing	repeated meetings as in repeated reading
Meaning		strong explicit learning	depth of processing through the use of images, elaboration, deliberate inferencing
Use	**grammar collocation**	implicit learning	repetition
	constraints on use	explicit learning	explicit guidance and feedback

Figure 2.2 Kinds of vocabulary knowledge and the most effective kinds of learning (Nation, 2013, p. 58)

Ellis (1994) argued that learning both the receptive and productive knowledge of the form aspect of a word highly relies on "implicit" learning, which is "strongly" influenced by repetition, involving "attention to the stimulus but does not involve other conscious operations" (p. 212). The meaning aspect, however, relies on "explicit" and "conscious" learning, which involves "a search for rules, or applying given rules" and is "strongly affected by the quality of the mental processing" (Nation, 2013, p. 58-59). Therefore, within a classroom context, to what extent learning word meaning can be enhanced by intentional and explicit vocabulary instruction seems to be an interesting area to explore.

The above discussion has related the different aspects of vocabulary knowledge and the most effective types of learning modality; the following sections 2.2.2 and 2.2.3 will begin briefly with a consideration of some aspects of second language acquisition, especially regarding implicit and explicit knowledge, followed by more specifically identifying how a word can be learnt by presenting two theoretical frameworks: the revised hierarchical model of lexical and conceptual representation in bilingual memory (Kroll & Stewart, 1994) and a psycholinguistic model of vocabulary acquisition in a second language (Jiang, 2000).

2.2.2 Depth of processing

The levels or depth of processing hypothesis was proposed by Craik and Lockhart (1972) as a framework for human memory research. The basic notions of this hypothesis are that:

> Episodic memory trace may be thought of as a rather automatic by-product of operations carried out by the cognitive system and that the durability of the trace is a positive function of "depth" of processing, where depth refers to greater degrees of semantic involvement (Craik & Tulving, 1975, p. 268).

The depth of processing hypothesis is further linked with vocabulary learning by Nation (2013). Arguably difficult words which normally require more effort to study would lead to deeper processing, therefore are better retained. However, this interpretation needs to be treated with caution, as there is evidence showing that words with lower learning burden, that is with lower amount of learning effort required (Nation, 2001), are also easy to learn and can reach good levels of retention. In addition, certain types of lexical items (i.e., cognates) are generally easy to acquire and can be retained fairly well. In general, the depth of processing hypothesis can still apply to vocabulary learning, especially regarding the learning condition, whereby vocabulary items acquired "under conditions involving thoughtful processing" are likely to reach good retention (Nation, 2013, p. 114).

2.2.3 The noticing hypothesis

In response to the long existing debate of whether language learning is a conscious or unconscious process, Schmidt (1990) argued that "the notion of consciousness is useful because it ties together such related concepts as attention, short term memory, control vs. automatic processing, and serial vs. parallel processing" (p. 131). The argument that Schmidt made is based on the two case studies that he conducted. The first is a Japanese adult English learner 'Wes' who migrated to the U. S. and began to learn English as a second language when he was already an adult. Schmidt observed that although Wes had developed a variety of language expertise which qualified him as a good language learner, he was persistently making mistakes on certain English grammatical structures. While

certain affective factors may not have explained why Wes' grammatical knowledge was much lower than his other language knowledge as he was clearly motivated in language learning, Schmidt (2012) assumed that there might be two reasons: One was "lack of aptitude", and the other was "over-reliance on an implicit learning strategy, learning through interaction alone, with little attention to language form and little conscious reflection about language structure" (p. 29).

Another possible reason for why Wes did not make much improvement in his English grammar may be due to the fact that adults and young learners learn language differently. Dekeyser (2003) argues that young language learners, who mainly relying on implicit learning from much contact and input in a naturalistic context, can make better progress in areas of language knowledge such as grammar which are difficult to acquire explicitly. By contrast, older learners, although having greater cognitive capacity, are already dominated by their native language. Unlike young learners who are able to implicitly learn from the context, older learners rely heavily on explicit learning from the instructed classroom context where large amounts of naturalistic input seem to be impossible.

The second case is based on Schmidt's own experience of learning Portuguese (Schmidt & Frota, 1986) during his five months' stay in Brazil where he had taken a five-week Portuguese class. The results showed that there was clearly advantages of classroom instruction for L2 learning that involved experiencing target language input to a large extent. However, based on comparing notes he made in class, class recordings, and tape-recorded L2 production and use after class, the authors found that certain L2 forms, although with highly frequent exposure, were not acquired unless these forms were consciously noticed and registered in the L2 input. A possible reason for this could be that instructed acquisition, even if it includes lots of target language input, is not the same as natural acquisition and that the former needs a more conscious, explicit approach.

Therefore, driven by these two case studies, Schmidt (1990, 1994) proposed the 'noticing' hypothesis which indicates that "input does not become intake for language learning unless

it is noticed, that is, consciously registered" (Schmidt, 2012, p. 27). The hypothesis was then framed in more detail as "consciousness as intention, consciousness as attention, and consciousness as awareness" (p. 30).

Regarding consciousness as intention, Schmidt (2012) argues that although incidental learning (e.g., people acquire vocabulary unconsciously through extensive reading) is useful and effective, intentionally and consciously paying attention to cues "that are not salient" and cues that make the learner aware of language features "that need to be processed differently from the way they are in the first language" (p. 30) is also necessary. Consciousness as attention "suggests that attention must be directed to whatever evidence is relevant for a particular learning domain, that is, that attention must be specifically focused and not just global" (p. 31). Schmidt argues that this consciousness as intention can also apply to vocabulary acquisition, whereby learners need to attend to both the word form and any possible meaning driven cues in the input.

Finally, consciousness as awareness indicates that awareness in language learning is necessary as the assumption that awareness and attention are the two sides of the same coin (Carr & Curran, 1994; Posner, 1994; Schmidt, 2001) makes it very difficult to separate these two. However, there are differences between instructed and natural language acquisition. "Awareness of abstract rules of grammar cannot be a prerequisite for learning, since native speakers clearly have some intuitive understanding of subtle points of grammar that they cannot verbalize" (Schmidt, 2012, p. 31), and this may also be true with some advanced L2 learners. In order to address these differences between instructed and natural acquisition, three hierarchical levels of awareness were proposed: Perception, Noticing and Understanding (Schmidt, 1990, p. 132). Schmidt (2012) argues that it is essential and certainly necessary for L2 learners to reach Noticing, the second level of awareness, where they can develop "conscious registration of attended specific instances of language" (p. 32). The higher level of Understanding where learners are required to generalise instances of language, however, is not necessary.

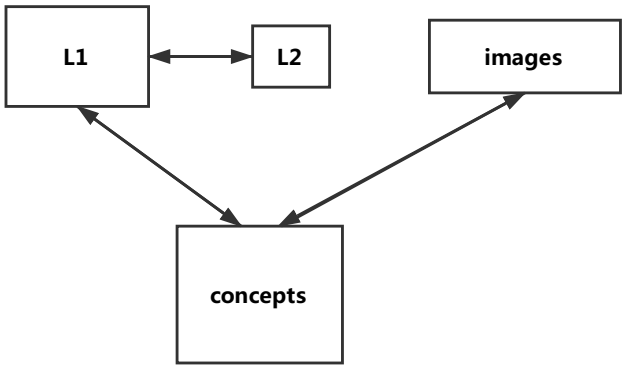
The ‘noticing’ hypothesis has, however, been mainly challenged by the argument that the hypothesis may suffer from a few methodological limitations because it was driven by two case studies (Tomlin & Villa, 1994). In addition, some kinds of learning do not require input, therefore do not need learners to be aware or to have attended (Gass, 1997). Finally, conscious registration with the input cannot be directly linked to acquisition as for L2 learning, much linguistic knowledge is not in the input (Carroll, 2006). However, the noticing hypothesis is still beneficial and has had a positive pedagogical impact on both second language and foreign language learning (Schmidt, 2012) as the hypothesis is supported by many empirical studies using a variety of methods including retrospective reports, stimulated recall interviews, and online measures such as note-taking, think-aloud protocols and eye-tracking (Godfroid, 2010; Godfroid, Housen, & Boers, 2010). In addition, the premise for the ‘noticing’ hypothesis is that in order to intake certain types of language input, the input should be recognised and attended to in the first place. The ‘noticing’ hypothesis, rather than being wrong, is not relevant to learning linguistic knowledge without input.

2.2.4 A theoretical framework for vocabulary learning

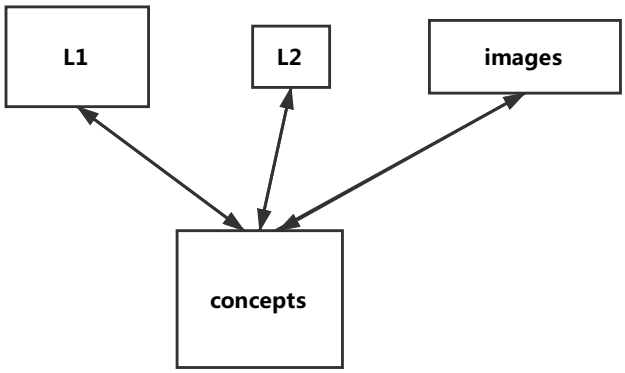
2.2.4.1 The revised hierarchical model of lexical and conceptual representation in bilingual memory

Before moving onto investigating how exactly L2 lexical knowledge may be represented in the mental lexicon from a theoretical viewpoint, it is worth looking at how early bilinguals process their L1 and L2 in general and to what extent are the two languages they acquired at an early stage of their life interconnected with each other. In order to address these two questions, it is important to understand the structure of bilingual memory (Kroll & Stewart, 1994). There has long been debate regarding whether the two languages that a fluent bilingual has acquired share a common memory system or are processed independently in two memory systems which correspond to each language (Kroll & Stewart, 1994; Snodgrass, 1984). Several language processing models (Kroll & Stewart, 1994; Potter, So, von Eckhardt, & Feldman, 1984) have been consequently proposed in order to address the issue whether bilinguals have a shared common memory system or two separate memory systems, and, more importantly, to provide initial insights into how bilinguals' two languages are interconnected.

The hierarchical model (see Figure 2.3) proposed by Potter et al. (1984) assumes that there are two levels of representation in bilingual memory which are hierarchically related. At the word association level, L2 words are strongly associated with their L1 translations, whereby no direct link is provided between L2 words and their conceptual representations. The processing of an L2 word has to go through its L1 translation before gaining access to the concepts. In contrast, the concept mediation level assumes that both words in the bilingual's two languages can directly link to the concepts.



Word association



Concept mediation

Figure 2.3 The hierarchical model of lexical and conceptual representation in bilingual memory (Kroll & Stewart, 1994, p. 150, adapted from Potter et al., 1984)

Potter et al. (1984) then conducted two experiments to test the two alternative levels of representation, one with a group of more-fluent Chinese-English bilinguals and the other with a group of less-fluent English-French bilinguals. Two main tasks were used in both the two experiments: translation of an L1 word into its corresponding L2 word and using an L2 word to name a picture. With reference to the above two alternative levels of representation (word association level and concept mediation level), Potter et al. (1984) hypothesised that the two tasks would require these bilinguals to go through different stages of processing and therefore result in different reaction times. Details for this are given in Figure 2.4.

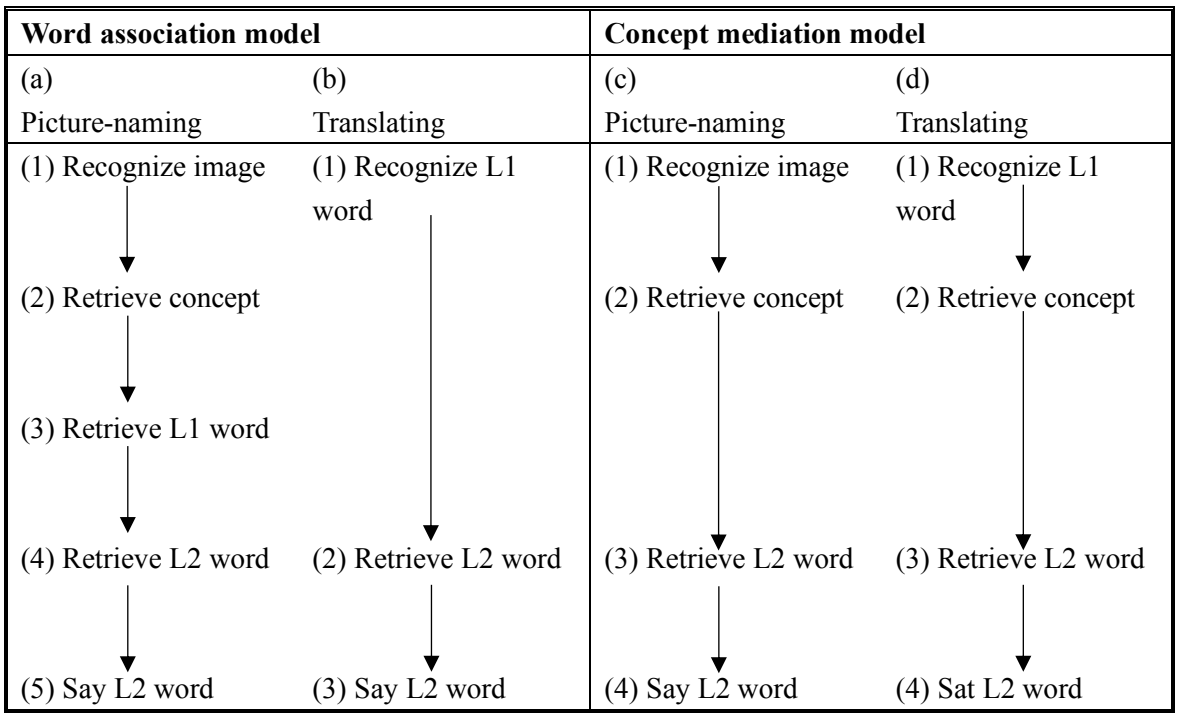


Figure 2.4 Two hypotheses about the processing sequence leading to production of a second language word in response to a picture or first language word (Potter et al., p. 26)

It can be seen from Figure 2.4 that if the word association model is correct, reaction times would be faster for translating into the L2 than picture naming in L2 as the word association model assumes that there is a direct link between bilinguals' two languages, therefore translating would bypass retrieving the concept. For the concept mediation model, however, the reaction time of translating would be approximately similar to that of picture-naming. Both the two tasks require bilinguals to retrieve concepts before they retrieve the corresponding L2 words.

Results indicated that both the more-fluent and less-fluent bilinguals had similar reaction times for doing both the two tasks, which supports the concept mediation model while contradicting the word association model. The results, however, were challenged by several other studies (Chen & Leung, 1989; Kroll & Curley, 1988) regarding whether the less-fluent English-French bilinguals in Potter et al's. (1984) study were indeed 'less-fluent' or they were just comparatively less fluent compared to the highly fluent bilinguals in their study. It is likely that these bilinguals, although less-fluent, may have passed the early L2 learning stage where their L2 is lexically associated with the L1.

Therefore, to test this idea further, Potter et al's. (1984) study was replicated by Kroll and Curley (1988). A broader range of bilingual participants were selected including those who had fewer than two years' experience of learning the L2. Results revealed that there were clear differences between the performance of the more-fluent and less-fluent bilinguals. While the former produced similar reaction times when completing the two tasks indicating that each of the bilinguals' two languages is directly represented in the concept, the latter reacted much faster when translating the L1 word into the corresponding L2 word than naming pictures in the L2, which confirms the word association hypothesis. The same results were also found by Cheng and Leung (1989).

However, there is still one underlining problem with the concept mediation hypothesis. In Kroll and Curley's (1988) study, they ran an additional experiment introducing a categorical factor. Participants were required to complete three tasks both in L1 and L2: naming words, translating words and naming pictures. For each task, they included two categories of target items, one semantically categorised (e.g., *suit, jacket, scarf*), the other semantically randomised (e.g., *lemon, dog, shoes*). They hypothesised that since in the fluent bilinguals, language is concept mediated, their performance would be facilitated if the words or the pictures were semantically categorised, which means they would react faster to items which are semantically categorised than those that are not.

The results were, however, surprisingly the opposite. These more-fluent bilinguals reacted more slowly when translating words into the L2 and naming pictures in the L1 when these were semantically categorised. Kroll and Stewart (1994) argued that "repeated access to concepts within the same semantic category should increase activation at the conceptual level and produce corresponding activation at the lexical level" and this "additional activation should produce competition among close alternatives, and hence interference, rather than facilitation" (p. 155).

In addition, Kroll and Stewart (1994) raised another issue about bilingual translation, whereby it is usually quicker for bilinguals to translate from L2 to L1 than from L1 to L2.

They hypothesised that translation from L2 to L1 is lexically associated which bypasses the retrieving of the concept, while translation from L1 to L2 requires concept mediation. This so-called "translation asymmetry" (p. 157) was, however, not revealed in either the word association or the concept mediation model mentioned above, suggesting that the models needed to be modified. Therefore, they proposed a revised hierarchical model of lexical and conceptual representation in bilingual memory (see Figure 2.5).

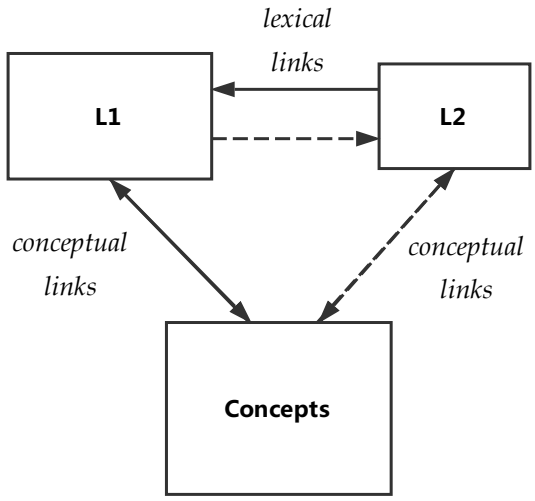


Figure 2.5 The revised hierarchical model of lexical and conceptual representation in bilingual memory (Kroll & Stewart, 1994, p. 158)

The model assumes that even with highly fluent bilinguals, L1 is still the dominant language, as a strong link between L1 lexical knowledge and its conceptual representations is already created when a language learner starts learning the L2 beyond early childhood (Kroll & Stewart, 1994, p. 158). At the initial stage of language learning, L2 words are "attached to this system by lexical links" with the L1. Direct conceptual links between L2 words and conceptual representations can also be created when further exposure is reached and bilinguals are becoming more fluent in L2. The lexical links from L2 to L1, however, do not disappear even if direct links are made between L2 and the concepts. In addition, according to this model, "both lexical and conceptual links are bidirectional", yet are different in strength. Since L2 words are initially associated with their correspondent L1 words, the lexical link from L2 to L1 is supposed to be stronger than from L1 to L2. Similarly, the link between L1 and its conceptual representations is stronger than the link

between L2 and its concepts.

Kroll and Stewart (1994) then designed an experiment to test the model and these hypotheses. The participants were 24 university students, who were highly fluent Dutch-English bilinguals. Although Dutch was their L1 and dominant language, they were rated in English speaking and reading to confirm their fluency in English. Two main tasks were used: word naming in L1 and in L2, and word translation from L1 to L2 and from L2 to L1. For both the two tasks, they also employed a categorical factor, whereby participants were presented with word lists which were either semantically categorised or randomised. An incidental vocabulary recall test took place at the end of the experiment.

The results confirmed what was predicted in the model. Initially, the participants named the Dutch (L1) words significantly faster than the English (L2) words, which indicates that although highly fluent in English, Dutch was still the dominant language for these bilinguals. Secondly, it took the participants significantly longer to translate from L1 to L2 than from L2 to L1 in general. When translating from L2 to L1, the reaction times for semantically categorised word lists and semantically randomised word lists were not significantly different, whereas for translating from L1 to L2, these bilinguals were significantly slower when they were presented with categorised word lists than with randomised word lists. This confirms the model's predictions that translating from L2 to L1 takes place via word association, while translating from L1 to L2 requires concept mediation.

The revised hierarchical model was, however, challenged by Brysbaert and Duyck (2010) who questioned the model from several different perspectives. Firstly, the direct empirical evidence for bilinguals' separate L1 and L2 lexicons and for their language selective access is quite limited. Secondly, there is evidence against the assumption predicted by the model that the lexical link is stronger from L2 to L1 than from L1 to L2. Thirdly, the direct link between L2 words and the concepts is stronger than what was predicted in the revised hierarchical model. Finally, it is still questionable whether all conceptual information is

language independent as proposed by the model. However, Kroll, van Hell, Tokowicz and Green (2010) pointed out that Brysbaert and Duyck may have "lost sight of the larger picture" (p. 379). The revised hierarchical model was intended to address "the way in which new lexical forms are mapped to meaning and the consequence of language learning history for lexical processing", which "cannot be accounted for solely within models of word recognition" (p. 373).

In summary, the revised hierarchical model merged the two alternative models proposed by Potter et al. (1984) into a single model of language processing, providing a clear picture of the structure of bilingual memory and how bilinguals process their two languages. In general, the model emphasised the important role of the L1 for L2 word processing, which shed light on the value of using the L1 for facilitating the acquisition of L2.

2.2.4.2 Jiang's (2000) psycholinguistic model of vocabulary acquisition in a second language

From a similar theoretical viewpoint, Jiang (2000) proposed a psycholinguistic model of vocabulary learning in L2, which hypothesised how L2 lexical knowledge is represented in the mental lexicon. Most research investigating L2 vocabulary learning tends to focus more on the processing and the learning, while ignoring how actually L2 vocabulary knowledge is represented in learners' mental lexicon, which is inevitably important within the whole learning process (Jiang, 2000). However, before moving onto the L2 lexical representation, it would be worthwhile exploring what exactly L1 vocabulary representation is like.

Drawing on Levelt (1989), Jiang (2000) proposed that L1 lexical representation contains knowledge of: "semantic, syntactic, morphological, and formal (phonological and orthographic) specifications about a lexical item" (p. 48), which can be divided into two components: the lemma and the lexeme. Lexical knowledge of morphology, phonology and orthography including morphological variants, spelling as well as pronunciation of a

word is represented in the lexeme component, while semantic and syntactic information of a certain word, e.g., meaning and part of speech, is represented in the lemma component (Figure 2.6).

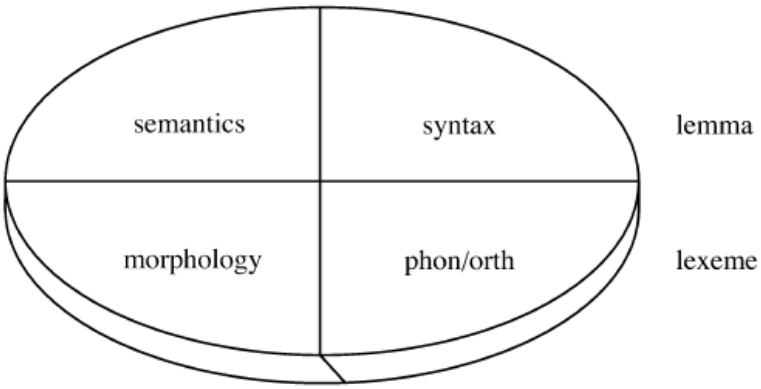


Figure 2.6 The internal structure of the lexical entry (Jiang, 2000, p. 48)

Jiang (2000) argued that, although L1 lexical representation includes different types of knowledge, these are strongly integrated with each other, so that L1 lexical representation is more of a process in which all types of knowledge are represented as a whole in L1 learners' mental lexicon. However, two constraints mean that L2 lexical representation strongly differs from that of L1. Firstly, compared to L1 learning, target language input for L2 learning is often insufficient and decontextualised, which suggests that it is quite difficult for L2 learners to integrate these different types of knowledge and represent them as a whole. Secondly, an L2 lexical entry may be strongly influenced by the existing L1 lexical system. Jiang (2000) indicated that L2 learners, especially adult L2 learners, may strongly rely on the existing L1 lexical system when learning L2 vocabulary. Hence, he then proposed three stages for the representation and processing of L2 lexical knowledge (see Figure 2.7).

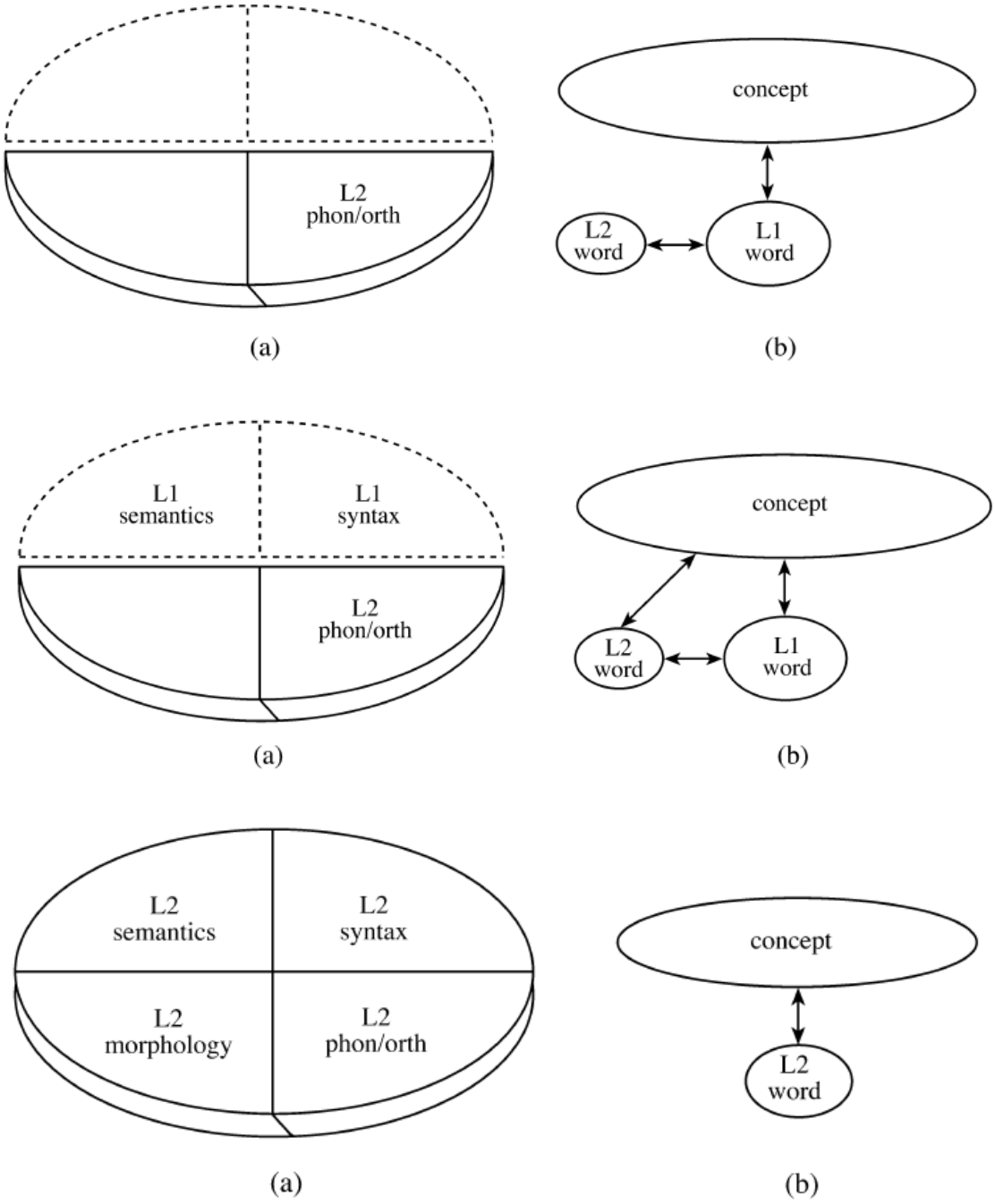


Figure 2.7 Three stages for L2 lexical representation and processing (Jiang, 2000, p. 51-53)

At the initial stage (formal stage), only the formal information (phonetic, form/spelling) of the target L2 word is represented in learners' minds, and the processing of the L2 word relies heavily on L1 translation. At the second stage (lemma-mediation), as learners' experience in the L2 develops, the link between the L2 word and its L1 translation is developed further: The L1 lemma information is attached to the L2 word forms. Jiang (2000) pointed out that the link between the L2 word and its conceptual representations is

fairly weak because "the lemma information is copied from L1, rather than created in the process of learning the L2 words, thus not highly integrated into the entry" (p. 52). At the final stage (L2 integration stage), when further exposure and use is achieved, both L2 lemma and lexeme are highly integrated and represented in L2 learners' mental lexicon.

This psycholinguistic model indicates the importance of L1 mediation within the process of L2 vocabulary learning, which has been confirmed by several empirical studies conducted by Jiang (2002, 2004a). The two similar studies involved a group of ESL learners and a group of English native speakers. Both groups were asked to complete a relevant semantic judgement task, in which they should judge whether the presented English word-pairs were related in meaning. Within the ESL group, learners responded significantly faster to the English word-pairs which had the same L1 translation (e.g., *problem – question*) than to those pairs which had different L1 translation (e.g., *painter – artist*). However, among native speakers, no such difference was confirmed. This can be regarded as an initial piece of evidence indicating that L1 mediation does exist, as was predicted in the L1 mediation stage that the L2 semantic information is highly copied from its L1 translation. Therefore, when participants were performing judgement tasks, L2 learners reacted faster to semantically-related word pairs than to word pairs which were less related or not semantically related.

More direct evidence of L1 as a mediation was further confirmed by Jiang (2004b) through a revised version of the sentence completion task used by Ijaz (1986). Participants were two groups of university students, one group including ten advanced Chinese ESL learners and another group including ten English native speakers. Six pairs of English words were selected. In each pair, the two words seemed to have the same Chinese translation, yet actually have different meanings for native English speakers (*doubt-suspect, insist-persist, safe-secure, accurate-precise, criterion-standard, complicated-complex*). Different sets of sentences corresponding to different word pairs were created. Within each set of sentences, each sentence could only be appropriately completed with one specific word from each word pair.

The ESL participants were also asked to rate the level of difficulty of the task itself from 1 (very easy) to 7 (very difficult) and were required to provide a brief written description in Chinese to tell the difference between the two words in each word-pair. Results show that native English speakers, reaching an average accuracy of 94%, significantly outperformed those Chinese ESL learners who only reached 65% in accuracy. In addition, results from the written descriptions indicated that in general, these ESL learners felt that it was difficult for them to describe the differences between the two words in each word-pair (seven out of 60 instances were successfully described). Moreover, there were differences in the level of difficulty to distinguish the two words within each pair. The rating results showed that two pairs (*doubt-suspect* and *insist-persist*) were easier than the other pairs. In general, results from these tasks indicated that these ESL learners were experiencing difficulties in distinguishing between words which are semantically-related in their L1. Hence, more direct evidence of L1 mediation was obtained.

However, Jiang's argument and empirical evidence of L1 as a mediation for L2 vocabulary processing does not equate to proposing that there is pedagogical value in using L1 for L2 vocabulary teaching (Tian, 2011). There is still a shortage of empirical studies to provide concrete evidence for the effectiveness of L1 use for L2 vocabulary teaching (Cook, 2001). Moreover, regarding using L1 for L2 instruction, two other problems may occur. First, since both plentiful and contextualised L2 input is needed for learners' L2 lexical representation, how much L1 for instructional purpose would be appropriate? Second, research results from Jiang (2004b) indirectly indicated that L1 use for L2 vocabulary instruction, such as directly providing L1 translation of a target L2 word, can sometimes cause problems, especially in teaching words of near synonymy. Thus, he argued that two conditions should be met for L1 use to be appropriate (p. 118):

> There is an indication of a mismatch between an L2 word and its L1 translation in meaning. Such an indication can be explicit, as in the form of overt correction from an instructor or interlocutor. Or it can be less direct, e.g. in an expression of confusion on the part of an interlocutor, or any other sign of unsuccessful communication.

> Information is available regarding exactly how the L2 word is semantically different from the L1 concept. This information may come from direct correction and explanation, or from context.

In addition, research evidence shows that L2 learners' lexical representation may stop at the second stage of the model even if they are exposed to L2 input of both high quality and high quantity. This may arise from "lexical fossilization" (Jiang, 2000, p. 54), whereby learners do not move from the second stage (L1 mediation) to the third stage (the L2 integration stage) suggested in the theory, potentially because "the presence of L1 lemma within the L2 lexical entry may block the integration of L2 lemma information" (Jiang, 2000, p. 55). This prompts the question, if the processing is stopped at the second stage, would clear contrasts between L1 and L2 lemma provided within lexical instruction motivate the processing to reach the third stage? More importantly, considering the possible limitations direct L1 translation may have, what kind of L1 information should be provided?

2.2.5 Incidental vocabulary learning

Having identified explanations for how bilinguals process their two languages and how L2 vocabulary knowledge represented in the mental lexicon, it is valuable now to explore possible ways for L2 vocabulary learning to occur from a pedagogical point of view.

2.2.5.1 Incidental vocabulary learning through reading

A definition given by Swanborn and de Glopper (1999) directly associated incidental vocabulary learning with natural reading, indicating that incidental vocabulary learning which happened under the modality of natural reading was a process of obtaining and retaining vocabulary meanings without noticing or focusing on any actual vocabulary. Here, natural reading modalities should be 1) reading situations which were familiar to language learners, 2) reading circumstances both inside and outside classroom and 3) adopting reading materials which were authentic. This definition provided by Swanborn

and de Glopper is, however, problematic, especially with reference to the noticing hypothesis (Schmidt, 1990, 1994) whereby in order to acquire certain types of language input, the input should be noticed. Therefore, incidental vocabulary learning can only happen on the basis that the unknown vocabulary items being noticed in the first place. A more appropriate definition for incidental vocabulary learning was provided by Laufer (2003) with the purpose of exploring some empirical evidence of whether most vocabulary could be acquired through reading, who argued that incidental vocabulary learning was more likely acquiring vocabulary "as a by-product of any activity not explicitly geared to lexical acquisition" (p. 574).

According to Schouten-van Parreren (1989), two different categories of target L2 vocabulary exist, the thousand most frequent words and less frequent words, with both contributing to vocabulary size. While many methods may be employed (e.g., short reading materials, listening texts, visual aids and real objects) to learn the most frequent words, less frequent words are likely to be more effectively acquired by extensive and intensive reading. In greater depth, Schouten-van Parreren (1989) put forward several theoretical arguments for why learners acquire less frequent vocabulary more successfully through reading meaningful texts than they do through more traditional methods, such as explicit vocabulary teaching or presenting lexical items in isolated sentences. First, if the lexical items are given as isolated items or in an alphabetically ordered word list, learners may find it difficult to remember these words or suffer from the process of memorising itself, as there are no contexts for these words so that the learning involvement load (Laufer & Hulstijn, 2001) could be very low. Second, isolated lexical items may lack linguistic reality since the specific meaning of a target lexical item is often accurately defined by the context in which it is embedded.

Slightly different from Schouten-van Parreren's view, four hypotheses were summarised by Laufer (2003) to identify how the process of incidental vocabulary learning occurs, including 'the noticing assumption', 'the guessing ability assumption', 'the guessing-retention link assumption' as well as 'the cumulative gain assumption' (p. 568). Initially,

language learners are assumed to spontaneously pay attention to unfamiliar vocabulary items while reading, which is then followed by actively inferring vocabulary meaning through both linguistic (contextualisation) and non-linguistic (world knowledge) clues. Moreover, with the help of successful meaning guessing, language learners may have certain possibilities, although quite low, to retain target new words. Finally, through repetitive exposure to certain lexical items, the outcomes of vocabulary retention are likely to be improved.

Considering whether L2 learners acquire much vocabulary incidentally through reading alone, Laufer (2003) illustrated several limitations within the learning process by challenging the four above assumptions. Firstly, the noticing assumption, in most cases, may not be reliable especially when reading and completing comprehension questions (Laufer, 2003). On the one hand, learners are likely to concentrate on the overall meaning of the given text in order to complete comprehension questions while overlooking any potential target words involved as well as their meanings (Min, 2008; Pigada & Schmitt, 2006). On the other hand, ignorance of new lexical items may occur because they are confused with other familiar ones which the learners have already acquired.

Secondly, the guessing assumption should not be taken for granted as it may highly depend on how much unknown information is included in the reading passage and whether the reading passage is confusing (Laufer, 2003). If the reading passage contains mostly unfamiliar words and challenges learners' knowledge of the world, or if no linguistic clues can be clearly identified, the possibility of successfully inferring the precise meaning of an individual word can be relatively low (Swanborn & de Glopper, 1999). In addition, compared with L1 learners who have knowledge of most vocabulary included in the text and can make good use of existing linguistic clues, L2 learners may have fewer opportunities to succeed.

Thirdly, counter-arguments have been constantly raised against "the effectiveness" (Min, 2008, p. 74) of incidental vocabulary learning through reading or whether the meaning of a

target word can be successfully guessed and whether there is a clear link between the success of the guess and the possibility of retaining the word (Laufer, 2003; Pigada & Schmitt, 2006). Initially, guessing may lead to faulty assumptions about the meanings of some words, which can result in ineffective vocabulary learning. Moreover, the possibility of long-term retention of a target word is largely reduced if the word meaning is easy to guess (Laufer, 2003). As a result, reading materials with moderate difficulty should be adopted in order to obtain maximum lexical retention. However, with regard to the limitations of the noticing and guessing assumption, effectiveness may not be guaranteed.

Finally, there is evidence that repeated exposures to a target word can have a positive influence on vocabulary retention (Laufer, 2003; Pigada & Schmitt, 2006), the ideal frequency of occurrence ranging from eight to ten (Horst, Cobb, & Meara, 1998; Nation & Wang, 1999; Pellicer-Sánchez, 2016; Pigada & Schmitt, 2006; Webb, 2007), yet "the efficiency" (Min, 2008, p. 74), in other words, how much exposure is adequate is still a matter of discussion. In addition, to maximise the frequency of encountering a certain word as well as to ensure the authenticity of target reading materials, L2 learners may be required to read a comparatively large number of texts, which is time-consuming and impractical in L2 classrooms.

2.2.5.2 Incidental vocabulary learning through listening

While several studies of L2 or foreign language vocabulary learning have been carried out in relation to reading (Horst, 2005; Pellicer-Sánchez, 2016; Zahar, Cobb, & Spada, 2001), there has been much less focus on how listening or oral texts may assist language learning or, to be more precise, lexical learning (Vidal, 2003, 2011). In order to address this, it seems to be vital to firstly identify how exactly the process of lexical learning works in relation to listening. Two phenomena, phonological memory (Service, 1992) and phonological loop (Baddeley, 1997), are important for this question. Initially, a target word is stored in short-term or working memory for quite a limited time period (Gathercole & Baddeley, 1993) when it is heard for the first time. Then, if the same word is rehearsed or

artificially repeated on several occasions, it is likely to be stored in long-term memory and eventually retained. However, this theory has been concluded from highly experimental situations (Vidal, 2003), where lexical items were emphasised during several rehearsal periods. Hence, it is still unclear whether this process can be replicated in natural communicative contexts or whether vocabulary can be acquired incidentally through listening texts. Moreover, considering the information stored in the working memory is quite limited and may last only for two seconds (Baddeley, 1997), even less information may be retained in the long-term memory for future retrieval (Tian, 2011).

To provide empirical evidence for whether learners could acquire vocabulary incidentally through listening, Vidal (2003) investigated incidental vocabulary learning through listening to EFL academic lectures. Within a period of four months, a group of learners listened to three academic lectures on different topics. Each lecture included 12 target vocabulary items. They were asked to focus on listening comprehension and were required to complete relevant comprehension tasks immediately after listening. At the same time, a vocabulary knowledge pre-test was conducted prior to the first listening session and there was a vocabulary post-test at the end of each session. An additional vocabulary delayed post-test was administered four weeks after the final listening comprehension session.

Vidal (2003) used a modified version of Paribakht and Wesche's (1993) vocabulary knowledge scale (VKS) to rate the vocabulary tests. Results indicated that the mean score of the immediately vocabulary post-test (30.41) was significantly higher than that of the pre-test (1.41), which revealed that learners can acquire a certain amount of vocabulary knowledge incidentally through listening to academic lectures. In addition, results from the delayed post-test confirmed that between 43% – 54% of the vocabulary knowledge gained at the post-test was retained four weeks after the final listening comprehension session. Moreover, findings also suggested that learners' performance in both short-term and long-term vocabulary learning outcomes can be influenced by their English language proficiency levels. Although higher proficiency level learners acquired more vocabulary knowledge than lower level learners at the immediate post-test, these higher level learners

forgot a larger proportion of the vocabulary knowledge gained four weeks later than those lower level participants whose vocabulary loss was relatively small.

It seems that incidental meaning recall enhanced either by listening or by reading is fairly low, yet even less meaning retention was confirmed through listening than through reading. For example, Brown, Waring and Donkaewbua (2008), in a small study of 35 pre to intermediate Japanese learners of English, found that while learners retained around four out of 28 words from texts in a reading plus listening modality or a reading-only format, only 0.56 of 28 words were retained from a text in a listening-only format, from which none were retained three months later.

However, a notable point is that in another study (Vidal, 2011), compared with vocabulary knowledge incidentally acquired through reading, knowledge learned from listening tended to have better long-term retention. Undergraduate students who registered at an English for specific purposes (ESP) course and whose major was tourism participated in the study. These students were from three intact classes and these classes were randomly assigned to a listening treatment condition (112 students), a reading treatment condition (80 students) as well as a control group (38 students). The TOFEL scores of the three groups were obtained and were used to allocate the participants in each group into four different proficiency levels for data analysis.

Three video recorded academic lectures and three correspondent readings were used for the intervention. Both the listening and the reading materials were selected from the same authentic source, with topics relevant to the participants' major. They were then modified to ensure that the materials used for both the two conditions contained the same target lexical items and the length and task difficulty of the materials within each treatment condition were equal. Altogether 36 target words were selected. In order to look at whether the different number of repetitions would have any impact on incidental vocabulary learning, the frequency of occurrence for the target words was controlled to be constant between the two treatment conditions ranging from one occurrence to six occurrences.

The participants were tested on their knowledge about the target items before the first intervention session. Three intervention sessions were administered within four weeks. During each session, the participants were either given 15 minutes to read one target reading material in the reading group or were required to listen to a 15 minute academic lecture if they were in the listening condition. They were then asked to complete corresponding comprehension questions to make sure their attention was on reading or listening comprehension. Immediately afterwards, they were tested on their knowledge about the target lexical items. A vocabulary delayed post-test was conducted one month after the final intervention session for both the two treatment groups. The control group, although they did not receive any reading or listening input, also received the three vocabulary tests at the same time points as the two intervention groups. The results from the control group were used to control for the confounding effect of repeated testing.

Results showed that the two treatment groups (both listening and reading) made significant gains from the pre-test to the immediate post-test as well as from the pre-test to the delayed post-test, which indicates that vocabulary knowledge can be acquired incidentally through reading or listening. The vocabulary test results for the control group, however, were not significantly different between any of the two test time points, showing that participants did not learn vocabulary just by taking the tests. In addition, at the immediate post-test, the vocabulary gains for the reading group were significantly higher than those for the listening group. Adding in the proficiency levels as a within-subject factor, results indicated that within each of the four proficiency levels, the reading group still outperformed the listening group, meaning that both high and low proficiency level learners acquired more vocabulary knowledge incidentally through reading than through listening.

The results for the delayed post-test were quite similar to the post-test. The vocabulary retention for the reading group was generally higher compared with the listening group. However, when looking more specifically at the data for each proficiency level, although vocabulary retention was better for reading than for listening among three lower-level

learners, there were no significant differences of the four weeks' vocabulary retention between the higher-level readers and listeners. As argued by Vidal (2011), it seems that "for higher proficiency students, vocabulary acquired through listening is more resistant to forgetting" (p. 238), possibly because for them such vocabulary may pass immediately into phonological memory. This then, according to Vidal (2011) leads to "stable long-term memory representations" (p. 244-245).

Regarding the impact of different repetitions on vocabulary acquisition, the two treatment conditions showed different patterns. On the one hand, for the reading group, vocabulary gains for words with three repetitions were much greater than for those with two repetitions. The gains were then facilitated gradually when the frequency of occurrence increased from three to five, yet there seemed to be no difference in gains between words receiving five and six exposures. On the other hand, in the listening condition, vocabulary gains started to increase gradually from words which received three repetitions to those which received five repetitions. The gains, however, grew rapidly from five exposures to six exposures. In addition, Vidal (2011) did a regression analysis to explore how much variance of the vocabulary gains can be explained by the factors including: type of word, frequency of word occurrence, type of word elaboration, and predictability from word form and parts. Results showed frequency of occurrence was the most important factor affecting the vocabulary gains in reading, explaining around 47% of the variance. However, in the listening context, frequency of occurrence was the least important factor, which only explained 24% of the variation in vocabulary gains. These findings suggest that perhaps more so than for incidental vocabulary learning through listening, words acquired incidentally through reading are more likely to be influenced by the different number of repetitions.

Three types of vocabulary knowledge were explored in a study by van Zeeland and Schmitt (2013a) which concentrated on incidental vocabulary learning through authentic L2 listening. They also explored possible correlations between lexical knowledge retention of a target word and its frequency of occurrence. Thirty postgraduate students (different

learning background, different L1 and mixed gender) at a UK university participated in this research, among whom 20 received listening input and completed an immediate post-test, while the other ten received the same listening input but completed a two-week delayed post-test. Four types of authentic listening materials including TV talk shows, TV interviews and informal lectures were selected, yet were amended by appropriately inserting target vocabulary to meet the purpose of different frequency of occurrence. 24 target lexical items were selected. These 24 items were then arranged into four occurrence categories (3, 7, 11, 15), six items in each occurrence category. Three tests were employed with a form recognition test at the beginning, followed by a grammar recognition test, and a meaning recall test at the end. The form recognition test required participants to write down relevant words while listening to the recording. In the grammar recognition test, on hearing the target vocabulary, students were asked to identify its word type (noun, verb and adjective). The meaning recall test simply asked learners to provide the meaning of a target item while a sentence including this item was heard.

The results showed that 29% of all 24 items were learnt immediately after the listening input and 19% of them were retained after a two-week delay. The highest level of knowledge gained was of word forms, followed by grammar knowledge, while least knowledge of meaning was obtained. However, most knowledge of form and grammar was lost after two weeks while most knowledge of meaning was retained after two weeks, which was the most difficult knowledge to acquire initially. In addition, although there was evidence showing short-term meaning acquisition of words which received 11 repetitions was significantly higher than those receiving only three or seven repetitions, this advantaged disappeared at two weeks' delayed post-test. Moreover, this research also explored how vocabulary knowledge learning differs according to word types, indicating that nouns receive the best retention for all three types of knowledge. Verbs received the same learning results for both form and meaning knowledge as adjectives did, yet slightly higher grammar knowledge retention than adjectives.

There are however limitations associated with the three studies reviewed above. The early study by Vidal (2003) did provide evidence for incidental vocabulary learning through listening. However, the amount of short-term learning and long-term retention for vocabulary meaning was still fairly small. Besides, the target lexical items tested included those in two categories, namely items which are "similar to Spanish" (the L1 of the participants in this study) and those which are "morphologically predictable" (p. 64). As the L1 for all participants in this study was Spanish, lexical items within these two categories either had corresponding L1 cognates or were easy to link to words which were familiar to the learners, so that it is likely that these items were those retained better by the learners. It is not clear, however, whether in instances where the learners' L1 is strongly different from the target language, vocabulary can still be learnt incidentally through listening.

From a broader view, Vidal's (2011) study more specifically compared the effects of reading input and listening input on incidental vocabulary learning, indicating that, in general, both short-term and long-term vocabulary gains were lower for listening than those for reading. In addition, the study revealed that more proficient learners tended to be more favoured by the listening condition as their long-term vocabulary retention was approximately the same as the higher-proficiency level learners in the reading group, although they were significantly outperformed by the latter at the immediate post-test. Moreover, results showed that incidental vocabulary learning through reading was more likely to be influenced by the frequency of occurrence, a factor which made less contribution to the listening condition.

Yet as argued by Vidal (2011), the learning for both the two conditions was fairly low regarding the vocabulary long-term retention. Future research needs to be implemented over a longer period of study, allowing participants to receive a higher amount of input in order to reach better learning outcomes. Another limitation is that although the two treatment groups experienced two different types of input, vocabulary knowledge was assessed through reading. It is likely that the reading group might be more favoured by this

test modality. Finally, it is questionable whether the target words were truly incidentally acquired as some of them had either implicit or explicit elaborations in the input.

Van Zeeland and Schmitt (2013a) identified several limitations for their research. First, the study only had a fairly small number of participants in both post and delayed post-tests. Second, the context in which the target items occurred was not appropriately controlled. Therefore, the meaning of some of the target lexical items were easily inferred from the context. Third, the exposure of all the lexical items took place in one listening session. Their study did however highlight some useful points for future research. First, relating to recognition and learnability of different word categories (noun, verb, adjective). Second, the study provided further evidence that incidental vocabulary learning could be enhanced by listening especially for word forms and grammar (parts of speech). Third, since meaning recall was improved the least through listening, other methods might need to be employed to assist meaning retention. Teacher induced intentional vocabulary instruction could provide possible solutions.

2.2.5.3 Summary

In general, the review of empirical evidence shows that existing research has tended to focus on vocabulary learning through written input. Few studies have explored how vocabulary knowledge can be acquired through oral input, recent exceptions being those by Brown et al. (2008), van Zeeland and Schmitt (2013a) and Vidal (2003, 2011). However, compared with incidental vocabulary learning through reading, acquiring vocabulary knowledge incidentally through listening is less effective. This may be because L2 reading "allows more attention to be focused on language form" (Vidal, 2011, p. 248) while in L2 listening gaining the overall gist is often more important than concentrating on specific details. Additionally, learners may find it harder to focus on individual items of lexis in more fleeting oral input than is the case for reading. Overall, supplementing incidental learning with explicit Focus-on-Form teaching thus seems potentially important for listening, perhaps more so than for incidental learning through reading.

2.2.6 Intentional vocabulary learning and incidental exposition plus explicit lexical Focus-on-Form

2.2.6.1 Intentional vocabulary learning and lexical Focus-on-Form

It is argued that intentional vocabulary learning and incidental vocabulary learning are strongly complementary and positively related (Schmitt, 2008), as both approaches may have some potential problems. On the one hand, a sufficient number of repetitions of the target vocabulary can be reached through incidental vocabulary learning as incidental learning requires exposure to extensive authentic materials. This is, however, difficult to address through intentional vocabulary learning alone (Tian & Macaro, 2012). On the other hand, doubts regarding the effectiveness and efficiency of incidental vocabulary learning mean that it needs to be supplemented with intentional teaching (Laufer, 2003; Min, 2008).

Hence, Schmitt (2008) claimed that, in order to effectively learn vocabulary at a productive level and simultaneously achieve better retention, intentional vocabulary learning should be promoted in addition to incidental learning. However, in reality explicit vocabulary instruction by language tutors in some foreign language learning contexts might occur less often than one might expect (Tang & Nesi, 2003); as a result, the amount of new vocabulary learnt without explicit instruction in the classroom context is likely to remain at a fairly low level.

Aiming to establish principles for vocabulary instruction, Schmitt (2008) suggested that three issues should be taken into account when teaching vocabulary intentionally. First, vocabulary learning tasks should be carefully designed to maximise learners' involvement in vocabulary learning (Laufer, 2009). Second, learners should receive a maximum amount of exposure to the target vocabulary. Third, in addition to meaning, other aspects of vocabulary knowledge should be enhanced (Brown et al., 2008; Horst et al., 1998; Laufer, 2009; Pigada & Schmitt, 2006; van Zeeland & Schmitt, 2013a).

Similarly, Laufer (2009) demonstrated that over the last three decades, the main research focus on pedagogical frameworks for vocabulary learning has moved away from incidental learning or indirect vocabulary learning through language input, to lexical Focus-on-Form. Unlike lexical Focus-on-Forms in which mainly decontextualized vocabulary teaching is employed (e.g., remembering word-lists), lexical Focus-on-Form advocates that both incidental and intentional learning can be enhanced by pedagogical activities that draw learners' attention to key information about a word, including its meaning, orthography and pragmatic aspects within a communicative language teaching classroom (Laufer, 2005). In addition, research has tended to concentrate more on effectiveness of word knowledge instruction, learning conditions, vocabulary activity engagement as well as target word recycling, while it tends to "attach less importance to the source of learning" (p. 341) (reading or listening input). Similarly, some relatively recent empirical research (Folse, 2006; Lee & Macaro, 2013) highlights the value of combining extensive language input, allowing for word recycling, with relevant communicative tasks in which vocabulary forms are emphasised

The effectiveness of lexical Focus-on-Form has been explored from a number of angles in the literature, as a systematic review by Laufer (2009) suggests. That review identifies seven categories of research areas within the empirical literature (p. 342):

A *Effects of text input, without lexical support (**A1**) and with lexical support (**A2**)*
B *Learning words in authentic and/or communicative tasks with Focus on Form*
C *Learning decontextualized vocabulary, or in minimal contexts with Focus on Forms*
D *Multiple exposures/rehearsals*
E *Acquiring partial knowledge of words*
F *Theoretical positions related to vocabulary learning*
G *Technology in vocabulary research and learning*

The following section will specifically look at the value of combining incidental vocabulary learning from authentic input and explicit lexical Focus-on-Form instruction.

2.2.6.2 Incidental learning plus explicit lexical Focus-on-form

Realising the possible shortcomings of considering incidental or intentional vocabulary learning separately, more recent research tends to concentrate on analysing the possible value of combining implicit learning and explicit lexical FoF within both reading and listening modalities (Hennebry et al., 2013; Laufer & Girsai, 2008; Sonbul & Schmitt, 2010; Tian, 2011; Tian & Macaro, 2012).

Sonbul and Schmitt (2010) concluded from their research within EFL learning contexts that learning outcomes of incidental vocabulary learning through reading plus direct vocabulary teaching afterwards were significantly superior to those from reading alone for three types of vocabulary knowledge including form recall, meaning recall and meaning recognition, with particular benefits for form recall.

Similarly, Laufer and Girsai (2008) compared vocabulary learning outcomes among EFL learners receiving three different types of instruction in an incidental reading modality: "meaning focused instruction (MFI), non-contrastive form-focused instruction (FFI), and contrastive analysis and translation (CAT)" (p. 694). The MFI group was given tasks which focused on the actual meaning comprehension of the reading materials and did not require extra effort focused on the target vocabulary. The FFI group was given tasks which concentrated on the target lexical items in the reading material. The CAT group was required to complete translation tasks in which they had to translate target lexical items from L2 to L1 and then from L1 to L2. Results showed that learners who received FFI and CAT tasks impressively outperformed their counterparts who received MFI tasks simply because MFI tasks only required learners to incidentally read in the target language and complete relevant reading comprehension questions, while FFI and CAT tasks, in addition to requiring learners to read incidentally, intentionally focused them on the target vocabulary.

Similar results have been obtained from listening. Tian and Macaro (2012), working with university EFL learners, found an advantage from intentional vocabulary instruction after

listening, where vocabulary meaning was provided through teacher codeswitching or English-only, compared with incidental lexical learning through listening alone. The relative advantages of this were further confirmed with younger learners by Hennebry et al. (2013), a study based on Sonbul and Schmitt's (2010) research, who claimed that direct teaching vocabulary by presenting meaning in the L1 or L2 after listening comprehension activities might result in more effective lexical retention than incidental learning through listening.

To sum up, the research literature suggests that there is pedagogical value in explicit vocabulary teaching within the framework of communicative language activities. As yet however there is still relatively little research evidence regarding which kind of explicit vocabulary teaching within such a framework would be the most effective.

2.2.7 Retention of collocations

Compared to research on the learning of single words, the amount of studies focusing on the recognition and retention of phrasal vocabulary, collocations or idioms is relatively small (Alali & Schmitt, 2012; Durrant & Schmitt, 2010; Pellicer-Sánchez, 2017; Schmitt, 2008). However, several studies have indicated that collocations can be incidentally recognised by language learners if they receive a large amount of input, either through listening or reading (Boers et al., 2006), although fewer improvements in learners' phrasal output have been established. Interestingly, empirical evidence from Schmitt's (2008) research indicated that learners who recognised collocations through extensive input might have better learning outcomes than those who were intentionally taught phrasal vocabulary through a descriptive grammar and translation course, and showed active use of formulaic sequences in oral output later on. Consequently, he argued that learning of collocations might not be appropriately addressed simply through intentional teaching.

Exploring the mechanics of how collocations are recognised and retained, Durrant and Schmitt (2010) investigated adult learners' retention of phrasal vocabulary from exposure.

The target collocations, adjective-noun combinations, were created by the researchers to make sure they followed the two criteria: First, they appeared with zero or low frequency in the British National Corpus; second, they were meaningfully combinable in plausible contexts. These collocations were then presented to learners in three different ways. The first 'single exposure', allowed participants to see the target collocation in a sentence context once for seven seconds; they were then asked to read aloud the sentence. The second, 'verbatim repetition', exposed learners to the same sentence containing the target collocation twice. The first time, they were given seven seconds to look at the sentence and read aloud, while the second time, only three seconds were allowed. In the final method, 'varied repetition', the target collocation was presented to learners twice in two different sentence contexts. Different from the verbatim repetition, at both time points learners had seven seconds to look at the sentence and read it aloud.

Results showed that compared to nouns encountered individually in sentences, nouns appearing in the form of adjective-noun collocations were remembered significantly better under all three conditions, which might indicate potential recognition of adjective-noun combinations. In addition, both repetition conditions brought about better retention than the single exposure condition. However, within the two repetition groups, the learning outcomes of the verbatim group were slightly superior to those of the varied group. Two possible reasons were proposed. First, target collocations might be better retained through repetitions within the same sentences than through repeated exposures within different contexts. Second, some cognitive pressure might have helped learners in the verbatim group to concentrate on relevant collocations and achieve better results, as learners in the verbatim group faced time-pressure compared to those in varied group who did not.

Despite the fact that Durrant and Schmitt's (2010) research showed the possible learnability of collocations, it suffered from the following limitations. First, only immediate post tests were adopted in this research. As a result, it did not investigate actual retention of certain collocations in terms of long-term memory. Second, as only one or two exposures were employed in this research, it is possible that a higher frequency of

exposure might result in better learning outcomes.

The study conducted by Durrant and Schmitt (2010) considered incidental learning, meaning that its relevance for intentional vocabulary instruction is limited. In order to more specifically investigate how collocations can be acquired intentionally, Alali and Schmitt (2012) conducted a study investigating the pedagogical effects of direct teaching of idioms compared with those of teaching single words. Six intervention sessions were conducted with 35 L2 English learners (whose first language is Arabic). The first three sessions allowed the teacher to directly teach single words, followed by a short period of grammar teaching. Within these three sessions, while the first session was purely grammar teaching, a review activity was combined with the grammar teaching in the second and the third session respectively. The review was oral based at the second session (reading aloud the target vocabulary items) yet was written based (filling the gaps with the target lexical items) in the third session. At the end of each session, there was an immediate vocabulary post-test. For the next three intervention sessions, all teaching procedures were identical, except that the vocabulary items intentionally taught were idioms.

The authors concluded from their results that all the teaching methods employed in the study led to positive outcomes for both single words and collocations, yet learning outcomes of collocations seemed to be lower than those of single words. In addition, this research also indicated that the written review activity was more effective than the oral review activity for both learning single words and idioms, which indicated that classroom activities should be carefully designed in order to enhance learning to a large extent. However, three limitations were identified (Alali & Schmitt, 2012). First, mainly written tests were adopted for practical reasons. Other testing modes, for example, oral testing might lead to better outcomes, as "it is a well-known fact that testing in the same mode as the input (e.g., oral input–oral measurement) leads to better scores than mixed input and testing modes (e.g., oral input–written measurement)" (p. 172). Second, because of the long period that the research lasted (six intervention sessions with the same group of learners), learners were likely to be aware of the potential experimental condition, which

could, to some extent, influence possible results. Third, the target individual words tested in this research were extracted from target collocations, so that exposure to single words was double compared to those of collocations.

In addition to these limitations, it is clear that the vocabulary teaching method adopted by Alali and Schmitt was mainly decontextualized vocabulary teaching, in other words, Focus-on-Forms, which is largely discouraged within communicative language teaching where Focus-on-Form is advocated instead (Laufer, 2009). Hence, it might still be questionable whether the results would be the same when adopting lexical Focus-on-Form in a communicative teaching context.

2.2.8 Summary

The review above first illustrated the different aspects of knowing a word, suggesting that the acquisition of word meaning may be effectively facilitated by explicit learning. Then from a theoretical point of view, for L2 vocabulary learning, there is a need to 'notice' the target lexical items. In addition, it is possible to argue that L1 seems to play an important role in the processing of L2 vocabulary learning and it is therefore valuable to investigate how the teacher's use of L1 can better help L2 learners to achieve better vocabulary learning outcomes. Moreover, the review also indicates that there is abundant research investigating incidental and intentional vocabulary learning through reading, yet few studies explored either incidental or intentional vocabulary learning through listening. Furthermore, more recent research tends to challenge the idea that either of these two approaches alone is sufficient for effective L2 vocabulary development, and argues for the combination of incidental exposure and intentional instruction, namely lexical Focus-on-Form. Furthermore, the review shows that there is still a lack of empirical studies exploring the pedagogical impact of incidental listening plus explicit lexical Focus-on-Form. Finally, even fewer studies have specifically investigated how collocations can be learnt. Therefore the present study will try to fill these gaps and address the above issues from a very specific angle.

2.3 The teacher's use of L1 for vocabulary instruction

The previous review indicates that the role of L1 in L2 vocabulary learning is theoretically supported by the revised hierarchical model (Kroll & Stewart, 1994) and the psycholinguistic model (Jiang, 2000). Besides, the 'noticing' hypothesis (Schmidt, 1990, 1994) indicates that for a target L2 item to be better retained on a long-term basis, it should be 'noticed' in the first place. However, it is still unclear, from a pedagogical perspective, whether these theories can be applied in a classroom context. The following section will discuss two possible methods for L2 vocabulary instruction in a classroom context with reference to the theoretical frameworks reviewed.

2.3.1 Teacher codeswitching and L2 vocabulary learning

It has been suggested by one previous piece of research that has been reported across different publications (Tian, 2011; Tian & Macaro, 2012)[1] that a specific type of L1 use within an L2 communicative framework, namely teacher codeswitching (CS), is a useful tool for L2 vocabulary instruction. But before defining teaching codeswitching, it is very important to explore naturalistic codeswitching briefly, which can be largely observed in bilingual individuals' daily speech and has been systematically investigated from both grammatical and syntactical perspectives.

Investigating grammatical features, bilingual codeswitching was systematically divided into three categories (see Figure 2.8) by Poplack (1980): "Inter-sentential switching", "'tag'-switching", and "Intra-sentential switching" (p. 615), among which intra-sentential switching is a more complicated phenomenon where codeswitched parts are naturally embedded in a sentence while keeping the grammatical structure of the original sentence logically. An example for this type of Mandarin-English codeswitching might be "*Hao,*

[1] Tian (2011) is based on her PhD study and Tian and Macaro (2012) is an article reporting one part of the study.

nimen liangge take turn, *shouhuoyuan he*… customer *zenmeshuo*? [OK, you two take turns, acting as a salesman and … how to say customer?]" (Bao, 2011, p. 45). Compared to intra-sentential switching, tag-switching is a less complicated one, and is simply a tag or single noun switching. An example provided by Bao (2011) is "*Zheshi yitiao* long pants. *Ni keyi shuo changku*, right? [This is a long pair of pants. You can say it long pants, right?]" (p. 45). The last type is called inter-sentential switching, in which a clear boundary exists between the two sentences in two different languages. An appropriate example for this could be "OK, very good, very nice, very nice, *feichang bang, xiake yihou ba ta tiezai heiban shangmian, qita tongxue keyi kan.* (I'll stick them [flash cards] on the blackboard and other students can have a look.)" (Qian, Tian, & Wang, 2009, p. 723).

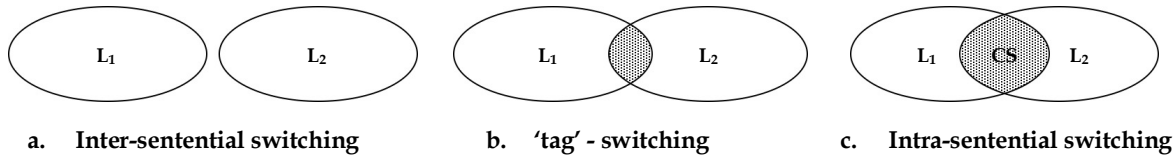


Figure 2.8 Representation of bilingual code-switching grammars (Poplack, 1980, p. 615)

In order to address syntactic features relating to codeswitching, Myers-Scotton (1993, 2001) proposed a Matrix Language Framework. This framework, particularly addressing intra-sentential codeswitching in which there is grammar contact between the two languages, is summarised by Tian and Macaro (2012). They state that in the Matrix Language Framework there is:

> a predominant language in the interaction – i.e. one that provides the morpho-syntactic structure – and an 'embedded language', i.e. one that occurs less frequently, is characterized by content words rather than function words, and is adopted for a number of purposes including the more effective communication of ideas through 'marked' (i.e. switched) lexical items" (Tian & Macaro, 2012, p. 369).

On the basis of Guo's (2007) research which highlighted the importance of intra-sentential CS and its positive roles in vocabulary instruction within Chinese EFL university contexts, Tian (2011) defined in her research that classroom CS is lexically-oriented intra-sentential CS which follows the Matrix Language Framework and is commonly adopted in a

communicative classroom. If through teacher codeswitching, L1 translation of a target L2 vocabulary was provided to L2 learners, this would aid the development of the second stage (L1 lemma mediation stage) in Jiang's (2000) psycholinguistic model of L2 vocabulary learning by providing a link between the L2 word and its concept. This argument is also supported by the revised hierarchical model of lexical and conceptual representation in bilingual memory (Kroll & Stewart, 1994), whereby L1 is still the dominant language at the initial stage of L2 learning and L2 words are lexically mediated by L1 before learners create a lemma with syntactic and semantic information from the L2. Hence, providing L1 translation through teacher codeswitching would reinforce the formal links between the words in the L1 and the L2. The meaning of the L2 word is accessed through the L1 translation and therefore the learning is facilitated.

The impact of teacher codeswitching within vocabulary instruction following listening was explored in a study involving 117 university-level learners of English in China by Tian (2011), aspects of which are also reported in Tian and Macaro (2012). Three groups of learners were studied: Two intervention groups received vocabulary instruction after listening to L2 texts, where vocabulary meaning was provided through teacher CS or L2; and a third (a control group) was exposed to listening texts without vocabulary instruction, taking part in discussions around listening strategies instead. The groups receiving intentional vocabulary instruction made significant gains in tests taken at the end of each teaching session. The Control group did not take the immediate vocabulary post-test. According to Tian and Macaro (2012), it was not appropriate to assess these learners' knowledge on those lexical items for which they had not received any type of instruction. However, they do recognise that such tests might have revealed to short-term vocabulary learning by the Control group. There was a significant decline in the scores for the intervention group at the delayed post-test, which was taken by all groups. Yet these were still significantly higher than at pre-test and long-term vocabulary retention was greater than achieved by the Control Group. Positive effects were also found for vocabulary instruction after listening comprehension activities in a study of high-school L2 learners of French in a study by Hennebry et al. (2013). In one of the tests taken by learners (although

not in others), such instruction also led to better vocabulary retention than was achieved with incidental learning through listening.

In both of these studies it was also found that teacher CS for vocabulary explanations led to significantly better learning outcomes than was achieved for those receiving L2-only explanations. The effects were not consistent across short and long-term learning, however. In Tian and Macaro (2012) meaning-recall tests given after each vocabulary instruction session, i.e., short-term learning, showed an advantage for teacher CS. Yet there was no advantage for longer term learning as assessed in delayed post-tests given between two and seven weeks after each teaching session. In Hennebry et al. (2013), however, teacher CS was more effective than L2-only explanations in the long-term, namely in the delayed post-test which assessed the learning of items encountered between one and four weeks previously. Indeed, in both studies the gap between when the vocabulary was taught and when it was assessed in the delayed post-test varied across items. Hence the interpretation of these results for long-term learning is rather problematic.

Another question that both studies looked at was whether L2 learners' language proficiency level may have had an impact on the vocabulary learning across the treatment conditions. Although Hennebry et al. (2013) found a correlation between learners' general language proficiency and their vocabulary learning outcomes, neither study found that learners of different levels may be favoured by one teaching condition over the other. In fact, both learners of high and low proficiency levels gained more vocabulary through the presentation of L1 equivalents rather than L2 definitions. Two hypotheses were presented to explain this. First, the difference in language proficiency among participants in this research might not be great enough to distinguish the effects of two teaching modes. Second, it is possible that the L2 definitions or paraphrases provided might have included some vocabulary or even grammar which was too far above learners' existing language level, making the processing of target lexical items more difficult. Thus, Hennebry et al. (2013) suggested that it is very important to explore qualitatively how learners try to understand the explanations they receive.

To sum up, although there is short-term value of using teacher-codeswitching for L2 vocabulary instruction compared with L2-only, it is questionable whether the superiority of teacher codeswitching may still exist regarding long-term retention. In addition, the participants in Tian and Macaro's (2012) study were university level EFL Chinese learners, while Hennebry et al. (2013) looked at learners of French at secondary school level. It would therefore be valuable to investigate further whether teacher codeswitching may have positive impacts on L2 vocabulary learning, especially within the context of secondary EFL teaching. Moreover, as raised by Hennebry et al. (2013), there is still a gap in investigating how language learners respond to different types of vocabulary instruction, to be more specific, what strategies they might employ to understand the vocabulary instruction they received.

2.3.2 Contrastive Focus-on-Form and L2 vocabulary learning

In contrastive Focus-on-Form, initially proposed by Laufer and Girsai (2008), cross-linguistic information about the target vocabulary is provided by the teacher through a communicative task. Cross-linguistic information goes beyond the giving of bilingual definitions of target words and instead refers to "the kind of instruction which leads to learners' understanding of the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system" (p. 696). CFoF is also theoretically supported by Jiang's (2000) psycholinguistic model; through CFoF, rich cross-linguistic information about the content of the L2 lemma can be delivered by the teacher. In addition, the teacher can also provide information about how the contents of the L1 and L2 lemma differs from each other. This may support learners in making progress towards the third stage (the L2 integration stage) proposed in Jiang's model, the point at which semantic, syntactic and morphological information is established within the L2 lemma. Moreover, CFoF is theoretically supported by Schmidt's (1990, 1994, 2012) 'noticing' hypothesis (Laufer & Girsai, 2008) in that the contrastive analysis can enhance the saliency of an L2 word for the learner and therefore the likelihood of it being learnt and retained.

Since no previous empirical study had explored the value of CFoF for vocabulary learning, Laufer and Girsai (2008) conducted a study to explore CFoF instruction through L2 reading plus task completion. Vocabulary learning outcomes, including single words and collocations, of two groups of learners were compared: a "non-contrastive form-focused instruction (FFI)" group and a "contrastive analysis and translation (CAT)" group (p. 694). The FFI group was given two tasks after reading relevant materials, including a meaning recognition task for the target lexical items and a text gap-fill task using the target lexical items. This was followed by the tutor providing general feedback and any necessary English-only vocabulary explanations. By contrast, after reading the L2 material, the CAT group was required to complete two sentence translation tasks. In the first task, they had to translate relevant sentences from L2 to L1. Each sentence, which originated from the reading material, contained one target lexical item. The second task required learners to translate relevant L1 sentences into the L2. These L1 sentences had the same meaning as the L2 sentences in the previous task. During the second task, the tutor was required to provide brief CFoF instruction on target lexical items to raise learners' awareness of the similarity and difference between the L2 target items and L1 translations. After learners completed the second task, feedback was given.

Both groups received a post-test on the day after the intervention. Vocabulary items were tested through both passive recall (translate L2 target items into L1) and active recall (translate L1 equivalents into L2 target items). An identical delayed post-test was administered one week later. Results from both post-test and delayed post-test showed that learners in the CAT group significantly outperformed those in the FFI group on learning outcomes for both single words and collocations, which might provide some indirect evidence to support the benefits of CFoF instruction over English-only. However, the teacher instruction and feedback might not be the only explanation for the CFoF group's higher level of performance. The difference in the reading tasks that the two groups received (gap-fill/meaning recognition vs. translation) might also have influenced the results. More importantly, it is likely that the CAT group might have been advantaged, as the format of the vocabulary recall test was quite similar to the reading comprehension

tasks they received, compared to the tasks given to the FFI group.

As a result, it may be premature to conclude that there are pedagogical advantages for CFoF, as only one piece of research has identified its benefits over non-CFoF for L2 vocabulary instruction. However, considering the possible value CFoF may have according to Jiang's model and the 'noticing' hypothesis, it still seems worthwhile for empirical studies to investigate whether explicit CFoF vocabulary instruction has a positive impact on learners' L2 vocabulary learning. In addition, CFoF has not been considered for vocabulary learning through listening; furthermore, a review of the literature indicates that there has been no study to date comparing all of CFoF, CS and L2 vocabulary explanations.

2.4 Language learner strategies and vocabulary learning

The above review specifically looked at two types of vocabulary instruction and their value with reference to the theoretical frameworks proposed. This, however, poses another question regarding how would L2 learners respond to the different types of vocabulary instruction and whether these responses would have any impact on their vocabulary learning. It is therefore useful to gain an overview of previous research exploring these issues, namely learning strategies or learner strategies (the two terms are often used interchangeably, as they are in this thesis).

It is important first of all to distinguish between strategies for language learning and for language use (Cohen, 1996, 1998). While the latter refers to mental activity undertaken in pursuit of understanding or using the language, as in completing a listening comprehension activity (Cohen, 1996), the former concerns the mental activity employed to commit new language to memory (Cohen, 1996). Schmitt (1997) further distinguishes between strategies for understanding words' meanings and strategies for committing words to memory (Cook & Mayer, 1983; Nation, 1990). According to Schmitt (1997), when an unknown word is encountered, two groups of strategies, "Determination Strategies" and

"Social Strategies", are usually employed by language learners (p. 206). Determination strategies involve using existing linguistic knowledge, contextualised information as well as reference materials to figure out the meaning of an unknown vocabulary item, whereas social strategies refer to activities like asking other people who may know the meaning of the word. These two groups of strategies can be labelled as "Discovery Strategies".

Strategies used for committing lexical information to memory, labelled as "Consolidation Strategies" can be categorised into four groups: Social Strategies, Memory Strategies, Cognitive Strategies and Metacognitive Strategies (Schmitt, 1997, p. 206). Social strategies here refer to activities such as group work which can be used to memorise the meaning of a word after it had been discovered. Memory strategies involve linking the meaning of an unknown word to existing knowledge, to images or grouping the new word with other familiar words to help memorise. Cognitive strategies here specifically refer to using repetitions and mechanical means to remember new words. Finally, metacognitive strategies refer to actions taken by learners themselves in order to control and evaluate their own learning in order to make it more efficient. Schmitt (1997) argues that although there are distinctions between the discovery strategy and consolidation strategy groups, some strategies can fall into the category of both groups. He developed a taxonomy in order to give an overview of vocabulary learning strategies.

Initially, a strategies list was compiled by Schmitt for a survey study of Japanese EFL learners. Three steps were taken to develop this: 1. A list of strategies was identified through consulting existing literature including vocabulary reference books and textbooks; 2. Additional strategies emerging from reports written by Japanese intermediate-level EFL learners about how they studied English vocabulary were added to the strategies list generated in the first step; 3. Several EFL teachers were asked to examine the strategy list generated after the second step and to add additional strategies based on their teaching experience. Altogether 40 strategies were identified in this strategies list for the survey study. In addition, based on the findings in the survey study, six additional strategies were added to the list. Moreover, after further consultations with literature on vocabulary,

introspection and conversations with other EFL teachers, 58 strategies were identified and made up the final taxonomy (See Table 2.1).

Table 2.1

A taxonomy of vocabulary learning strategies (Schmitt, 1997, p. 207-208)

Strategy Group	Strategy	Use %	Helpful %
Strategies for the Discovery of a New Word's Meaning			
DET	Analyse part of speech	32	75
DET	Analyse affixes and roots	15	69
DET	Check for L1 cognate	11	40
DET	Analyse any available pictures and gestures	47	84
DET	Guess from textual context	74	73
DET	Bilingual dictionary	85	95
DET	Monolingual dictionary	35	77
DET	Word lists	--	--
DET	Flash cards	--	--
SOC	Ask teacher for an L1 translation	45	61
SOC	Ask teacher for paraphrase or synonym of new word	42	86
SOC	Ask teacher for a sentence including the new word	24	78
SOC	Ask classmates for meaning	73	65
SOC	Discover new meaning through group activity	35	65
Strategies for Consolidating a Word Once it has been Encountered			
SOC	Study and practice meaning in a group	30	51
SOC	Teacher checks students' flash cards or word lists for accuracy	3	39
SOC	Interact with native speakers	--	--
MEM	Study word with a pictorial representation of its meaning	--	--
MEM	Image word's meaning	50	38
MEM	Connect word to a personal experience	37	62
MEM	Associate the word with its coordinates	13	54
MEM	Connect the word to its synonyms and antonyms	41	88
MEM	Use semantic maps	9	47
MEM	Use 'scales' for gradable adjectives	16	62
MEM	Peg Method	--	--
MEM	Loci Method	--	--
MEM	Group words together to study them	--	--
MEM	Group words together spatially on a page	--	--
MEM	Use new word in sentences	18	82
MEM	Group words together within a storyline	--	--
MEM	Study the spelling of a word	74	87

MEM	Study the sound of a word	60	81
MEM	Say new word aloud when studying	69	91
MEM	Image word form	32	22
MEM	Underline initial letter of the word	--	--
MEM	Configuration	--	--
MEM	Use Keyword Method	13	31
MEM	Affixes and Roots (remembering)	14	61
MEM	Part of Speech (remembering)	30	73
MEM	Paraphrase the words meaning	40	77
MEM	Use cognates in study	10	34
MEM	Learn the words of an idiom together	48	77
MEM	Use Physical action when learning a word	13	49
MEM	Use semantic feature grids	--	--
COG	Verbal repetition	76	84
COG	Written repetition	76	91
COG	Word lists	54	67
COG	Flash cards	25	65
COG	Take notes in class	64	84
COG	Use the vocabulary section in your textbook	48	76
COG	Listen to tape of word lists	--	--
COG	Put English labels on physical objects	--	--
COG	Keep a vocabulary notebook	--	--
MET	Use English language media (songs, movies, newscasts, etc.)	--	--
MET	Testing oneself with word tests	--	--
MET	Use spaced word practice	--	--
MET	Skip or pass new word	41	16
MET	Continue to study word over time	45	87

-- Strategy was not included on the initial list used in the survey

In addition to the list of these 58 strategies, based on the survey data with a large sample of participants (600), the above table also reports to what extent the Japanese EFL learners used these strategies and to what extent they thought these strategies were helpful. Schmitt's (1997) taxonomy is indeed valuable as it gives a relatively broader picture about what the vocabulary learning strategies are based on empirical evidence. However, it should be noted that the data were collected from a relatively single language learning background. The 600 participants were all Japanese EFL learners, although they ranged from junior-high school, high school, university to adult learners. It is therefore still

questionable whether these strategies may be used by learners from other language backgrounds.

The attention learners give to different aspects of language in oral input may also influence learning. For example, discussing whether, when listening, a sample of L2 learners of French focused more on verbs or nouns, Graham and Santos (2012) argue that learners' attentional focus is determined by specific task goals. Consequently, this focus may be altered by modifying task demands. Thus they found that when completing a recall task where they were required to note down what they had understood from a listening passage, participants were likely to concentrate on so-called key words, which were mostly nouns already familiar to them. However, in the case of a transcription task, learners tended to focus more on the form of the task rather than understanding the meaning of the listening passage, leading to higher levels of verb recognition.

In addition to focusing on verbs or nouns while listening, learners may also apply different strategies when listening to teachers' explanations of vocabulary items, which may to some extent influence vocabulary learning outcomes (Hennebry et al., 2013; Macaro, 2014). These strategies may also differ when different types of vocabulary instruction are adopted, which may have further impacts on both short-term and long-term learning outcomes.

To examine this issue, Macaro (2014) explored the strategies learners of French employed to understand their teacher's target language vocabulary explanations, as part of the larger study reported in Hennebry et al. (2013) in which L2 only explanations were provided by the teacher following listening to a passage. These vocabulary explanations were video recorded. As soon after the last session of vocabulary explanations as possible, stimulated recall interviews (SRI) were conducted with learners. During the interview, they viewed on a laptop the PowerPoint slides used by their teacher in the session, together with a short video clip in which the target word was explained by the class teacher in French. Then, they were asked to recall what they were thinking about when they saw the slide and when

they heard the teacher's explanation. Each participant was asked about six target lexical items.

The analysis of interview transcripts, indicated that the range of strategies learner employed in response to the teacher's vocabulary explanation was quite restricted. Searching for L1 cognates was the dominant strategy employed by most learners. If the target word was cognate with English, most learners were likely to ignore the explanation provided by the teacher. Interestingly, even if the target French word did not have a cognate in English, some learners still tried to find some clues from the target word itself in order to search for an English cognate. Moreover, during the SRI, many learners reported simply gave up trying to understand the teacher's vocabulary explanation, rather than asking the researcher to repeat the video clip, even if they were allowed to do so.

Although the range of language learner strategies identified by Macaro (2014) is quite narrow, this research did highlight a valuable area to investigate in the area of L2 vocabulary acquisition. In addition, as indicated as a limitation, this study was not able to more specifically compare strategic behaviour among learners at different language proficiency levels, which may provide some valuable insights as to whether higher level learners may employ a wider combination of strategies to understand the teacher's vocabulary explanation. Thus, considering within the Chinese EFL context where the L1, Chinese, has no real cognates with L2, English, is it possible that Chinese EFL learners might employ a wider range of different strategies. Additionally, the strategies employed may differ according to the type of vocabulary explanations provided.

2.5 Vocabulary intervention and L2 listening comprehension

One final issue to consider regarding incidental vocabulary learning is that in addition to potentially enhancing vocabulary knowledge, it enables learners to develop other important skills such as listening and reading comprehension (Nation & Webb, 2011). Therefore, it is worth exploring whether learners experiencing different forms of intentional vocabulary

instruction after listening to oral texts show different rates of progress in listening, a question which previous research has not explored.

In a study not about listening but about reading, but of relevance because of its comprehensive nature, Wright and Cervetti (2016) conducted a systematic review of the impact of different types of vocabulary instruction on written text (reading) comprehension. 36 peer-reviewed journal articles, written in English and dated from 1965-2015, were selected and reviewed in their study. There were two main selection criteria. First, the type of the vocabulary intervention used in the study design should either be direct teaching vocabulary meaning or teaching word-learning strategies. Second, the type of comprehension measure used should either be general reading comprehension or comprehension of written texts including the target lexical items used for the vocabulary intervention.

Results confirmed that there was strong empirical evidence indicating that vocabulary interventions had a positive impact on comprehension of passages that included the vocabulary items taught in the intervention. However, regarding the studies which measured learners' general reading comprehension, there was no evidence showing that vocabulary interventions involving teaching one or two word-learning strategies would influence learners' general reading comprehension. In addition, there was relative little empirical evidence showing that the direct teaching of word meaning (no matter how much vocabulary knowledge was taught or how long the intervention was) could have a positive impact on general reading comprehension. In addition, more 'passive' forms of direct teaching (i.e., where the teacher just gave the dictionary definition) were less successful than direct teaching which required more active processing by learners. The review, however, was focused on reading comprehension only. In view of the differences between reading comprehension and listening comprehension, it is still worth to further explore the impact of different types of vocabulary instruction on learners' general listening comprehension.

The question of 'optimal L1 use' within this intentional vocabulary instruction is also worthy of exploration. Using L1 for vocabulary instruction (e.g., providing L1 meaning of the target lexical item through teacher codeswitching), on the one hand, may facilitate vocabulary learning. On the other hand, however, L1 use may hinder other aspects of learning. For example, L2 learners may not bother to use linguistic inferencing to guess the meaning of a target L2 word, as they know that the L1 translation will be provided. This therefore calls for more research to "document the effects of L1 use on L2 learning" and to gain better insights into the "cost-benefits of Lexical Focus-on-Form by the teacher" raised by Tian and Macaro (2012, p. 371).

The influence of factors such as lexical and grammatical knowledge, language processing speed and cognitive abilities on language comprehension is also an important area, with lexical knowledge having received particular research attention. Several studies provide evidence of a strong link between vocabulary breadth and listening comprehension, indicating L2 learners with a higher level vocabulary knowledge are likely to be more successful listeners.

For example, Andringa, Olsthoorn, van Beuningen, Schoonen and Hulstijn (2012) compared individual differences between native and non-native Dutch speakers in their Dutch listening comprehension. Two groups of participants were recruited, one was 121 native speakers of Dutch (native group) while the other was 113 Dutch L2 speakers (non-native group). The range of age, gender, educational backgrounds was approximately equal between the two groups.

A series of mainly computer-based tasks was designed based on the factors which were hypothesised to have an impact on Dutch listening proficiency: language knowledge, processing speed, working memory as well as IQ level. Vocabulary knowledge was proposed to contribute to the language knowledge factor. Participants were also given a test of listening comprehension, which was the dependent variable, including five listening

texts selected from a nation-wide Dutch as a second language listening proficiency test. Multiple choice comprehension questions accompanied the texts.

The participants' receptive vocabulary knowledge was assessed through a vocabulary size test, using multiple choice questions. For each question, a Dutch vocabulary item was presented and the participants were given five choices to select from. The first four choices included one correct choice which represented the correct meaning of the vocabulary item and three wrong choices. The last choice was 'I really don't know', giving participants the opportunity to give up. Altogether 60 items were selected, mainly from Hazenberg and Hulstijn's (1996) vocabulary test. However, as this test was designed for non-native speakers, several more difficult items were added based on their test design and selection criteria.

Correlations were run between each of the two tasks. Results showed that for the non-native group, there were a strong correlation between the vocabulary test and the listening comprehension test (.68) while a moderate correlation (.35) was confirmed between the two tests in the native group. The data were then submitted to multisample structural equation modelling (SEM). For the non-native group, language knowledge explained 95% of the variance, while for the native group, it explained 79%. Although language knowledge was tested through three aspects: vocabulary knowledge, grammatical accuracy and segmentation accuracy in this study, it was still shown that vocabulary knowledge is an important factor that determines L2 listening proficiency.

The study, however, suffers from a few limitations. First, the vocabulary test for both the two groups had a fairly low reliability (.51 for natives and .42 for non-natives) and the reliability of the listening comprehension test for the non-native group was also fairly low (.66). Although it was argued by Andringa et al. (2012) that "this is not necessarily problematic, as measurement error is partialed out in the extracted factors" in SEM (p. 65), it can still be a problem regarding the vocabulary test. In addition, as discussed above, a fifth "I don't know" option giving the participants the opportunity to pass or give up was

used in the vocabulary test; this can sometimes be problematic as there is empirical evidence (Zhang, 2013) showing that "the I don't know option not only reduced the number of guesses but also discouraged partial knowledge" (p. 790).

The strong correlation between vocabulary knowledge and L2 listening comprehension found by Andringa et al. (2012) might lead one to expect growth in vocabulary and improvements in listening comprehension to coincide. Other studies, however, have found weaker correlations between vocabulary knowledge and listening comprehension suggesting that listening proficiency is not solely nor even largely determined by lexical knowledge. The study conducted by Mecartty (2000) specifically looked at the role of lexical and grammatical knowledge in reading and listening comprehension among English learners of Spanish as a foreign language. The study involved 154 university learners of Spanish at late beginner' level. Half of these students were assigned to a reading comprehension group and the others were allocated to a listening comprehension group. Both groups were tested on their vocabulary knowledge. The reading group completed a reading comprehension task while the listening group was required to do a listening comprehension task.

Vocabulary knowledge was measured through two tasks: a word-association and a word-antonym task. For the word-association task, participants were asked to select a Spanish word and a corresponding English word with an equivalent meaning to the Spanish word, whereas the word antonym task required the participants to choose an English word with the opposite meaning of the target Spanish word. Initially, 80 lexical items were selected from three Spanish textbooks. Then, six Spanish instructors were asked to categorise these items into three levels: beginning, intermediate and advanced. Items which were not able to clearly fall into each of the three levels were discarded. In addition, the remaining items were piloted to further ensure that they were not too difficult or easy for the participants. Finally, 12 items were selected for the two tasks respectively.

The input materials for listening and reading comprehension were identical. Two passages were selected from popular Spanish magazines and modified to an appropriate level. Each text was approximately 500 words in length. To ensure that the task difficulty for the listening comprehension was equal to that for the reading comprehension, the speech rate for the listening recording was relatively slow to make sure that there was "an abundance of prosodic information closer to a presumed reading rate" (Mecartty, 2000, p. 332). The recording for each listening passage was approximately four minutes. For each passage, sixteen multiple choice questions were designed to assess comprehension, including eight questions for global understanding and eight questions for local understanding. The comprehension questions were identical between the two modalities. Both the two groups completed the vocabulary knowledge test in approximately ten minutes. The reading group was given twenty-five minutes to read one passage and answer correspondent comprehension questions. They were not allowed to look again at the passage when answering the questions. For the listening group, participants heard each passage twice and were allowed to take notes. After that, they were asked to answer the comprehension questions.

Results revealed that both listening and reading comprehension were significantly correlated with lexical knowledge, yet the correlation between reading comprehension and lexical knowledge (.50) was stronger than between listening comprehension and lexical knowledge (.38). Results from multiple regression analyses indicated that for reading comprehension, vocabulary knowledge explained approximately 25% of the variance. For listening comprehension, however, only 14% of the variance was explained by lexical knowledge. Mecartty (2000) argued that although based on the results, lexical knowledge can be an important factor that determines the success of listening and reading comprehension, yet "the fact that lexical knowledge contributes different amounts to the total variance in comprehension (or why the contribution is unequal) is an intriguing finding" (p. 337). It is likely that compared with listening comprehension, lexical knowledge is more of an important factor for reading comprehension, which leads to

further consideration of other factors that may contribute to the comprehension process, especially for listening comprehension.

The relationship between the different types of linguistic knowledge and listening proficiency is reflected partly in John Anderson's (2015) model, one of the most widely cited L2 listening models. In that model, language comprehension is divided into three phases: "perception", "parsing" and "utilization" (Anderson, 2015, p. 313). These three phases to some extent overlap with each other. Initially in the perception phase, the phonemes within the listening input are segmented while the speech is heard and encoded. At the second parsing phase, as the speech is broken into phrasal units, with syntactic knowledge being used to gain understanding of the whole sentence, meaningful mental representations are created. Finally, information from the perception and parsing phases is related to information stored in long-term memory in the form of schemata during the final phase of utilization.

This final phase is central to the construction of coherent meaning from the text. While lexical knowledge is very important for perception and parsing, it plays a less central role in utilization where other sources of knowledge come into play. For example, Goh (2000) analysed the listening problems of Chinese learners of English with reference to the three phases of Anderson's model. 40 University-level EFL students participated at the study. Qualitative data were collected from three main sources. First, all the 40 students were asked to keep a weekly dairy for the listening course. They were required to write about the content of the listening course and how they understood what they heard and the potential problems that they experienced during the listening course. Second, semi-structured group interview was conducted with 17 of the 40 participants. The main purpose of the interview was to figure out the participants' knowledge about the task of learning to listen to English. Within the 40 participants, 23 students were also asked to do an immediate retrospective verbalisation with the purpose of understanding how they felt immediately after the listening course. The language used in the three sources was mainly

English. The students were, however, allowed to use Chinese when they found it difficult to express themselves in English.

From these different data sources five common listening comprehension problems (Goh, 2000, p. 60) were identified.

> Quickly forget what is heard
> Do not recognise words they know
> Understand word but not the intended message
> Neglect the next part when thinking about meaning
> Unable to form a mental representation from words heard

The problem of "understand[ing] the words but not the intended message" within the utilization phase in Anderson (2015)'s model was reported by 21 of the 40 students. The problem, specifically related to vocabulary knowledge, indicated that more than half of the participants were "unable to get the full meaning of the message even though they had understood the literal meaning of the words" (Goh, 2000, p. 60).

In addition, the data from the verbal reports were compared between two groups of learners, one with more advanced L2 listening proficiency and the other with less advanced proficiency, in order to explore whether there were any group differences in problems faced when listening. The results indicated that the problem of "understand[ing] the words but not the intended message" was only reported by participants with high listening proficiency. Goh (2000) argued that even though these learners' listening proficiency was high, they might still lack schemata to comprehend the input. In addition, insufficient contextual information during listening might deprive them of the ability to make appropriate linguistic inferences and therefore prevent them from understanding the intended meaning.

Regarding the low-level listeners, issues with inferencing were not mentioned, most likely because these participants with lower listening proficiency "hardly ever got beyond the perception or parsing phase because of limited proficiency and inadequate processing capacity" (Goh, 2000, p. 14). It was likely that these participants might suffer from speech

recognition or segmentation problems which then prevented them from moving onto a higher-level of processing.

Likewise, learners with similar levels of vocabulary knowledge may use that knowledge more or less effectively when listening. Graham, Santos and Vanderplank (2008) compared French L2 learners' strategy use and their French listening comprehension. 15 French L2 learners whose L1 is English were selected from three schools in England. Approaches were taken to ensure that these learners had similar French learning background and had approximately the same level of French contact during each week. Their listening proficiency was assessed at two time points, with six months' gap. At each time point, learners were given two audio recordings in French to listen to. Each recording consisted of four different conversations between two people. The participants were allowed to listen to each recording twice and, at the same time, were required to write a recall protocol in English about what they understood. Steps were taken beforehand to make sure that these conversations were of the same level of length, task difficulty and speech rate.

The recall protocols were then rated independently by two researchers on the basis of agreed rating procedures. The scores rated by the two researchers were then compared to ensure they were not significantly different with each other and were strongly correlated. For each participant, a total listening score was given by adding the test scores at the two time points. After that, the 15 participants were assigned to either a top listening proficiency group or a bottom listening proficiency group. Two participants, one from each proficiency group, were selected for a case study.

For these two participants, their lexical and grammatical knowledge was assessed through a French vocabulary recognition test (X-Lex, Merea & Milton, 2003) and a grammaticality judgement test, respectively. Results showed that the two participants had similar high-level of both grammar and vocabulary knowledge. The two students' listening strategy use was then explored through verbal reports given while they were listening to two additional French recordings. Through a more specific qualitative analysis of the verbal reports, the

high listening proficiency learner tended to "acknowledge the provisional nature of his interpretations" of the listening input when he was in doubt. He also "double-checked and questioned his interpretations, thus employing a number of metacognitive strategies". The student with lower listening proficiency, however, reported "largely prediction of lexis, writing visual prompts and selective attention, which for her meant listening out for particular words" (Graham et al., p. 66).

It can be seen that, although the two participants had similar levels of vocabulary knowledge, they showed different levels of listening proficiency. Possible explanations for this can be drawn from their strategy use when listening. While the higher-level learner tended to use strategies to gain a more global understanding of the listening input, the lower-level learner showed selective attention on particular single words, focusing on local information, and therefore was less successful in listening.

Both Goh (2000) and Graham et al.'s. (2008) studies had limitations. The two studies used mainly a qualitative approach based on interviews, diaries and self-report verbal protocols. However, as it was argued by Graham et al. (2008), such a qualitative approach is necessary especially when the study was trying to look at the learners' behaviour when listening in the L2, whereby how strategies are used seems to be more important than which ones and how many or often.

In summary, this section indicates that vocabulary knowledge can be related to L2 listening comprehension to some extent and it is an important factor that could explain listening proficiency, yet it is still remains doubtful whether such knowledge is the only factor that determines the success of L2 listening. In addition, even if learners' vocabulary knowledge can benefit from intentional vocabulary teaching following listening activities, it is unclear whether other aspects of language learning (notably listening itself) can be improved at the same time.

2.6 Chapter summary

In this chapter, a review of literature identifies several research gaps within the area of L2 vocabulary learning.

1. Compared to research investigating both intentional and incidental vocabulary learning through written input, quite few studies have explored pedagogical impacts of vocabulary learning through incidental oral input plus explicit lexical Focus-on-Form.

2. Regarding lexical Focus-on-Form, little research has empirically investigated how teacher-led vocabulary instruction can better assist L2 learners' vocabulary learning, to be more specific, which type of vocabulary instruction would be more effective?

3. Regarding the different types of vocabulary instruction, although some empirical evidence has confirmed the positive impact of the teacher's use of L1, the evidence was not unequivocal especially from a long-term view.

4. Based on some important existing theoretical frameworks for L2 vocabulary learning, the positive role of the L1 as a mediation for L2 vocabulary learning seems fairly strong. However, it is still debatable whether simply providing L1 definitions could be effective for learning L2 vocabulary. Meanwhile, little research has explored the pedagogical impact of providing vocabulary instruction through CFoF.

5. Previous research tended to discuss the learning of different types of vocabulary items as a whole, yet investigations into different levels of learnability of different types of vocabulary items, including collocations, is an area which few studies have addressed.

6. Few studies have specifically looked into language learner strategies used in understanding teacher's instruction, especially L2 learners' strategic behaviour in response to different types of vocabulary explanations provided by the teacher.

7. Although L2 vocabulary knowledge can be improved through intentional vocabulary instruction within a listening modality, whether such different types of instruction may influence other aspects of language learning, i.e., listening proficiency, has not been explored before.

2.7 Research questions

1. What is the impact of three different types of vocabulary instruction during listening activities (teacher codeswitching; target language explanations; and contrastive Focus-on-Form) on vocabulary learning and retention?

2. What is the impact of each type of instruction on learners' listening comprehension?

3. Does the impact of each type of instruction vary according to learners' English language proficiency level?

4. Does the impact of each type of instruction vary according to different word classes (nouns, verbs, adjectives), and for collocations?

5. Does the number of repetitions influence vocabulary retention?

6. What learning strategies do learners use in response to the instruction for the target items offered by the teacher in each of the three experimental conditions?

CHAPTER 3 METHODOLOGY

3.1 Research questions

1. What is the impact of three different types of vocabulary instruction during listening activities (teacher codeswitching; target language explanations; and contrastive Focus-on-Form) on vocabulary learning and retention?

2. What is the impact of each type of instruction on learners' listening comprehension?

3. Does the impact of each type of instruction vary according to learners' English language proficiency level?

4. Does the impact of each type of instruction vary according to different word classes (nouns, verbs, adjectives), and for collocations?

5. Does the number of repetitions influence vocabulary retention?

6. What learning strategies do learners use in response to the instruction for the target items offered by the teacher in each of the three experimental conditions?

3.2 Research design

In the area of social science, the idea that there are two paradigms, namely qualitative and quantitative, might be viewed as an oversimplification, as the main focus of the two are on the "data rather than foundational beliefs and assumptions" (Willis, 2007, p. 8). Thus, it is suggested that there are three paradigms, namely postpositivism, critical theory and interpretivism (Cupchik, 2001; Guba, 1990). Willis (2007) also argues that there are family resemblances within paradigm groups. Qualitative research might fall into the paradigm of critical theory and interpretivism, while quantitative research would always fall into the paradigm of postpositivism. However, for mixed methods research which combines quantitative research and qualitative research, an alternative paradigm, pragmatism, is suggested (Teddlie & Tashakkori, 2009). Pragmatism, which offers a different worldview to those of postpositivism and interpretivism, "sidesteps the contentious issues of truth and reality, accepts, philosophically, that there are singular and multiple realities that are open to empirical inquiry and orients itself toward solving practical problems in the 'real world'" (Feilzer, 2010, p. 8). Each paradigm can be characterised by the type of questions it poses: ontological, epistemological, and methodological (Guba, 1990, p. 18):

> Ontological: What is the nature of the "knowable"? Or, what is the nature of "reality"?
> Epistemological: What is the nature of the relationship between the knower (the inquirer) and the known (or knowable)?
> Methodological: How should the inquirer go about finding out knowledge?

Although Willis (2007) argues that ontology and epistemology are two fundamental issues within each paradigm, it seems that these two issues are of less interest for the present study as it was designed within a paradigm of pragmatism. Therefore, the attention here is to focus on what are the best methods to answer the research questions.

It has long been argued that there are merits for adopting either quantitative or qualitative methodology, yet the two are not mutually exclusive (Best & Kahn, 1998), and many discussions tend to advocate that combining the two methodologies together would be a reasonable idea for conducting social research (Punch, 1998). Moreover, when combined,

the two methodologies could compensate for each other, which might reduce the potential weakness of adopting either one alone (Bryman, 1988). Hence, Johnson, Onwuegbuzie, and Turner (2007) defined mixed methods research as "the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for broad purposes of breadth and depth of understanding and corroboration" (p. 133).

Two issues emerge from the definition. First, according to this definition, mixed methods research should include both qualitative research and quantitative research. Second, methodology "is not restricted to data collection and analysis but extends to other aspects of the research process such as the perspective of the researcher" (Richards, Ross, & Seedhouse, 2012, p. 304). Finally, mixed methods research can also be related to the issue of 'methodological triangulation', meaning that mixed methods research can "reap the benefits of two paradigms and minimizing the drawbacks of each" since methods either in qualitative or in quantitative paradigm have strengths and weaknesses (Hussein, 2009, p. 4). Specifically regarding the case of the present study, qualitative data may further help to explain the quantitative data collected through the classroom intervention.

Regarding the six research questions listed above, the present study adopted a mixed methods approach combining qualitative research and quantitative research. There are two reasons for adopting this approach. On the one hand, whether learners' vocabulary knowledge and L2 listening comprehension can be improved through the intervention could be addressed most effectively by collecting quantitative data. On the other hand, qualitative research was adopted to gain some insights into learners' strategic behaviour within the vocabulary instruction. To collect quantitative data for the first five research questions, the present study employed a quasi-experimental design. To address the last research question with its focus on learners' perspectives, qualitative data were collected through stimulated recall interviews. The following sections will explore the specific methods employed in detail.

3.3 Experimentation

In order to address the first five research questions for the present study, a quasi-experimental design was employed. According to Shadish, Cook and Campbell (2002), several unique terms can be clearly defined regarding modern experimentation (p. 12):

> Experiment: A study in which an intervention is deliberately introduced to observe its effects.
> Randomized experiment: An experiment in which units are assigned to receive the treatment or an alternative condition by a random process such as the toss of a coin or a table of random numbers.
> Quasi-Experiment: An experiment in which units are not assigned to conditions randomly.

From these definitions, it is clear that the difference between the randomised experiment and the quasi-experiment is that the former allows the researcher to randomly assign units into different conditions while the latter does not. However, more distinctions between the two methods are clarified by Silver-Pacuilla, Brown, Overton and Stewart (2011), who argue that "while a true experiment includes (1) pre-post test design, (2) a treatment group and a control group, and (3) random assignment of study participants, quasi-experimental studies lack one or more of these design elements" (p. 17).

For practical reasons, i.e., gaining access to potential senior secondary school teaching facilities, obtaining consent from all parties and collecting data within one school term, the original plan of the present study was to adopt a pre-post test design study without a control group (see Figure 3.1). In this study design, data on selected participants' level of performance are obtained at the pre-test before the actual intervention takes place. After the intervention, similar data can be collected through the post-test. The group of selected participants who receive the intervention is called the treatment group. Finally, through investigating the difference between data collected from the pre-test and from the post-test, the impact of the intervention can be assessed.

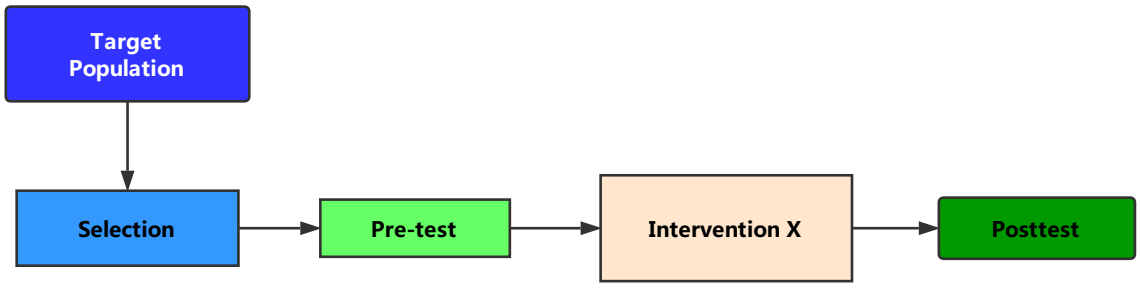


Figure 3.1 Study design of a pre-post test without a control group (Silver-Pacuilla, et al., 2011, p. 17)

However, the literature review indicated that there was still value in investigating incidental vocabulary learning through listening. As a result, for the main study it was decided to add in one group (a control/comparison group), which enabled the study design to explore the value of incidental vocabulary learning through listening, without explicit vocabulary teaching. A pre-post test design study with a control group was adopted (see Figure 3.2).

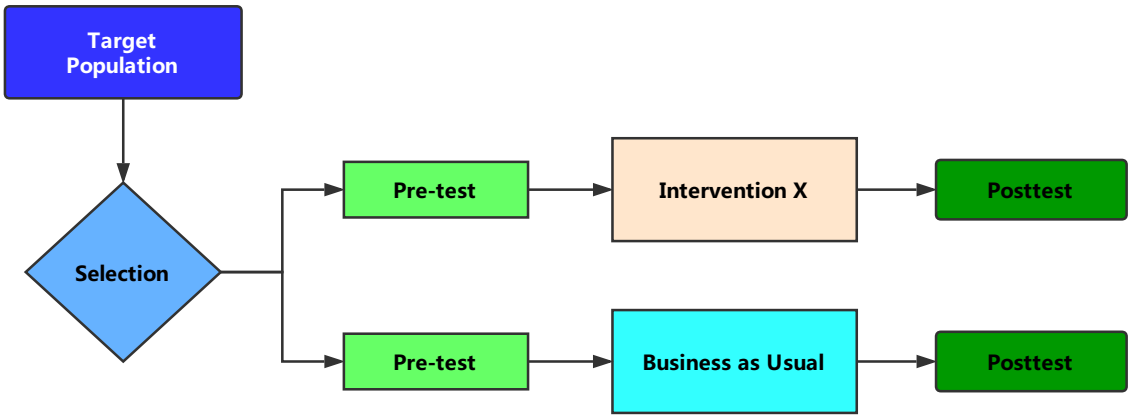


Figure 3.2 Quasi-experimental study design of a pre-post test with a control group (Silver-Pacuilla et al., 2011, p. 17)

This study design is similar to the design presented in Figure 3.1. Participants' level of performance was measured through a pre-test before the experimentation. The same data were then collected again after the experimentation. Finally, through investigating the difference between data collected from the pre-test and from the post-test, the impact of the intervention can be assessed. The reason for collecting pre-post test data from the control

group is to rule out other factors which may affect the difference in the pre-test and post-test results. For example, in the case of the present study, the improvement of English vocabulary knowledge might be due to the participants being accustomed to the vocabulary tests used or gaining access to the target vocabulary items outside the intervention sessions, rather than being due to their learning from the vocabulary explanations offered by the teacher in the intervention.

The main problem with this study design is that it does not randomly assign the participants into a treatment group or a control group as an experimental study does, which may "create less compelling support for counterfactual inferences. For example, quasi-experimental control groups may differ from the treatment condition in many systematic (non-random) ways other than the presence of the treatment." (Shadish, Cook, & Campbell, 2002, p. 14). Steiner, Wroblewski and Cook (2009) argue that "random assignment remains the most reliable technique for justifying causal inference. It provides the logically most valid and efficient causal counterfactual. Consequently, results are more credible than those from other quasi- or nonexperimental methods." (p. 90).

However, randomised control-group pre-post test design is still rare (Steiner et al., 2009) and may not be practical (Shadish et al., 2002), especially in classroom-based research investigating the impact of teaching strategies on learning outcomes. In this research area, keeping learners intact in their original class or group could reduce "the reactive effects of the experimental procedure and, therefore, improves the external validity of the design" (Dimitrov & Rumrill, 2003, p. 160). In the present study, measures were taken to overcome lack of randomisation and details are given in Section 3.4.

Since the present study tried to compare the impact on both short-term and long-term vocabulary learning of three types of vocabulary instruction for EFL students, there were three treatment groups as well as one control group. Besides, in addition to a pre-test before the intervention, a post-test and a delayed post-test were employed to explore short-term and long-term vocabulary learning. Moreover, in order to analyse whether the number

of repetitions influence vocabulary acquisition, a final delayed post-test was employed. As a result, the above study design was slightly modified (see Figure 3.3).

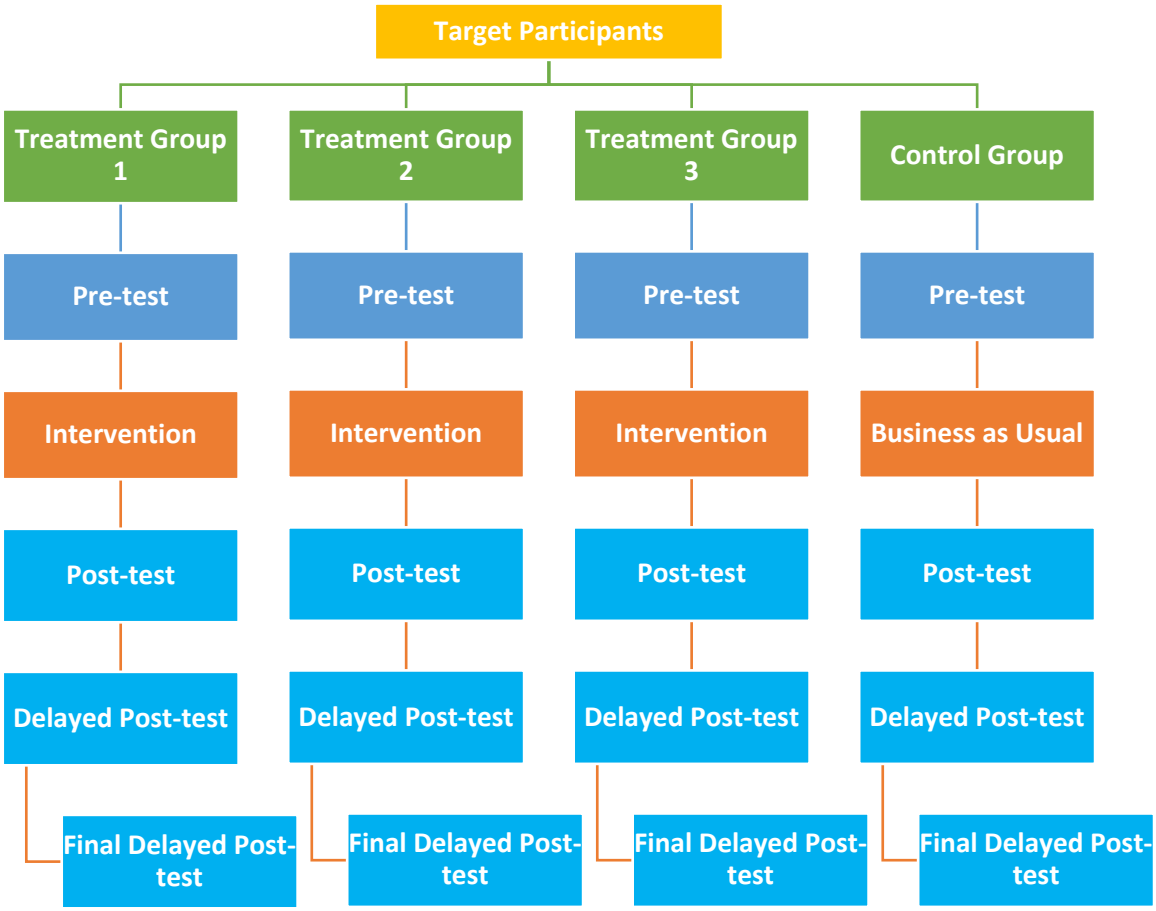


Figure 3.3 Study design of pre-, post-, delayed post- and final delayed post-tests with three treatment groups and one control group

3.4 Participants

The target participants for this study were first-year senior secondary school EFL learners, aged 15-16 years. They were considered to be the most appropriate participants firstly because they had received at least six years of EFL education (three years in primary school and three years in junior secondary school). Their general level of English enabled them to understand the listening passages adopted in the intervention. Besides, one class of

learners received vocabulary instruction in English-only. Their previous English study experience could help them to cope with this situation. Secondly, they had experienced at least three years of English-only teaching in junior secondary school. Thirdly, compared to students in the second or final year of senior secondary school study, they faced fewer exam pressures and were not likely to be studying simply for examination purposes, meaning data collected from students in the first-year might be more reliable.

Thus, 152 first-year students from four different classes in a public senior secondary school in China were chosen. The school was selected through convenience sampling, as a school that the researcher knows and that was willing and able to participate in the study. Normally, students have to take the District High-School Entrance Examination (DHEE) before starting senior secondary school education. Each school will have its own passing score for DHEE and will admit students on the basis of their aspirations and scores. Students who are accepted by the same school will have been randomly assigned to each class according to their DHEE scores, so that each class will have approximately the same mean examination scores and total number of students.

The present study decided to keep the participants in their original classes rather than randomly allocating them into different experimental groups according to their English language proficiency level as Tian and Macaro (2012) did. In that study, participants were university-level students. University students do not have a fixed ‘class’ as they have optional modules to choose and they are already familiar with having lectures and seminars with students from other majors. For the senior secondary school learners in the present study, however, it would have been too disruptive to their learning to move them from their normal learning environment and would threaten the ecological validity of the study.

Although these senior secondary school learners had been randomly assigned by the school to each class based on their DHEE scores, one notable issue is that, in addition to English, there are other subjects tested in DHEE. Therefore, it is unclear whether the average English language proficiency for each class was approximately the same. In order to

overcome this issue, steps were taken to collect baseline data about these participants' general English language proficiency before the intervention from three sources: 1) their English examination scores from two recent school English tests, a test which included listening, reading, writing, speaking as well as grammar and vocabulary tests; 2) results from a general English vocabulary knowledge test, adapted from McLean, Kramer and Beglar (2015), administered by the researcher; and 3) results from a modified IELTS English listening comprehension practice test, also administered by the researcher. Details and issues relating to the general vocabulary test and the general listening comprehension test will be further explored in Section 3.5.2. In addition, for the purpose of the analysis, these test results were further used to classify participants into four overall proficiency levels (high, mid-high, mid-low and low level).

The four classes were randomly assigned to three treatment groups, namely a second language (L2) group (38 students), a teacher codeswitching (CS) group (40 students) and a contrastive Focus-on-Form (CFoF) group (37 students), as well as a control group (37 students). All four classes/groups were instructed by the researcher himself, who has three years of senior secondary school EFL teaching experience and has Chinese as a first language.

3.5 Data collection procedure

Prior to the pilot study, it was originally planned to collect data from three experimental groups, divided into two phases and lasted four months in total. In this original design, each phase started with a pre-test, followed by three sessions of intervention as well as a delayed post-test. An additional delayed post-test was conducted within four weeks of the completion of the second phase. The stimulated recall interview was conducted after the final delayed post-test.

For the first phase, the first pre-test session included a general listening comprehension test and a general vocabulary test. The aim of employing these two tests was to provide

additional evidence for learners' language proficiency level. Besides, a vocabulary pre-test was also administered at this stage, testing learners' pre-knowledge of vocabulary items taught through the first three intervention sessions. In the following three weeks (week 2-4), each treatment group received one intervention session a week. At the end of each teaching session, there was a vocabulary post-test testing learners' short-term vocabulary learning outcomes. In week 6, there was a delayed vocabulary post-test, in which long-term retention of vocabulary items from the first three sessions was tested.

For the second phase, the second pre-test session was administered in week 7. Learners were only tested on their pre-knowledge of vocabulary items from the second three intervention sessions. Thereafter, the procedure was exactly the same as in the first phase until the second delayed test which was administered at week 12. The final delayed post-test examined learners' long-term vocabulary retention for vocabulary items from all six intervention sessions at week 15. Finally, in week 16, stimulated recall interview sessions were conducted with 12 learners. Figure 3.4 shows how exactly the procedure worked.

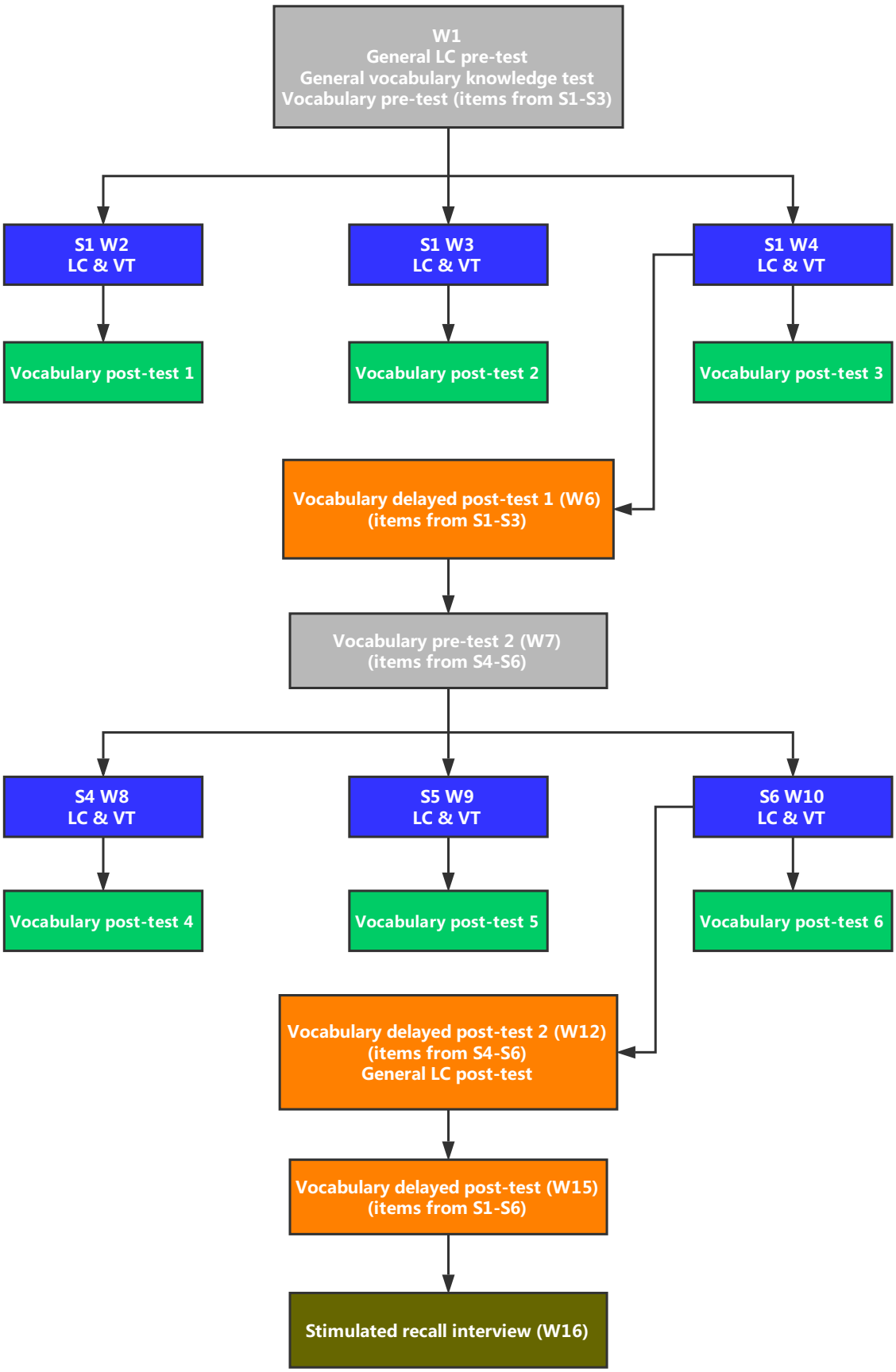


Figure 3.4 Pre-pilot data collection procedure

Note. **W**-Week; **S**-Session; **LC**-Listening comprehension; **VT**-Vocabulary teaching

However, as revealed in the pilot phase (see Section 3.7), several problems were identified with this original data collection procedure. First, although potential participants were first-year senior secondary school students, they were still likely to face exam pressures. During the pilot phase, although only two intervention sessions were conducted, the planned procedure had to be changed as the students were required to focus on preparing the final-term examinations and were not able to participate. Therefore, it was likely that the proposed procedure for the main study would be challenging and may not be practical for implementation as it lasted 16 weeks and required frequent meeting with potential participants, namely on 12 occasions, during this period.

Second, as discussed above, in order to rule out other factors which might have influenced the results of the intervention, a control group needed to be added to the pre-pilot plan. Third, the original procedure had two vocabulary delayed post-tests, one placed one week after the third intervention session and the other placed two weeks after the final intervention session. Thus, the length of delay for each intervention session was not equal, between two to four weeks. As revealed in the pilot phase, it was likely that items which had a shorter delay would be advantaged.

Taken into account the potential problems of this pre-pilot procedure, it was modified as appropriate for the main study. As shown in Figure 3.5, the whole procedure lasted three months, one month less than a normal school term, starting with a baseline test session in week 1, followed by six sessions of intervention and two sessions of vocabulary review from week 4 to week 11. It was not possible to start the first intervention session immediately after week 1 as there were eight days' national holiday between week 2 to 3, which would have interrupted the data collection procedure as the six intervention sessions and two review sessions needed to occur in eight consecutive weeks. The final delayed post-test session was conducted at week 12 after the completion of the last review session. Finally, the stimulated recall interview was done in week 13, one week after the final delayed post-test session. The main advantage of this study design was that the data collection procedure was complete one month before the end of the school term in China,

so that the procedure was not likely to be affected by the school's intense examination review sessions.

In the baseline test session in week 1, all four groups, including the three treatment groups and the Control group, received a general vocabulary knowledge test as well as a general listening comprehension test. The aim of employing these two tests was to provide additional evidence for learners' language proficiency level. The vocabulary pre-test was combined with the general vocabulary knowledge test, adopting the same format, in order to have a general idea about participants' pre-knowledge of the target lexical items. In the following six weeks (week 4-9), each treatment group received one intervention session a week, which was video recorded with consent from all parties. The intervention session allowed the teacher to administer the listening comprehension tasks as well as teach the vocabulary items. A vocabulary post-test assessing learners' short-term vocabulary learning outcomes then followed. At the end of each session, there was a review activity where selected target lexical items were revisited.

Between the third intervention session and the sixth intervention session inclusive, each session had an additional vocabulary delayed post-test, which assessed long-term vocabulary retention for target items from the session two weeks previously. This delayed post-test was combined with the vocabulary post-test for the current session. The delayed post-tests for the fifth intervention session were conducted at the beginning of the first review session in week 10, while the delayed post-test for the sixth intervention session was administered at beginning of the second review session in week 11. In this way, each delayed post-test was administered exactly two weeks after the correspondent intervention session. In addition to conducting the vocabulary delayed post-test, each review session allowed the teacher to review selected target lexical items. This meant that different vocabulary items received a different number of repetitions (see Table 3.3 and 3.4).

The Control group also received six matched sessions and two review sessions. However, while the treatment groups received vocabulary teaching, the Control group was taught

mainly about the target language culture and given background information relating to the listening passages. In addition, there was no review activity for the Control group in the sessions. At the same time, instead of having review activity in the review sessions, the teacher asked them to give group presentations.

The final vocabulary delayed post-test took place in week 12 with the three treatment groups only, examining whether the different numbers of repetition influenced the vocabulary retention. The Control group did not receive this vocabulary final delayed post-test since its purpose was to explore the impact of repetitions, which the Control group had not received during the teaching sessions. Also in week 12, all groups received a listening comprehension post-test, reassessing their listening proficiency. Finally, in week 13, stimulated recall interview sessions were conducted with 12 learners from the three treatment groups, which were audio recorded for further data analysis. Figure 3.5 shows how exactly the final procedure worked. Details for the research instruments mentioned above will be given in the following sections.

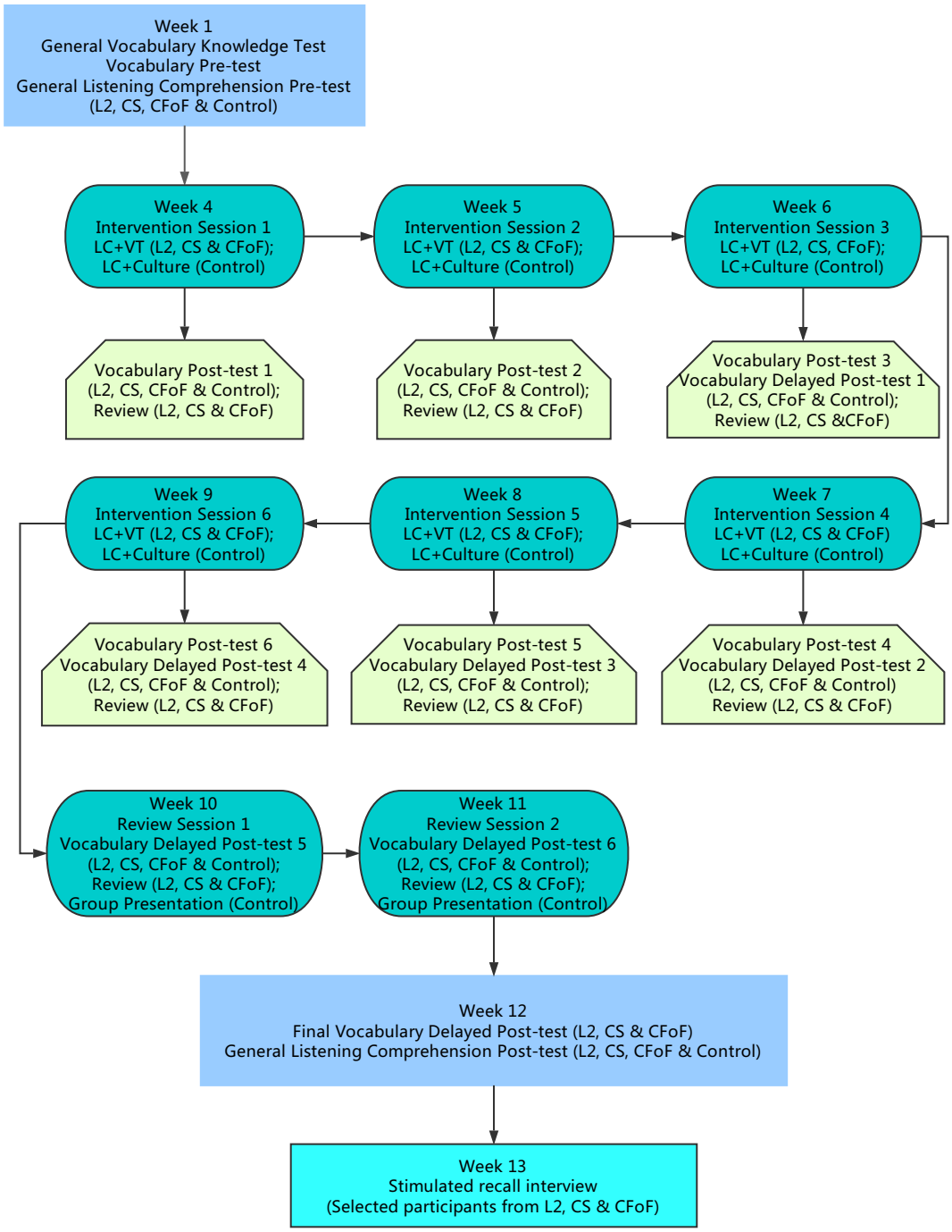


Figure 3.5 Final data collection procedure

Note. **LC** - Listening comprehension; **VT** - Vocabulary teaching; **L2** - English only; **CS** - Teacher codeswitching; **CFoF** - Contrastive Focus-on-Form

3.6 Research instruments

3.6.1 General vocabulary knowledge test and general listening comprehension test

Since the present study did not randomly assign the participants into different treatment conditions, a general vocabulary knowledge test as well as a general listening comprehension test were employed in order to provide additional evidence for the participants' English language proficiency level before the intervention. Considering the primary aim of the study is to investigate vocabulary learning outcomes through oral input, a written form of vocabulary test may not have been appropriate. Thus, the present study adopted a listening vocabulary levels test (LVLT) designed by McLean et al. (2015), yet modified as appropriate. The LVLT is a newly developed aural vocabulary size test, which aims to test the first five 1000-word frequency levels and the AWL for both pedagogical and research purposes. Within each 1000 frequency level, 24 vocabulary items are tested. An additional 28 items are focused on academic lexical items. 30 minutes are required for testing all 150 items. An example of the original Japanese version for testing each item is given in the Figure 3.6 (English translations of items a-d added here for clarity):

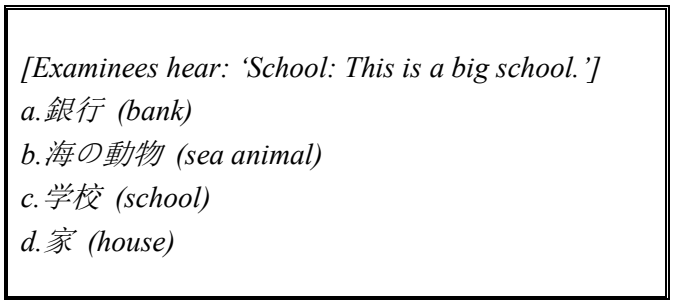


Figure 3.6 An example item of the aural vocabulary levels test (Japanese version, McLean et al., 2015, p. 743)

This test has been further edited and expanded to the first eight 1000-Word Frequency Bands by McLean, Ishii, Stoeckel, Bennett and Matsumoto (2016). However, at the time of the classroom intervention being administered, the edited version was not yet available, meaning it could not be adopted in the present study. Besides, the main modifications that the authors made to the previous version was to change some of the inappropriate answer

choices given for certain target lexical items. For example, they changed some of the answer choices (in Japanese) which are loanwords from English or which differ in part of speech or grammatical function with the target item being tested. However, an examination of these changes suggested that these were not problematic when translating the Japanese version into Chinese to be used at the present study.

In general, one underlying problem with this type of vocabulary knowledge test (using multiple choices questions to test vocabulary size) is that it cannot rule out the possible percentage of successful guessing by the test takers (Stewart & White, 2011; Wesche & Paribakht, 1996). Therefore, the test results could sometimes be an overestimation of vocabulary knowledge. Gyllstad, Vilkaitė and Schmitt (2015) conducted a study to investigate issues with guessing and sampling rates relating to assess vocabulary size through multiple choice formats. Participants first took the Vocabulary Size Test (VST) (Nation & Beglar, 2007) and then were interviewed about their knowledge of the target item tested in the VST. The data collected from the VST were compared with those from the interview. Results revealed "a clear tendency for scores on the MCT to be proportionally higher than scores from the interview measure" (Gyllstad et al., 2015, p. 296). Between approximately 11% – 26% of the vocabulary items being recognised in the multiple choice tasks in the VST were not reported in the interview procedure, which calls for extra caution when interpreting and using the results from this type of test. In addition, they also questioned whether it is enough that only ten samples were selected and tested in each 1000 frequency level in Nation and Beglar (2007)'s VST as, based on their results, the estimated vocabulary size would become more accurate if the sampling rates were higher.

However, there are still advantages in employing the LVLT in the present study. First, the sampling rates are higher in the LVLT. Unlike the VST which only tests ten sample items within each 1000 frequency band, 24 sample items were selected and tested in the LVLT. Second, it is quite easy to implement and could test large numbers of vocabulary items in a fairly short time. Third, test scores are easy to calculate because they are from multiple

choice questions. Finally, the purpose of adopting this test was to provide partial evidence for the participants' general English language level rather than to give a wholly accurate picture of participants' vocabulary size.

Considering the actual participants' language proficiency level, the present study only tested the first three 1000-word frequency levels. At the same time, as the listening passages to be used in the study included a certain amount of academic words, the AWL was also tested. Moreover, the original Japanese version was translated into a Chinese version. The example of the translated Chinese version being used at the present study is given in Figure 3.7 below (English translations of items a-d added here for clarity). Finally, since the present study did not test the fourth and the fifth 1000-word frequency levels, the participants were given 25 minutes in the present study to complete the test rather than 30 minutes.

[Examinees hear: 'School: This is a big school.']
a.银行 (bank)
b.海洋动物 (sea animal)
c.学校 (school)
d.家 (house)

Figure 3.7 An example item of the aural vocabulary levels test (Chinese version)

It was originally decided to employ the listening comprehension test used in the National University Entrance Examination (NUEE). The test includes two sections. There are five short dialogues in the first section. One multiple choice comprehension question is asked for each dialogue. The second section includes four long dialogues and a monologue. For each long dialogue or monologue, three to five multiple choice comprehension questions are asked. In total, there are 20 questions for this listening test. There were several potential reasons for using this test. First, its level is appropriate for the participants as it is specifically designed for senior secondary school EFL learners. Second, the test is widely used for national-based examinations and therefore is likely to be reliable. Third, the score is easy to calculate as only multiple choice questions are used in the test.

It was, however, decided to abandon this test for the main study. There were two reasons for doing this. On the one hand, it is relatively easy for the potential participants to get access to this test as the test materials are already available in some English practice books. Besides, for teaching and practising purposes, some of the materials are used in the senior secondary school EFL classroom context. As revealed in the pilot phase, one participant pointed out that he had encountered some of the listening materials before having the listening test. On the other hand, test takers' listening proficiency can be overestimated as the only question type used in this listening test is multiple choice questions. It cannot rule out the possibility of guessing. In the pilot phase, the participants gave answers to all the twenty questions in the listening test, yet they also admitted that they did not know the answers for some of the questions but just picked one choice randomly.

Therefore, in the main study, the general listening comprehension test was based on an IELTS (International English Language Testing System) listening practice test. Relevant test materials are available online. There are several reasons for adopting this type of test. First, this test is specifically designed for learners who are not English native speakers. Second, the test is widely adopted for testing English language proficiency levels, and has been established as valid and reliable (Breeze & Miller, 2012). Third, potential participants are not likely to have prior access to these test materials, as the purpose of taking this test is to study abroad or for immigration purposes. Fourth, in addition to multiple choice questions, there are several other question types used in the IELTS listening test, including form filling, sentence completion, matching, map filling and completing notes. It is likely that listening knowledge is tested from different perspectives so that the possibility of guessing can be minimised.

The standard format of this test includes four sections. The first and the third sections are dialogues while the second and the fourth sections are monologues. From the perspective of task difficulty, the first two sections are easier than the last two sections, as the former, based on a survival context, is about daily life and cultural issues while the latter, set in the academic context, is focused on doing research and listening to academic lectures.

However, taking the general language proficiency level of the participants in the present study into account, using all four sections might be challenging for them. Thus, only the first two sections of the IELTS listening practice test were employed.

Before the first section of the listening test, there is a short set of instructions which briefly introduces the listening test. At the beginning of the first listening section, there is an example, where the test takers can see a sample listening comprehension question and its correct answer while they hear a short dialogue between a male and a female speaker about this sample question. After that, a short instruction regarding how to answer the sample comprehension question is given. These sets of instruction give the test taker a general idea about how to cope with the listening tasks. For each listening section, ten comprehension questions are given. These ten questions are divided into two parts, normally between four to six questions in each part. The test takers are given approximately ten seconds to read the questions in each part before they hear the relevant listening input and start to answer the comprehension questions. The listening test lasts 10-15 minutes.

In order to address the second research question exploring the impact of each type of vocabulary instruction on learners' listening comprehension, two IELTS listening comprehension tests were employed, one at pre-test and one at post-test. Steps were taken to ensure that the two listening comprehension tests were of comparable level of difficulty, matching each test on 1) vocabulary frequency level; 2) length and speech rate; and 3) the number of global or local questions assessed. Details for these are given in Table 3.1.

Table 3.1

Summary of the two listening comprehension tests

		Listening comprehension pre-test	Listening comprehension post-test
Vocabulary frequency level (Cumulative token %)	**K1**	85.63	85.79
	K2	91.22	91.16
	K3	95.25	94.62
Word count		1190	1073
Speech rate (words/min)		138	146
Global questions (%)		30	30
Local questions (%)		70	70

3.6.2 Listening comprehension material and target vocabulary items

Since the primary aim of the study was to investigate L2 vocabulary learning through oral input, it was important to select and modify appropriate listening material for teaching target vocabulary items. Selection and modification of input material, either oral or written texts, designed to examine learners' vocabulary learning should follow four principles: "students' interests, instructional appropriateness, as well as linguistic difficulty and text length" (Min, 2008, p. 83). In order to enhance learners' interest, input materials should be authentic, in which language used can be more natural and lively and texts on a variety of topics are likely to be involved (Min, 2008; Tian, 2011). Discussing reading, but making comments also applicable to listening material, Schouten-van Parreren (1989) identified several advantages of authentic reading materials in facilitating L2 learners' incidental vocabulary learning, arguing that compared to readers designed for foreign language learners, L2 beginners might find more enjoyment and achieve better learning outcomes through reading books which were specifically designed for native speakers of the same age level.

In addition, since all input texts were used for the classroom intervention in this study, it was important to take instructional appropriateness into account, which means the input materials should be relevant to the specific learning contexts as well as corresponding to original textbooks aimed at learners' language proficiency level (Min, 2008; Tian, 2011). Moreover, adopting input texts which are instructionally appropriate can, to some extent, reduce the possibility of learners' awareness of the experimental condition (Tian & Macaro, 2012), which can maximise the likelihood of collecting accurate data.

However, for the present research, it was difficult to fully satisfy the above two principles when selecting target listening materials. Considering that oral texts selected for this research should be instructionally appropriate at EFL secondary school level, it was not appropriate to adopt materials which are specifically designed for native speakers, as L2 and L1 learning situations at senior secondary school level are quite different. Hence, the listening materials were mainly selected from the L2 English textbooks designed for senior secondary school learners, yet modified as appropriate.

Moreover, extra modifications were made to the linguistic difficulty of each input to satisfy basic comprehension purposes for learners. It is suggested that in order to achieve fairly easy comprehension of a given written text, a minimum percentage of familiar vocabulary within the text is likely to be 98% (Hsueh-Chao & Nation, 2000; Schmitt, 2008). The same figure of lexical coverage is also suggested in the listening modality (Nation, 2006). However, later studies suggest the figure can be lower than this. Stæhr (2009) calculated that participants needed to know 89% of the words in his listening passages in order to comprehend them. With the purpose of testing high school EFL vocabulary retention in a reading modality, Min (2008) adopted written texts within which the percentage of new vocabulary was below 3%. Slightly different, Tian and Macaro (2012) employed oral texts with a maximum 5% unknown words in order to investigate EFL vocabulary learning of university first-year English-majors. Similarly, listening passages of 95% lexical coverage were used by van Zeeland and Schmitt (2013b) when testing undergraduate level L2 learners' retention of three types of vocabulary knowledge.

Considering that the participants of the present study were first-year senior secondary school EFL learners, whose language proficiency level may be lower than participants in previous studies, it was decided that the percentage of unfamiliar lexical items in each listening passage should be around 5%. Additionally, since the present study aims to investigate how easily lexical items from different word classes can be learnt and explore the retention of collocations, further modification was made to ensure all input materials adopted contain unfamiliar words of different types as well as a certain number of collocations.

In addition to linguistic difficulty, another noteworthy issue is the length of input materials. Tian (2011) employed listening texts of approximately 256 words each on average to investigate the EFL vocabulary learning of first-year university English-majors. Each intervention session in Tian's study lasted two hours, which could cover two to three listening texts. However, in a reading modality, Min (2008) adopted reading materials which had an average length of 706 words. As the present study differed from the above two conditions in terms of the target participants' proficiency level and teaching environment, and the time limitation for each intervention session was 45 minutes, only one listening text, approximately 250 words in length, was adopted for each session.

The final issue regarding the oral input adopted for the classroom intervention is its speech rate, which is crucial especially when considering task difficulty of the listening comprehension material (Brindley & Slatyer, 2002). After examining the mode of speech delivery of British English in different genres, Tauroza and Allison (1990) concluded that the average speech rate ranged from 150 to 170 words per minute in radio monologues; 125 to 160 in academic lectures; 160 to 210 in interviews and 190 to 230 in conversations. In the latter genre, the range achieved 160 to 230 words per minute in moderately fast interactions. Thus, considering the participants' language proficiency level while ensuring the authenticity of the oral input adopted, the average speech rate for the listening materials was controlled at between 150-175 words per minute. The mode of these listening passages were monologues.

Based on the requirements outlined earlier, the original listening materials were selected from a senior secondary school English text book called '*New Senior English for China*' (Liu, Gong, Zheng, Brooks, Allingham, Sjoquist, & Miller, 2007). There are three reasons for choosing this textbook. First, it was not used by the senior secondary school participating in the present study, which can avoid some potential contamination. Second, compared to other EFL textbooks, it is more widely used in China and has a high reputation for its usefulness from a pedagogical perspective. Third, it is a textbook recommended by the secondary school English language teaching curriculum. Selection and modification of relevant materials were based on the following criteria: 1) All materials are relevant to learners, and authentically written and cover a wide range of topics. 2) Each listening passage should contain at least 95% of familiar words. As the listening passages were adopted from the senior-secondary school textbooks, unknown lexical items were already selected and highlighted in the textbooks based on the senior-secondary school English curriculum (MoE, 2003), showing that the passages met the 95% criterion. 3) The 5% unfamiliar lexical items should include single words of different classes (noun, verb, adjective) and a certain number of collocations. 4) Each listening passage should be approximately 250 words in length. 5) The speech rate for each listening passage should be controlled at between 150-175 words/min. A summary of the listening passage is given in Table 3.2 and the passages themselves with correspondent comprehension questions appear in Appendix A.

Table 3.2
Summary of listening passages

Session no.	**Topic**	**Word count**	**Speech rate (words/min)**	**Percentage of unknown vocabulary**
S1	A student of African wildlife	236	162	5%
S2	A master of nonverbal humour	257	171	4%
S3	Communication: No problem?	257	154	4%
S4	Showing our feelings	258	152	4%
S5	Theme parks – Fun and more than fun	255	175	4%
S6	The noodle harvest	264	155	5%

For each listening passage, three corresponding comprehension questions were designed, with one question focusing on the global understanding of the listening passage and two questions addressing local understanding of the details. The selection of target vocabulary items was based on the modified listening materials. All listening passages were analysed using both '*Classic Web Vocabulary Profiler*' (see http://www.lextutor.ca/vp/eng/) and '*Compleat Web English Vocabulary Profiler*' (see http://www.lextutor.ca/vp/comp/).

The former profiler automatically categorises each individual word in the listening passages into four frequency bands: K1 words (1-1000), K2 words (1001-2000), Academic word list (AWL) words as well as Off-List words, while the latter categorises each individual word into 26 frequency bands, from K1 words to K25 words (24000-25000) as well as Off-List words. Considering that the requirement for taking the English test in DHEE is that students should acquire approximately 1800 English words mainly from K1 and K2 level, the target lexical items were mainly selected from K2 words, K3 words as well as AWL words.

In addition, as lexical items were highlighted for teaching purposes in the English textbooks used for the selection of the listening passages, these lexical items were also taken into account when choosing target items. Moreover, a selection of target lexical items also fulfilled the requirement of including different types of single words and a certain number of collocations. Furthermore, the position of the target lexical items within each sentence was also an important issue to take into account (Graham & Santos, 2012). They were positioned initially, medially, and finally in sentences, to control for order/recency effects.

Originally, altogether 60 target lexical items were selected, including 17 nouns, 13 verbs, 13 adjectives and 17 collocations. However, further modifications had to be made to the listening passages and these target lexical items for the main study. The results from the pilot phase (see Table 3.10, Section 3.7.4) revealed that only one learner out of 18 participants had no prior knowledge of the target vocabulary items, while half of the

participants correctly provided meanings for at least three lexical items. Although these learners for the pilot study might have slightly higher language proficiency levels than participants for the main study, it is still possible that some of the target lexical items might too easy for learners in the main study. Hence, those items which were already known to most learners in the pilot study were substituted with more difficult ones. Altogether 60 target lexical items were selected, including 18 nouns, 13 verbs, 12 adjectives and 17 collocations. These items were presented and highlighted in the listening passages in Appendix A.

There was, however, another problem with selecting target lexical items. As observed in the pilot study, the participants tended to only focus on the vocabulary explanations from the teacher, while paying less attention to listening comprehension and did not try to figure out the answers for the comprehension questions in the second intervention session. It was likely that these participants were aware that they would be tested on their vocabulary knowledge after the first intervention session and they were used to the examination-oriented learning modality in the senior secondary school. In addition, as revealed in the pilot phase that there were still around ten minutes left at the end of each intervention session, which meant it was possible to add in more content. Therefore, in addition to the 60 target vocabulary items, using the same two vocabulary profilers, four to six additional vocabulary items were selected within each listening passage. The selection criteria were in line with those used for choosing these target lexical times discussed above. These additional items were set as distractors and were treated equally with the ten target items for each intervention session. It was believed that in this way, even if the participants concentrated on remembering the meaning of the vocabulary items, they were not able to know which items were the focus of the study.

3.6.3 Three types of vocabulary instruction

To address the first research question, the study design had three treatment groups: an L2 group, a CS group and a CFoF group. All three groups were treated equally, except for the

vocabulary explanations they received after listening to the passages. However, a noteworthy point here is that the research design needed to ensure that each treatment group received the same amount of vocabulary explanation for a specific vocabulary item, so that no treatment group was advantaged or disadvantaged.

Thus, for the L2 group, learners were provided with the English-only meaning for the target vocabulary item. At the same time, in order to help learners to understand the meaning, an additional sentence exemplifying the target lexical item was given. Slightly different from the L2 group, the CS group received the meaning of the target lexical item through teacher codeswitching, so that the learners in this group received the L1 (Chinese) meaning of the target L2 vocabulary item from the teacher. They also received an additional sentence exemplifying the target item. For both the L2 and the CS groups, they were required to understand the sentence by themselves first. The teacher then briefly explained the example sentence to both groups in the L2.

Finally, for the CFoF group, the learners were firstly given the L1 meaning of the target vocabulary item, plus an additional explanation was provided in the L2 in order to compare and contrast the L2 word and its L1 version, especially when there was a mismatch between them. There were mainly two ways that the L2 word did not match the L1 translation. First, it could be that the L1 translation did not match the grammatical structure involved in using the L2 word. An example for this is shown below in Figure 3.8. Second, it could also be that the L1 translation matched more than one L2 word, from different parts of speech. Figure 3.9 shows an example for this. One thing should be noted here is that although mismatch was found for the target 60 lexical items, there was variability in the degree of this mismatch and their suitability for contrastive analysis.

Listening material: This character was a social failure but was loved for his optimism and determination to ***overcome*** all difficulties.

L2 explanation: Here, overcome is a verb, which means 'succeed in a struggle against something'. So 'overcome all difficulties' means succeed in a struggle against all difficulties. Another example for this word can be 'I would overcome any weakness, any despair, any fear.'.

CS explanation: Here, overcome is a verb, which means '战胜'. So 'overcome all difficulties' means 战胜 all difficulties. Another example for this word can be "I would overcome any weakness, any despair, any fear.'.

CFoF explanation: Here, overcome is a verb, which means '战胜' which is equivalent to 'fight to win'. Therefore, '战胜' can be used to in relation to people. However, we only use 'overcome something' in English. So 'overcome all difficulties' means '战胜' all difficulties.

Figure 3.8 Example One for each type of vocabulary instruction

Listening material: Making a fist and shaking it almost always means that someone is angry and ***threaten***ing another person.

L2 explanation: Here, threaten is a verb, which means 'to tell someone that you will kill or hurt them if they do not do what you want'. Another example for this word can be 'They threatened to kill him unless he did as they asked.'.

CS explanation: Here, threaten is a verb, which means '恐吓'. So, 'threatening another person' means '恐吓 another person'. An example sentence for this word can be 'They threatened to kill him unless he did as they asked.'.

CFoF explanation: Here, threaten is a verb, which means '恐吓'. In Chinese, '恐吓' can be used both as a noun and a verb. In English, however, 'threat' is the noun version for 'threaten'. They are different words.

Figure 3.9 Example Two for each type of vocabulary instruction

3.6.4 Vocabulary test for the classroom intervention

A vocabulary pre-test was employed with all four groups before the first intervention session. However, a noteworthy point here is what is called pre-test sensitisation, which occurs when "subjects have made the connection between the object of the pretest and the nature of the treatment". In this case, subjects "may not respond with their true feelings" and they "will tend to bias their answers to make them more palpable to either the researcher or the public at large" (Bonate, 2000, p. 44). This issue was detected in the pilot phase of the present study. Due to time restrictions, the first intervention session took place immediately after the vocabulary pre-test (see Section 3.7.2) for the data collection procedure of the pilot study). One participant mentioned in the pilot interview that since he heard some target vocabulary items once in the vocabulary pre-test, he paid extra attention to these items when they were heard and explained during the intervention session.

In order to minimise the possibility of pre-test sensitisation, the vocabulary pre-test for testing the pre-knowledge of 60 target lexical items was combined with the general vocabulary knowledge test (the listening vocabulary levels test) and adopted the same format, in which participants heard the researcher reading out the target lexical item and an example sentence including the target item. They were then required to choose from the four options to give the most appropriate Chinese meaning for this item. In this way, it was believed that the participants would not know which items in the test were the focus of the study. In addition, at the main study, the first intervention session was placed three weeks after the baseline test session where the vocabulary pre-test took place (see Figure 3.3).

There are two limitations regarding the design of the vocabulary pre-test which need to be acknowledged. Firstly, as it was mixed with the general vocabulary knowledge test, the results of this test could be overestimated because of guessing (Gyllstad et al., 2015; Stewart & White, 2011; Wesche & Paribakht, 1996). Secondly, the format of the vocabulary pre-test was not identical to the vocabulary post-test and delayed post-test, since the former was meaning recognition and the latter was meaning recall. Laufer and

Goldstein (2004) in their study of 435 high school and university L2 learners indicated that meaning recognition was an easier task than meaning recall. Therefore, the main problem with the design of the vocabulary pre-test was that it might overestimate participants' pre-knowledge of the target items. However, it was still decided to use this study design as it was an appropriate method to hide the target lexical items in order to avoid possible pre-test sensitisation. In addition, compared with underestimating participants' pre-knowledge of the target items, overestimating their pre-knowledge was relatively easier to be addressed. Therefore, several steps were taken to minimise these limitations.

Firstly, an additional question was added to the vocabulary post-test collecting information on whether the participants had already known the target items before the intervention. Gyllstad et al. (2015) suggested that "external validity evidence (…) can provide a useful, and perhaps necessary, complement to other validation approaches in order to get a more comprehensive picture of the characteristics" (p. 300). Secondly, further modifications were made to the results of the pre-test after the post-test data were collected. Details regarding these steps will be further discussed in the Section 3.8.1.

A vocabulary post-test and delayed post-test were employed in order to compare the participants' vocabulary learning outcomes before and after the intervention. In addition, a final delayed post-test was administered two weeks after the delayed post-test to explore whether the number of repetitions influence vocabulary acquisition. The details for the timescale and implementation procedures for these tests are given in the previous Section 3.5. The format of the vocabulary post-, delayed post- as well as the final delayed post-test in the present study used a modified version of what was employed in Tian (2011), developed according to Paribakht and Wesche's (1993) vocabulary knowledge scale. Tian's (2011) test has three main features (see Figure 3.10). Firstly, only productive knowledge of the target vocabulary meaning is tested. Secondly, participants are allowed to provide the meaning either in L1 or L2, which ensures that all groups are treated equally. Besides, allowing the participants to provide the meanings either in L1 or L2 can provide indirect evidence for their preference regarding the type of vocabulary instruction. Thirdly,

participants were asked to indicate how confident they were about the meanings they provided through an incremental rating system.

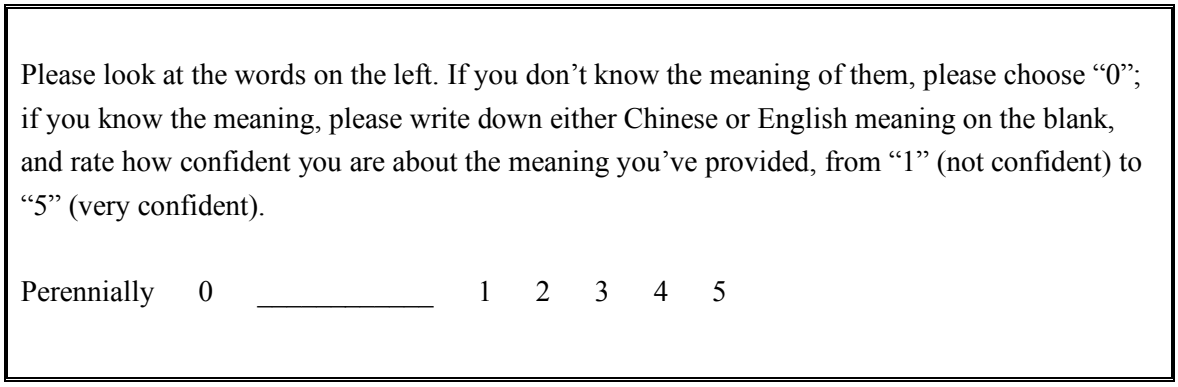


Figure 3.10 Test format designed by Tian (2011, p. 155)

Tian's (2011) test design was modified as appropriate in the present study. First, since the primary aim of the research is to investigate vocabulary learning through listening, the vocabulary test was orally based. Hence, learners were not able to see the actual vocabulary item; instead they heard the teacher reading it out. Additionally, as discovered in the pilot phase, simply reading out the target lexical items may have limitations, as the participants may confuse a specific target item with its potential heterography (e.g., true & chew). Thus, an additional sentence including the target item was also read out by the teacher to provide additional evidence for the participants to distinguish between the target item and its heterography. Moreover, since the format of the vocabulary pre-test was slightly different from the post-test and the delayed post-test, participants were required to provide additional information on whether they had already known the target items before the intervention in the vocabulary post-test. See Figure 3.11 and 3.12 for the sample of the vocabulary post-test and delayed post-test adopted in the present study.

Please listen to the teacher reading out the words. For each word, you will also hear a sentence including the word. From 1 (not confident) to 5 (very confident), please circle a number to show the degree to which you feel confident about the exact meaning of the given words and then write down their meaning in the blank space either in Chinese or in English. If you do not know the word, you can circle 0, without providing any answers. If you know a specific word, please also specify whether you knew this word before having this class by circling YES or NO on the right.

							Did you know this word before today?
0	_____	1	2	3	4	5	Please circle a response: YES NO

(Participants will hear: Chew, I can chew it.)

Figure 3.11 Vocabulary post-test for the present study

Please listen to the teacher reading out the words. For each word, you will also hear a sentence including the word. From 1 (not confident) to 5 (very confident), please circle a number to show the degree to which you feel confident about the exact meaning of the given words and then write down their meaning in the blank space either in Chinese or in English. If you do not know the word, you can circle 0, without providing any answers.

0	_____	1	2	3	4	5

(Participants will hear: Chew, I can chew it.)

Figure 3.12 Vocabulary delayed post- and final delayed post-test for the present study

3.6.5 Teaching procedure

In the present study, the overall teaching procedures for the three intervention groups, L2 group, CS group as well as CFoF group, were the same within six intervention sessions and two review sessions, differing only in how information about the target items was given. Since the main purpose of the present study was to explore different types of vocabulary instruction within a communicative EFL teaching context, the teaching procedures for the six intervention sessions involved implementing listening comprehension tasks, followed by teacher-student interactions geared towards meaning

comprehension of the relevant listening passages, within which knowledge of the target vocabulary was delivered.

With the purpose of imitating a normal EFL teaching context to a large extent, at the beginning of each intervention session, there was a warming up activity as well as a pre-listening activity. The warming up activity included general interactions between the instructor and the whole class which were intended to help the participants to focus their attention on the lesson as well as to become familiar with the instructor who was not their usual EFL tutor. Through the pre-listening activity, participants received a certain amount of preliminary background information about the upcoming listening passage, similar to the steps followed by Tian (2011). Then, each participant was given a handout which included three multiple choice listening comprehension questions and had two minutes to read all the questions before the instructor played the listening passage for the first time. The listening passage was played once only. After listening, all participants were given two minutes to complete all questions and all handouts were then collected immediately by the instructor. The reason for doing this was to concentrate participants' attention on the comprehension of meaning rather than on any unfamiliar vocabulary items during the listening passage itself.

In the following stage, through teacher-learner interactions, participants were firstly asked questions about the main idea of the listening passage in L2. They were asked to form groups and have a brief discussion with their group members. After that, the listening passage was replayed once sentence by sentence. More specific comprehension questions were asked in L2 at this stage in order to focus learners' attention on listening comprehension. At the same time, the meaning of each sentence was explained by the instructor. Wherever a target lexical item was met, the instructor provided additional explanations for the target lexical item. Additional explanations were also given for the items which were set as distractors. The L2 group received English-only explanations. The CS group was provided with L1 explanations of the target lexical item in the form of teacher codeswitching, and the CFoF group received L1 explanations through comparing

and contrasting L2 word and its L1 translation. All target items and the distractors were presented by PowerPoint to make sure all participants had access to the spelling of these items.

After explaining all items in detail, the meaning of the whole listening passage was reviewed in English-only by the instructor. At the same time, L2, CS and CFoF explanations for both the target items and the distractors were repeated again. Then, the whole listening passage was played again to make sure all participants had fully understood it. After that, answers to the listening comprehension questions were given to the learners with reference to specific listening input relating to the question. For each intervention session, the participants heard the listening passage three times in total. After that, a vocabulary post-test was employed to test potential achievement on target lexical learning. Starting from the third intervention session, there was a two-week delayed post-test which was combined with the vocabulary post-test for the current session. At the end of each intervention session, selected target lexical items were reviewed in order to meet the requirement of different numbers of repetition. Each intervention session lasted 45 minutes. All six intervention sessions were video recorded with consent from all parties.

In order to make sure that the participants focused on the listening comprehension and did not take notes for the target vocabulary items and their explanations, each intervention session was placed in a multi-media classroom which is not the normal classroom where the participants have their lessons. They were not allowed to bring any notes or books to the multi-media classroom and pens were provided to them by the instructor. All participants were informed not to write down any lexical items, rather to concentrate on the comprehension of the meaning of the listening passages, and this was monitored by the instructor during each teaching session. See Figure 3.13 for an example of one intervention session for the treatment groups.

There were two reasons for having two review sessions after the six intervention sessions. Firstly, the implementation of the delayed post-test for the fifth and the sixth intervention

sessions required two extra sessions. Secondly, the review sessions enabled the researcher to review more target lexical items to meet the requirement of different number of repetitions. Therefore, at the beginning of each review session, a vocabulary delayed post-test was administered. Then, selected target lexical items were reviewed.

Listening passage 3 - A Master of Nonverbal Humour

1. Greeting the students (5 minutes) (English-only)

Self-introduction

Brief questions about learners' attitudes towards learning English

2. Introducing the topic (5 minutes) (English-only)

a. Do you know the words comedy and comedian?

b. Can you name some famous comedians in China and around the world? Who is your favourite comedian?

c. Showing pictures of several comedians through PowerPoint and providing learners with some background information of relevant comedians. (Stephen Chow, HK; Mr. Bean, UK)

d. Showing the picture of Charlie Chaplin and asking learners question: who is he? Where is he from? Have you ever seen any of his movies?

3. Listening comprehension task (10 minutes)

a. Giving out questions for the following listening comprehension task

b. Playing the listening material

c. Asking learners to finish relevant question

d. Collecting all answer sheets

4. Vocabulary explanation (20 minutes)

a. Playing the listening material sentence by sentence

b. Presenting the lexical items through PowerPoint. This is where the treatment element comes in either L2, CS or CFoF

c. After finishing the whole passage explanation, quickly going over the whole listening passage using the outlined vocabulary items on the PowerPoint

d. Playing the listening passage for the third time and checking the answers for the comprehension questions with the students.

5. Vocabulary post-test (5 minutes)

a. Giving out the vocabulary post-test and delayed post-test and collecting it

b. Reviewing selected target lexical items

c. Saying goodbye to learners and ending the lesson

Figure 3.13 Teaching procedure of one intervention session for the treatment groups

The Control group listened to the same passages as the intervention groups, yet the teaching procedures for them were slightly different from the three treatment groups. Before listening to the target materials, participants in this group also received a warming up activity as well as a pre-listening activity, just like the three treatment groups did, to help them become familiar with the background information relating to the listening passage. They then received the three listening comprehension questions and were given time to read them. After that, they heard the listening passage once and completed the comprehension questions. The answer sheets were collected by the instructor immediately after they completed the comprehension task.

To ensure that the Control group had as many focused exposures to the target items as the intervention groups, they were allowed to listen to the listening passage sentence by sentence twice, matching the two rounds of vocabulary explanations received by the treatment groups. During this period, instead of receiving vocabulary instruction, they were taught about cultural and background information relating to the listening passage, focusing on the general meaning of the listening passage. Hence, like the intervention groups, the Control group heard the passage three times in total. Thereafter, in order to counterbalance the repetitions that the three treatment groups had, the Control group went over the passage again. At the end of each session, they received a vocabulary post-test. A two-week vocabulary delayed-post test was also administered starting from the third session. See Figure 3.14 for an example of one intervention session for the Control group. In addition, the Control group also received two review sessions. Within these sessions, they initially had a vocabulary delayed post-test, followed by a group presentation activity organised by the researcher.

Listening passage 3 - A Master of Nonverbal Humour

1. Greeting the students (5 minutes) (English-only)

Self-introduction

Brief questions about learners' attitudes towards learning English

2. Introducing the topic (5 minutes) (English-only)

a. Do you know the words comedy and comedian?

b. Can you name some famous comedians in China and around the world? Who is your favourite comedian?

c. Showing pictures of several comedians through PowerPoint and providing learners with some background information of relevant comedians. (Stephen Chow, HK; Mr. Bean, UK)

d. Showing the picture of Charlie Chaplin and asking learners question: Who is he? Where is he from? Have you ever seen any of his movies?

3. Listening comprehension task (10 minutes)

a. Giving out questions for the following listening comprehension task

b. Playing the listening material

c. Asking learners to finish relevant question

d. Collecting all answer sheets

4. Culture teaching (20 minutes) (English-only)

a. Playing the listening material sentence by sentence twice

b. Teaching culture and background information about the listening passage

c. Going over the answers for the listening comprehension questions in full with further meaning explained about the passage

5. Vocabulary post-test (5 minutes)

a. Giving out the vocabulary post-test and delayed post-test and collecting it

b. Saying goodbye to learners and ending the lesson

Figure 3.14 Teaching procedure of one intervention session for the Control group

As discussed before, for the three treatment groups, there was a review activity at the end of each intervention session and review session in order to meet the requirement of different number of repetitions. During the review activity, the learners first heard the teacher reading out a target lexical item. They were then asked to recall and say aloud the meaning of the word by themselves. After that, the meaning of each item was confirmed by the teacher. Details of the target items reviewed at each session are outlined in the Table 3.3 and of how many repetitions different words had are outlined in the Table 3.4.

Table 3.3

Details of the review activity in each session

Teaching session	Details of the review activity
Intervention Session 1	Review all words from Intervention Session 1
Intervention Session 2	Review all words from Intervention Session 2
Intervention Session 3	Review 5 words from Session 1 and 5 words from Session 3
Intervention Session 4	Review 5 words from Session 1 and 5 words from Session 2
Intervention Session 5	Review 5 words from Session 1 and 5 words from Session 3
Intervention Session 6	Review 5 words from Session 1 and 5 words from Session 2
Review Session 1	Review 5 words from Session 1 and 5 words from Session 2
Review Session 2	Review all words from Session 1 to Session 6

Table 3.4

Details of how many repetitions different words had

Number of repetitions	Vocabulary items
9 repetitions	5 words from Intervention Session 1: ***moustache, stiff, overcome, hut, chew***
7 repetitions	5 words from Intervention Session 2: ***explore, output, insect, crop, anxious***
5 repetitions	5 words from Intervention Session 3: ***outspoken, prominent, alter, argue, entertainment***
3 repetitions	All the other words: ***leather, convincing, spoil, sauce, weird, shelter, inspire, represent, association, dormitory, curious, approach, dashing, crossroads, theme, various, wander, fantasy, swing, tourism, ancient, inner, function, frown, threaten, subjective, rank, similarity***

3.6.6 Stimulated recall interview

In the present study, three groups of learners received three types of vocabulary instruction. It is possible that how learners tried to understand the input they received, as well as how engaged they were in the task of trying to understand the passages and explanations they heard (Schmitt, 2008; Tseng & Schmitt, 2008), would have influenced the degree to which they retained the vocabulary presented. Gaining insights into learners' strategic behaviour

as they listen to the teacher's explanations, and into their attitudes towards the instruction, may help to illuminate the results from the quantitative data. These insights were sought through stimulated recall.

Stimulated recall, as a type of research method, has been widely adopted in different areas of educational research including physical education (Lyle, 1999), nursing education (Daly, 2001) and language education (Gass & Mackey, 2000). Lyle (2003) defined stimulated recall as "an introspection procedure in which (normally) videotaped passages of behaviour are replayed to individuals to stimulate recall of their concurrent cognitive activity" (p. 861). In addition to presenting the video episode to the participants, the participants would normally be required to answer several well-structured, yet mostly open-ended, interview questions immediately after or during the watching of the video episode.

More specifically, through exploring the possible value of stimulated recall for second language research, Gass and Mackey (2000) propose that it can be a useful tool to investigate learners' cognitive process and language learner strategies. They indicate that when retrieving second language lexical items, learners may go through several steps of unconscious cognitive process or employing different learner strategies, which may not easily be examined through test-oriented quantitative instruments, so that stimulated recall could be a useful tool to collect this kind of data. However, the stimulated recall method has potential limitations especially regarding its reliability and validity. Hence, Gass and Mackey (2000) also suggest that the recall process should be administered as soon as possible after the video prompts and the researcher must ensure that the design for the actual procedures for creating the recall should always follow the original aim of the research study. Moreover, the validity issue can be addressed as long as "the questions/prompts do not alter the cognitive process being employed at the time of the event" (Lyle, 2003, p. 865), in other words, any further questions asked should not be too intrusive.

Lyle (2003) concluded that the limitations of stimulated recall can be minimised if the recall design is carefully structured. Previous research has confirmed that stimulated recall can be useful when analysing "learning strategies, knowledge management, teacher behaviour, teacher training, mentoring behaviour and decision-making" (p. 874). Considering that classroom teaching combines both "complexity, uniqueness, multi-variable goals" and "non-deliberative decision-making and interactive behaviours" (p. 874), it is possible that adopting stimulated recall for classroom teaching research would have significant advantages.

In the present study, through stimulated recall interview, qualitative data were collected from learners from the three experimental groups. In order to avoid influencing the research results, the stimulated recall interview session was placed after the final delayed post-test (see Section 3.5). The original plan was to select two learners from each treatment condition, one more advanced learner and one less advanced learner, to participate in the stimulated recall interview. For each learner, the focus of the interview was on how these learners understood the teacher's instruction for one word and one collocation during the intervention.

However, this procedure had to be modified as it was rather unclear whether enough qualitative data could be collected through this. As discovered in the pilot phase, some of the students were somehow reluctant to express their opinions. They were silent at certain stages of the interview and even refused to say anything for specific questions, although they were encouraged to speak by the researcher many times. Therefore, for the main study, within each treatment group, four learners representing the four different proficiency levels were selected to participate. Altogether 12 learners were individually interviewed. During the interview sessions, the learners were asked about how they tried to understand the explanations given by the teacher for five lexical items, including two nouns, a verb, an adjective as well as a collocation, after watching a video clip showing the teacher explaining the target lexical item, which the learner had experienced during the intervention. A sample of stimulated recall questions used for the L2 group are listed below

in Figure 3.15. Considering the participants' proficiency level, the language used in the interview was Chinese (English translations in the figure below are given for clarification). A few sub-questions were added in when interviewing based on learners' response. The sub-questions were mainly used to confirm the answer or ask for clarifications.

Watch a video of teaching the target word 'convincing'

1. 当你第一次听到 convincing 这个单词的时候，你感觉怎样？
(How did you feel when you first heard the sentence containing the word 'convincing'?)
2. 听到该单词后，你接下来做了什么？
(What did you do then?)
3. 当老师在解释 convincing 含义的时候，你脑海里在想什么？
(What was in your mind when the tutor was explaining the meaning of 'convincing' to you?)
4. 对于老师所提供的 convincing 的英文解释，你感觉怎样？
(How did you feel about the English-only explanations of 'convincing' provided by the tutor?)
5. 因为当时你并不能写下这个单词以及它的意思，你是如何应对老师所给出的该单词的英文解释的？
(Since you were not allowed to write down the word and its meaning, how did you cope with the English explanation provided by the tutor?)
6. 在接受了老师这种解释单词的方法之后，你对于未来是否可以正确使用 convincing 这个词有什么看法吗？
(How did you feel about the possibility to successfully use 'convincing' in the future after receiving this type of vocabulary instruction?)
7. 就帮助你去理解听力材料这一方面来说，你对于老师所提供的对 convincing 这个单词的解释有什么看法？
(What was your opinion about the vocabulary instruction for the word 'convincing' that you received in relation to understanding relevant listening materials?)
8. 总的来说，你对于老师对该单词解释的方式感觉怎么样？
(In general, how did you feel about the way that the tutor explained the word to you?)

Figure 3.15 A sample of stimulated recall interview questions for the L2 group

3.7 The pilot phase

The aim of the pilot study was 1) to rehearse the intervention sessions to see whether the planned teaching procedure was practical and 2) to ensure the validity of relevant listening comprehension materials and test instruments.

3.7.1 Participants and group allocation (the pilot phase)

Three groups of first-year EFL learners, six for each group (L2, CS and CFoF), from a public senior secondary school were invited to participate in the pilot study. They were from three different classes, yet were taught by the same English tutor. Although they had already studied in the senior secondary school for eight months, these students were believed to be the best participants for the pilot study because their language proficiency levels were similar to those of the potential participants in the main study, albeit a little higher, so that their performance in the pilot study could provide valuable evidence for the validity of relevant test instruments.

Learners from the same class were assigned to the same treatment group. Within each group, there were six learners. It was not considered appropriate to reallocate learners into different treatment groups according to their English language proficiency levels, for two reasons. Firstly, it was felt to be much easier to successfully adopt a communicative language teaching approach if learners' original learning environment was maintained. Secondly, reallocating learners into different treatment groups would not be allowed in the main study. In order to have evidence for learners' English language proficiency level, details of the participants' performance levels in their most recent English examinations were obtained.

The mean test scores from two recent English language examinations for all 18 participants are given in Table 3.5. Descriptive statistics on English language test scores for each group are given in Table 3.6. The maximum possible score for the test is 110. The participants were then categorised into four language levels: high (87.0-110), middle-high (82.0-86.9), middle-low (76.0-81.9) and low (0-75.9). Descriptive statistics on test scores for each treatment group are presented in Table 3.6. Since in the pilot study there were only six participants in each treatment group, non-parametric tests were used in order to explore the group difference before the intervention. Using a Kruskal-Wallis test, no statistically significant differences were found between the English language test scores for different

groups [$X^2(2) = 0.82$, $p > .05$]. However, it should be noted that as it was not appropriate to reallocate learners into different treatment groups, there were no low level learners in the CS group. Moreover, learners' language levels were not equally distributed among all three groups.

Table 3.5

Summary of English language test scores and proficiency levels (the pilot phase)

No.	Treatment group	English language test score	Proficiency level
1	CS	85.25	Middle-high
2	CS	83.75	Middle-high
3	CS	80.00	Middle-low
4	CS	87.25	High
5	CS	80.25	Middle-low
6	CS	82.50	Middle-high
7	CFoF	83.25	Middle-high
8	CFoF	89.75	High
9	CFoF	73.50	Low
10	CFoF	61.75	Low
11	CFoF	80.25	Middle-low
12	CFoF	87.00	High
13	L2	94.50	High
14	L2	77.75	Middle-low
15	L2	86.50	Middle-high
16	L2	74.50	Low
17	L2	80.00	Middle-low
18	L2	73.25	Low

Table 3.6

Descriptive statistics on English language test scores for each group in the pilot phase

Treatment group	Mean	*SD*	Median	Range
CS	83.17	2.84	83.13	7.25
CFoF	79.25	10.26	81.75	28.00
L2	81.08	8.08	78.88	21.25

Note. The maximum score of the English language test is 110.

3.7.2 Procedure

The original plan for the pilot study involved the researcher initially testing each of the three treatment groups on both vocabulary pre-knowledge of the twenty target items for the two intervention sessions and listening comprehension skills. Then, each group received two intervention sessions in the following two weeks. At the end of each intervention session, there was a vocabulary post-test for the current session. One week after the second intervention session, each group received a vocabulary delayed post-test which was used to explore the long-term retention of these twenty target items and a listening comprehension post-test. During the week between the second intervention and the delayed post-tests, one learner within each treatment group was selected to participate in a stimulated recall interview.

However, because of practical reasons and accessibility to relevant teaching facilities, the first intervention for each group had to be conducted immediately after the vocabulary and listening pre-test, which was different from the original plan in which the first intervention should be administered one week after the pre-tests. In addition, the second intervention for the CFoF and L2 groups had to be delayed so that the vocabulary and listening delayed post-tests for these two groups were conducted only one day after the second intervention. Table 3.7 below shows the specific procedures for the pilot study.

Table 3.7

Procedure for the pilot study

Date	CS	CFoF	L2
Monday 20th April, 2015	Briefing and introducing the researcher	Briefing and introducing the researcher	Briefing and introducing the researcher
Wednesday 22nd April, 2015	Vocabulary pre-test; Listening pre-test; First intervention Vocabulary post-test		
Thursday 23rd April, 2015		Vocabulary pre-test; Listening pre-test; First intervention Vocabulary post-test	
Saturday 25th April, 2015			Vocabulary pre-test; Listening pre-test; First intervention Vocabulary post-test
Wednesday 29th April, 2015	Second intervention; Vocabulary post-test		
Wednesday 6th May, 2015	Vocabulary delayed post-test; Listening post-test		
Monday 11th May, 2015		Second intervention; Vocabulary post-test	
Tuesday 12th May, 2015			Second intervention; Vocabulary post-test
Wednesday 13th May, 2015	Stimulated recall	Stimulated recall	Stimulated recall
Thursday 14th May, 2015		Vocabulary delayed post-test; Listening post-test	Vocabulary delayed post-test; Listening post-test

3.7.3 Data analysis and results for the pilot study

To address the first research question, i.e., what is the impact of three different types of vocabulary instruction during listening activities (teacher codeswitching; target language

explanations; and contrastive Focus-on-Form) on vocabulary learning and retention, descriptive statistics for the three treatment groups on their performance in the vocabulary pre-test, post-test as well as delayed post-test are presented in Table 3.8.

Table 3.8
Descriptive statistics (%) for the vocabulary tests

Treatment group		Pre-test	Post-test	Delayed Post-test
CS	**Mean**	18.33	63.33	40.00
	SD	9.31	9.83	12.65
	Range	25.00	25.00	30.00
	Min	10.00	50.00	25.00
	Max	35.00	75.00	55.00
CFoF	**Mean**	9.17	65.83	35.00
	SD	7.36	17.72	21.45
	Range	20.00	45.00	45.00
	Min	0.00	40.00	15.00
	Max	20.00	85.00	60.00
L2	**Mean**	10.83	43.33	38.33
	SD	4.92	10.80	8.76
	Range	10.00	30.00	25.00
	Min	5.00	30.00	25.00
	Max	15.00	60.00	50.00

Considering that in the vocabulary pre-test, the participants were familiar with some of the target lexical items, it would be more reliable to analyse gain scores at the vocabulary post-test and delayed post-test. See Table 3.9 for descriptive statistics for vocabulary gain scores from the pre-test to the post-test and from the pre-test to the delayed post-test.

Table 3.9

Descriptive statistics (%) for gain scores from the pre-test to the post-test and from the pre-test to the delayed post-test

Treatment group		Post-test	Delayed Post-test
CS	**Mean**	45.00	21.67
	SD	6.32	9.31
	Range	15.00	25.00
	Min	40.00	15.00
	Max	55.00	40.00
CFoF	**Mean**	56.67	25.83
	SD	16.33	17.15
	Range	40.00	40.00
	Min	35.00	10.00
	Max	75.00	50.00
L2	**Mean**	32.50	27.50
	SD	11.29	10.37
	Range	30.00	25.00
	Min	15.00	10.00
	Max	45.00	35.00

As there were only six participants in each group and the data were not normally distributed, a Kruskal-Wallis test was employed to explore any differences among the three groups for both the post-test and the delayed post-test gain scores. Test results show that there was a statistically significant difference among the three groups for the vocabulary post-test [$X^2(2) = 6.21$, $p < .05$], but not for the delayed post-test [$X^2(2) = 0.53$, $p > .05$]. To determine where the significant difference was located for the vocabulary post-test, the Mann-Whitney test was used as a post-hoc test. The results indicate the CS group did not differ significantly from the CFoF group, nor the CS group from the L2 group, but that the CFoF group ($Mdn = 8.67$) significantly outperformed the L2 group ($Mdn = 4.33$), $U = 5.00$, $p < .05$. However, if a Bonferonni correction is applied to correct for multiple comparisons, reducing the alpha level to .016 (.05/3), the difference between the short-term vocabulary knowledge retention for the CFoF group and the L2 group is no longer statistically significant.

In order to answer the fourth research question: Does the impact of each type of instruction vary according to different word classes (nouns, verbs, adjectives), and for collocations, the Kruskal-Wallis test was used to explore whether the three different forms of instruction led to different learning outcomes for different word classes (nouns, verbs, adjectives) and for collocations in the vocabulary post-test as well as in the delayed post-test across the three treatment groups. The only significant difference found across the three treatment groups was for collocations in the vocabulary post-test [$X^2(2) = 13.50$, $p < .01$]. However, this difference may be only short-term as there was no significant difference for scores from the three groups for the vocabulary delayed post-test. The follow-up Mann-Whitney test was conducted to locate differences between the groups for collocations. Although no significant difference was found between the CS group and the CFoF group, the CS group (*Mdn* =9.50) performed significantly better than the L2 group (*Mdn* = 3.50), $U = 0.00$, $p < .01$, and the CFoF group (*Mdn* = 9.50) also significantly outperformed the L2 group (*Mdn* = 3.50), $U = 0.00$, $p < .01$.

3.7.4 Implications for the main study

The pilot phase did reveal a few potential problems with the original study design and relevant research instruments. First, the listening comprehension test had to be replaced by a more reliable and valid test. Details for this were discussed in Section 3.6.1. Second, results for the vocabulary pre-test in the pilot phase indicated that some of the target lexical items might be too easy for these first-year senior secondary school learners. As shown in the Table 3.10 below, three out of the eighteen participants in the pilot phase had known 20% of the target items before the intervention, and at least thirteen learners knew at least 10% of the target items. Further modifications had to be made to the listening passages and these target lexical items for the main study.

Table 3.10
Results for the vocabulary pre-test

	Score (%)	Frequency	Valid Percent
Vocabulary Pre-test	0.00	1.00	5.56
	5.00	4.00	22.22
	10.00	4.00	22.22
	15.00	5.00	27.78
	20.00	3.00	16.67
	35.00	1.00	5.56

Third, there was possible pre-test sensitisation discovered in the pilot phase, which then required extra modifications to be made to the design of the pre-test and the data collection procedure. Fourth, additional lexical items could be added into each intervention session (see Section 3.6.2 for details). The predicted time for each intervention session was 45 minutes, which is also the exact time period for a senior secondary school English lesson. However, the real situation in the pilot study was that each session could be completed within 35 minutes. In these 35 minutes, in addition to the target lexical items, a certain amount of non-target vocabulary was also added and explained by the researcher. As a result, although the main study would require extra time to hand out and collect vocabulary and listening tests as there would be far more learners in each treatment group, there could still be time for explaining more lexical items.

Fifth, for conducting the actual classroom intervention, it seemed to be quite difficult to adhere to the target language all the time even under the L2-only condition. The reason for this could be that some of the prepared explanations for relevant lexical items might be too difficult for the learners to understand. In order to address this, to ensure that the target language was used exclusively for the L2-only condition, further modifications were made for those difficult explanations. Besides, additional information, which may help the participants to understand, was added into the PPT document for each intervention session.

Finally, the amount of information obtained from the stimulated recall interview was relatively small. The original plan for the pilot study was to interview two learners in each

group. For each interview, two lexical items were used as stimuli. However, owing to time restrictions, only one individual interview session was done within each treatment group and each interview session focused on only one lexical item. It was found that the students being interviewed seemed to be nervous and reluctant to speak. It might be that they had only met the researcher for three time and they tended to treat the researcher as a teacher which then made them unwilling to express their opinions. In addition, it was possible that these participants were too afraid to say things that they felt were wrong, although they were assured prior to the interview that the interview was not a test and there was no correct or wrong answer for each question. Therefore, the plan for the stimulated recall interviews was modified for the main study, including more participants and interviewing about more lexical items. In addition, more sub-questions, probing answers given, were inserted to facilitate the main interview questions and to help learners to recall more deeply. Details for this were discussed in Section 3.6.6.

3.8 Data analysis

As discussed in the research design section, both quantitative and qualitative data were collected in the present study. Quantitative data were collected from the general vocabulary test, two listening comprehension tests, two school English tests, the vocabulary pre-test, six vocabulary post-tests, six vocabulary delayed post-test and the final vocabulary delayed post-test. Qualitative data were collected from the twelve stimulated recall interviews. After collecting the quantitative data, the first thing was to identify a marking scheme for each test instrument to make sure that all participants were treated equally for each test item.

3.8.1 Marking scheme

The general vocabulary knowledge test included four sections: 24 vocabulary items were tested in the first three sections respectively while 28 items were tested in the final section. Hence, there were altogether 100 test items in this test. The marking scheme for this test

was that only totally correct answers were accepted, with one mark for each, so that the maximum score for this test was 100. The scores for the two school English tests were obtained from the participants' class teacher. As it was not possible to retrieve the original test papers, the marking scheme used by the teachers who marked the tests was consulted. The scheme was exactly the same as that for the general vocabulary test. There were 100 test items and for each correct item, one mark was given. The maximum score was 100.

Since the format of the two listening comprehension tests was exactly the same, the same marking scheme was applied to both tests. However, one notable point was that combined with the general vocabulary knowledge test, the school English test, the listening comprehension pre-test scores were used to provide additional evidence for the participants' language proficiency level before the intervention. Hence, the marking scheme for the listening comprehension test should follow that for the general vocabulary test. There were 20 items in each listening comprehension test, therefore for each correct item, one mark was given. The maximum score for each listening comprehension test was 20. Unlike the general vocabulary knowledge test which only used multiple choice questions, there were four types of questions in the listening comprehension test: notes completion, table filling, sentence completion and multiple choice. Regarding the first three question types, the test takers were required to give the exact information according to what they heard by completing the notes, filling the tables or completing the sentences. Thus, extra caution was required when judging these answers.

The original plan was to follow the marking scheme of a standard IELTS test, where spelling mistakes are strictly not allowed. However, considering the general proficiency level of the participants at the present study, it was decided to treat each case of misspelling carefully. In the case where the spelling of the comprehension question's answer was given in the listening input (e.g., the speaker spelt the name of a specific train station in the listening input and the question asked is to write down the name of the train station), the misspelling was not allowed. Where the listening input gave no indication of spelling, and where any misspelling did not impede comprehension or alter the meaning of

the answer, it was accepted (e.g., the correct answer should be ‘motivation’, yet the test taker wrote down ‘motivetion’). The two listening comprehension tests were marked by the researcher twice, with a gap of six months in between. Results showed that the intra-rater reliability was high. Over 98% of the items were marked the same between the two time points. The items marked differently at two time points were sent to a second researcher for advice. Final agreement was made on the marking of every single item.

The marking scheme for the vocabulary post-, delayed post- and final delayed post-test was exactly the same. Correspondent to the six intervention sessions, there were six vocabulary post-tests and six delayed post-tests. The first thing to do was to combine these post-tests and delayed post-tests respectively. To ensure each group was treated equally, the participants were allowed to give the meaning of the target word either in L1 or in L2. Therefore, for each correct L1 or L2 meaning, one mark was given. The total score for each vocabulary test was 60.

As discussed earlier, the design of the vocabulary pre-test had a few limitations as the format of this test was not identical to that of the vocabulary post- and delayed post-test. Therefore, modifications were made when marking the vocabulary pre-test. Initially, the marking scheme followed that for the general vocabulary knowledge test, whereby one mark was given for each correct choice. Then, for each participant, the pre-test results were compared with the post-test. In cases where a participant selected the correct meaning for a particular vocabulary item in the pre-test but did not provide the correct answer for the same item in the post-test, the score obtained for this particular item in the pre-test was deleted. In addition, participants were asked an additional question in the post-test regarding whether they knew a particular vocabulary item before the intervention. This information was used to double confirm the modifications made.

3.8.2 Preparation for analysing the quantitative data

For analysing the quantitative data, the SPSS statistical package (Version 23) was adopted. Field (2013) argues that "when the data are quantitative this involves both looking at your data graphically … to see what the general trends in the data are, and fitting statistical model to the data" (p. 19). This indicates that before going into exploring which type of test to use for analysing quantitative data, it will be necessary to explore the frequency distribution or histogram, which will present a general tendency of the data collected. At the same time, it is also important to look into the centre of a distribution (e.g., the mode, the median and the mean). Then, the next stage will be deciding which specific statistical tests to use with the data collected.

Thus, for the present study, quantitative data collected from the three treatment groups were initially analysed descriptively. Then, specific statistical tests were adopted in order to explore any differences in the vocabulary test results among the three treatment groups. One question to be addressed was whether to use parametric or non-parametric tests. Field (2013) indicates that in order to decide which statistical model to fit in, it is important to investigate whether the distribution of data collected conforms to or violates the two assumptions: "normality of something or other" and "homoscedasticity/homogeneity of variance" (p. 165), looking at the results of a Kolmogorov-Smirnov or Shapiro-Wilks test for normality and Levene's test possibly supplemented with the variance ratio (Hartley's F_{max}). If the data conform to the two assumptions, parametric statistical tests can be adopted. Otherwise, non-parametric tests might be appropriate. Details for the selection of appropriate statistical tests and the assumptions of using correspondent tests will be further discussed in the Findings Chapter.

3.8.3 Preparation for analysing the qualitative data

For analysing the qualitative data, the twelve interview sessions were initially transcribed into written form and then translated into English. Regarding the coding of these

transcripts, a few issues were noted. Creswell (2009) argued that there might be several potential problems with analysing qualitative data. 1) Coding might strongly rely on first impressions. 2) Data that conflict with what the researcher expects to find might be ignored and too much attention paid to data that confirm what the researcher expected. 3) New data emerging at a later stage might be treated inappropriately.

Therefore, it was decided that the first thing to do with was to read these transcripts thoroughly several times and have a general idea about them, without any hypotheses or expectations. Notes were taken when the same information occurred more than once. Then, based on existing literature on vocabulary learning strategies and L2 listening strategies, an initial code book was drawn with reference to the notes taken from reading the transcripts. After that, following Macaro's (2014) procedures, the first round of initial coding was conducted by the researcher himself with three sample transcripts, one randomly selected from each treatment condition, concentrating on the last research question regarding the identification of participants' initial strategic reaction to the target lexical items when presented aurally as well as what strategies did they use in response to the teacher's vocabulary instruction. In cases where new codes merged, they were added to the initial code book.

Considering the limitations that Creswell (2009) raised and in an effort to maximise reliability, the present study involved a second round of initial coding with the three selected sample transcripts by a second researcher whose first language is Chinese independently. The reason for selecting a researcher who is fluent in Chinese for this second round of coding was mainly because the stimulated recall interviews were conducted in Chinese and the transcripts were then translated into English. For the second researcher, in addition to reading the English version of transcripts, she was asked to consult the original interview recording as much as possible especially in the case when she felt uncertainty with the English transcripts. The second researcher was also asked to note down the places where she was not able to match transcripts totally with any existing codes in the code book.

Then, the coding results from the second round were compared with those from the first round. Inter-rater reliability between two researchers' coding results was checked. Details for this are given in the Findings Chapter. A discussion took place in instances where the two researchers coded differently. In places where the two researchers were not able to reach an agreement were sent to a third researcher for advice. Finally, the initial code book was modified based on the discussion and rules for coding were applied to the coding all the transcripts.

3.9 Validity and reliability of the study design

Efforts were made to ensure that the design of the present study was valid and reliable. Firstly, all listening materials were selected from authentic English textbooks. Although a few modifications were made to these listening passages in order to fit the study design, a native speaker of English was consulted over these modifications. Besides, the six modified listening passages were recorded by three English native speakers to further ensure their authenticity. Secondly, there might be potential limitations with the general vocabulary knowledge test as only the first three sections and the academic word section were used in the present study. However, the main purpose of using this test was not to explore the precise vocabulary size for each participant, but was rather to provide additional evidence for learners' English language proficiency level.

Thirdly, potential limitations might also be posed by the two listening comprehension tests as only the first two sections of a standard IELTS listening test were used. However, taking participants' proficiency level into account, the primary reason for doing this was to make sure that the task was not too difficult for them. As outlined in the Findings Chapter, the mean score for the two listening comprehension tests was still relatively low even though the participants only completed the first two sections. In addition, the four sections in an IELTS listening test were independent of each other. Therefore, using part of the test was not likely to affect its validity and reliability.

Fourthly, extra efforts were made to prevent the participants from copying each other during the tests. As discussed above, the participants were invited to a multi-media classroom where they usually have group activities and presentations. Although they were already familiar with the classroom, the layout of the classroom was slightly changed by the researcher, making sure that there was a certain distance between each seat. In addition, the participants were told the tests were part of the classroom activities rather than the examinations they usually had in the school. They were also ensured that the test results would only be used for research purposes.

Finally, efforts were also made to ensure that the data collected from the stimulated recall interviews were valid and reliable. The interview sessions were placed at the end of the data collection procedure. Before that, the interviewees had met the researcher for ten times, so that they were familiar with the researcher. Additionally, the participants had full control of the laptop when watching the video clip showing the researcher teaching them a particular vocabulary item. They were allowed to stop or reply the video by themselves in order to help them to recall. Moreover, the interview started immediately after they finished watching the video clip.

3.10 Ethical issues

There were several reasons why ethical issues should be deeply considered in the present study. First, participants for the present study (as well as the pilot study) were first-year senior secondary school students who are aged 15/16. Second, the intervention sessions were video recorded. Third, some of these participants were invited to take part in an individual interview which was audio recorded.

Before conducting the main study, an information sheet and consent form in both English and Chinese were given to the participants, the head teacher, the class teachers as well as the participants' parents (as was done for the pilot study). They were given the following information: 1) the details of the study, 2) why they had been selected, 3) what would

happen if they take part, 4) risks and benefits for taking part, 5) what would happen to the data they provide, 6) who had reviewed the study, as well as 7) what would happen if they change mind, which they would be free to do at any time. After they understood all the above information, they were asked to tick relevant options in the consent form to give consent to the following actions: 1) completing relevant tests and tasks and attending classroom intervention, 2) providing details of English test results for students participating, 3) video-recording the intervention sessions, 4) taking part in a stimulated recall interview and 5) audio-recording the interview sessions.

All parties were informed that they could withdraw their consent at any time point and participants were allowed to stop at any stage of the research. In addition, they were also given the contact information of the researcher's supervisor and were allowed to complain or enquire about any additional information. Moreover, in order not to disadvantage any one group, following the completion of the intervention, relevant teaching materials were provided to the original class teachers. Participants in any one treatment group had access to those materials used by the other two groups.

Extra attention was paid to the Control group of the main study who did not receive any vocabulary treatment yet were tested several times on their vocabulary knowledge. However, it was still believed that the testing procedures for this group was ethical for several reasons. First, although no vocabulary treatment was given to the Control group, they received the listening comprehension task and heard the listening passages including the tested vocabulary items three times during each teaching session. Second, results from the immediate vocabulary post-test revealed that the Control group also made significant progress which indicated that they were also benefiting from their vocabulary knowledge even if they were not taught about this. Third, they were provided with the teaching materials used for all the three treatment groups immediately after the listening post-test when the intervention completed. Ethical approval documents are given in Appendix J and relevant information sheets and consent forms are attached in Appendix K.

CHAPTER 4 FINDINGS

4.1 Introduction

As stated in the Methodology chapter, the implementation of the main research study lasted three months. Both quantitative and qualitative data from four classes of senior secondary school EFL learners were collected before, during and after the intervention. Within the whole procedure, participants were required to complete two listening comprehension tests and five vocabulary knowledge tests, including a general vocabulary knowledge test, a vocabulary pre-test, post-test, delayed post-test and a vocabulary final delayed post-test. The vocabulary post-test and the delayed post-test were divided into six parts to fit in the teaching sessions. Qualitative data were collected from 12 selected learners through stimulated recall interview at the end of the data collection procedure. This chapter will initially summarise details of the quantitative data collected from the main study (see Table 4.1). Then, the five research questions will be restated and findings for each research question will be presented.

Table 4.1
Summary of the quantitative data collected from the main study

Tests	**Groups**			
	L2	**CS**	**CFoF**	**Control**
General vocabulary knowledge test	✓	✓	✓	✓
General listening comprehension pre-test	✓	✓	✓	✓
Vocabulary pre-test	✓	✓	✓	✓
Vocabulary post-test 1	✓	✓	✓	✓
Vocabulary post-test 2	✓	✓	✓	✓
Vocabulary post-test 3 and delayed post-test 1	✓	✓	✓	✓
Vocabulary post-test 4 and delayed post-test 2	✓	✓	✓	✓
Vocabulary post-test 5 and delayed post-test 3	✓	✓	✓	✓

Vocabulary post-test 6 and delayed post-test 4	✓	✓	✓	✓
Vocabulary delayed post-test 5	✓	✓	✓	✓
Vocabulary delayed post-test 6	✓	✓	✓	✓
General listening comprehension post-test	✓	✓	✓	✓
Vocabulary final delayed post-test	✓	✓	✓	

Note. The Control group did not receive the vocabulary final delayed post-test as its purpose was to explore the impact of repetitions, which they had not received during the teaching sessions.

4.2 Restating the research questions

1. What is the impact of three different types of vocabulary instruction during listening activities (teacher codeswitching; target language explanations; and contrastive Focus-on-Form) on vocabulary learning and retention?

2. What is the impact of each type of instruction on learners' listening comprehension?

3. Does the impact of each type of instruction vary according to learners' English language proficiency level?

4. Does the impact of each type of instruction vary according to different word classes (nouns, verbs, adjectives), and for collocations?

5. Does the number of repetitions influence vocabulary retention?

6. What learning strategies do learners use in response to the instruction for the target items offered by the teacher in each of the three experimental conditions?

4.3 Reliability

Before moving onto analysing the quantitative data, it was important to check the reliability of each test instrument used in the present study. The primary aim of checking the reliability is to make sure that test items used in a specific instrument all measure the same thing (Bland & Altman, 1997). A valuable way of doing this is to look at whether there is internal consistency among these test items using Cronbach's alpha (Cronbach, 1951).

First, as the present study used four sections of the listening vocabulary levels test (1000-3000 frequency levels and academic word list), for each participant, the test score for each section was calculated. These four sections were treated as four test items and their scores were submitted to SPSS as four variables respectively. Results indicated that the scale had a relatively low level of internal consistency, as determined by a Cronbach's alpha of .63. However, this was not regarded as hugely problematic. There were two reasons for this. On the one hand, the present study used only part of this vocabulary levels test, which might be a possible explanation for the fairly low internal-consistency. On the other hand, the current procedure for checking the reliability treated the four sections as four test items. However, there were altogether 100 test items within these four sections. It was likely that the Cronbach's alpha would be increased if using these 100 test items to form the scale.

Second, each listening comprehension test using the first two sections of an IELTS test included 20 comprehension questions. Hence, these 20 questions were treated as 20 test items in order to assess the reliability. Initially, for each participant, the original test paper was revisited. Each correct answer was marked as one in SPSS, while each wrong or blank answer was marked as zero. The scale had a relatively low level of internal consistency (Cronbach's alpha) for the listening comprehension pre-test (.62) as well as for the listening comprehension post-test (.60). This was however consistent with alphas found in IELTS research reports for the first two sections of the test, namely around .67 (Breeze & Miller, 2012).

Third, as required by the fourth research question, does the impact of each type of instruction vary according to different word classes (nouns, verbs, adjectives), and for collocations, the scores for nouns, verbs, adjective and collocations were calculated respectively within each of the three vocabulary tests (pre-test, post-test, delayed post-test). Therefore, the scores for the three word classes and for collocations were treated as four test items. The scale had a moderate level of internal consistency (Cronbach's alpha) for the vocabulary pre-test (.76), while it had a good level of internal consistency for the post-test (.94) and the delayed post-test (.92). Regarding the vocabulary final delayed post-test, the scores for the items which received nine, seven, five, three repetitions respectively were treated as four test items. The scale also had a moderate level of internal consistency (Cronbach's alpha) for the vocabulary final delayed post-test (.74).

4.4 Participants and baseline tests

As a quasi-experimental design was adopted, instead of randomly allocating the participants into different groups, the present study randomly assigned the four classes to three treatment groups, namely English-only (L2) group (38 students), Teacher Code-switching (CS) group (40 students) and contrastive Focus-on-Form (CFoF) group (37 students), as well as a Control group (37 students). At the beginning, these 152 learners all agreed to participate in the present study, with consent from all relevant parties. However, there were a few learners, within each group, who missed the fourth and the fifth intervention session due to having compulsory school activities at the time of the tests. It was therefore decided to delete data collected from these learners.

There were two reasons for doing this. First, for these learners, their vocabulary post-test and delayed post-test results would be underestimated as they missed two vocabulary post-tests and two delayed post-tests. It was not appropriate to include their data for analysing the first, third and fourth research question. Second, they did not receive the same amount of listening input, vocabulary instruction and review (culture teaching for the Control group) as the other learners in their groups who took part in all the six sessions. Therefore,

their scores might be artificially depressed if their data were included for analysing the second and the fifth research questions. As a result, data collected from 137 students were used at the main study, including 35 students from the L2 group, 36 students from the CS group and 33 students from the CFoF and the Control group respectively.

In addition, there were a few learners who did not take the final vocabulary delayed post-test and the listening comprehension post-test. It was decided to keep their data when analysing the first, third and fourth research questions which did not require data from these two tests. However, when addressing the second and fifth research questions, their data were deleted. Details will be given below when analysing relevant research questions.

Before the intervention sessions for the three treatment groups or the listening and culture sessions for the Control group, all four groups were required to complete two baseline tests, including a general vocabulary knowledge test (Listening vocabulary levels test) and a general English listening comprehension test (IELTS listening test). In addition, two of the participants' most recent school English tests (listening, reading, writing, speaking, grammar and vocabulary) scores were obtained. Firstly, descriptive statistics for the four groups on their performance in the two school English tests are given in Table 4.2.

Table 4.2

Descriptive statistics for the two school English tests

Group (N)		School English test A	School English test B
L2 (35)	**Mean**	81.50	73.94
	SD	9.02	10.13
	Range	40.00	39.00
	Min	55.00	51.00
	Max	95.00	90.00
CS (36)	**Mean**	78.65	70.92
	SD	7.91	8.71
	Range	35.00	41.00
	Min	58.00	52.00
	Max	93.00	93.00
CFoF (33)	**Mean**	73.70	68.11
	SD	9.67	8.15
	Range	55.00	38.00
	Min	37.00	48.00
	Max	92.00	86.00
Control (33)	**Mean**	75.26	69.70
	SD	9.73	8.14
	Range	43.00	31.00
	Min	47.00	53.00
	Max	90.00	84.00

Note. The maximum score for the school English test is 100.00.

One notable issue with the school English test was that it was not possible to revisit the test papers meaning that the researcher was not able to check the reliability (using Cronbach's alpha) of this test. In order to overcome this issue, it was firstly decided to check the correlation between the two school English tests. As there were five outliers detected in the data for the school English test A, a Spearman's rank-order correlation was run to assess the relationship between the two school English tests. Preliminary analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot. There was a strong positive correlation between the two tests, $r_s(135) = .71$, $p < .001$.

Thereafter, for each participant, the mean of the two school English test scores was calculated to formulate an average school English test score. It was decided to use these

average scores. The main reason for doing this was that, as the two school English tests were strongly correlated with each other and the test format and marking scheme was the same between the two tests, using the average score would be more reliable than using either one test score alone.

Hence, for each participant, the three baseline tests scores (general vocabulary knowledge test, general English comprehension test and the average school English test score) were combined to give a composite test score. The procedure for formulating this composite score was to follow what was did by Tian (2011), adding up the raw scores of the three baseline tests. This procedure was believed to be appropriate, given as discussed in the Methodology Chapter, the marking scheme for the three baseline tests was the same, i.e., one point was given to each correct answer. Descriptive statistics for the four groups' performance in the three baseline tests are presented in Table 4.3. A separate table (Table 4.4) gives the descriptive statistics for the participants' composite test scores.

Table 4.3

Descriptive statistics for the three baseline tests

Group (N)		General vocabulary knowledge test	Listening comprehension pre-test	Average school English test
L2 (35)	**Mean**	50.43	4.29	77.72
	SD	8.06	2.27	9.18
	Range	35.00	8.00	39.50
	Min	31.00	1.00	53.00
	Max	66.00	9.00	92.50
CS (36)	**Mean**	50.31	3.58	74.78
	SD	6.60	2.03	7.59
	Range	31.00	9.00	35.00
	Min	35.00	0.00	58.00
	Max	66.00	9.00	93.00
CFoF (33)	**Mean**	49.18	3.75	70.90
	SD	8.69	2.52	8.46
	Range	43.00	11.00	45.50
	Min	29.00	0.00	42.50
	Max	72.00	11.00	88.00
Control (33)	**Mean**	46.94	3.24	72.48
	SD	7.69	2.65	7.74
	Range	33.00	9.00	30.50
	Min	28.00	0.00	54.00
	Max	61.00	9.00	84.50

Note. The maximum score for the school English test and the general vocabulary knowledge test is 100.00, while for the listening comprehension pre-test is 20.00.

Table 4.4

Descriptive statistics for the composite score (combination of the three tests in Table 4.3)

Group (N)	Composite test				
	Mean	***SD***	**Range**	**Min**	**Max**
L2 (35)	132.44	15.86	58.50	100.50	159.00
CS (36)	128.67	13.14	68.00	98.00	166.00
CFoF (33)	123.84	16.18	65.50	93.50	159.00
Control (33)	122.66	11.63	42.50	102.00	144.50

Note. The maximum score is 220.00.

The composite test scores were submitted to a one-way ANOVA in SPSS (Version 23) to determine if the four groups were significantly different. Several assumptions should be met before using this statistical test: 1) There should be no outliers in each group's data; 2) the data should be approximately normally distributed in each group; 3) there should be homogeneity of variances (Field, 2013; Larson-Hall, 2010).

Initially, a boxplot was generated (see Figure 4.1). Outliers with values greater than 1.5 box-lengths but less than 3 box-lengths from the edge of the box were detected in the CS and CFoF group (there were none in the L2 and Control group). The general vocabulary knowledge and the listening comprehension tests of relevant outliers were revisited to make sure there were no entry or measurement errors. However, it was not possible to revisit the school English test. It was decided to keep the outliers and run tests without outliers to compare the results. Results from a Shapiro-Wilk test ($p > .05$) showed that the data were normally distributed among all groups. In addition, the assumption of homogeneity of variances was met, as ascertained by Levene's test of homogeneity of variances ($p = .30$).

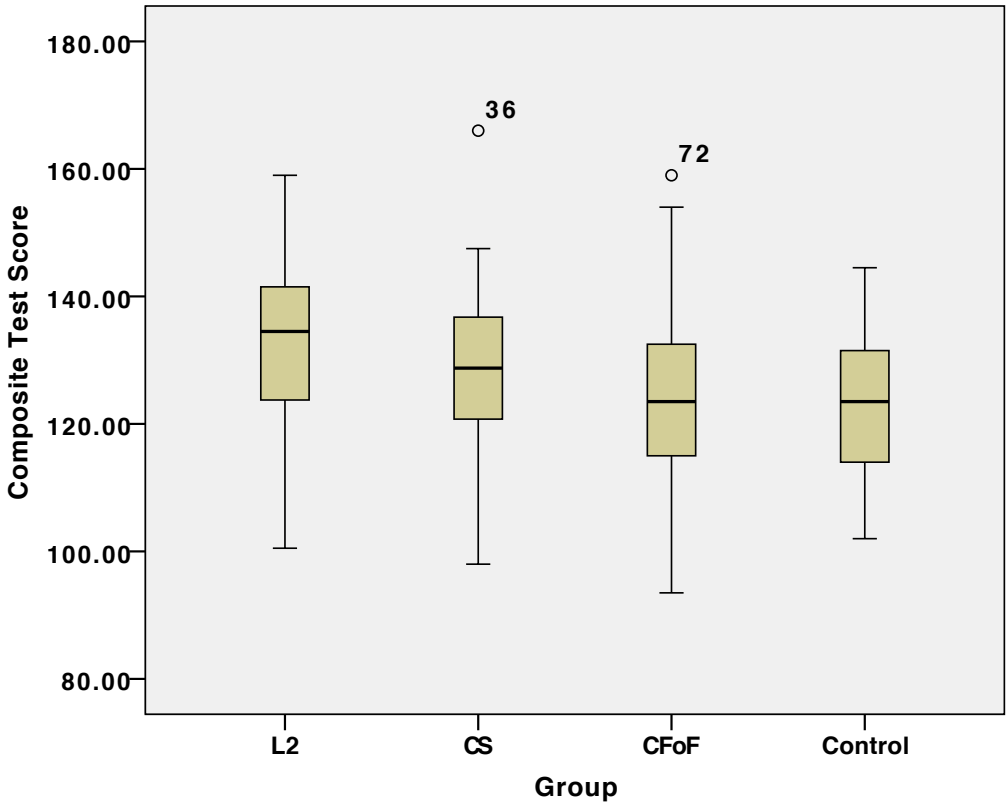


Figure 4.1 Boxplot for the composite test score for each group

Results from the one-way ANOVA indicated that there were significant differences among the four groups, $F(3, 133) = 3.39$, $p = .020$, $\eta_p^2 = .071$. The means plots are presented in Figure 4.2. The composite test scores decreased from the L2 group to the CS group, the CFoF group and the Control group, in that order. Post-hoc tests with Bonferroni correction revealed that the L2 group significantly outperformed the Control group in the composite test scores ($p = .027$), but no other group differences were significant ($p > .05$).

As the assumption of no existing outliers was violated, the outliers were then removed and the data were analysed again using a one-way ANOVA. The assumption of homogeneity of variances was met, as ascertained by Levene's test of homogeneity of variances ($p = .24$). Slightly different from the results generated with the outliers, in addition to finding significant differences between the L2 group and the Control group ($p = .023$), the L2 group significantly outperformed the CFoF group ($p = .027$). Therefore, it is quite safe to argue that there were significant differences among the four groups on the composite test

scores before the intervention, the differences were significant between the L2 group and the Control group, and might exist between the L2 group and the CFoF group. Note was taken of this limitation and it was decided to take it into account when analysing data for the first research question by including baseline test scores as covariates.

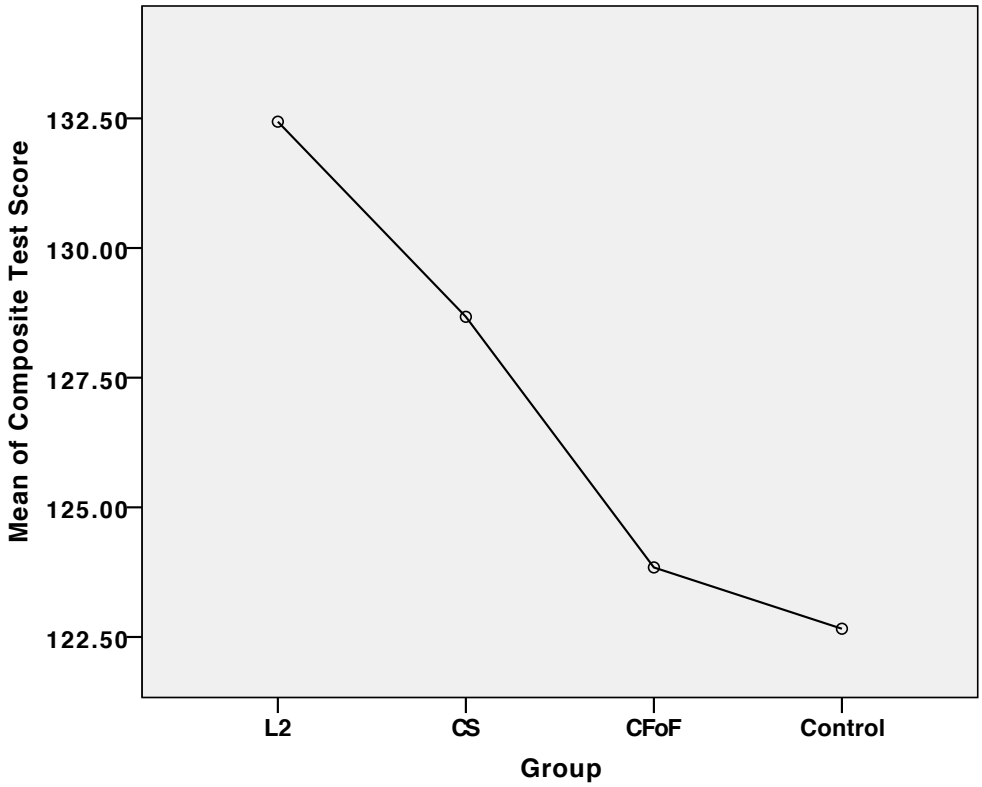


Figure 4.2 The means plots for the composite test scores for each group

4.4.1 Allocating proficiency levels across the treatment conditions

The scores for the composite test presented in Table 4.4 were used to assign these learners to proficiency levels. Initially, the four groups of learners were arranged in ascending order of their composite test scores. They were then divided into four proficiency level groups based on their scores: low, mid-low, mid-high and high. Since there were altogether 137 learners, it was not possible to equally divide them into four groups. In addition, learners who had the same composite test score should be in the same proficiency level group. Therefore, the number of learners allocated to each of the four proficiency levels varied

slightly across groups. Descriptive statistics for the composite test scores of the four proficiency levels are presented in Table 4.5.

Table 4.5
Descriptive statistics for the composite test scores of the four language proficiency levels

Proficiency levels (N)	**Composite test**				
	Mean	***SD***	**Range**	**Min**	**Max**
Low (34)	107.69	6.76	24.00	93.50	117.50
Mid-Low (35)	123.25	2.60	9.00	118.00	127.00
Mid-High (35)	131.93	2.46	9.00	127.50	136.50
High (33)	145.73	7.58	29.00	137.00	166.00

Note. The maximum score is 220.00.

Another notable issue here is that, although learners were assigned to different proficiency levels according to the combined general language test scores, it is not safe to argue these proficiency level groups are significantly different. Therefore, it was decided to employ relevant statistical tests to determine whether there were differences among the four different proficiency level groups. A Kruskal-Wallis test was adopted for two reasons. First, the data were not normally distributed across all the four proficiency level groups. Second, there were outliers in the high proficiency level group.

Distributions of the combined general language test scores were similar for all groups, as assessed by visual inspection of a boxplot. Median scores were statistically significantly different between the different proficiency levels, $\chi^2(3) = 127.53$, $p < .001$. After applying a Bonferroni correction for multiple comparisons (reducing the alpha level to .008), pairwise comparisons (Mann-Whitney U) were then performed indicating that there were significant differences between each of the two proficiency levels. Details for these are given in Table 4.6.

Table 4.6

Results for the pairwise comparisons between the proficiency levels

	High (*Mdn* = 169.00)	**Mid-High** (*Mdn* = 151.00)	**Mid-Low** (*Mdn* = 138.00)	**Low** (*Mdn* = 120.00)
High				
Mid-High	$p = .002$			
Mid-Low	$p < .001$	$p = .002$		
Low	$p < .001$	$p < .001$	$p = .002$	

The main purpose of categorising learners into the proficiency levels was to answer the third research question, exploring whether the impact of each type of instruction varies according to learners' English language proficiency level. Besides, in order to explore what learning strategies do learners use in response to the vocabulary instruction offered by the teacher in each of the three experimental conditions, the present study invited four learners in each treatment group, representing four proficiency levels, to participate in a stimulated recall interview.

However, there was one underlining problem with assigning learners to the proficiency levels across the four groups. As the present study did not randomly assign participants into groups using the baseline tests, significant differences were found between the L2 group and the Control group on the composite test scores. If the outliers were removed, significant differences were also detected between the L2 group and the CFoF group. As a result, the distribution of the four proficiency levels was quite different across the four groups (see Table 4.7).

Table 4.7

The distribution of proficiency levels in each treatment group

Group (N)	Proficiency Level	Frequency	Valid Percent
L2 (35)	**High**	14.00	40.00
	Mid-High	9.00	25.78
	Mid-Low	6.00	17.11
	Low	6.00	17.11
CS (36)	**High**	9.00	25.00
	Mid-High	12.00	33.33
	Mid-Low	8.00	22.22
	Low	7.00	19.44
CFoF (33)	**High**	6.00	18.18
	Mid-High	8.00	24.24
	Mid-Low	8.00	24.24
	Low	11.00	33.33
Control (33)	**High**	4.00	12.12
	Mid-High	6.00	18.18
	Mid-Low	13.00	39.39
	Low	10.00	30.30

4.4.2 Re-allocating proficiency levels within each treatment condition

It should be noted that one aim of the third research question was to explore whether there were significant group differences between the four proficiency levels on both short-term and long-term vocabulary learning within each type of treatment. Large differences between the number of participants in each proficiency level would have potential impacts on the statistical test results. Therefore, in addition to allocating learners into the four proficiency levels across the treatment conditions, it was decided to re-allocate the participants into proficiency levels within each treatment group, so that the number of participants in each proficiency level was approximately the same within each group.

Hence, the learners in the four groups were rearranged respectively according to the ascending order of their composite test scores. They were then divided into four proficiency level groups, so that the number of learners in each level within each treatment

condition was approximately the same. Table 4.8 gives the rearranged distribution of the proficiency levels in each group. Descriptive statistics for the composite test scores of the four proficiency levels within each group are presented in Table 4.9.

Table 4.8
The rearranged distribution of proficiency levels in each treatment group

Group (N)	Proficiency Level	Frequency	Valid Percent
L2 (35)	**High**	10.00	28.57
	Mid-High	8.00	22.86
	Mid-Low	9.00	25.71
	Low	9.00	25.71
CS (36)	**High**	9.00	25.00
	Mid-High	9.00	25.00
	Mid-Low	9.00	25.00
	Low	9.00	25.00
CFoF (33)	**High**	9.00	27.27
	Mid-High	8.00	24.24
	Mid-Low	8.00	24.24
	Low	8.00	24.24
Control (33)	**High**	9.00	27.27
	Mid-High	8.00	24.24
	Mid-Low	8.00	24.24
	Low	8.00	24.24

Table 4.9

Descriptive statistics for the composite test scores of the four language proficiency levels within each group

Group (N)	Proficiency Level	Range	Min	Max	Mean	*SD*
L2 (35)	**High**	17.50	141.50	159.00	150.15	6.44
	Mid-High	6.00	134.50	140.50	137.31	2.23
	Mid-Low	8.50	123.50	132.00	128.44	3.26
	Low	20.50	100.50	121.00	109.91	8.70
CS (36)	**High**	29.00	137.00	166.00	144.47	8.74
	Mid-High	7.50	129.00	136.50	132.56	2.57
	Mid-Low	7.50	121.00	128.50	125.50	2.40
	Low	22.50	98.00	120.50	112.17	7.12
CFoF (33)	**High**	26.50	132.50	159.00	142.78	9.93
	Mid-High	8.00	123.50	131.50	128.28	2.74
	Mid-Low	8.00	115.00	123.00	119.00	3.32
	Low	21.00	93.50	114.50	102.94	7.43
Control (33)	**High**	13.00	131.50	144.50	136.50	4.69
	Mid-High	7.50	123.50	131.00	126.16	2.31
	Mid-Low	8.50	114.00	122.50	119.69	2.93
	Low	9.50	102.00	111.50	106.56	3.55

Note. The maximum score is 220.00.

The next step was to employ relevant statistical tests to determine whether there were differences among these four reallocated proficiency levels. The original plan was to adopt a one-way ANOVA. Initially, a boxplot was generated (see Figure 4.3). No outliers were detected in each proficiency level group. Results from Shapiro-Wilk test ($p > .05$) showed that the data were normally distributed among all groups. However, the assumption of homogeneity of variances was not met, as ascertained by Levene's test of homogeneity of variances ($p < .001$). Therefore, a one-way Welch ANOVA was carried out.

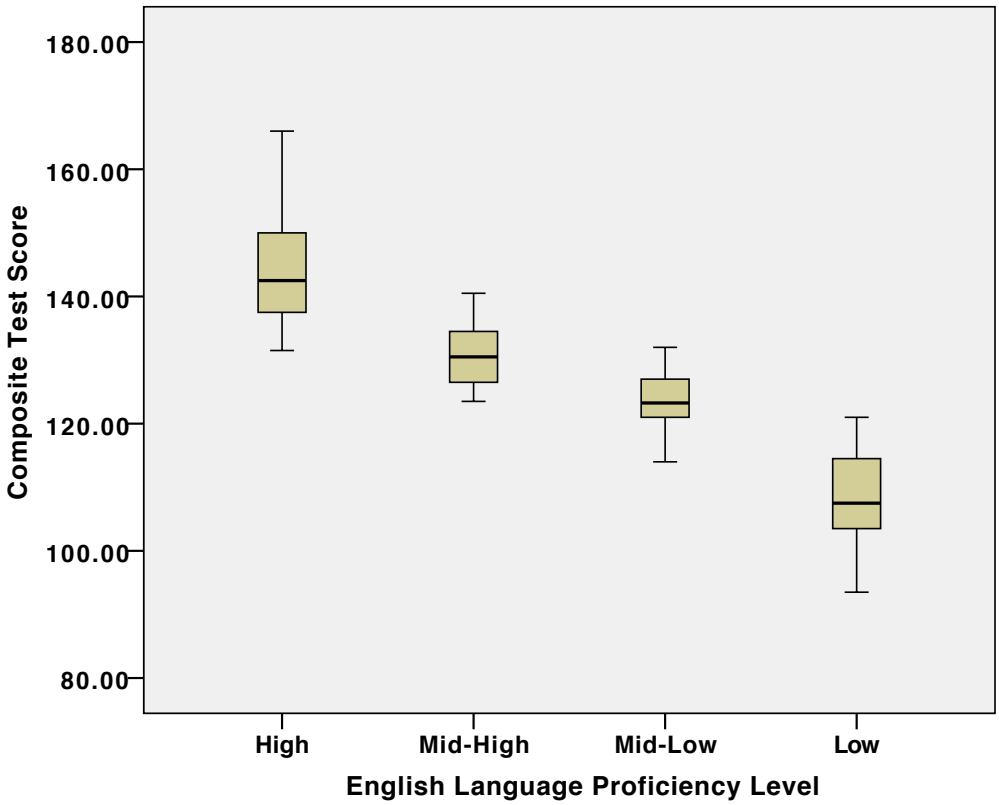


Figure 4.3 Boxplot for the composite test score for proficiency level

Results from the one-way Welch ANOVA indicated that there were significant differences among the four proficiency levels, $F(3, 72.61) = 124.15$, $p < .001$. The means plots are presented in Figure 4.4. The composite test scores decreased from the high proficiency level to mid-high proficiency level, mid-low proficiency level and low proficiency level, in that order. Games-Howell post hoc analysis was employed as the assumption of homogeneity of variances was not met, which revealed that there were significant differences between each two of the four proficiency levels. Details are given in Table 4.10.

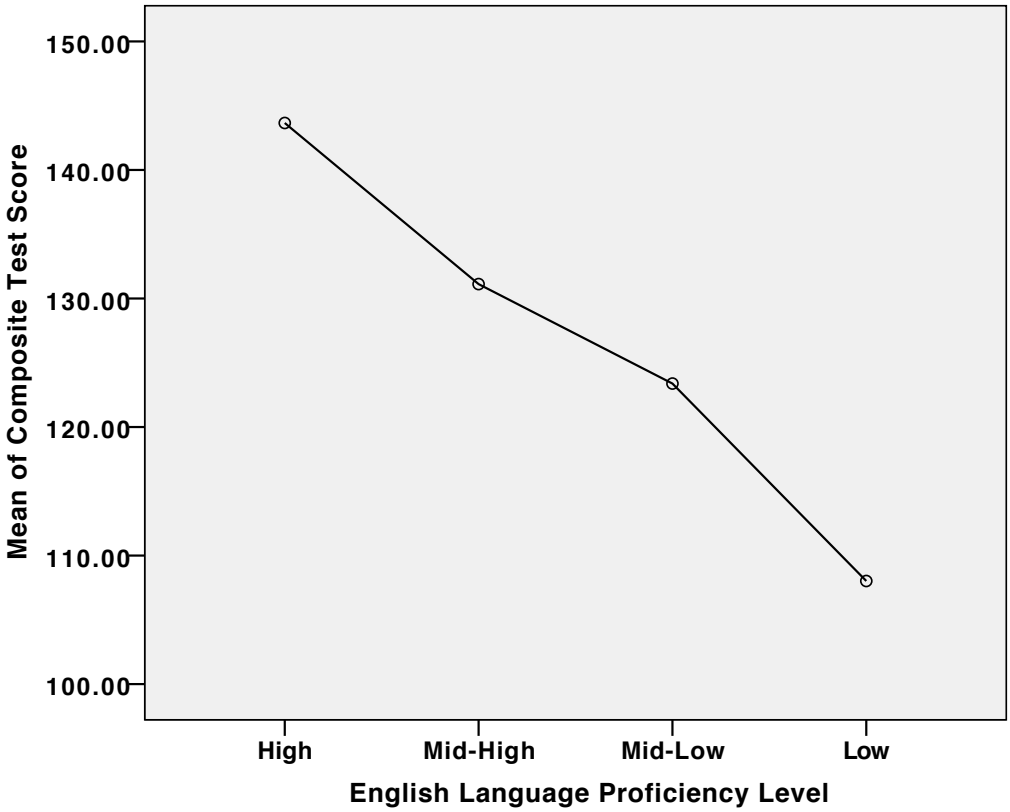


Figure 4.4 The means plots for the General language test scores for each proficiency level

Table 4.10
Results for the Games-Howell post hoc analysis

	High	Mid-High	Mid-Low	Low
High				
Mid-High	$p < .001$			
Mid-Low	$p < .001$	$p < .001$		
Low	$p < .001$	$p < .001$	$p < .001$	

In the findings reported below, post-hoc effect sizes are interpreted as follows: *d* - small = .20, medium = .50, large = .80; *r* – small = .10, medium = .30, large = .50 (Field, 2013).

4.5 Short-term and long-term vocabulary learning

In order to answer the first research question, quantitative data collected from the vocabulary tests were submitted to a two-way mixed ANOVA test in SPSS (Version 23) for further analysis. A two-way mixed ANOVA is a type of repeated-measures ANOVA which explores mean differences between groups that have been split on two independent variables (Larson-Hall, 2010). The two independent variables are called two factors in ANOVA, where one factor is a within-subjects factor and the other factor is a between-subjects factor. The main difference between the within-subjects factor and the between-subjects factor is that the former includes "related groups (also known as repeated measures), such as time (e.g., pre- and post-intervention)", while the latter includes "independent groups (e.g., gender: male/female)" (Laerd Statistics, 2015, p. 1). Therefore, in addition to exploring any significant differences among independent groups over time, the primary aim of conducting the two-way mixed ANOVA is to explore whether there is a two-way interaction between the two factors on the dependent variable (Field, 2013; Larson-Hall, 2010).

However, in order to use mixed ANOVA, several assumptions should be met:

1) The study design should have one dependent variable which is assessed at the continuous level; one between-subjects factor should consist of at least two unrelated categories and one within-subjects factor which should have at least two related categories (Larson-Hall, 2010);
2) "There should be no significant outliers in any cell of the design";
3) The "dependent variable should be approximately normally distributed for each cell of the design" (Laerd Statistics, 2015; p. 7);
4) There should be homogeneity of variance and covariance on the dependent variable between the groups of the between-subjects factor (Field, 2013; Larson-Hall, 2010);
5) There should be sphericity, which "measures whether differences between the variances of a single participant's data are equal" (Larson-Hall, 2010, p. 336).

It was decided that this statistical test was appropriate, as the study design and the measurements adopted met the requirement of the first assumption. In the present study, the means of the vocabulary test scores collected from the four different groups were compared at three time points: pre-test, post-test and delayed post-test, where the dependent variable is the vocabulary test scores, the within-subjects factor is the three test time points and the between-subjects factor is the four different groups. Through running this test, the mean differences of the vocabulary tests at different test time points as well as among different groups could be explored. In addition, this test could also explore whether there is an interaction between the test time points and the group difference.

Before moving onto exploring whether the data met the remaining assumptions, descriptive statistics for the vocabulary pre-test, post-test and delayed post-test were produced for all groups, as shown in Table 4.11, with scores for all six post-tests aggregated and all six delayed post-tests aggregated. One thing that should be clarified here is that, in order to address the limitations of the vocabulary pre-test discussed in the Methodology chapter, the scores of the vocabulary pre-test were modified as appropriate before analysing.

In the case when a participant selected the correct meaning for a particular vocabulary item in the pre-test but did not provide the correct answer for the same item in the post-test, the score obtained for this particular item in the pre-test was deleted. Additionally, since in the post-test, participants were asked to provided information on whether they had known a specific vocabulary item before having the intervention, evidence retrieved from this was used to doubly confirm the modification of the pre-test. Following these rules, participants in each group were treated equally regarding the modification of their pre-test scores. All participants were affected due to the fact that recognition precedes recall in vocabulary learning, therefore had to have their scores adjusted. Descriptive statistics were then performed on both the original vocabulary pre-test scores and the modified pre-test scores. Details for these are given in Table 4.12.

Table 4.11

Descriptive statistics for the vocabulary tests

Group (N)		Pre-test	Post-test 1-6 TOTAL	Delayed Post-test 1-6 TOTAL
L2 (35)	**Mean**	5.06	27.80	16.20
	SD	2.55	8.35	8.06
	Range	10.00	38.00	32.00
	Min	0.00	7.00	4.00
	Max	10.00	45.00	36.00
CS (36)	**Mean**	5.58	44.28	16.69
	SD	2.37	7.46	8.35
	Range	11.00	26.00	31.00
	Min	1.00	30.00	5.00
	Max	12.00	56.00	36.00
CFoF (33)	**Mean**	3.69	50.15	28.61
	SD	1.94	6.47	9.36
	Range	8.00	29.00	35.00
	Min	0.00	30.00	14.00
	Max	8.00	59.00	49.00
Control (33)	**Mean**	1.06	7.82	1.33
	SD	1.14	5.06	1.63
	Range	4.00	22.00	8.00
	Min	0.00	0.00	0.00
	Max	4.00	22.00	8.00

Note. The maximum score for the vocabulary test is 60.00

Table 4.12

Descriptive statistics for the original vocabulary pre-test and the modified vocabulary pre-test

Group (N)		Range	Min	Max	Mean	*SD*
L2	OVP	12.00	4.00	16.00	10.29	3.30
(35)	MVP	10.00	0.00	10.00	5.06	2.55
CS	OVP	19.00	2.00	21.00	11.64	4.85
(36)	MVP	11.00	1.00	12.00	5.58	2.37
CFoF	OVP	7.00	3.00	10.00	6.00	1.44
(33)	MVP	8.00	0.00	8.00	3.69	1.94
Control	OVP	7.00	2.00	9.00	5.57	1.73
(33)	MVP	4.00	0.00	4.00	1.06	1.14

Note. The maximum score for the vocabulary pre-test is 60.00. OVP – Original vocabulary pre-test. MVP – Modified vocabulary pre-test

With the purpose to test whether the data fit the second assumption, making sure that there are no outliers, a boxplot was generated for the vocabulary test scores for each group at each test time point. The boxplot is given in Figure 4.5.

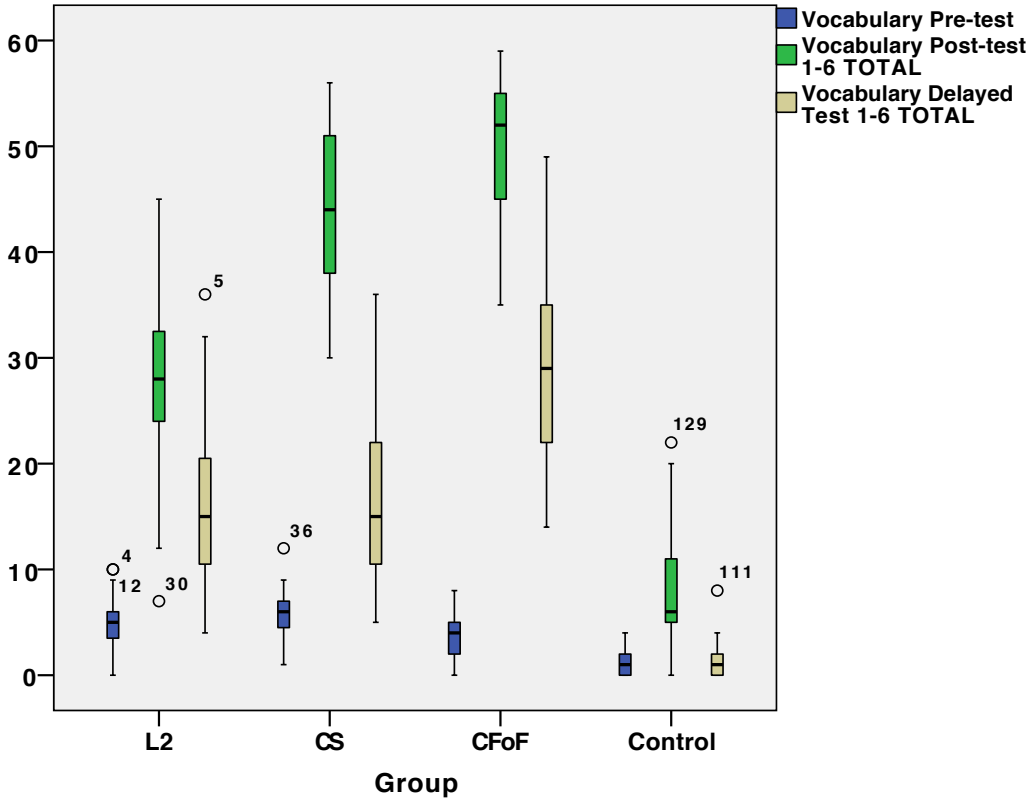


Figure 4.5 Boxplot for the vocabulary test scores for each group at three test points.

The boxplot shows that there were no outliers in the data for the CFoF group. However, outliers of values greater than 1.5 box-lengths but less than 3 box-lengths from the edge of the box were found in the L2, CS and Control group. Several steps were taken to address this. First, the relevant participants' original test scores were recalculated, indicating that there were no data entry errors. Second, details of each of the relevant written test results were revisited, which then eliminated the possibility of measurement errors. As a result, the seven outliers were genuinely unusual values.

It was decided to keep the outliers in the whole date set for several reasons. Initially, although these genuinely unusual data points may violate one or more assumptions of the two-way mixed ANOVA, they are not invalid data. It is argued that the specific data cannot

be removed simply because they do not fit the statistical model's assumptions (Draper & Smith, 1998; Faraway, 2015). Secondly, excluding any outliers which are neither entry nor measurement errors would have potential impacts on further data analysis (Laerd Statistics, 2015).

Thirdly, deleting existing outliers would simply result in the emergence of other outliers (Larson-Hall, 2010). Fourthly, removing outliers would mean the data themselves are not independent, which violates a basic assumption required by all statistical tests (Huber, 1981; Larson-Hall, 2010; Wilcox, 1998). Finally, since the vocabulary test results for analysing the first research question are only part of the data collected at the present study, deleting one outlier means removing all data collected from this outlier, which may have further impacts when analysing other research questions.

The normality of the vocabulary test scores for each group at each time point was assessed by Shapiro-Wilk's test, indicating that the test scores were normally distributed for the three treatment groups at each time point ($p > .05$), yet were not normal for the Control group ($p < .05$). In addition, results from the Levene's test of homogeneity of variances and Box's M test showed that the assumption of homogeneity of variance ($p < .05$) and the assumption of homogeneity of covariance ($p < .001$) were not met. It is clear that, at this stage, if the data from the Control group were included and analysed through a two-way mixed ANOVA, several important assumptions would be violated.

However, there is no alternative non-parametric tests for a two-way mixed ANOVA (Larson-Hall, 2010), which means it is not possible to run test comparisons. In addition, a two-way mixed ANOVA is the most appropriate statistical test, as it could determine if a two-way interaction exists which is of most interests in the present study. Therefore, it was decided at this stage to leave out the data collected from the Control group in order to use the two-way mixed ANOVA.

One thing that should be clarified here is that the data collected from the Control group were not deleted. They were temporarily removed from the whole dataset in order to use

the two-way mixed ANOVA which could explore the mean differences among the three treatment groups at three time points, so that the first research question could be addressed. However, the primary aim for collecting data from the Control group was to control other factors which might have influenced the vocabulary learning for the three treatment groups received classroom intervention. It is important to provide evidence for this through comparing the data from the Control group and from the three treatment groups. Hence, gain scores from the pre-test to the post-test and from the pre-test to the delayed post-test were calculated respectively. They were then submitted to relevant statistical tests and test results were used to compare with those from the two-way mixed ANOVA. Details for these are given in Section 4.5.1 and 4.5.2.

As the data for the experimental groups met the first three assumptions, they were submitted to the two-way mixed ANOVA. Results indicated that there was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2) = 4.50$, $p = .11$. The two-way mixed ANOVA indicated that there were main effects of both test time point: $F(2, 202) = 1292.73$, $p < .001$, $\eta_p^2 = .93$, and treatment group: $F(2, 101) = 40.29$, $p < .001$, $\eta_p^2 = .44$. A significant Test Time Point x Treatment Group interaction was also found: $F(4, 202) = 1592.17$, $p < .001$, $\eta_p^2 = .55$, meaning the groups progressed differently over time. The details of the interaction are presented in Figure 4.6. The two-way mixed ANOVA was then followed by Bonferroni corrected post-hoc tests, indicating that all groups made significant pre-delayed test gains ($p < .001$). At Time 2, all groups had significant gains ($p < .001$). At Time 1, the CS and L2 were not significantly different ($p = 1.00$); they were at post-test ($p < .001$, $d = 2.08$), to the advantage of the CS group, but not at delayed post-test ($p = 1.00$). The CFoF group was significantly ahead of the L2 group at Time 2 ($p < .001$, $d = 3.08$) and Time 3 ($p < .001$, $d = 1.45$), although not at Time 1 ($p > .05$). The CFoF group also significantly outperformed the CS group at Time 2 ($p = .003$, $d = .88$) and 3 ($p < .001$, $d = 1.38$), yet was outperformed by the CS group at Time 1 ($p = .003$, $d = .87$). In all cases effect sizes were large, particularly at Time 2.

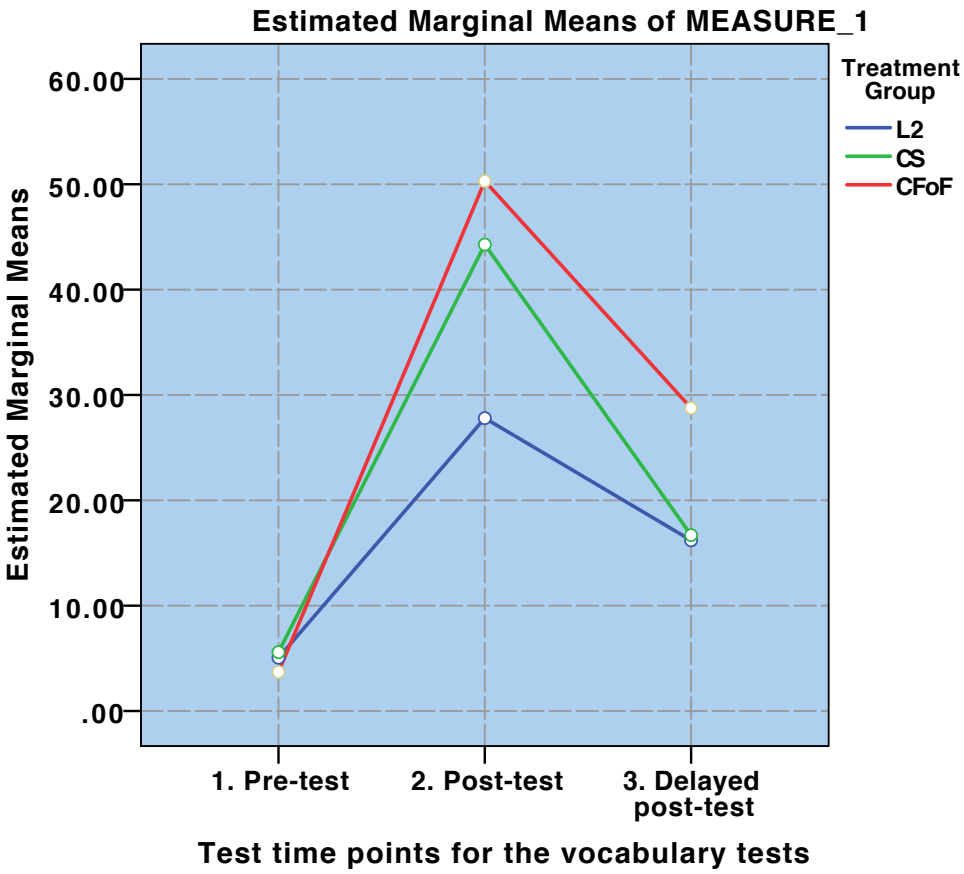


Figure 4.6 Profile plot for the vocabulary Test Time Point x Treatment Group interaction

Since the present study did not randomly allocate the participants into different groups, it was recognised that there were group differences among the three treatment groups before the intervention. Therefore, it was considered advisable to run the two-way mixed ANOVA with the general vocabulary test, the pre-listening test and the average school English test scores as covariates and compare the results with those from the tests without covariates. Similar results were generated for both the tests with and without covariates. Therefore, only the tests results from the two-way mixed ANOVA without covariates are presented here.

In addition, one important issue should be noted here is that the pre-test scores were modified in order to overcome the issue that recognition precedes recall in vocabulary learning. However, due to the fact that recall and recognition reflect different types of knowledge, original pre-test scores revealed partial knowledge of the target vocabulary

items before the intervention. Hence, it was considered worthwhile to run the two-way mixed ANOVA using the original pre-test scores as well. Similar results were produced regarding the performance of the three treatment groups.

The next stage was to compare the vocabulary learning outcomes from the three treatment groups with those from the Control group in order to rule out other factors which may affect the difference in the pre-test and post-test results. Gain scores for the vocabulary post-test and delayed post-test were calculated. Table 4.13 shows the descriptive statistics for gain scores from the pre-test to the post-test and from the pre-test to the delayed post-test.

Table 4.13

Descriptive statistics for gain scores from the pre-test to the post-test and from the pre-test to the delayed post-test

Group (N)		Pre to Post-test	Pre to Delayed Post-test
L2 (35)	**Mean**	22.74	11.14
	SD	7.46	6.84
	Range	35.00	29.00
	Min	4.00	1.00
	Max	39.00	30.00
CS (36)	**Mean**	38.69	11.11
	SD	7.15	7.93
	Range	29.00	29.00
	Min	25.00	0.00
	Max	54.00	29.00
CFoF (33)	**Mean**	46.61	25.06
	SD	5.41	9.01
	Range	26.00	31.00
	Min	29.00	11.00
	Max	55.00	42.00
Control (33)	**Mean**	6.76	0.27
	SD	4.96	0.88
	Range	22.00	4.00
	Min	0.00	0.00
	Max	22.00	4.00

Note. The maximum score is 60.00.

4.5.1 Differences in vocabulary gain scores from the pre-test to the post-test between groups

A one-way ANOVA was then conducted to determine if there were differences in gain scores from the pre-test to the post-test between groups. There were outliers of values greater than 1.5 box-lengths but less than 3 box-lengths from the edge of the box within the L2, CFoF and Control group, inspected by the box-plot given in Figure 4.7. It was decided to keep the outliers and reasons were the same as those given when analysing data using the two-way mixed ANOVA.

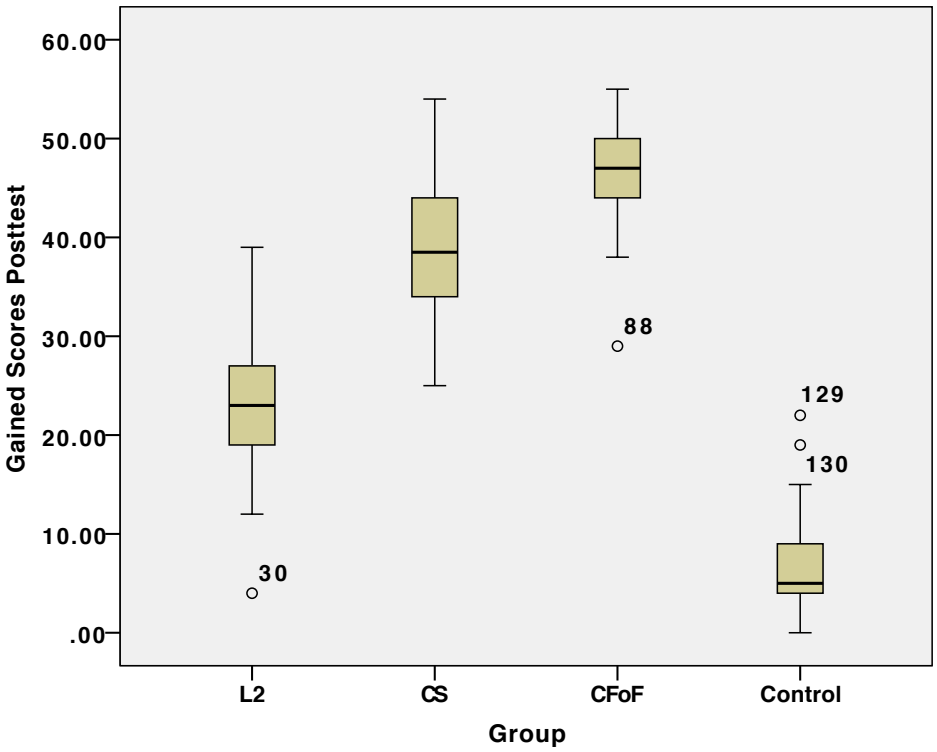


Figure 4.7 Boxplot for the gain scores from the pre-test to the post-test for the four groups

The data for the three treatment groups were normally distributed, assessed by Shapiro-Wilk's test ($p > .05$), yet were not normal for the Control group ($p < .05$). Further inspection of the histogram of the data from the Control group indicated that the distribution of the data was approximately normal. Since the one-way ANOVA is a fairly robust test, it was decided to carry on conducting the test. The assumption of homogeneity of variances was met, as assessed by Levene's test of homogeneity of variances ($p > .05$).

The vocabulary gain scores from the pre-test to the post-test were significantly different among the four groups, $F(3, 133) = 256.90$, $p < .001$, $\eta_p^2 = .85$. The means plots are given in Figure 4.8.

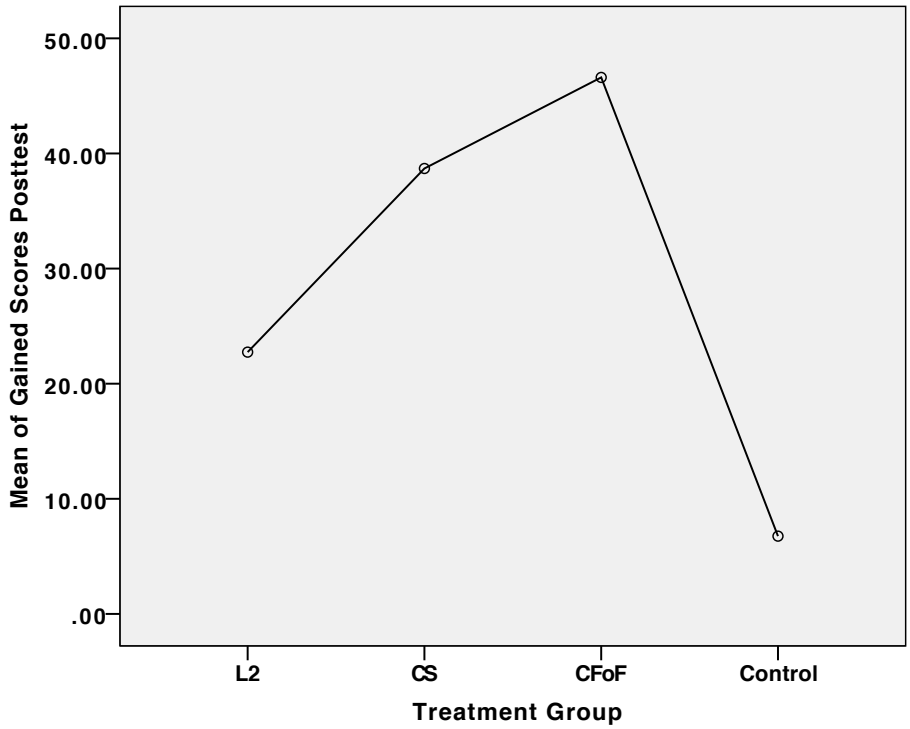


Figure 4.8 The means plot for the gain scores from the pre-test to the post-test

Post-hoc tests with Bonferroni correction revealed that the gain scores of the Control group were significantly lower than the L2 group ($p < .001$, $d = 2.51$), the CS group ($p < .001$, $d = 5.15$) as well as the CFoF group ($p < .001$, $d = 7.68$), in all cases with very large effect sizes. Post-hoc tests also revealed statistically significant differences among the three treatment groups. The gain scores of the CFoF group were significantly higher than both the L2 group ($p < .001$, $d = 3.65$) and the CS group ($p < .001$, $d = 1.24$), and the CS group significantly outperformed the L2 group ($p < .001$, $d = 2.19$). These results confirmed those of the two-way mixed ANOVA (involving just the three treatment groups), although the two tests are not exact alternatives (Larson-Hall, 2010).

4.5.2 Differences in vocabulary gain scores from the pre-test to the delayed post-test between groups

Kruskal-Wallis test was employed to determine whether there were significant differences between the Control group and the three treatment groups on the gain scores from the pre-test to the delayed post-test. There are several reasons for this. First, many outliers were detected in the data of the Control group. Second, both the data of the L2 group and the Control group were not normally distributed, assessed by Shapiro-Wilk's test ($p < .05$) and through observing the relevant histogram. Third, the results from the Levene's test of homogeneity of variances ($p < .05$) indicated that the assumption was violated. Although there are other post-hoc tests available when the assumption of homogeneity of variances is not met, it is better if a non-parametric test is employed as the data also had problems with normality and outliers.

Distributions of the gain scores at the delayed post-test were not similar for all groups, as assessed by visual inspection of a boxplot. The results from the Kruskal-Wallis test revealed that the mean ranks of gain scores from the pre-test to the delayed post-test were statistically significantly different between groups, $\chi^2(3) = 92.33$, $p < .001$. After a Bonferroni correction for multiple comparisons (reducing the alpha level to .008), pairwise comparisons (Mann-Whitney U) were then performed indicating that there were significant differences between the Control group (mean rank = 17.14) and the L2 group (mean rank = 50.87) ($U = 4.50$, $z = -7.24$, $p < .001$, $r = .85$), the Control group (mean rank = 19.42) and the CS group (mean rank = 49.28) ($U = 80.00$, $z = -6.55$, $p < .001$, $r = .74$) as well as between the Control group (mean rank = 17.00) and the CFoF group (mean rank = 50.00) ($U = .00$, $z = -7.30$, $p < .001$, $r = .86$). As was found through the two-way mixed ANOVA, the pre to delayed post-test gain scores of the CFoF group (mean rank = 48.09) were significantly higher than both the L2 group (mean rank = 21.69) ($U = 129.00$, $z = -5.51$, $p < .001$, $r = .67$). In addition, the CFoF group (mean rank = 48.27) significantly outperformed the CS group (mean rank = 22.83) ($U = 156.00$, $z = -5.27$, $p < .001$, $r = .63$), yet the difference between the L2 group and the CS group was not significant ($p > .05$).

These results confirmed those of the two-way mixed ANOVA (involving just the three treatment groups).

As the data for the Control group were not normally distributed, a non-parametric Friedman test was used to explore whether the Control group made significant progress or not at the post-test and the delayed post-test. Results indicated that there were significant differences between the three test time points, $\chi^2(2) = 53.96$, $p < .001$. Pairwise comparisons (Wilcoxon Signed Ranks Test, with Bonferroni correction reducing the alpha level to .016) indicated that there were significant differences between the pre-test (*Mdn* = 1.00) and the post-test (*Mdn* = 6.00) ($z = -4.87$, $p < .001$, $r = .85$) as well as between the post-test (*Mdn* = 6.00) and the delayed post-test (*Mdn* = 1.00) ($z = -4.97$, $p < .001$, $r = .87$), but not between the pre-test and the delayed post-test ($p = .07$).

To summarise the findings for the first research question, all four groups, including the three treatment groups and the Control group, made significant short-term vocabulary gains. The three treatment groups also made significant long-term vocabulary gains. However, the vocabulary learning of the Control group did not significantly improve after two weeks' delay. Regarding both short-term and long-term vocabulary acquisition, the three treatment groups significantly outperformed the Control group. Within the three treatment groups, the CFoF group's performance was significantly superior to the L2 and CS group on both short-term and long-term vocabulary learning. In addition, the CS group significantly outperformed the L2 group on short-term vocabulary retention. However, for long-term vocabulary acquisition there were no significantly statistical differences between the L2 group and the CS group.

4.6 Vocabulary intervention and listening comprehension

In order to address the second research question: What is the impact of each type of instruction on learners' listening comprehension, the data for the two listening comprehension tests were submitted to SPSS for further analysis. Descriptive tests were

conducted and results for these are given in Table 4.14. As two learners, one from the L2 group and one from the CFoF group, did not participate in the listening comprehension post-test, results for their listening comprehension pre-test were removed from the whole dataset.

Table 4.14
Descriptive statistics (%) for the listening comprehension pre-test and listening comprehension post-test

Group (N)	**Listening comprehension pre-test**				**Listening comprehension post-test**			
	Mean	***SD***	**Min**	**Max**	**Mean**	***SD***	**Min**	**Max**
L2 (34)	21.03	11.27	5.00	45.00	27.94	8.80	5.00	40.00
CS (36)	17.91	10.17	0.00	45.00	27.50	12.04	5.00	50.00
CFoF (32)	18.91	12.81	0.00	55.00	21.25	11.29	0.00	40.00
Control (33)	16.21	13.23	0.00	45.00	33.33	15.29	0.00	55.00

It was decided to employ a two-way mixed ANOVA for addressing this research question as the means of the listening comprehension test scores collected from the three treatment groups and the Control group were compared at two time points: pre-test and post-test. In addition to exploring the mean differences of the listening comprehension tests between the two test time points as well as among the four groups, the two-way mixed ANOVA could also explore whether there is an interaction between the test time points and the group difference.

Initially, a boxplot, given in Figure 4.9, was generated to see whether there were any outliers for the results of the two listening comprehension tests for each group.

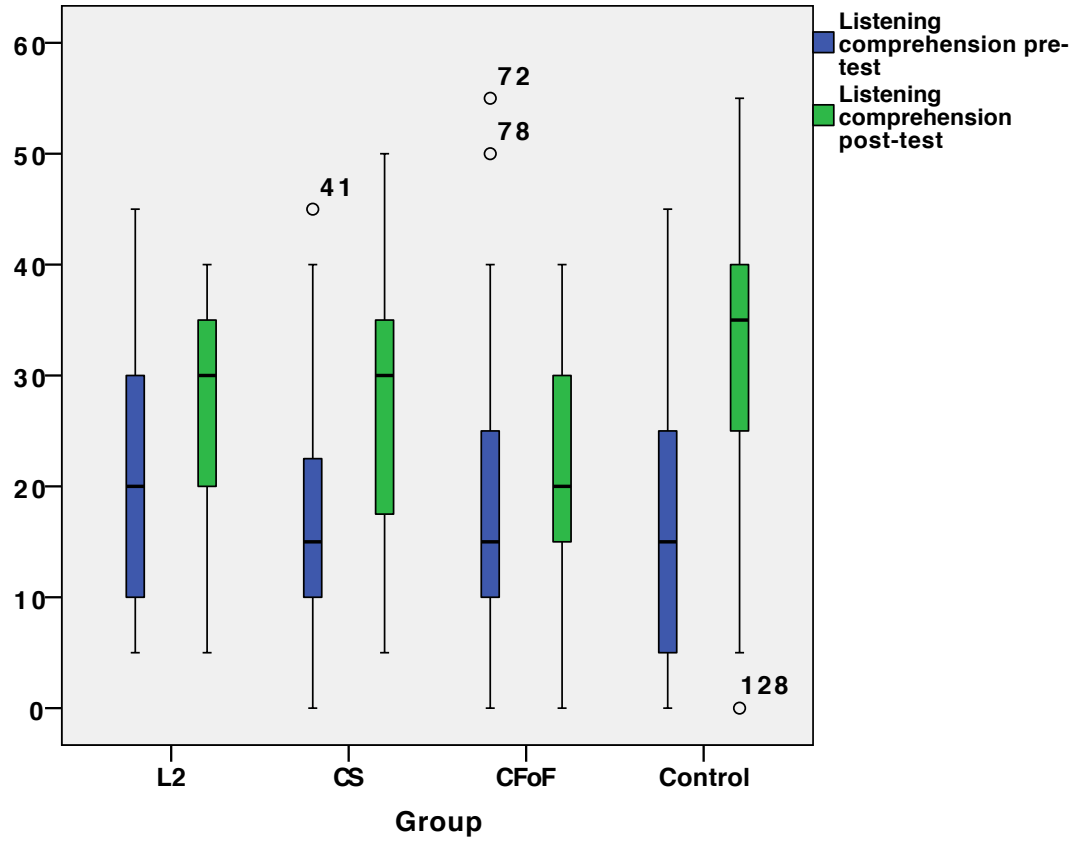


Figure 4.9 Boxplot for the listening comprehension test scores for each group at the two test points.

Four outliers of values greater than 1.5 box-lengths but less than 3 box-lengths from the edge of the box were found at the two listening comprehension tests, one from the CS group and two from the CFoF group at the pre-test, and one from the Control group at the post-test. As the relevant participants' original test scores were revisited indicating that there were no data entry or measurement errors, it was decided to conduct the tests both with and without outliers and compare the results.

Shapiro-Wilk's test indicated that the data for the listening comprehension post-test were normally distributed ($p > .05$), yet were not normal for the listening comprehension pre-test ($p < .05$). Further inspection of the histogram of data for the listening comprehension pre-test indicated that the distribution of the data was approximately normal. Since the two-way mixed ANOVA is a fairly robust test, it was decided to carry on conducting the test. Results from the Levene's test of homogeneity of variances and Box's M test showed that

the assumption of homogeneity of variance ($p > .05$) and the assumption of homogeneity of covariance ($p > .05$) were met.

The two-way mixed ANOVA indicated that there were main effects of test time point: $F(1, 131) = 57.36$, $p < .001$, $\eta_p^2 = .31$, but not of treatment group: $F(3, 131) = 1.58$, $p = .20$, $\eta_p^2 = .04$. A significant Test Time Point x Treatment Group interaction was also found: $F(3, 131) = 6.59$, $p < .001$, $\eta_p^2 = .13$, meaning the groups progressed differently over the two test points. The details of the interaction are presented in Figure 4.10.

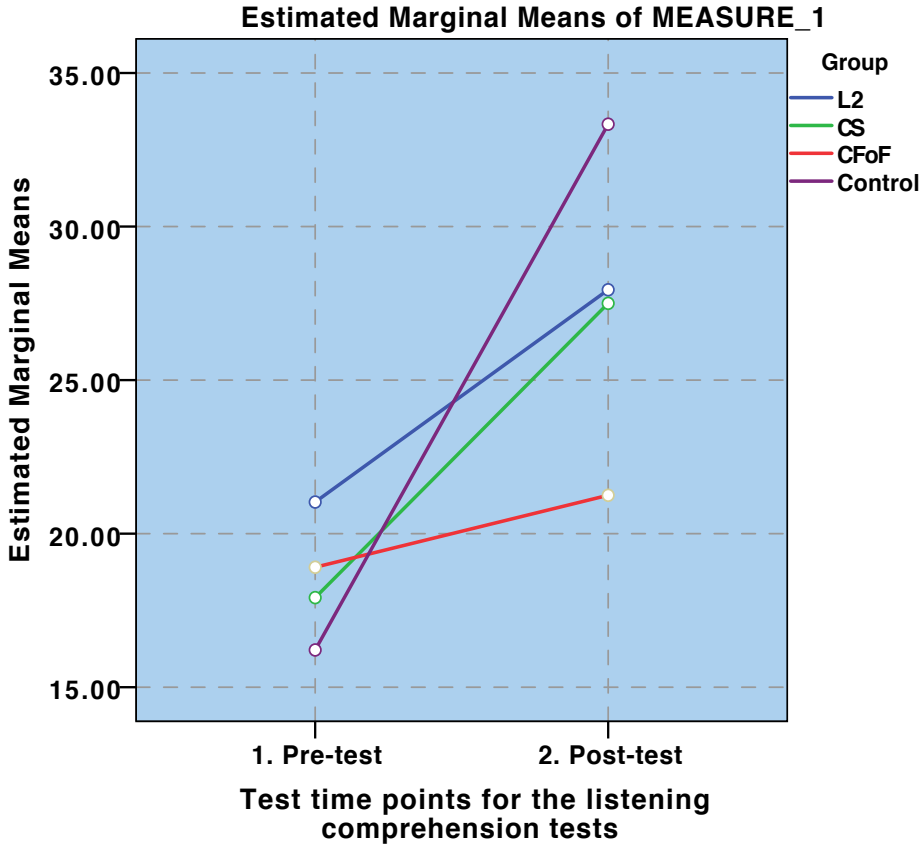


Figure 4.10 Profile plot for the listening comprehension Test Time Point x Treatment Group interaction

The two-way mixed ANOVA was then followed by Bonferroni corrected post-hoc tests, indicating that three groups made significant pre-post gains: the L2 group ($p = .004$, $d = .68$), the CS group ($p < .001$, $d = .86$) and the Control group ($p < .001$, $d = 1.20$). The CFoF group, however, did not ($p = .80$). There were no significant differences between any two groups at the listening comprehension pre-test ($p > .05$). At the post-test, however, the

CFoF group was significantly outperformed by the Control group ($p = .001$, $d = .90$). No significant differences were detected among the L2, CS and Control group at Time 2 ($p > .05$). At the same time, test comparisons were run without the outliers. Similar results were obtained.

As discussed in the Methodology Chapter, the present study employed the first two sections of an IELTS listening test. The task difficulty of the two sections within each test was different. The second section was more difficult than the first section. The above analysis looked at the mean differences of the two listening comprehension tests as a whole, which did not provide a clear picture of how the four groups performed in each section nor explore whether there were any significant group differences regarding one particular section of the listening test. It is, therefore, worth analysing the mean differences between the pre- and the post-test for each section separately.

One issue noticed before doing this analysis was that the same section of the two listening tests should be at approximately the same difficulty level. Initial inspection suggested that the two tests were quite similar. In both tests, Section One included ten local questions (listening for details), while Section Two had six global questions (listening in context) and four local questions. In addition, for Section One, the question type of the ten comprehension questions in both tests was filling blanks. For Section Two, there were four multiple choice questions and six filling the blank questions in the pre-test, while six multiple choice questions and four filling the blank question in the post-test. A factor analysis was then calculated to corroborate the similarity observed between the two tests, showing that the factors were well balanced across the two tests.

The next step was to calculate the raw scores for each section within each test. Descriptive statistics for this are presented in Table 4.15.

Table 4.15

Descriptive statistics for each section of the listening comprehension pre-test and listening comprehension post-test

Group (N)		Range	Min	Max	Mean	*SD*
L2 (34)	Pre-test Section 1	4.00	0.00	4.00	1.82	1.06
	Post-test Section 1	5.00	1.00	6.00	2.97	1.17
	Pre-test Section 2	8.00	0.00	8.00	2.47	1.69
	Post-test Section 2	5.00	0.00	5.00	2.62	1.46
CS (36)	Pre-test Section 1	4.00	0.00	4.00	1.22	1.10
	Post-test Section 1	5.00	0.00	5.00	2.47	1.28
	Pre-test Section 2	5.00	0.00	5.00	2.31	1.49
	Post-test Section 2	6.00	0.00	6.00	3.17	1.40
CFoF (32)	Pre-test Section 1	6.00	0.00	6.00	1.94	1.72
	Post-test Section 1	6.00	0.00	6.00	2.69	1.58
	Pre-test Section 2	5.00	0.00	5.00	1.84	1.14
	Post-test Section 2	4.00	0.00	4.00	1.56	1.27
Control (33)	Pre-test Section 1	5.00	0.00	5.00	1.56	1.52
	Post-test Section 1	7.00	0.00	7.00	3.82	1.73
	Pre-test Section 2	4.00	0.00	4.00	1.88	1.49
	Post-test Section 2	5.00	0.00	5.00	3.09	1.68

Note. The maximum score is 10.00

The data for pre-test Section One and post-test Section One were not normally distributed and many outliers were detected. Therefore, the data were split to compare groups. A Wilcoxon Signed Ranks Test was used to explore whether each group made significant progress from pre-test Section One to post-test Section One. The four groups all made significant gains. Details are given in Table 4.16 below.

Table 4.16

Results for the Wilcoxon Signed Ranks Test

Group (N)		Mean Rank	Wilcoxon Signed Ranks Test
L2 (34)	Pre-test Section 1	12.25	$z = 3.57$, $p < .001$, $r = .61$
	Post-test Section 1	14.88	
CS (36)	Pre-test Section 1	7.00	$z = 3.93$, $p < .001$, $r = .66$
	Post-test Section 1	15.22	
CFoF (32)	Pre-test Section 1	9.45	$z = 2.08$, $p = .04$, $r = .37$
	Post-test Section 1	16.03	
Control (33)	Pre-test Section 1	7.80	$z = 4.23$, $p < .001$, $r = .73$
	Post-test Section 1	18.11	

A Kruskal-Wallis test was then performed to explore whether groups were significantly different at the two test time points. The results revealed that at pre-test, the four groups were not significantly different ($p = .10$). The mean ranks of the post-test Section One scores, however, were significantly different between groups [$X^2(3) = 14.74$, $p = .002$]. After applying a Bonferroni correction for multiple comparisons (reducing the alpha level to .008), pairwise comparisons (Mann-Whitney U) were then performed indicating that there were significant differences between the CS group (mean rank = 27.43) and the Control group (mean rank = 44.04) ($U = 321.50$, $z = -3.47$, $p = .001$, $r = .41$), as well as between the CFoF group (mean rank = 27.13) and the Control group (mean rank = 39.50) ($U = 340.00$, $z = -2.66$, $p = .008$, $r = .32$).

The data for the pre-test Section Two and post-test Section Two were normally distributed as assessed by Shapiro-Wilk's test ($p > .05$) and no outliers were found in both the two sections. Therefore, they were submitted to a two-way mixed ANOVA. Results indicated that there was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. There were main effects of both test time point: $F(1, 132) = 9.08$, $p = .003$, $\eta_p^2 = .06$, and treatment group: $F(1, 101) = 5.36$, $p = .002$, $\eta_p^2 = .11$. A significant Test Time Point x Treatment

Group interaction was also found: $F(3, 132) = 4.34$, $p = .006$, $\eta_p^2 = .09$, meaning the groups progressed differently over time. The details for this interaction are given in Figure 4.11.

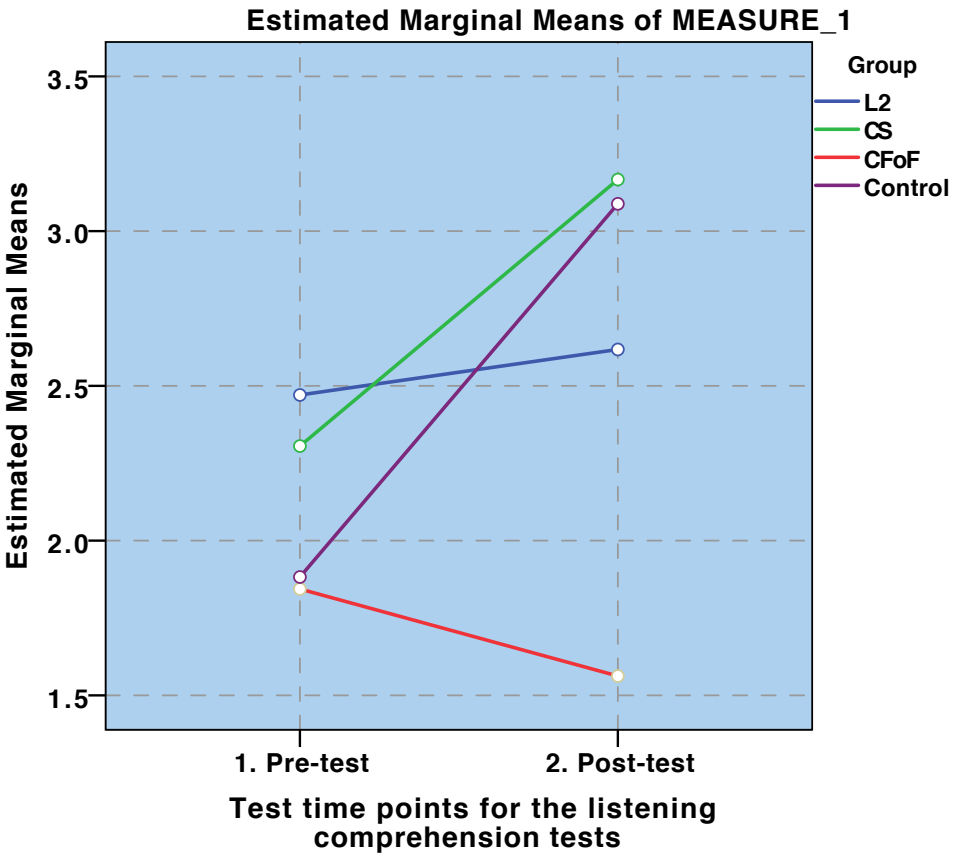


Figure 4.11 Profile plot for the Test Time Point x Treatment Group interaction for the listening comprehension test Section Two

The two-way mixed ANOVA was then followed by Bonferroni corrected post-hoc tests, indicating that the CS ($p = .006$, $d = .59$) and the Control group ($p < .001$, $d = .76$) made significant pre-delayed test gains, while the L2 and the CFoF group did not. At the pre-test, the four groups were not significantly different from each other. At post-test time two, however, the CFoF group was outperformed by the L2 ($p = .02$, $d = .77$), the CS ($p < .001$, $d = 1.20$) as well as the Control ($p < .001$, $d = 1.03$) group. No other group differences were detected.

To sum up, for Section One, all groups made significant pre-post gains. Although they were not significantly different at the pre-test, the Control group outperformed the CS and

the CFoF group at the post-test. Regarding Section Two, only the CS and the Control group made significant progress from the pre-test to the post-test, while the L2 and the CFoF did not. While there were no significant differences between the four groups at the pre-test, the CFoF group was significantly outperformed by the L2, CS and Control group at the post-test.

In general, it can be seen from the results of the statistical analysis that, after six sessions of listening comprehension and vocabulary teaching for the three treatment groups and six sessions of listening comprehension and culture study for the Control group, the L2, CS and Control group significantly improved in their listening comprehension ability. However, significant improvement was not confirmed with the CFoF group. In addition, although the four groups did not differ from each other in listening comprehension ability before having the sessions, the L2, CS and Control group performed significantly better than the CFoF group did on the listening comprehension post-test after the experimental sessions.

4.7 English language proficiency level and vocabulary acquisition through different types of instruction

In order to gain insights into whether the impact of each type of instruction varies according to learners' English language proficiency level, namely addressing the third research question of the present study, participants were assigned to four proficiency levels according to their performance on the general vocabulary knowledge test, the listening comprehension pre-test as well as the average school English test scores. There were two ways of interpreting data in order to address this research question. First, within each proficiency level, whether there were any significant differences of vocabulary learning between the three treatment conditions. Second, within each treatment condition, whether the vocabulary learning of the four proficiency levels were significantly different.

4.7.1 Vocabulary learning differences between treatment conditions within each proficiency level

Initially, learners were allocated to different proficiency levels across the treatment groups. However, as discussed in Section 4.4, there was one underlining problem with this. As there were group differences at the baseline tests, the distribution of learners in each proficiency level was quite different within each group. In order to minimise the effects of this issue, it was decided to merge the two higher levels (high, mid-high) and two lower levels (mid-low, low) respectively to form one high proficiency level and one low proficiency level. Thereafter, the data were split to files, representing the two new proficiency levels. The next step was to compare the vocabulary learning differences between treatment groups within each proficiency level.

Since the data were not normally distributed and there were many outliers in the high proficiency level, a Kruskal-Wallis test was adopted. The gain scores from the pre-test to the post-test and from the pre-test to the delayed post-test were calculated for each treatment group within each proficiency level. Details for these are given in Table 4.17.

Table 4.17

Descriptive statistics for the gain scores at the vocabulary post-test as well as at the delayed post-test for each treatment group within each proficiency level

Proficiency Level	Group (N)	Gain scores	Min	Max	Mean	*SD*	Median
High	L2 (24)	P-T	15.00	39.00	25.29	6.21	24.50
		D P-T	4.00	40.00	12.46	7.34	11.00
	CS (22)	P-T	25.00	48.00	38.77	6.44	38.50
		D P-T	0.00	27.00	9.95	7.27	8.50
	CFoF (14)	P-T	39.00	55.00	49.29	3.89	50.00
		D P-T	11.00	42.00	27.64	10.08	29.50
Low	L2 (11)	P-T	4.00	32.00	17.19	7.12	19.00
		D P-T	1.00	15.00	8.27	4.71	7.00
	CS (14)	P-T	25.00	54.00	38.57	8.40	38.00
		D P-T	0.00	29.00	12.93	8.84	13.50
	CFoF (19)	P-T	29.00	54.00	44.63	5.60	44.00
		D P-T	12.00	36.00	23.16	7.87	26.00

Note. The maximum score is 60.00. P-T means post-test, and D P-T means delayed post-test.

The gain scores were then submitted respectively to a Kruskal-Wallis test to detect whether there were any significant differences between different treatment groups within each proficiency level. Within each proficiency level, distributions of the gain scores from the pre-test to the post-test between different treatment groups were similar, as assessed by visual inspection of a boxplot. Results showed that median gain scores at the post-test were significantly different between the three treatment groups within the high proficiency level, $\chi^2(2) = 43.45$, $p < .001$ as well as within the low proficiency level, $\chi^2(2) = 26.11$, $p < .001$. Median gain scores at the delayed post-test were also significantly different between the three treatment groups within the high proficiency level, $\chi^2(2) = 21.65$, $p < .001$ as well as within the low proficiency level, $\chi^2(2) = 18.27$, $p < .001$.

Subsequently, after applying a Bonferroni correction for multiple comparisons (reducing the alpha level to .016), pairwise comparisons were performed using the Mann-Whitney U test. Details for these pairwise comparisons are presented in Table 4.18.

Table 4.18

Results for pairwise comparisons of the gain scores at the post-test (above the diagonal) and at the delayed post-test (below the diagonal) between treatment groups within each proficiency level

Prof. Level	Group	L2	CS	CFoF
High	L2	--	$U = 43.00$, $z = -4.87$, $p < .001$*, $r = .72$	$U = .50$, $z = -5.08$, $p < .001$*, $r = .82$
	CS	$U = 212.00$, $z = -1.15$, $p = .25$	--	$U = 17.00$, $z = -4.45$, $p < .001$*, $r = .74$
	CFoF	$U = 38.00$, $z = -3.94$, $p < .001$*, $r = .64$	$U = 24.00$, $z = -4.23$, $p < .001$*, $r = .70$	--
Low	L2	--	$U = 3.50$, $z = -4.03$, $p < .001$*, $r = .81$	$U = 1.00$, $z = -4.46$, $p < .001$*, $r = .81$
	CS	$U = 50.00$, $z = -1.48$, $p = .15$	--	$U = 71.00$, $z = -2.26$, $p = .024$
	CFoF	$U = 11.50$, $z = -4.01$, $p < .001$*, $r = .73$	$U = 56.00$, $z = -2.81$, $p = .004$*, $r = .49$	--

Note. * - *p*-value is statistically significant. Prof. - Proficiency

To summarise the findings here, regarding the short-term vocabulary acquisition of the high proficiency level learners, both the CS and CFoF group outperformed the L2 group. In addition, the high proficiency level learners in the CFoF group also outperformed their counterparts in the CS group. Regarding the low proficiency level learners' short-term vocabulary learning, the CS and CFoF group were significantly superior to the L2 group, yet there were no significant differences between the CFoF and the CS group. Regarding long-term vocabulary retention of both the high and low proficiency level learners, while the L2 and CS group were not significantly different, the CFoF group significantly outperformed the L2 and CS group.

4.7.2 Vocabulary learning differences between different proficiency levels within each treatment condition

In order to address the vocabulary learning differences between the proficiency levels within each treatment condition, the proficiency levels were rearranged within each treatment group. Thereafter, the data were split into files in SPSS, representing each treatment group and then the performance of learners from each proficiency level, within

each group, was explored. As there were only small numbers of participants in each cell of the data and not all data cells were normally distributed, the gain scores from the vocabulary pre-test to the post-test as well as from the pre-test to the delayed post-test were calculated for each proficiency level within each group. Descriptive statistic for this are presented in Table 4.19.

Table 4.19

Descriptive statistics for the gain scores at the vocabulary post-test as well as at the delayed post-test for each proficiency level within each group

Group (N)	Proficiency level (N)	Gain scores	Min	Max	Mean	*SD*
L2 (35)	High (10)	P-T	15.00	39.00	24.20	6.92
		D P-T	4.00	30.00	12.00	7.99
	Mid-High (8)	P-T	18.00	37.00	25.75	7.48
		D P-T	6.00	26.00	13.25	7.78
	Mid-Low (9)	P-T	20.00	32.00	25.78	4.18
		D P-T	4.00	20.00	11.78	5.70
	Low (8)	P-T	4.00	20.00	14.50	5.45
		D P-T	1.00	15.00	7.25	4.89
CS (36)	High (9)	P-T	34.00	48.00	40.56	5.00
		D P-T	8.00	27.00	13.44	6.80
	Mid-High (9)	P-T	25.00	47.00	38.00	7.45
		D P-T	3.00	18.00	8.56	5.41
	Mid-Low (9)	P-T	25.00	51.00	39.11	8.84
		D P-T	0.00	29.00	10.11	10.90
	Low (9)	P-T	29.00	54.00	37.11	7.57
		D P-T	0.00	25.00	12.33	8.05
CFoF (33)	High (9)	P-T	39.00	55.00	49.22	4.74
		D P-T	17.00	42.00	30.56	8.62
	Mid-High (8)	P-T	44.00	52.00	48.63	2.92
		D P-T	11.00	38.00	22.38	9.15
	Mid-Low (8)	P-T	29.00	51.00	44.50	6.85
		D P-T	16.00	36.00	25.25	7.25
	Low (8)	P-T	38.00	54.00	43.75	4.98
		D P-T	12.00	34.00	21.37	9.35
Control (33)	High (9)	P-T	2.00	12.00	6.22	3.27
		D P-T	0.00	3.00	0.33	1.00
	Mid-High (8)	P-T	3.00	11.00	5.75	2.82
		D P-T	0.00	0.00	0.00	0.00

Mid-Low (8)	P-T	3.00	22.00	11.63	6.72
	D P-T	0.00	4.00	0.50	1.41
Low (8)	P-T	0.00	7.00	3.50	2.56
	D P-T	0.00	1.00	0.25	0.46

Note. The maximum score is 60.00. P-T means post-test, and D P-T means delayed post-test

The gain scores were then submitted respectively to a Kruskal-Wallis test to detect whether there were any significant differences between different proficiency levels within each treatment group. Within the L2, CS and Control group, distributions of the gain scores from the pre-test to the post-test between different proficiency level groups were similar, as assessed by visual inspection of a boxplot. Results showed that median gain scores at the post-test were significantly different between the different levels of proficiency within the L2 group, $\chi^2(3) = 13.63$, $p = .003$; CS group, $\chi^2(3) = 7.89$, $p = .048$; as well as the Control, $\chi^2(3) = 8.79$, $p = .032$. However, the differences were not statistically significant at the delayed post-test within each of the three groups. Within the CFoF group however, no statistically significant differences were confirmed either at the post-test or delayed post-test.

Subsequently, after applying a Bonferroni correction for multiple comparisons (reducing the alpha level to .008), pairwise comparisons were performed using a Mann-Whitney U test. Results indicated that within the L2 group, there were significant differences in median gain scores at the post-test between the High (*Mdn* = 23.00) and Low (*Mdn* = 14.50) proficiency level ($p = .007$, $r = .64$), and between the Mid-High (*Mdn* = 23.50) and Low (*Mdn* = 14.50) proficiency level ($p = .007$, $r = .67$) groups, as well between the Mid-Low (*Mdn* = 25.00) and Low (*Mdn* = 14.50) ($p = .001$, $r = .82$), but not between any other proficiency levels. The effect sizes were large in all cases. However, within both the CS and the Control group, the median gain scores at the post-test were not significant different between any proficiency levels.

To sum up, within the CS and CFoF group, participants of the four proficiency levels progressed similarly at both the post-test and the delayed post-test, meaning that no learners of a particular proficiency level benefited more or were disadvantaged for both

short-term and long-term learning under the two types of vocabulary instruction. The situation was the same for the L2 group for long-term learning. However, for short-term vocabulary acquisition, the higher proficiency learners benefited more from the L2 condition than the lower proficiency level did.

4.8 The type of vocabulary instruction and different word classes and collocations

With the purpose to address the third research question, investigating whether the impact of each type of instruction varies according to different word classes (nouns, verbs, adjectives), and for collocations, 60 target lexical items including 18 nouns, 13 verbs, 12 adjectives as well as 17 collocations were selected across the six listening passages adopted for the classroom intervention. Participants in the present study were tested on their knowledge on these 60 items four weeks before, immediately after and two weeks after the intervention. Results were calculated according to the three word classes and for collocations at the three test time points. Regarding the fourth research question, there are two issues which are of most interest in the present study. First, within each treatment group, whether there are any differences in both short-term and long-term learning of collocations and the three word classes. Second, whether the learning outcomes of a specific word class or the collocations are different cross the three treatment groups. Therefore, the data were analysed through two different methods.

Assessing boxplots indicated that there were many outliers across each cell of the research design. In addition, the data were not normally distributed, examining by the Shapiro-Wilk's test of normality ($p < .05$). It was decided to use non-parametric tests to answer this research question. Initially, since the results of the vocabulary pre-test indicated that some participants had already known some of the vocabulary items before the intervention, gain scores from the pre-test to the post-test as well as from the pre-test to the delayed post-test were calculated according to the three word classes and for collocations. Then, descriptive

statistics were calculated to give an overview of this. These are given in Tables 4.20 and 4.21.

Table 4.20

Descriptive statistics (%) for the gain scores from the pre-test to the post-test according to the three word classes and for collocations

Group (N)		Noun	Verb	Adjective	Collocation
L2 (35)	**Mean**	33.49	31.43	37.38	48.07
	SD	13.44	15.54	15.31	18.95
	Min	11.11	0.00	8.33	0.00
	Max	66.67	69.23	66.67	88.24
CS (36)	**Mean**	58.18	53.63	62.96	80.55
	SD	12.91	19.93	14.56	13.10
	Min	33.33	15.38	33.33	47.06
	Max	88.89	92.31	91.67	100.00
CFoF (33)	**Mean**	76.60	64.34	82.32	84.85
	SD	11.52	13.16	13.62	11.86
	Min	38.89	30.77	41.67	47.06
	Max	100.00	84.62	100.00	100.00

Table 4.21

Descriptive statistics (%) for the gain scores from the pre-test to the delayed post-test according to the three word classes and for collocations

Group (N)		Noun	Verb	Adjective	Collocation
L2 (35)	**Mean**	18.89	12.53	14.05	26.22
	SD	15.72	12.80	12.91	14.29
	Min	0.00	0.00	0.00	0.00
	Max	55.56	53.85	41.67	64.71
CS (36)	**Mean**	15.74	14.53	12.73	31.37
	SD	15.96	15.36	12.03	13.92
	Min	0.00	0.00	0.00	11.76
	Max	61.11	46.15	41.67	64.71
CFoF (33)	**Mean**	46.80	31.47	31.82	50.45
	SD	20.46	16.02	16.73	18.78
	Min	11.11	7.69	8.33	17.65
	Max	72.22	61.54	75.00	88.24

Initially, the data were split according to treatment groups. Then, gain scores at the post-test were submitted to a Friedman test to determine if there were differences between the short-term learning outcomes of the different types of vocabulary item within each treatment group. Results indicated that gain scores at post-test were significantly different among the different types of vocabulary item within the L2 group [$\chi^2(3) = 23.49$, $p < .001$], the CS group [$\chi^2(3) = 55.87$, $p < .001$] as well as the CFoF group [$\chi^2(3) = 45.65$, $p < .001$]. Pairwise comparisons (Wilcoxon signed-rank test) were performed with a Bonferroni correction for multiple comparisons (reducing the alpha level to .008). Details of the pairwise comparisons where significant differences were confirmed are given in Table 4.22.

Table 4.22

Pairwise comparisons for the gain scores from the pre-test to the post-test according to the three word classes and for collocations within each treatment group

Group (N)	**Vocabulary type 1**	**Vocabulary type 2**	***z***	***p*-value**	***r***
L2 (35)	Adj. (*Mdn* = 33.33)	C. (*Mdn* = 52.94)	-2.75	= .006	.46
	V. (Mdn = 30.77)	C. (*Mdn* = 52.94)	-4.31	< .001	.73
	N. (*Mdn* = 33.33)	C. (*Mdn* = 52.94)	-3.64	< .001	.62
CS (36)	V. (*Mdn* = 53.85)	C. (*Mdn* = 82.35)	-5.20	< .001	.87
	N. (*Mdn* = 58.34)	C. (*Mdn* = 82.35)	-5.09	< .001	.85
	V. (*Mdn* = 53.85)	Adj. (*Mdn* = 58.33)	-3.07	= .002	.51
	Adj. (*Mdn* = 58.33)	C. (*Mdn* = 82.35)	-4.46	< .001	.74
CFoF (33)	V. (*Mdn* = 69.23)	N. (*Mdn* = 77.78)	-4.16	< .001	.72
	V. (*Mdn* = 69.23)	Adj. (*Mdn* = 83.33)	-4.84	< .001	.84
	V. (Mdn = 69.23)	C. (*Mdn* = 88.24)	-4.96	< .001	.86
	N. (*Mdn* = 77.78)	C. (*Mdn* = 88.24)	-3.33	= .001	.58

Note. N. = noun, V. = verb, Adj. = adjective, C. = collocation

In order to compare the long-term learning outcomes of the different types of vocabulary item within each treatment group, gain scores from the pre-test to the delayed post-test were submitted to the Friedman's test. Results showed that gain scores at the delayed post-test were significantly different among the different types of vocabulary item within the L2

group [$\chi^2(3) = 20.71$, $p < .001$], the CS group [$\chi^2(3) = 48.10$, $p < .00$] as well as the CFoF group [$\chi^2(3) = 48.32$, $p < .001$]. Pairwise comparisons (Wilcoxon signed-rank test) were performed with a Bonferroni correction for multiple comparisons (reducing the alpha level to .008). Details of the pairwise comparisons where significant differences were confirmed are given in Table 4.23.

Table 4.23

Pairwise comparisons for the gain scores from the pre-test to the delayed post-test according to the three word classes and for collocations within each treatment group

Group (N)	**Vocabulary type 1**	**Vocabulary type 2**	***z***	***p*-value**	***r***
L2 (35)	V. (*Mdn* = 7.69)	C. (*Mdn* = 29.41)	-4.27	< .001	.72
	V. (*Mdn* = 7.69)	N. (Mdn = 11.11)	-2.87	= .004	.48
	Adj. (*Mdn* = 8.33)	C. (*Mdn* = 29.41)	-3.87	< .001	.65
CS (36)	V. (*Mdn* = 7.69)	C. (*Mdn* = 29.41)	-4.66	< .001	.78
	N. (*Mdn* = 11.11)	C. (*Mdn* = 29.41)	-4.40	< .001	.73
	A. (*Mdn* = 8.33)	C. (*Mdn* = 29.41)	-5.22	< .001	.87
CFoF (33)	V. (*Mdn* = 30.77)	N. (*Mdn* = 50.00)	-4.03	< .001	.70
	Adj. (*Mdn* = 25.00)	N. (*Mdn* = 50.00)	-3.52	< .001	.61
	Adj. (*Mdn* = 25.00)	C. (*Mdn* = 47.06)	-4.47	< .001	.78
	V. (*Mdn* = 30.77)	C. (*Mdn* = 47.06)	-4.62	< .001	.80

Note. N. = noun, V. = verb, Adj. = adjective, C. = collocation

To summarise the results here:

1) Within the L2 group, both the short-term and long-term learning of collocations was better than the learning of verbs and adjectives. The superiority of the collocation retention over that of nouns was also confirmed for short-term learning. Additionally, retention of nouns outperformed retention of verbs for long-term learning.

2) Within the CS group, both the short-term and long-term learning of collocations significantly outperformed the learning of the three word classes. There was also a short-term advantage for learning adjectives over verbs.

3) Within the CFoF group, there was better retention of collocations and nouns compared with verbs for both short-term and long-term acquisition. In addition, the learning of collocations was better than learning of nouns in the short-term, while was better than adjectives in the long-term. Furthermore, the short-term learning of adjectives was superior to the learning of verbs. Finally, the long-term learning of nouns was better than the learning of the adjectives.

With the purpose of addressing whether the short-term learning outcomes of a specific word class or the collocations are different across the three treatment groups, a Kruskal-Wallis test was adopted. Distributions of the gain scores for nouns, verbs, adjectives as well as collocations at the post-test were similar for all treatment groups, as assessed by visual inspection of a boxplot. Median scores were statistically significantly different between the three treatment groups on the gain scores of nouns [$\chi^2(2) = 69.77$, $p < .001$], verb [$\chi^2(2) = 42.66$, $p < .001$], adjectives [$\chi^2(2) = 63.97$, $p < .001$] and collocations [$\chi^2(2) = 56.37$, $p < .001$]. After a Bonferroni correction for multiple comparisons (reducing the alpha level to .016), pairwise comparisons were then performed using Mann-Whitney U test. Details for this are presented in Table 4.24.

Table 4.24

Pairwise comparisons for the gain scores from the pre-test to the post-test across the three treatment groups according to the three word classes and for collocations

	Group 1	Group 2	*U*	*z*	*p*	*r*
	L2 (*Mdn* = 33.33)	CS (*Mdn* = 58.34)	125.00	-5.84	< .001*	.69
N.	L2 (*Mdn* = 33.33)	CFoF (*Mdn* = 77.78)	17.00	-6.90	< .001*	.83
	CS (*Mdn* = 58.34)	CFoF (*Mdn* = 77.78)	151.00	-5.36	< .001*	.64
	L2 (*Mdn* = 30.77)	CS (*Mdn* = 53.85)	247.50	-4.34	< .001*	.52
V.	L2 (*Mdn* = 30.77)	CFoF (*Mdn* = 69.23)	72.00	-6.24	< .001*	.75
	CS (*Mdn* = 53.85)	CFoF (*Mdn* = 69.23)	401.00	-2.35	= .019	/
	L2 (*Mdn* = 33.33)	CS (*Mdn* = 58.33)	154.00	-5.53	< .001*	.65
Adj.	L2 (*Mdn* = 33.33)	CFoF (*Mdn* = 83.33)	25.50	-6.81	< .001*	.82
	CS (*Mdn* = 58.33)	CFoF (*Mdn* = 83.33)	199.00	-4.80	< .001*	.57
	L2 (*Mdn* = 52.94)	CS (*Mdn* = 82.35)	83.50	-6.31	< .001*	.75
C.	L2 (*Mdn* = 52.94)	CFoF (*Mdn* = 88.24)	50.00	-6.50	< .001*	.78
	CS (*Mdn* = 82.35)	CFoF (*Mdn* = 88.24)	467.00	-1.55	= .12	/

Note. N. = noun, V. = verb, Adj. = adjective, C. = collocation; * - *p*-value is significant
/ - Effect size was not calculated as p-value was not significant.

The Kruskal-Wallis test was also employed to explore whether the long-term learning outcomes of a specific word class or the collocations were significantly different across the three treatment groups. Assessing by visual inspection of a boxplot, distributions of the gain scores for nouns, verbs, adjectives as well as collocations at the delayed post-test were similar for all treatment groups. Median scores were significantly different between the three treatment groups on the gain scores of nouns [$\chi^2(2) = 36.31$, $p < .001$], verbs [$\chi^2(2) = 24.68$, $p < .001$], adjectives [$\chi^2(2) = 27.99$, $p < .001$] and collocations [$\chi^2(2) = 28.35$, $p < .001$] at the delayed post-test. After a Bonferroni correction for multiple comparisons (reducing the alpha level to .016), pairwise comparisons were then performed using Mann-Whitney U test. Details for these are presented in Table 4.25.

Table 4.25

Pairwise comparisons for the gain scores from the pre-test to the delayed post-test across the three treatment groups according to the three word classes and for collocations

	Group 1	**Group 2**	***U***	***z***	***p***	***r***
	L2 (*Mdn* = 11.11)	CS (*Mdn* = 11.11)	533.00	-1.13	= .26	/
N.	L2 (*Mdn* = 11.11)	CFoF (*Mdn* = 50.00)	175.00	-4.96	< .001*	.60
	CS (*Mdn* = 11.11)	CFoF (*Mdn* = 50.00)	147.50	-5.40	< .001*	.65
	L2 (*Mdn* = 7.69)	CS (*Mdn* = 7.69)	602.50	-.33	= .74	/
V.	L2 (*Mdn* = 7.69)	CFoF (*Mdn* = 30.77)	208.00	-4.59	< .001*	.55
	CS (*Mdn* = 7.69)	CFoF (*Mdn* = 69.23)	265.00	-4.00	< .001*	.48
	L2 (*Mdn* = 8.33)	CS (*Mdn* = 8.33)	604.50	-.30	= .76	/
Adj.	L2 (*Mdn* = 8.33)	CFoF (*Mdn* = 25.00)	234.00	-4.27	< .001*	.51
	CS (*Mdn* = 8.33)	CFoF (*Mdn* = 25.00)	193.50	-4.89	< .001*	.58
	L2 (*Mdn* = 29.41)	CS (*Mdn* = 29.41)	522.00	-1.26	= .21	/
C.	L2 (*Mdn* = 29.41)	CFoF (*Mdn* = 47.06)	182.00	-4.88	< .001*	.59
	CS (*Mdn* = 29.41)	CFoF (*Mdn* = 47.06)	248.50	-4.18	< .001*	.50

Note. N. = noun, V. = verb, Adj. = adjective, C. = collocation; * *p*-value is significant
/ - Effect size was not calculated as p-value was not significant.

In summary, for short-term learning of the three word classes and for collocations, both the CS and CFoF group outperformed the L2 group. The CFoF group was superior to the CS group in the learning of nouns and adjectives, yet this was not true for the learning of verbs and collocations. For long-term learning, the CFoF group was significantly better than both the L2 group and the CS group in the retention of the three word classes and for collocations. The L2 group and the CS group were not significantly different in the learning of the three word classes nor for learning collocations.

4.9 The number of repetitions and vocabulary retention

In addition to exploring whether the three types of vocabulary instruction differ in the impact on vocabulary learning and retention, the present study also explored whether the number of repetitions influences vocabulary retention. In order to do this, a short review activity was implemented within each intervention session. In addition, there were two extra review sessions one week and two weeks after the final intervention session. The details of the review activity were discussed in the Methodology Chapter, Section 3.6.5. A final delayed post-test was conducted one week after the second review session which collected the main evidence for the impact of the different number of repetitions.

One thing that should be noted is that the vocabulary items reviewed were only single words. There were five words in the nine, seven and five repetitions groups respectively and the remaining 28 single words were repeated three times. Therefore, when exploring this research question, the gain scores from the pre-test to the final delayed post-test for the single words repeated nine times, seven times, five times as well as three times were calculated respectively. The descriptive statistics of the gain scores from the pre-test to the final delayed test of the different repetition groups are given in Table 4.26.

One limitation should be mentioned. Owing to the restriction of time limits for each intervention session, the actual time spent on the review activity was fairly short. Therefore, only a small number of vocabulary items could receive a high amount of repetitions. As mentioned above, there were five words in the nine, seven and five repetitions groups respectively. However, the number of single words repeated three times was 28, which was much larger than the number of words receiving higher repetitions. This may have potential drawbacks for the statistical tests adopted when analysing data for this research question, because of the unequal size of groups of words.

Table 4.26

Descriptive statistics (%) of the gain scores from the pre-test to the final delayed of the different number of repetitions groups

	Nine Repetitions	Seven Repetitions	Five Repetitions	Three Repetitions
Mean	59.81	23.08	22.50	16.07
SD	29.03	26.41	24.56	14.29
Range	100.00	100.00	80.00	60.71
Min	0.00	0.00	0.00	0.00
Max	100.00	100.00	80.00	60.71

Since the data for all the four groups were not normally distributed and there were many outliers within the three repetitions group, it was decided to use non-parametric tests. The data were submitted to a Friedman test to determine if there were differences between the different repetition groups. Results indicated that gain scores at the final post-test were significantly different among the different repetition groups, $\chi^2(3) = 142.79$, $p < .001$. Pairwise comparisons (Wilcoxon Signed Ranks Test, with Bonferroni correction reducing the alpha level to .008) indicated that there were significant differences between the nine repetitions group and the other three groups. The nine repetitions group (*Mdn* = 60.00) significantly outperformed the seven repetitions (*Mdn* = 20.00) ($z = -7.84$, $p < .001$, $r = .78$), five repetitions (*Mdn* = 20.00) ($z = -8.16$, $p < .001$, $r = .81$) as well as the three repetitions group (*Mdn* = 10.71) ($z = -8.62$, $p < .001$, $r = .86$). The effect sizes were large in all cases. However, no other group differences were confirmed.

Since the review activity was conducted within each treatment group, one underlying question is whether the results would be the same if the data were analysed within each treatment group. Therefore, the data were split in SPSS in order to compare the results of the three treatment groups. The results were similar across the three treatment groups and to the results considering the treatment groups as a whole. There were significant differences among gain scores for the different repetition groups within the L2 group [$\chi^2(3) = 53.82$, $p < .001$], the CS group [$\chi^2(3) = 56.30$, $p < .001$] as well as the CFoF group [$\chi^2(3) = 49.03$, $p < .001$]. The results for the post-hoc analysis were also similar across the three

treatment groups and similar to the results regarding the three treatment groups as a whole, in which the nine repetitions group significantly outperformed the other three groups.

In summary, the retention for the single words receiving nine repetitions were significantly higher than the retention for their counterparts receiving seven, five or three repetitions. The situation was the same across the three treatment groups as well as within each treatment condition. However, there were only five single words receiving seven or five repetitions, yet 28 items received three repetitions. The difference in group sizes, however, means that the findings relating to repetitions need to be interpreted with caution.

4.10 Qualitative analysis of learners' strategies within the different types of vocabulary instruction

The previous sections of the Findings Chapter presented the results of the quantitative analysis of the data collected from the quasi-experimental design, which revealed what the participants learned through the classroom intervention. However, in addition to exploring the impact of the three types of vocabulary instruction on both short-term and long-term vocabulary learning, the present study aimed to gain insights into why the vocabulary learning differed across the three treatment groups, especially investigating whether the strategies used were the same for the learners across the three groups. It is clear that this could not be addressed by analysing the quantitative data collected.

Therefore, a stimulated recall interview was used to collect qualitative data after the quasi-experimental study. It was believed that through qualitative analysis, it would be possible to explore what were the strategies adopted by the participants within the three types of vocabulary instruction, in order to gain insights into the possible reasons for why the learning outcomes of the three treatment conditions differed.

4.10.1 Participants for the stimulated recall interviews

Within each treatment group, four learners representing the four different proficiency levels were selected to participate. Hence, altogether 12 learners were individually interviewed. Table 4.27 gives an overview of these learners' composite test scores and their proficiency levels and Table 4.28 presents an overview of these learners' vocabulary test scores.

Table 4.27
Overview of the selected learners' combined general English language test scores and proficiency levels

Group	Student Code	Composite test score	Proficiency Level
L2	**L2S1HS**	154.25	High
	L2S2HS	136.50	Mid-high
	L2S3LS	123.50	Mid-Low
	L2S4LS	101.50	Low
CS	**CSS5HS**	166.00	High
	CSS7HS	135.00	Mid-high
	CSS6LS	127.00	Mid-Low
	CSS8LS	116.00	Low
CFoF	**CFoFS11HS**	150.00	High
	CFoFS9HS	130.00	Mid-high
	CFoFS10LS	119.50	Mid-Low
	CFoFS12LS	108.50	Low

Note. The maximum score is 220.00.

Table 4.28

Overview of the selected learners' vocabulary test scores

Group	Student Code	Vocabulary pre-test	Vocabulary post-test	Vocabulary delayed post-test
L2	**L2S1HS**	10.00	33.00	22.00
	L2S2HS	6.00	24.00	13.00
	L2S3LS	3.00	24.00	10.00
	L2S4LS	3.00	7.00	4.00
CS	**CSS5HS**	12.00	56.00	29.00
	CSS7HS	7.00	50.00	25.00
	CSS6LS	5.00	39.00	5.00
	CSS8LS	7.00	36.00	7.00
CFoF	**CFoFS11HS**	4.00	59.00	33.00
	CFoFS9HS	5.00	45.00	35.00
	CFoFS10LS	6.00	49.00	33.00
	CFoFS12LS	5.00	52.00	17.00

Note. The maximum sore for each vocabulary test is 60.00.

The overview of the selected learners' vocabulary test scores indicated that for both the L2 and CS group learners, the three vocabulary test scores in general decreased from the high proficiency level learner to the low proficiency level learner in that order. However, this was not true for the CFoF group. The four learners in the CFoF group all made large vocabulary gains at the post-test. At the delayed post-test, the low proficiency level learner only retained approximately one third of the vocabulary knowledge she gained at the post-test, but the other three learners all retained more than half of the knowledge they gained immediately after the intervention. In addition, taking into account their pre-test scores, the two lower level learners in the CS group did not retain any vocabulary item two weeks after the intervention, which was surprising. Besides, the two lower level learners in the L2 group had the same score at the vocabulary pre-test. Moreover, the high proficiency level learner's pre-test score was lowest among the four learners in the CFoF group. These observations, however, should be interpreted with caution and may in part be due to the modifications that were carried out on the pre-test scores because the format of the vocabulary pre-test was different from that of the post- and delayed post-test.

4.10.2 Stimuli

Altogether six vocabulary items were selected as stimuli for the stimulated recall interviews, including two nouns (*association*, *tourism*), one verb (*threaten*), one adjective (*convincing*) as well as two collocations (*figure out*, *at ease*). There were two reasons for this. On the one hand, selecting different types of vocabulary items as stimuli could explore more strategies as the participants may adopt different strategies for learning different types of vocabulary items. On the other hand, six stimuli were appropriate as based on the experience of the pilot phase; this number enabled the present study to collect enough qualitative data within restricted time limits for each interview.

4.10.3 Transcribing and coding

The 12 stimulated recalled interviews, lasting approximately 20 minutes each, were initially transcribed and translated from Chinese into English. The next thing to do was to systematically examine the 12 transcripts and code them appropriately. One notable issue here is that there are different approaches regarding coding verbal protocols in the field of learner strategies which are listed in Figure 4.12.

Bottom-up	Top-down
Inductive	Deductive
Descriptive	Conceptual
Grounded approach	Start-list approach
Emergency of theory from data	Code-and-retrieve
Emic	Etic
Participant-centred	Researcher-centred

Figure 4.12 Approaches to the analysis of verbal protocols (Gu, 2014, p. 75)

It is argued that there are merits for using approaches from both of the two sides presented in Figure 4.12 to analyse verbal protocols (Miles & Humberman, 1994). In order to explore

which strategies were adopted by the participants during the classroom intervention, the primary aim of the present study was to discover as many meaningful instances as possible purely from the 12 transcripts, which is a bottom-up approach. However, it is impossible and arguably unhelpful to ignore existing research findings and the relevant research question posed in the present study is in itself hypothesis-driven by a theoretical background (Gu, 2014). Therefore, it was decided to combine both bottom-up and top-down approaches for coding in the present study.

The first thing to do was to identify initial codes, including those emerging from the data and those from previous studies, and summarise them in a code book for further coding. Three transcripts, one from each treatment group, were randomly selected. They were investigated in great detail and meaningful instances relating to strategies were highlighted from three perspectives: 1. Strategies used while listening to the sentence including unfamiliar vocabulary items; 2. Vocabulary learning strategies in response to the teacher's vocabulary explanations; 3. Vocabulary learning strategies used to remember the target lexical items.

Then, these highlighted instances were labelled based on codes from existing research findings, drawing in particular on Vandergrift and Goh (2012) and Schmitt (1997). Instances where no existing codes were available were labelled as new codes. Names and definitions were given to these new codes. As a result, the initial code book was drawn up including the name, definition and example from the transcripts for each code (see Appendix H). The initial code book was sent to a second researcher for examination to ensure that each specific code corresponded to the example suggested. Places where disagreements existed were commented on and the code book was sent back to the original researcher for revision. In this way, a final code book was created. The final code book combined strategies for listening comprehension with those for learning vocabulary, with reference to Vandergrift and Goh's (2012) L2 listening comprehension strategies and Schmitt's (1997) vocabulary learning strategies taxonomy. In order to make the following

discussion clearer, a list of the identified codes and their correspondent definition is presented in Table 4.29 below.

Thereafter, using the code book, a sample coding was conducted with the three selected transcripts by both the researcher and the second researcher separately. The coding results showed that there were 18 out of 108 instances where the two researchers coded differently. Therefore, approximately 83.33% of instances coded independently by the two researchers were the same. After that, a discussion was held to investigate the 18 instances where the two researchers coded differently. Agreements were made on coding 10 out of these 18 instances. Eight places where agreements could not be reached were sent to a third researcher for advice. Final agreements were made regarding the coding of every single instance across the three selected transcripts. In this way, basic guidelines were confirmed and used by the researcher to code all the 12 transcripts subsequently. A coded sample transcript of a stimulated recall interview is given in Appendix I.

Table 4.29
Code book (Based on Schmitt, 1997; Vandergrift & Goh, 2012, p. 277-284)

No.	Strategy	Definition
	Part 1	**L2 listening comprehension strategy**
1	**Linguistic inferencing***	"Using known words in an utterance to guess the meaning of unknown words."
2	**Matching with familiar words**	Matching with known or familiar words to assist in understanding and/or task completion.
3	**Monitoring***	"Checking, verifying, or correcting one's comprehension or performance in the course of a task."
4	**Translation***	"Rendering ideas from one language to L1 in a relatively verbatim manner."
5	**Selective attention on nearest information**	Attending to language input or situational details within a fairly short input distance from the target item to assist in understanding and/or task completion.
6	**Linguistic contextualisation***	"Relating a word or a phrase heard to an environment where the word has appeared before."
7	**Deduction***	"Consciously applying learned or self-developed rules to understand the target language."

	Part 2	Vocabulary learning strategies
8	**Attend to spelling**	Try to spell the unfamiliar words.
9	**Give up**	Think of nothing, blank in mind.
10	**Selective attention on known words**	Attending to known or familiar words in the language input that assist in understanding and/or task completion.
11	**Ignoring**	Ignore the vocabulary explanation provided.
12	**Summarisation**	Use a simple L1 word to summarise the English explanation.
13	**Searching**	Search for other familiar words which have similar meanings.
14	**Confirming**	Refer back to the original sentence to confirm the meaning.
15	**Making judgement**	Making judgement on the vocabulary explanation provided.
16	**Waiting**	Waiting for explanation, verification, rephrasing, or examples about the language and/or task.
17	**Using phonological knowledge**	Use the pronunciation of the vocabulary item to remember.
18	**Repetition**	Repeat the vocabulary item and its meaning verbally for several times.
19	**Force**	Force oneself to remember the vocabulary meaning provided.
20	**Segment word**	Segment a word according to its pronunciation to assist remembering.
21	**Remember as a collocation**	Remember the meaning of a collocation which includes the target single word.
22	**Making connections**	Connect to familiar words to remember.
23	**Rely on examples**	Use an example sentence or phrase including the new vocabulary item.
24	**Keyword strategy**	Generate an L1 word "that is similar in sound or appearance to the new foreign language word; and development of an interactive image involving that keyword and the definition of the new word". (Lawson & Hogben, 1998, p. 179)
25	**Remember part of a collocation**	Remember part of a vocabulary item to assist remembering of the whole collocation.

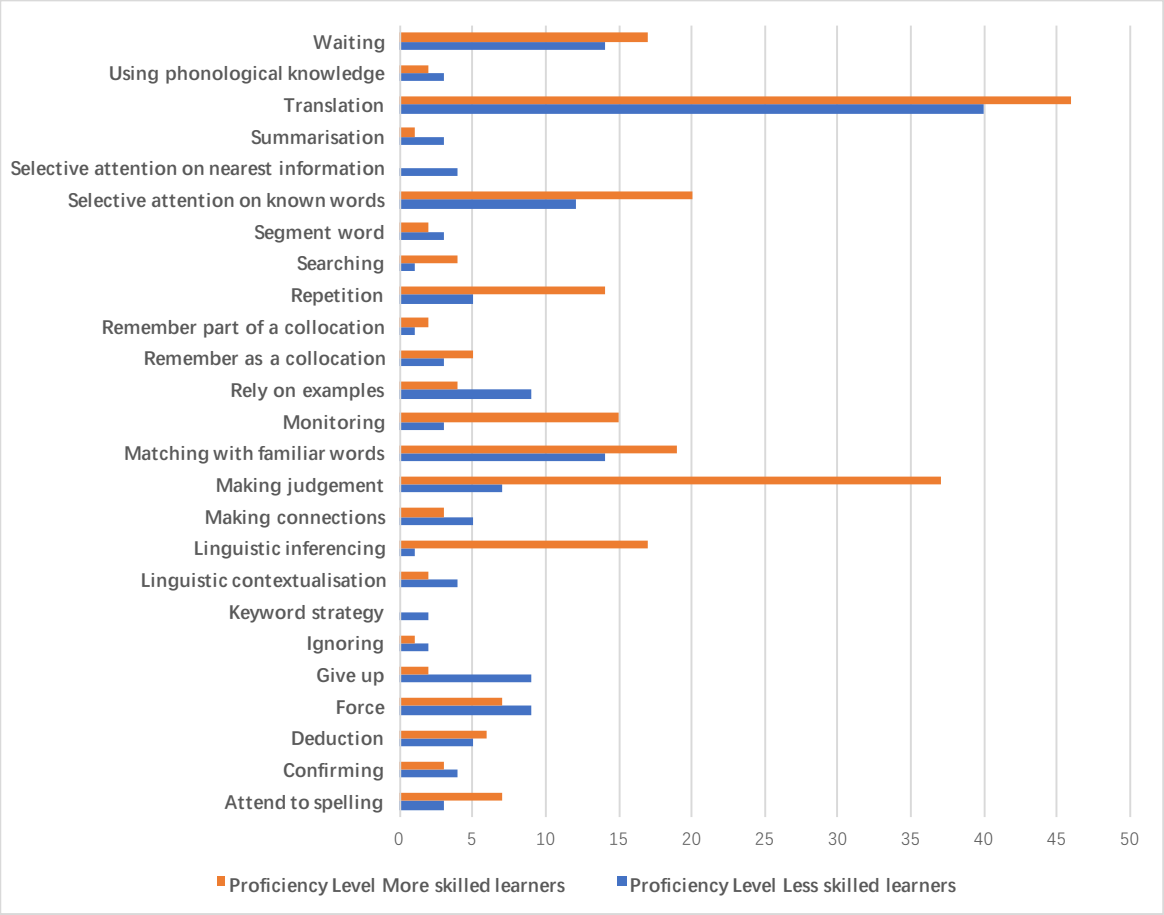
Note. * - listening comprehension strategies developed by Vandergrift & Goh (2012), p. 277-284

4.10.4 Quantitative overview

This section begins with a quantitative overview of the qualitative data in order to provide a general picture of learners' strategy use regarding how they understood the unfamiliar vocabulary items in the listening passage and the corresponding vocabulary instruction provided by the researcher plus any strategies they used to remember the new vocabulary.

Initially, the 12 interview transcripts were transferred into one NVivo file and were coded according to the agreed code book. Secondly, each interview transcript was labelled to show both general language proficiency level and treatment condition of the learner. One notable issue was that, in each treatment group, four learners were interviewed including two more skilled learners (high, mid-high levels) and two less skilled learners (low and mid-low levels). In order to give a clearer picture for this quantitative overview, it was decided to combine the coding of the two more skilled learners and the two less skilled learners respectively within each treatment group. Thirdly, using the Query function in NVivo 11, the number of instances of different types of strategies reported by less skilled and more skilled learners and reported by learners from the three treatment conditions were obtained. These results are given in Graphs 4.1 and 4.3.

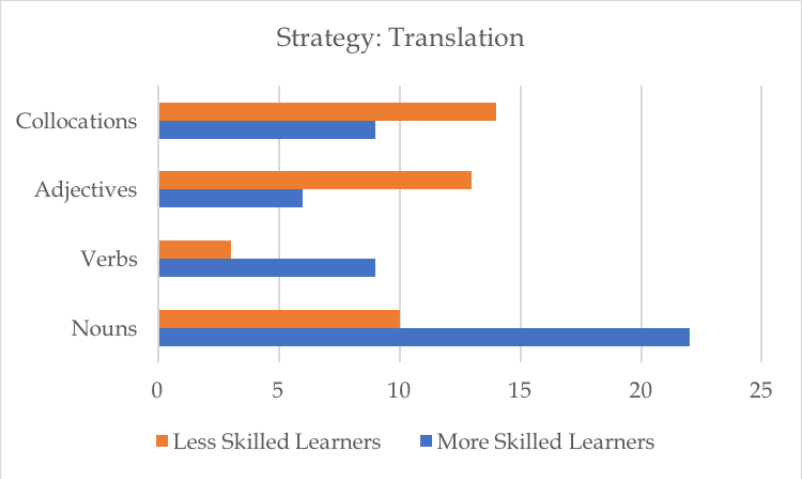
4.10.4.1 Strategy use by proficiency levels



Graph 4.1 The number of instances of different types of strategies reported by less skilled and more skilled learners during stimulated recall interviews

It can be seen from Graph 4.1 that there are differences between less skilled and more skilled learners in strategies used for the four types of vocabulary items. In general, more skilled learners reported a greater number of instances of different types of strategies than less skilled learners. On the one hand, the strategy of *selective attention on nearest information* was only used by less skilled learners. They also reported more instances of the strategies of *rely on examples* and *give up* than more skilled learners. On the other hand, more skilled learners tended to use strategies such as *translation*, *selective attention on known words*, *repetition*, *monitoring*, *making judgement*, *linguistic inferencing* as well as *attend to spelling* more frequently than less skilled learners did.

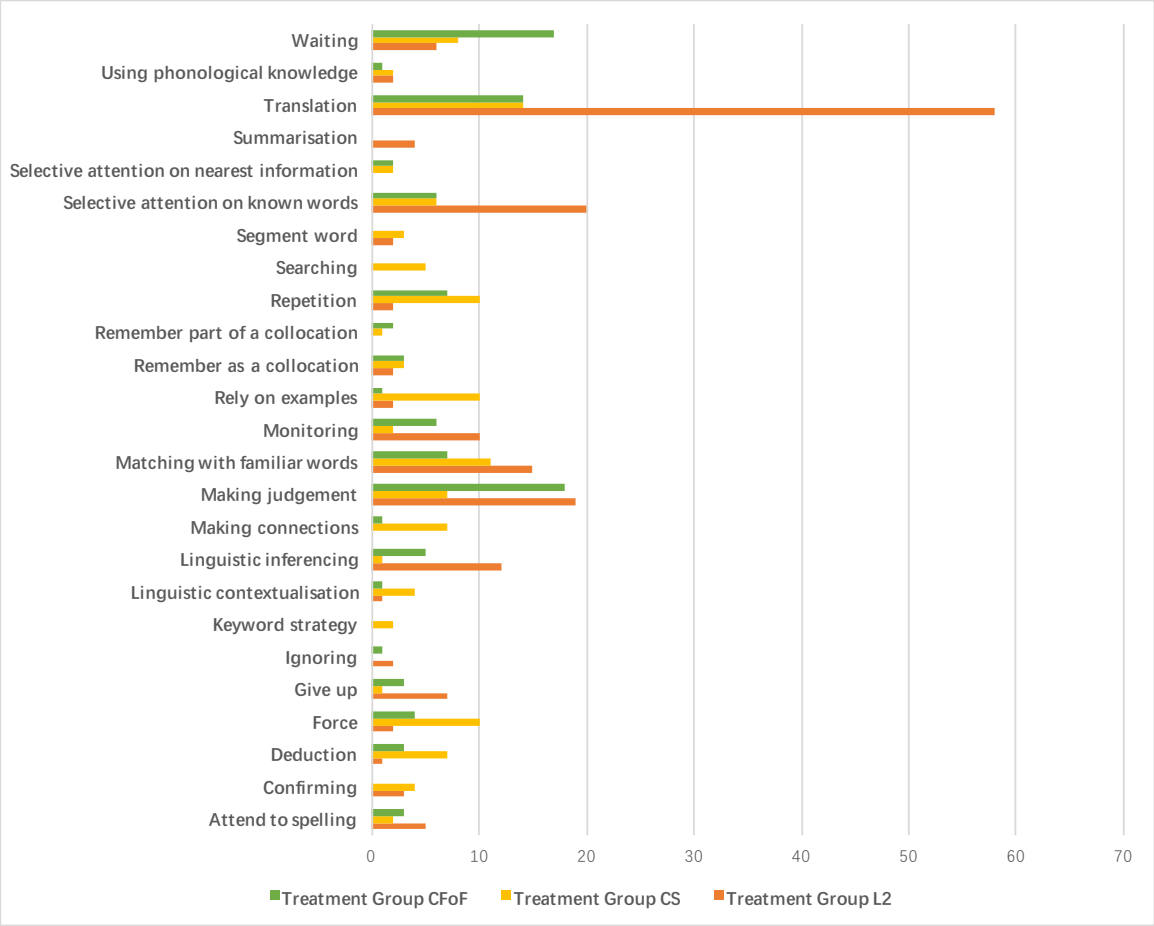
Graph 4.1, however, gives no insights into whether strategy use is different across different word classes. Therefore, the number of different strategy used for learning each word class and for collocations were calculated respectively. The patterns were generally similar as looking at the four types of lexical items as a whole. However, although *translation* was reported more frequently by more skilled learners than less skilled learners for learning nouns and verbs, the results were reversed when learning adjectives and collocations. Details for this are given in Graph 4.2.



Graph 4.2 The number of instances of using Translation reported by less skilled and more skilled learners for learning the three word classes and collocations during stimulated recall interviews

4.10.4.2 Strategy use by treatment conditions

From another perspective, Graph 4.3 reveals the number of instances of different types of strategies reported by learners from the three treatment conditions.

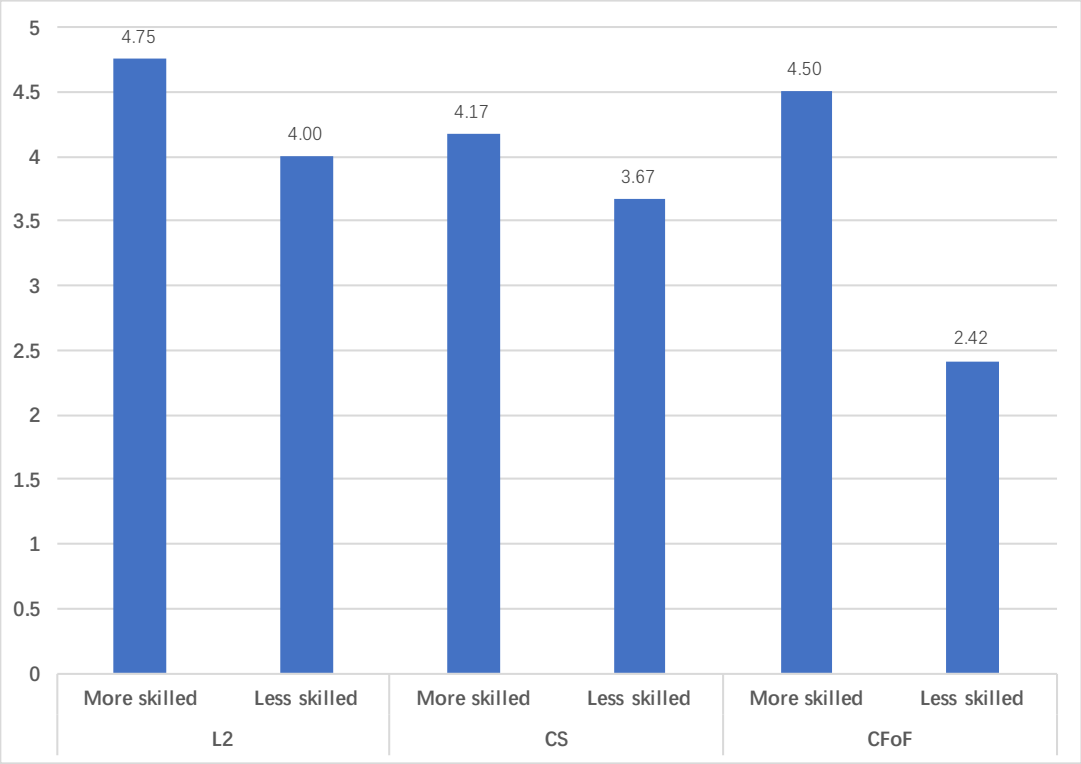


Graph 4.3 The number of instances of different types of strategies reported by learners from the three treatment conditions during stimulated recall interviews

It can be seen from Graph 4.3 that the total number of instances of different types of strategies reported by the L2 group was greater than the number of instances reported by the CS and the CFoF group. The graph also indicates that learners from the L2 group used strategies such as *translation*, *selective attention on known words* and *linguistic inferencing* more frequently than learners from the other two groups did. In addition, the number of instances of using *repetition*, *making connections*, and *force* were reported most frequently by the participants from the CS group among the three groups. Moreover, learners from the CFoF group tended to use the strategy *waiting* (for the teacher's explanation) more

frequently than the other two groups did. Finally, *summarisation* was only adopted by the L2 group, while *keyword strategy* and *searching* was adopted only by the CS group.

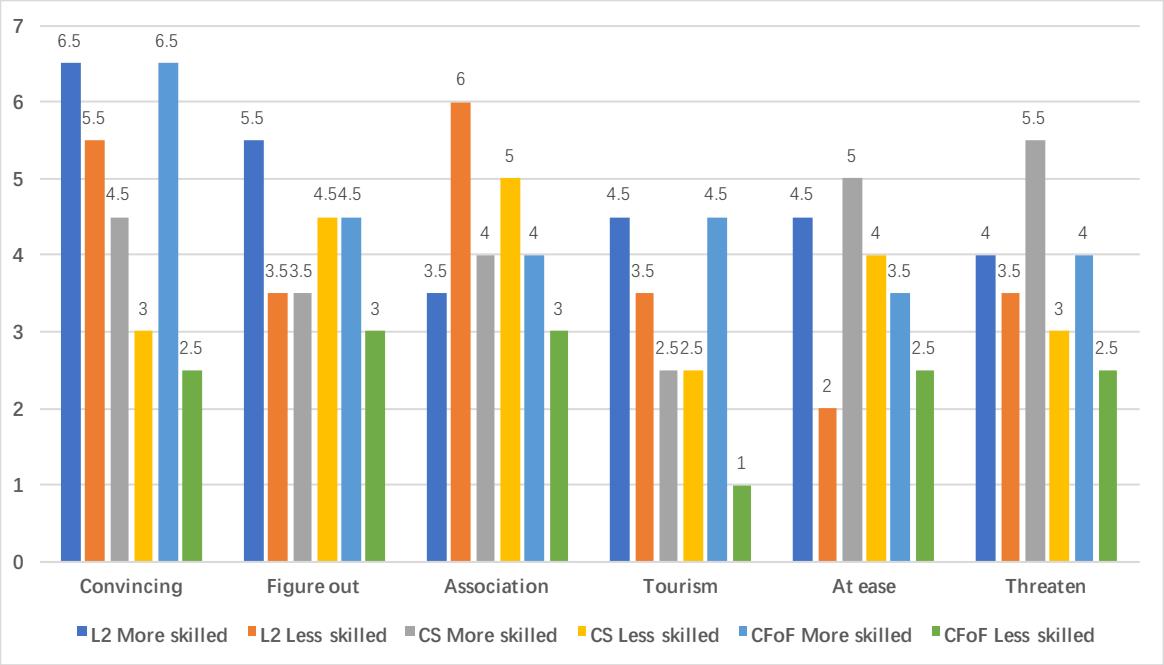
In addition to giving a quantitative overview of the number of instances of different types of strategies reported, it is valuable to look at the mean number of different strategies that more and less skilled learners from each treatment group used to understand and learn the target vocabulary items. Therefore, the number of different strategies employed by each of the 12 learners was calculated respectively. Then, within each treatment group, the mean number was calculated for the two more skilled learners and the two less skilled learners respectively. Results for this are presented in Graph 4.4.



Graph 4.4 Mean number of different strategies used by treatment group and proficiency level

Since participants were interviewed about their approaches to learning six vocabulary items, it is worth looking more specifically at the mean number of different strategies used for all vocabulary items, by treatment group and proficiency level. In order to do that, the number of different strategies employed for each vocabulary item was counted for each

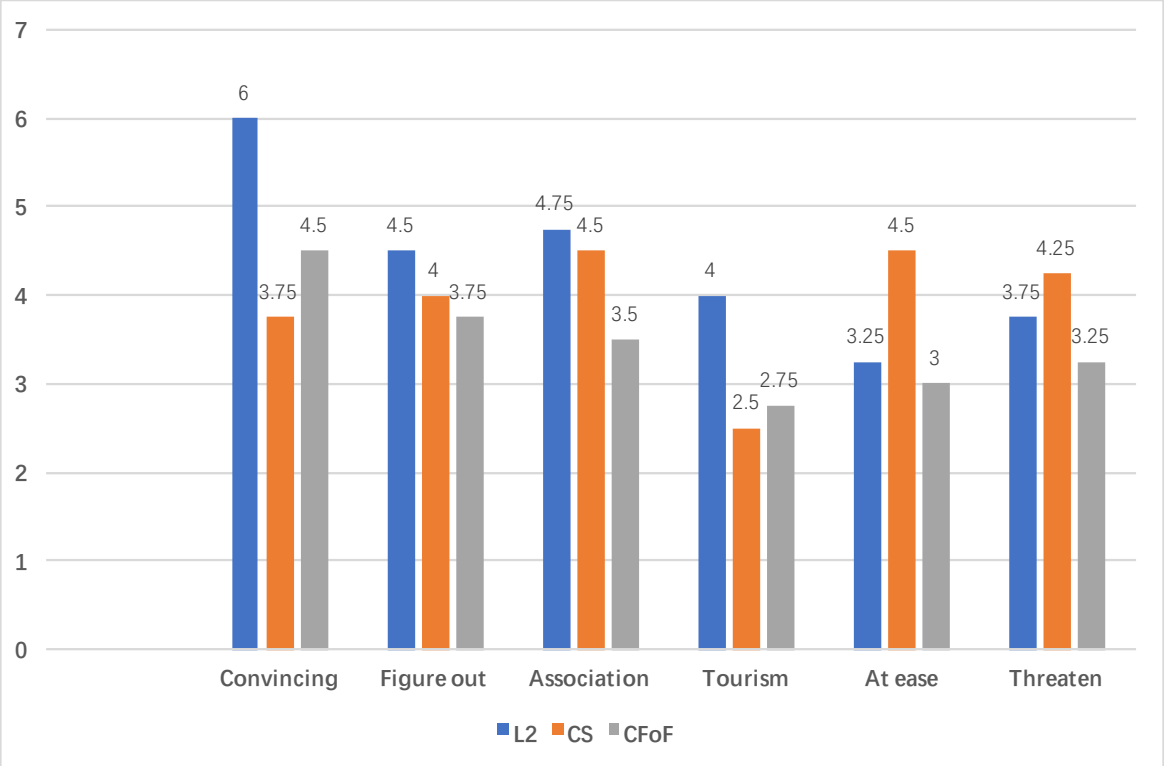
interview participant. After that, the mean number of different strategies used by more skilled and less skilled learners for learning each vocabulary item was obtained within each treatment group. These results are given in Graph 4.5.



Graph 4.5 Mean number of different strategies used per vocabulary items, by treatment group and proficiency level

It can be seen from Graph 4.4 and 4.5 that, in general, more skilled learners used more types of strategies than less skilled learners, except for learning the vocabulary item, *association*, where it is the other way around for the L2 and CS group. The differences between more and less skilled learners seem to be especially marked for the CFoF group, as it is shown in Graph 4.4 that the less skilled learners had an especially low mean compared with the more skilled learners in the CFoF group. The more qualitative analysis below explores these differences in more detail.

Another area of interest was to look at the differences in the mean number of different strategies used for each vocabulary item just across the treatment conditions, not separating out by proficiency level. These are shown in in Graph 4.6.



Graph 4.6 Mean number of different strategies used for vocabulary items by treatment group

Although Graph 4.4 indicated that the L2 group used more types of strategies than the other two groups, it can be seen from Graph 4.6 that for learning two vocabulary items, *at ease* and *threaten*, the CS group employed a wider range of strategies than the L2 and the CFoF group. In general, the CFoF group used fewer types of strategies than the other two groups, except for learning *convincing* and *tourism*, where it used more strategies than the CS group while fewer than the L2 group. These differences will be further explored qualitative analysis.

4.10.5 Qualitative analysis

The quantitative analysis gives an overview of the numerical differences in the frequency of each strategy employed between more skilled and less skilled learners as well as among learners from the three treatment groups. It also compares the mean number of different strategies used for vocabulary items, by treatment group and proficiency level. However, the quantitative analysis "cannot capture how a given strategy is used or the particular combinations of strategies used to build meaning. Neither can it capture the effective use

of a strategy, such as the accuracy of an inference, an appropriate connection to prior knowledge or the depth of summarization" (Vandergrift, 2003, p. 477). Therefore, the previous quantitative analysis was supplemented by a qualitative analysis looking specifically at how learners used strategies.

In the sections below, interview transcripts will be quoted in a few places. On the occasions where specific Chinese words were used by the learners during the stimulated recall interview, meaning that the *translation* strategy was used, the Chinese words are presented but followed by corresponding English translations in brackets for clarification. [e.g., I guessed that it would have something to do with '旅游 (tour)'.].

Section 4.10.5.1 first reports the qualitative findings of the different patterns or combinations of strategy use across the three treatment groups. These patterns are combinations of L2 listening comprehension strategies and vocabulary learning strategies. From a different perspective, Section 4.10.5.2 reports differences of strategy use between more and less skilled learners. Finally, in more detail, Section 4.10.5.3 specifically looks at the differences of strategy use within the three types of vocabulary instruction in order to provide more evidence for the quantitative findings relating to vocabulary learning.

4.10.5.1 Differences in strategy use across the three treatment conditions

Results from the previous quantitative overview revealed that learners from different treatment groups tended to employ different types of strategies and the frequency of using particular types of strategy was also different across the three treatment conditions. The quantitative overview, however, did not reveal how exactly the strategies were used, i.e., whether there were any particular patterns of strategy use within one particular treatment group. Further qualitative analysis with the 12 interview transcripts found that there were two patterns of strategy use in the L2 group and the four participants from the CS group had one particular combination of strategy use. There was also one strategy use pattern shared by the four learners from the CFoF group.

In the L2 group one pattern was the combination of *matching with familiar words* and *translation*, while the other pattern was combining *selective attention on known words*, *translation* and *making judgement*. The first of these patterns is shown in the following example. The segment below was taken from the listening passage '*Theme park*', which was used in the fifth intervention session.

> **Listening segment 1:** With all these attractions, no wonder ***tourism*** is increasing.
>
> **Vocabulary instruction:** ***Tourism***, which is a noun, means 'the industry of sightseeing and holidays'. An example sentence for this word can be 'The tourism in China is developing.'.

R: When you first heard the sentence, did you catch the word *tourism*?
L2S1HS: Yes, I did.
R: What did you do then?
L2S1HS: I guessed that it would have something to do with '旅游 (tour)'.

Here, L2S1HS, a more skilled learner in the L2 group, firstly used the combination of *matching with familiar words* and *translation*. When she first heard the sentence, she realized that there was an unknown vocabulary item. Although she was not able to see the sentence written down, what she did was to match the unfamiliar item with a familiar item '旅游 (tour)'. At the same time, there was an on-going process of translating the English word 'tour' to its L1 meaning '旅游 (tour)' in her mind. She continued:

R: So how did you figure out it was actually ***tourism***?
L2S1HS: With the help of your English explanation. I noticed that in your explanation, there was 'industry', and I realised that it was '工业 (industry)'. I supposed the meaning of the word might be '旅游业(tourism)'.
R: What else were you thinking?
L2S1HS: I was actually not that comfortable with the English explanation you provided in the class and I am still thinking that your explanation does not match the word itself.

She then combined *selective attention on known words*, *translation* and *making judgements* in order to understand the vocabulary instruction provided. When received the L2 explanation, her attention was focused on 'industry' which was familiar to her. She then translated 'industry' into the L1. Instead of understanding the remaining part of the

explanation, she linked '工业 (industry)' with '旅游 (tour)' which she had guessed previously and successfully acquired the L1 meaning of ***tourism***, '旅游业'.

She was, however, still struggling with the L2 explanation provided and moved to *making judgements* on that. The potential reason for this can be that the student was trying to translate the L2 explanation into L1 Chinese, but then found out that the direct translation of that explanation did not make sense to her. This provided more evidence that the student was *making judgements* as she was questioning things that did not seem quite right rather than just accepting them. It was also realised that L2S1HS responded to the teacher's vocabulary instruction in a fairly active way. Although the L2 meaning, in general, might be a bit difficult for her, she was making an effort by using her existing vocabulary knowledge to figure out the L2 meaning.

Similar patterns were found for L2S3LS, a less skilled participant in the L2 group. The listening segment below two was taken from the listening passage '*A student of African wildlife*', which was used in the third intervention session.

> **Listening segment 2:** She also discovered how chimps communicate with each other, and her study of their body language helped her ***figure out*** their social system.
>
> **Vocabulary instruction:** *Figure out* means 'understand sb./sth. by thinking'. An example sentence for this collocation can be 'Can you figure out this difficult question?'.

R: Did you recognise ***figure out*** when you heard this sentence in my class?
L2S3LS: Yes, it seems that the pronunciation is very familiar to me…maybe I have heard about this collocation before.
R: What did you do when I was explaining this collocation to you?
L2S3LS: I tried to translate the English meaning that you provided into Chinese…but I failed. I knew that in your explanation, 'understand' should be a familiar word to me, but I couldn't remember the exact Chinese meaning for 'understand'…your explanation is too long…very difficult to me.

The student initially tried to match ***figure out*** (target lexical item) with a potential familiar vocabulary item which has similar pronunciation (*matching with familiar words*). When provided with the L2 explanation, she tried to translate the L2 explanation word by word

into Chinese (*translation*). She failed, however, as she forgot the L1 translation of the word 'understand' in the L2 meaning, although, as reported by her, 'understand' was a familiar word to her (*selective attention on known words*). At the same time, she also felt that the L2 meaning provided by the teacher was too long and difficult for her, most likely because she had difficulty in translating the L2 meaning into the L1 and her existing vocabulary knowledge was too low for her to understand the explanation. Therefore, she treated the vocabulary explanation in a fairly passive way and was likely to *give up* at the end.

Regarding the CS group, one type of strategy combination was found. The four participants in the CS group tended to combine *waiting*, *force* and *repetition* in order to understand the vocabulary instruction provided and remember the target lexical item. Listening segment three below was taken from the listening passage '*Showing our Feelings*', which was employed in the sixth (final) intervention session.

Listening segment 3: Making a fist and shaking it almost always means that someone is angry and ***threaten***ing another person.

Vocabulary instruction: ***threaten*** is a verb, which means '恐吓'. An example sentence for this word can be 'They threatened to kill him unless he did as they asked.'

R: What did you do when I played the listening segment including the word ***threaten***?
CSS7HS: I remember that I actually did not clearly recognise what was in the listening segment. Therefore, what I did was to wait for you to explain.
…
R: What were you thinking after I explained this word to you?
CSS7HS: I recognised that this word was unfamiliar to me and I forced myself to remember it.
R: How exactly did you force yourself to remember?
CSS7HS: I repeated both the pronunciation of the word and its Chinese meaning many times.
R: You repeated them? In your mind or orally?
CSS7HS: Orally in a very low voice. In my opinion, reading out the word and its meaning can help me remember, especially when they were repeated orally for many times.

The transcript segment above was from a more skilled learner, CSS7HS, from the CS group. After the listening material was played by the teacher, the student failed to catch any information and did not understand the sentence. Therefore, she used the strategy *waiting*

and waited for the teacher to provide more information. As this was the sixth intervention session, it was likely that the student was quite certain that the teacher would provide explanations afterwards, especially that the vocabulary explanations were given in L1. It was also likely that the student was expecting and intentionally paid attention to receive the L1 explanation while being less focused on the L2 listening comprehension.

The student then *force*d herself to remember the word by reading out the pronunciation of the word and its corresponding Chinese meaning several times (*repetition*). It was noticed that although the student relied mainly on waiting for the L1 explanation and forced herself to remember by repetition, she did all these in a fairly active way. She seemed to be quite confident that forcing herself to repeat the pronunciation and L1 meaning of the target word was an effective way for her to remember new vocabulary items.

The above combination of strategy use was also reported by CSS8LS, a less skilled learner from the CS group. The following segment four was also taken from the listening passage '*Showing our feelings*', employed at the final intervention session.

Listening segment 4: Its function is to show happiness and put people ***at ease***.

Vocabulary instruction: ***At ease*** is a collocation which means '自由自在，放松'. An example sentence for this can be 'He felt completely at ease.'.

R: Did you recognised the collocation ***at ease*** when you heard it for the first time?
CSS8LS: Yes, I did.
R: What did you do then?
CSS8LS: Nothing.
R: Just recognised it and did nothing?
CSS8LS: Yes, and waited for you to explain.
…
R: So what did you do to cope with my explanation?
CSS8LS: Forced myself to remember.
R: How exactly did you force yourself to remember?
CSS8LS: I linked the Chinese meaning and the English word and read them out for a few times

The student successfully recognised ***at ease*** (the target lexical item) from the listening passage. As she reported, she did nothing then but waited for the L1 explanation which would normally be provided by the teacher immediately afterwards (*waiting*). Regarding how she understood the explanation provided, she recalled that she forced herself to remember (*force*) by using the *repetition* strategy reading out the target vocabulary item and its L1 meaning several times.

In the CFoF group, one strategy use pattern was reported by the four interviewees, which was *selective attention on known words*, *waiting* and *translation*.

> **Listening segment 3**: Making a fist and shaking it almost always means that someone is angry and ***threaten***ing another person.
>
> **Vocabulary instruction**: Here, ***threaten*** is a verb, which means '恐吓'. In Chinese, '恐吓' can be used both as a noun and a verb. In English, however, 'threat' is the noun version for 'threaten'. They are different words.

R: What did you do when you heard the sentence including *threaten*?
CFOFS11HS: I recognised that there was an unfamiliar word to me and I linked this word (***threaten***) with 'angry' in the same sentence.
…
R: What was on your mind at the time when I was explaining the word to you?
CFOFS11HS: You gave us another word *threat* which means '威胁'. I followed your instruction and thought about *threaten*, which means '恐吓'. The two Chinese meanings are connected … The two words have a similar Chinese meaning but are different in forms.

When listening to the sentence including ***threaten*** (the target word), CFOFS11HS, a more skilled learner, firstly linked the target word with 'angry' which is a familiar word to her (*selective attention on known words*). Although it was not reported by her, she might also have paid attention to the grammatical structure of the sentence as there was an 'and' between the target vocabulary item ***threaten***(ing) and the word 'angry' she recognised. She then waited for and listened to the teacher's explanation (*waiting*). During this period, she was thinking a lot about the individual vocabulary items, linking ***threaten*** with 'threat', which was quite a deep level of processing of individual lexical items. There was also a potential process of using the *translation* strategy as the participant was trying to compare

the Chinese meaning of the two words, which was not the explanation provided by the teacher.

Another instance of this pattern can be seen in CFOFS10LS, a less skilled learner from the CFoF group.

Listening segment 4: Its function is to show happiness and put people ***at ease***. **Vocabulary instruction**: ***At ease*** means '自由自在，放松'. The meaning of this collocation mostly came from 'ease' which is related to 'easy'. There is no need to direct translate the word 'at'.

R: Did you recognise the collocation ***at ease*** when you heard it for the first time?
CFOFS10LS: I firstly recognised 'at'.
R: What about *ease*?
CFOFS10LS: I recognised that it was an unfamiliar word.
R: Did you recognise *at ease* as a collocation?
CFOFS10LS: Yes, and the collocation is unfamiliar to me.
R: What was on your mind when I was explaining the collocation to you?
CFOFS10LS: In your explanation, you mentioned that *ease* was related to the word 'easy'. So I first thought about the Chinese meaning of 'easy' and then thought about the meaning of ***at ease***.

Here, CFOFS10LS, who is a less skilled participant from the CFoF group, reported a similar pattern of strategy use. When asked about whether she recognised the target collocation, she first reported that she recognised 'at' which was a familiar word to her. There was a potential use of *selected attention on known words* strategy in the first instance. What she did then was to wait for the CFoF explanation to be provided by the teacher (*waiting*). In the explanation, the teacher mentioned another word 'easy' but did not give the meaning of the word. Therefore, the student used the *translation* strategy, thinking about the Chinese meaning of 'easy', and then followed the explanation linking this word to the target collocation.

In summary, through a deep qualitative analysis of the interview data across the three treatment conditions, there were certain patterns of strategy use shared by the learners within the same treatment condition. Two patterns were discovered in the L2 group, one

was the combination of *matching with familiar words* and *translation*, while the other was combining *selective attention on known words*, *translation* and *making judgement*. In the CS group, there was only one type of strategy combination shared by the four learners: *waiting*, *force* and *repetition*. There was also one strategy use pattern reported by the four learners in the CFoF group, which was *selective attention on known words*, *waiting* and *translation*.

Although the use of the same pattern was confirmed by both more skilled and less skilled learners within the same group, the above analysis showed that more skilled learners used these strategies in a more active manner, compared with less skilled learners in the L2 and the CS group. In the L2 group, the more skilled learner held a questioning attitude towards the L2 explanation provided, yet the less skilled learner felt that the L2 explanation was too difficult and tended to give up trying to understand the explanation in the end. Similarly in the CS group, although both the two learners forced themselves to repeat and remember the meaning of the target lexical item, the more skilled learner was confident in doing this and regarded this as an effective way of learning, yet this was not reported by the less skilled learner. In the CFoF group, while both the two learners paid much attention to the individual lexical items appearing in the teacher's vocabulary explanation, the more skilled learner seemed to involve a deep processing of these items, yet the less skilled learner did not. These differences of strategy use between more and less skilled learners are further explored in Section 4.10.5.2 below.

4.10.5.2 Differences in strategy use between more skilled and less skilled learners

Results from the previous quantitative overview of learners' use of strategies revealed that more skilled learners used a large number of different strategies than less skilled learners and these differences were especially marked for the L2 and the CFoF group. Therefore, it is worthwhile exploring these differences further through qualitative analysis.

The following exemplifies the patterns found firstly in the L2 group, from a more skilled learner (L2S1HS) and then a less skilled learner (L2S3LS), when asked about how they understood the explanation for the target word ***convincing***. Listening segment five below was taken from the listening passage '*A master of nonverbal humour'* which was used for the first intervention session.

Listening segment 5: The acting is so ***convincing*** that it makes you believe that it is one of the best meals he has ever tasted!

Vocabulary instruction: ***Convincing*** is an adjective, which means 'that makes somebody feel certain or sure'. An example sentence for how to use this word can be 'Ling Dan has a convincing success in 2012 Olympic Games.'.

More skilled learner:

R: What were you thinking when I was explaining this word to you?

L2S1HS: I tried to guess the meaning of the word.

R: How did you guess?

L2S1HS: I followed the L2 meaning you provided and guessed … I could only understand part of the L2 meaning. Some of the words in it were familiar to me.

R: What else were you thinking?

L2S1HS: I was not able to understand the full meaning of the L2 explanation you gave.

R: Were there any particular difficulties?

L2S1HS: When I was listening to the sentence including ***convincing*** in the listening passage, I might have already had a general meaning of this word in mind. However, after listening to your explanation, I sort of felt that something was wrong.

R: What do you mean by something was wrong?

L2S1HS: (silence)

R: Did you feel confused or maybe uncertain with the L2 meaning?

L2S1HS: Yes.

R: Since you were not allowed to write down anything in my class, did you do anything to remember it? If yes, what did you do?

L2S1HS: I forced myself to remember it in my mind.

R: In what way that you forced yourself to remember?

L2S1HS: I tried to divide the word into different parts and remembered it through its pronunciation.

…

R: I can see this is how you remember the spelling, but did you do anything to remember the meaning of the word ***convincing*** in my class?

L2S1HS: I initially used the previous way to remember the spelling. Then I tried to match the word itself with the meaning that you provided.

Less skilled learner:

R: So what were you thinking or what did you do when I was explaining this word to you?

L2S3LS: (silence)

R: You can't understand?

L2S3LS: Yes. It's difficult for me … I know some of the words in the L2 explanation, for example, I know "sure".

R: Ok. What did you do then?

L2S3LS: '做什么事让人感到很确定' (do something to make people feel sure).

R: So you actually translated? the L2 meaning into Chinese?

L2S3LS: ah…yes, yes, yes.

R: What about the example sentence I gave you?

L2S3LS: I can't understand the Chinese meaning of the example sentence.

R: Ok. it is too difficult?

L2S3LS: yes.

R: since you were not allowed to write down anything, what did you do to remember the word and its meaning?

L2S3LS: I did nothing because I couldn't understand the exact meaning that you provided … I think, in order to know the meaning of the word, I should firstly understand every single word in your L2 explanation

L2S1HS used 11 different strategies in response to the teacher's explanation for the target word ***convincing***. She initially used *linguistic inferencing* in order to guess the meaning of the target word from the L2 meaning. In the meantime, she used *selective attention to known words* to focus on the familiar words in the L2 meaning. It was possible that she was also *making judgements* about the explanation provided as she felt that she was not able to fully understand the L2 meaning. At the same time, she was holding a questioning attitude when she was not able to fully understand the L2 meaning provided by the teacher. There was also a potential process of using *translation* strategy as what L2S1HS meant by 'was not able to understand' could be that she was not able to translate the L2 meaning into L1.

Monitoring was also used by L2S1HS as she mentioned that she had a general idea about the meaning of the target word when listening to the passage, yet was not sure about whether the meaning she had was correct after listening to the teacher's vocabulary explanation. Regarding what L2S1HS did to remember ***convincing***, she initially segmented the word (*segment word*) *using phonological knowledge* and remembered the spelling of the word (*attend to spelling*). The series of procedure she followed was fairly active as these were not instructed or required by the teacher. She then *forced* herself to remember the meaning through *repetition* by linking the word itself and its meaning provided by the teacher.

However, compared with L2S1HS, much fewer different strategies were adopted by L2S3LS. L2S3LS reported using only four different strategies. When the L2 meaning was provided by the teacher, L2S3LS felt that the explanation was too difficult for her (*making judgements*). She then focused on the familiar words in the L2 meaning (*selective attention on known words*) and tried to translate the L2 meaning word by word into Chinese (*translation*). Although she had a rough L1 meaning of the word, she did not do anything (*give up*) to remember the word as, reported by her, she could not understand the exact L2 meaning. In addition to this numerical difference, compared with L2S1HS who showed a much more questioning, active, deeper learner approach, L2S3LS was quite passive in

response to the teacher's vocabulary instruction and, as reported by her, 'did nothing' to remember the target lexical item.

Such differences between the number and manner of use of different strategies employed by more skilled learners and less skilled learners were also confirmed in the CFoF group. The two transcripts below are from CFOFS11HS (a more skilled learner) and CFOFS10LS (a less skilled learner). They were recalling how they understood the teacher explaining the target word ***association***. The listening segment six was taken from the listening passage '*Communication, no problem*', which was used at the fourth intervention session.

Listening segment 6: Yesterday, another student and I, representing our university's student ***association***, went to the Capital International Airport to meet this year's international students. **Vocabulary instruction**: Here, in the structure 'n. + ***association***', it means '协会'. However, it has a different meaning '联合' when used as '***association*** with sb./sth.'.

More skilled learner

R: Did you recognise the word ***association*** when you first heard the sentence? What did you do?

CFOFS11HS: Yes. I recognised this word. I felt I heard the word before but couldn't remember the Chinese meaning of the word… I tried to guess the meaning of the word and also tried to guess the meaning of 'student association'.

R: What were you thinking when I was explaining the word to you?

CFOFS11HS: I felt that the first part of your explanation was useful. However, the second part was not useful.

R: Ok, why? … Is it too much for you?

CFOFS11HS: Yes. I could simply remember the first structure because there was 'student association' in the listening passage. However, I kind of felt that the second structure was irrelevant in this occasion.

R: I see. Did you do anything to help you remember?

CFOFS11HS: I remembered 'student association' as '学生会' (student association).

R: You remembered this word by remembering it as a collocation?

CFOFS11HS: Yes.

R: The meaning of ***association*** that I gave you was '协会' (association), but not '会' (meeting). Why did you not remember 'student association' as '学生协会' (student association)?

CFOFS11HS: Actually I did do so in the first place, but I felt that I was more comfortable with the Chinese translation '学生会' (student association).

Less skilled learner

R: Did you recognise the word ***association*** when you first heard the sentence? What were you thinking?

CFOFS10LS: Yes, I did. I felt that I knew this word before, yet I cannot remember the meaning.

R: So what did you do then?

CFOFS10LS: I tried to search for the Chinese meaning of the word in my mind.

R: Did you succeed?

CFOFS10LS: No.

R: What did you do when I was explaining this word to you?

CFOFS10LS: (silence)

R: Just listen?

CFOFS10LS: Yes.

R: Did you do anything else?

CFOFS10LS: (silence)

R: So how did you cope with my explanation?

CFOFS10LS: I felt the word was too long and too difficult to remember.

R: Too long? You mean the spelling of the word is difficult to remember?

CFOFS10LS: Yes.

R: So you really focused on the spelling?

CFOFS10LS: Yes.

CFOFS11HS reported that she used six different strategies for the vocabulary explanation for ***association***. Initially, she successfully recognised the word and realised that she might know this word before. It was likely that she was trying the match the word with a familiar word which has similar pronunciation as ***association*** (*matching with familiar words*). What she did then was search for the Chinese meaning of this word (*translation*) in her mind and guessing the meaning of the word (*linguistic inferencing*). This suggests that the student was quite active when encountering unfamiliar lexical items as she was trying to guess the meaning of the word by herself although she was quite certain that the meaning would be explained by the teacher afterwards. After receiving the teacher's explanation, she *made judgements* on the explanations provided. She felt that the second part of the vocabulary instruction [***association*** has a different meaning '联合(work with)' when used as 'association with sb./sth.'] was not helpful to her. Again, the student showed a fairly active manner in response to the vocabulary explanation provided. She was trying to figure out whether or not the explanation was useful to her.

Regarding what she did in order to remember the meaning of the word, CFOFS11HS recalled that she remembered the word as a collocation since the target word appeared as part of a collocation, 'student association', in the listening passage (*remember as a collocation*). There was again a process of using *translation* strategy as the student remembered the Chinese meaning of this collocation. When translating the collocation 'student association' into Chinese, the student slightly changed the direct L1 translation from a less authentic Chinese collocation (学生协会) to a more authentic Chinese collocation (学生会) as she felt more comfortable with the latter (*monitoring*).

CFOFS10LS, however, only used four different strategies. Similar to CFOFS11HS, CFOFS10LS also tried to match ***association*** with the word she knew before (*matching with familiar words*) and tried to recall its L1 meaning (*translation*). Afterwards, she did nothing but waited for the teacher to explain (*waiting*). Finally, regarding how to understand the teacher's explanation and remember the word, CFOFS11HS reported that

she felt the word itself is too long and difficult to remember. She was likely to *attend to spelling* at this stage.

Additionally, compared with her counterpart CFOFS11HS in the same group, CFOFS10LS was relatively passive when approaching the listening task and teacher's explanation. It seems that she was not attending to the teacher's vocabulary explanation as, reported by her, she was thinking about the spelling of the word during that period. In addition, her manner was also passive regarding remembering the word itself as she felt the spelling of the word was too long for her.

In summary, the qualitative analysis confirmed that there were differences in the number of different strategies used between the more skilled learners and the less skilled learners within the L2 and the CFoF group. More skilled learners tended to use more different strategies than less skilled learners. However, in addition to the difference in the number, learners of different proficiency levels also had a different approach to the task, whereby the more skilled learners were more active while the less skilled learners were fairly passive.

4.10.5.3 Differences in strategy use in response to the teacher's vocabulary explanations across the three treatment conditions

The above section 4.10.5.1 explored the patterns of strategy use in general. It was valuable to more specifically look at strategy use for learners in each group, which might provide possible explanations for the quantitative data collected through classroom intervention.

The transcript below was from a more skilled learner, CFOFS9HS from the CFoF group.

Listening segment 4: Its function is to show happiness and put people ***at ease***. **Vocabulary instruction**: ***At ease*** means '自由自在，放松'. The meaning of this collocation mostly came from 'ease' which is related to 'easy'. There is no need to direct translate the word 'at'.

R: So what was on your mind when I was explaining this word to you?
CFOFS9HS: You also mentioned that 'ease' came from 'easy', which gave me a deep impression because this confirmed what I was thinking when I heard the sentence. Then, I used the Chinese meaning of 'easy' to help me remember the meaning of this collocation.

As it can be seen from the transcript, the learner was first linked 'ease' and 'easy' through the help with teacher's explanation. She then felt that the explanation had given her a deep impression about the word itself (*making judgements*). Thereafter, she employed the *translation* strategy to translate the word 'easy' and used the Chinese meaning of 'easy' to help her remember the meaning of ***at ease***. It seems that the learner paid much attention to the individual words 'easy' and 'ease' when receiving the vocabulary explanation and the method she employed to remember this collocation was to using the Chinese meaning of the individual word 'easy' rather than using any clues from the listening passage.

This selective attention to individual words in the explanation provided was also confirmed by the other two learners in the CFoF group: CFOFS10LS and CFOFS11HS. The former also paid much attention to 'easy' when understanding the vocabulary explanation for ***at ease***, and the latter focused on the individual word 'threat' to help her understanding the meaning of target word ***threaten***. Relevant transcripts and interpretations were given in the above Section 4.10.5.1.

Learners in the L2 group, on the other hand, although they also paid attention to the target lexical item itself and certain known vocabulary items within the L2 meaning in the first place, the learners were trying to understand the L2 meaning as a whole.

Listening segment 5: The acting is so ***convincing*** that it makes you believe that it is one of the best meals he has ever tasted!

Vocabulary instruction: ***Convincing*** is an adjective, which means 'that makes somebody feel certain or sure'. An example sentence for how to use this word can be 'Ling Dan has a convincing success in 2012 Olympic Games.'.

R: What were you thinking when I was explaining this word to you?
L2S2HS: In your explanation, there is a word "sure". That is '确信 (sure)'. So the meaning of ***convincing*** is '使某人感到确信 (that makes sb. feel sure)'.
R: What else were you thinking?
L2S2HS: Actually I felt that I had gathered a general Chinese meaning of this word from your English explanation, so that I can manage to understand it.

The above transcript was from a more skilled learner, L2S2HS, from the L2 group. When receiving the L2 explanation provided by the teacher, she first paid attention to the familiar word 'sure' in the L2 explanation (*selective attention on known words*). She then employed the *translation* strategy trying to translate every single word in the L2 meaning into Chinese. Finally, she felt that she had understood the meaning of the target item because she was able to summarise the L2 meaning using an L1 word (*summarisation*). It was likely that although L2S2HS was paying attention to an individual word 'sure' in the L2 meaning initially, the reason why she did this was to help her generalise and translate the whole L2 explanation into Chinese, which then led her to have a more global understanding of the vocabulary explanation provided.

Similar to the L2 group, the learners in the CS group also focused more on a global understanding rather than on individual lexical items.

Listening segment 3: Making a fist and shaking it almost always means that someone is angry and ***threaten***ing another person.

Vocabulary instruction: ***threaten*** is a verb, which means '恐吓'. An example sentence for this word can be 'They threatened to kill him unless he did as they asked.'.

R: What did you do to cope with my vocabulary explanation?
CSS5HS: In the listening passage you gave us, there was 'angry'. And you said that this word is a verb, so I guessed it might have something related to 'anger' and I was right. Also, I felt that the example sentence that you provided was very easy to remember, so I actually used the example sentence to help me remember ***threaten***.

The above transcript was from CSS5HS, a more skilled learner from the CS group. In order to understand the teacher explanation, what she did was to refer back to the listening segment including the target lexical item. Then, she used *linguistic contextualisation*,

linking ***threaten*** to the word 'angry' in the original listening passage. At the same time, she also tried to guess the meaning through *linguistic inferencing*. Finally, she found the additional sentence provided by the teacher in order to exemplify the target word was helpful (*making judgements*) and used this sentence to help her remember (*rely on examples*).

Therefore, similar to L2S2HS, CSS5HS also initially paid attention to individual word (i.e., angry) when trying to understand teacher's explanation. However, what she did afterwards was to use the grammatical structure of the sentence including the target item to guess the meaning. In addition, rather than focusing mainly on the L1 meaning provided, she used the example sentence to help her to remember the target word. It was likely that CSS5HS was trying using more global resources rather than restricting on local information.

In summary, the learners from different treatment conditions employed different strategies. Correspondent to these different strategies used, learners in different treatment conditions tended to attend to different things. While learners from the L2 and the CS group were able to use global information, learners in the CFoF group tended to focus more locally on individual lexical items.

CHAPTER 5 DISCUSSION

5.1 Introduction

Within the context of communicative language teaching, the present study, employing a quasi-experimental design, investigated the impact of different types of vocabulary instruction on short-term and long-term vocabulary learning of 137 Chinese senior secondary school EFL learners. These learners were assigned to three treatment groups and a control group. A classroom intervention was conducted. During each intervention session, all four groups were required to complete a listening comprehension task. While the intervention groups, after listening, were taught the meaning of the target lexical items which appeared in the listening comprehension passage, the Control group was taught about culture and background information relating to the passage. Learners' vocabulary knowledge of these target items was tested through a pre-test (four weeks before the first intervention session), a post-test (immediately after each intervention session), and a delayed post-test (two weeks after each intervention session).

While the original focus and interests of the present study were mainly about the impact of vocabulary instruction on vocabulary learning, two additional research questions emerged before data collection. On the one hand, what is the potential impact of different types of vocabulary intervention on learners' general listening comprehension? On the other hand, what learning strategies do learners use in response to the instruction for the target items offered by the teacher in each of the three experimental conditions? These two questions were believed to be important and findings for these would supplement what was found in the first research question and give a clearer picture of the vocabulary intervention. In this chapter, the findings for these research questions will be discussed with reference to the theories and the empirical background presented in Chapter Two.

5.2 What is the impact of the three types of vocabulary instruction on vocabulary learning and retention?

This study compared the impact of three different types of vocabulary instruction during listening activities (teacher codeswitching; target language explanations; and contrastive Focus-on-Form) on vocabulary learning and retention. Results reveal a clear short-term pedagogical advantage for vocabulary instruction with teacher CS compared with L2-only explanations. On average, while approximately 23 out of 60 target lexical items were immediately recalled by the L2 group, roughly 39 out of 60 items were acquired by the CS group immediately after the intervention. This advantage disappeared, however, for two weeks' long-term retention of the target lexical items. Learners in both the CS and the L2 group retained around 16 out of 60 lexical items at the delayed post-test, a finding that corresponds with that of Tian (2011) where there was only short-term advantage of using teacher CS over L2-only vocabulary instruction. In their study, on average, the CS group acquired around 100 out of 170 vocabulary items, which was significantly higher than the NCS (L2-only) group who recalled on average 83 out of 170 target items at the immediate post-test. However, at the delayed post-test, both the CS and the NCS group only retained approximately 19 out of 170 lexical items.

The short-term advantage of teacher CS provides further evidence for the revised hierarchical model of lexical and conceptual representation in bilingual memory (Kroll & Stewart, 1994), whereby the L1 remains the dominant language for language learners who start learning an L2 beyond early childhood, as a strong link between L1 vocabulary knowledge and its conceptual representations is already created. Before being represented as a meaningful mental concept, L2 words need to be attached to L1 translations through lexical links in the first place. Thus for learners in the CS group, the meaning of the L2 word is accessed through the L1 translation and therefore the learning is facilitated. Learners in the L2 group, however, were not able to or might experience difficulty creating such a link through L2-only explanations, and therefore were disadvantaged for short-term learning.

Similarly, as was proposed in Jiang (2000)'s psycholinguistic model, by giving the L1 translation of the target L2 lexical item and thus creating a link between the L2 word and its concept, teacher CS seems to have supported learners' movement on the path towards the second stage of Jiang's model of vocabulary acquisition (L1 lemma mediation stage). By contrast, learners in the L2 group may have experienced difficulties in moving beyond the initial stage of learning, and hence were outperformed by the CS group in the short-term.

The disappearance of the CS versus L2 advantage two weeks after the intervention in the present study, while concurring with that of Tian and Macaro (2012), contradicts what was found in Hennebry et al. (2013), where there was a long-term advantage of using direct L1 translation over L2-only for vocabulary instruction with young learners of L2 French. However, it should be noted that in Hennebry et al. (2013), the vocabulary post-test was placed between one and four weeks after the intervention. One possible explanation could be that in their study, the L1 translation group retained more vocabulary items taught on the final intervention session (one week before the test) than the L2-only group did, which made them significantly outperform the L2-only group in the vocabulary post-test. However, the vocabulary knowledge was tested after a longer interval in both Tian and Macaro (2012) (between two and seven weeks after the vocabulary instruction) and the present study (two weeks).

Therefore, the balance of evidence indicates that providing L1 translations for vocabulary instruction, namely teacher codeswitching, does not have long-term benefits (more than one week) when compared with L2-only instruction. Again with reference to Jiang (2000, 2004b) this may be because the link between the L2 word and its conceptual representations is fairly weak at the L1 lemma mediation stage, as the lemma information is largely copied from the learner's L1 rather than being "created in the process of learning" (Jiang, 2000, p.52).

In addition, a third approach, contrastive Focus-on-Form, was explored in the present study, going beyond what was previously investigated. Learners receiving CFoF instruction had better outcomes for both short-term and long-term learning than learners experiencing teacher CS or L2-only explanations. This compares with the findings for reading in Laufer and Girsai's (2008) study. There, L2 learners who received CFoF instruction retained significantly more vocabulary items than their counterparts who received L2-only vocabulary explanations at both immediate post-test and one-week delayed post-test.

A further added dimension for the present study comes from showing that giving cross-linguistic information in vocabulary explanations is more beneficial than direct L1 translation through teacher CS as well as being more beneficial than L2-only vocabulary explorations. Furthermore, the advantage of CFoF over both teacher CS and L2-only persisted on a long-term basis (two weeks after the vocabulary instruction), possibly underscoring the importance of 'noticing' (Schmidt, 1990, 1994, 2012). Learners in the present study (aged 15-16), although having greater cognitive capacity than younger learners, may have already been dominated by their native language. Compared with younger counterparts who are able to implicitly learn from the input, when it is plentiful, these learners rely heavily on explicit learning from the instructed classroom context where a large amount of naturalistic input seems to be impossible. Therefore, through CFoF, a more conscious and explicit approach, cross-linguistic information is provided, which makes an L2 word more salient to these learners and hence aids short and long-term acquisition.

The superiority of CFoF over teacher CS also corresponds to what was found by Wright and Cervetti's (2016) review, whereby more 'passive' forms of direct teaching (i.e., where the teacher just gave the dictionary definition) were less successful than direct teaching which required more active processing by learners. Wright and Cervetti relate their findings to the 'depth of processing' hypothesis (Craik & Lockhart, 1972) and considered comparing and contrasting word meanings as one type of 'active' processing. It is likely

that, through CFoF instruction, learners were able to compare and contrast the L2 word and its L1 translation, which involved a deeper processing of the target L2 word. This deeper processing then led to a better vocabulary retention.

Lexical fossilization, whereby learners do not move from the second stage (L1 mediation) to the third stage (the L2 integration stage) (Jiang, 2000, 2004b), is a possible explanation for the lack of long-term advantages for teacher CS. This may be because "the presence of L1 lemma within the L2 lexical entry may block the integration of L2 lemma information" (Jiang, 2000, p. 55). Furthermore, if learners can access a word's meaning easily through the L1, they may be less likely to extract further information about a word through contextualised input, a necessary step in the movement to the L2 integration stage (Jiang, 2000, 2004b). In the latter, semantic, syntactic and morphological information are established within the L2 lemma and integrated into learners' mental lexicon. The superior short and long-term vocabulary learning for the CFoF group suggests that this kind of teaching helps learners move nearer towards the L2 integration stage.

As discussed in the Methodology chapter, for the target 60 lexical items, there was variability in the degree of mismatch between L1 and L2 meanings and their suitability for the CFoF treatment. This may have had potential impacts on findings; the superiority of the CFoF treatment may be due to the fact that their learning of a small subset of lexical items which have a strong mismatch between L1 and L2 meanings was significantly better than the L2 and CS groups. It therefore should be noted that instead of using two-way mixed ANOVA, generalised linear mixed-effects modelling may be a more preferable statistical test, which can handle both participants and test items effects.

The L2 group, in general, at first sight seems to be the most disadvantaged within the three treatment groups, as they were outperformed by the CS for short-term learning while were significantly below the CFoF for both short-term and long-term vocabulary acquisition. One thing, however, needs to be pointed out. The vocabulary forgetting rate of the L2 group from the post-test to the delayed post-test was much lower than that of the CS group,

while slightly higher than the CFoF group. Around 50% of the vocabulary knowledge recalled immediately after the intervention was retained two weeks later by the L2 group. The CS group, however, only retained 28% of the knowledge gained at the post-test. This is correspondent with what was found by Tian (2011). In her study, learners in the non-codeswitching group (who received L2-only vocabulary explanations) retained 23% of the knowledge gained while the CS group who only retained 17% of the knowledge gained. The forgetting rate in the present study, in general, was lowest for the CFoF group. Approximately 53% of the vocabulary knowledge gained was retained at the delayed post-test.

It should be noted that vocabulary gains calculated were absolute gains, which did not take participants' pre-knowledge of these vocabulary items into account. Therefore, results may be affected in a way that vocabulary gains for learners with higher pre-test scores can be relatively smaller, yet can be relatively higher for those with lower pre-test scores. In cases like this, relative gains are preferable. However, it is believed that the findings of the present study were largely unaffected by this issue. As the mean scores of the vocabulary pre-test were approximately similar between the L2 and CS group and both the L2 and CS group had a higher pre-test score than the CFoF group, it was likely that the vocabulary gains for the L2 and CS group were relatively overestimated while they were underestimated for the CFoF group. Therefore the superiority of the CFoF treatment over the L2 and CS is still convincing.

One possible reason for why learners in the L2 group forgot much less than the CS from the post-test to the delayed post-test could be due to the fact that learners in the L2 group had to work harder to figure out the L2-only explanations. The qualitative analysis of strategies used by learners in each of the three experimental conditions showed that the most frequently reported strategies used by the L2 group was *translation*. It is possible that these learners, in addition to trying to understand the L2-only explanations, also made efforts to translate these explanations into the L1, which led them to a deeper processing level with reference to the 'depth of processing' hypothesis (Craik & Lockhart, 1972).

Thus, it is likely, once they succeed in translating, they had a deeper impression of the vocabulary item itself and its correspondent meaning. It could also be possible that for these learners, they tried to remember both L2 meaning and its L1 translation. They then therefore had a greater chance to retain what they acquired compared with their counterparts from the CS group, who were given the L1 meaning by the teacher.

Regarding the performance of the Control group, results show that at post and delayed post-test they scored significantly lower than the three treatment groups. This confirms the previous empirical evidence in both reading (Laufer & Girsai, 2008; Sonbul & Schmitt, 2010) and listening contexts (Hennebry et al., 2013; Tian & Macaro, 2012), indicating that in order to improve learners' vocabulary knowledge to a large extent, incidental vocabulary learning needs to be supplemented by lexical Focus-on-Form (Laufer, 2005, 2009; Schmitt, 2008).

However, the Control group still made significant pre to post-test vocabulary gains. On average, learners in the Control group gained approximately six out of 60 lexical items at the post-test. This indicates that vocabulary knowledge can be acquired incidentally through listening, confirming what was found by previous empirical studies (Brown, et al., 2008; Vidal, 2003, 2011; van Zeeland & Schmitt, 2013a). However, incidental learning through listening within the present study was found to be very unstable on a long-term basis. The Control group learners' performance dropped dramatically at the delayed post-test compared with the post-test, bringing it back to a level not significantly different from the pre-test. This contradicts what was found by Vidal (2011) and van Zeeland and Schmitt (2013a), both investigating incidental vocabulary learning through listening among university level L2 learners, whereby longer-term vocabulary meaning retention was confirmed. One possible reason may be attributable to the relatively low language proficiency of the senior secondary school EFL learners in the present study, compared with the university-level participants in Vidal (2011) and van Zeeland and Schmitt (2013a). Like the lowest proficiency students in Vidal (2011), whose vocabulary gains through listening were relatively small and short-term compared with high proficiency students,

learners in the present study may have experienced difficulties with speech segmentation, isolating and focusing on specific lexical items. Hence their longer term learning may have been challenged in the incidental listening condition.

One thing should be noted is that, for these learners, the six listening passages may have been too challenging. When designing the study, the way in which the percentage of lexical coverage of the listening passages was calculated was based on the senior-secondary school English curriculum and the percentage of items that the textbook highlighted as likely to be unknown. In reality, however, when the learners' vocabulary size was measured later before conducting the intervention, it became evident that that the listening passages were too challenging for these learners. This may have posed extra difficulty for the learning of the Control group who did not receive any vocabulary intervention.

5.3 What is the impact of the vocabulary intervention on learners' general listening comprehension?

Turning to the second research question, in order to explore whether the vocabulary intervention would influence learners' general listening comprehension, two listening comprehension tests were used, one at the baseline test session while the other took place after the vocabulary intervention and review. Findings indicate that vocabulary learning through oral input can bring improvement in listening proficiency, supporting the argument that other language skills in addition to vocabulary knowledge can be improved through incidental vocabulary learning (Nation & Webb, 2011). Significant pre to post-test improvement in listening was found for two of the intervention groups, the CS and L2 groups, but also for the Control group, who received six sessions of listening comprehension and culture study but no vocabulary explanations. In addition, the Control group, who made the least progress in vocabulary learning, showed the greatest improvement in listening. This is contrary to what one might expect from studies that have found a strong and significant relationship between vocabulary knowledge and listening comprehension (e.g., Andringa et al., 2012).

Furthermore, the listening comprehension scores of the CFoF group did not improve after the intervention, and were significantly below those of the Control group at post-test (with no significant differences at pre-test). Again this is surprising, given that the CFoF group made the greatest gains in vocabulary. This is, however, in line with what was found in a recent systematic review (Wright & Cervetti, 2016) in the reading modality. The review, in general, indicates that studies conducting vocabulary interventions (either through directly teaching vocabulary meanings or through teaching word-learning strategies) did not have a positive impact on learners' general reading comprehension. As suggested by Wright and Cervetti (2016), possible reason could be that the length of the vocabulary interventions conducted in the studies they reviewed was rather short. In order to have an impact on general comprehension, more longitudinal vocabulary interventions, teaching larger amount of vocabulary items in greater depth, seem to be necessary.

The findings regarding listening comprehension gains suggest that growth in vocabulary knowledge was not the only factor or the most important factor that led to the improvement in listening, which further confirms what was discovered in Goh (2000) where half of the listeners in the study reported that they still experienced difficulty in understanding the full meaning of listening passages although they literally understood meaning of the single words. The listening gains by the Control group may have arisen precisely because they received no vocabulary explanations and hence were pushed towards seeking global understanding of the passage rather than a detailed focus on individual vocabulary items. While such an approach may have disadvantages for vocabulary acquisition (Vidal, 2011), it is arguably beneficial for the development of strategic behaviour in listening. This may then have improved the ability of the Control group to construct coherent meaning from the listening passage. This may be true too for the L2 group, who were required to understand the L2 explanations they received without recourse to L1 translations.

The CS group made as much progress in listening as the L2 group, however, who were given direct translations for vocabulary items contained in the listening passages. One might have expected this approach to limit their development of listening strategies or a

focus on global understanding. Yet global understanding may still have been encouraged by another aspect of the vocabulary instruction that was common to both the CS group and the L2 group, namely the giving of an additional sentence exemplifying the target lexical item after the codeswitched or L2 explanations. Learners in both groups then had to understand this additional sentence, potentially focusing them more on global understanding than on the individual lexical item. The qualitative analysis of learners' strategies within each type of the instruction provided further confirmation of this. One learner from the L2 group initially focused on the known words in the L2 meaning and used these known words to guess the meaning of the L2 explanation and summarise it into Chinese. Similarly, one learner from the CS group used known words in the listening input and use linguistic contextualisation to guess the meaning of the unknown word. She then used the example sentence provided to remember the meaning of the unknown word.

By contrast, the CFoF group was given no additional exemplifying sentence after the vocabulary explanations, just cross-linguistic information comparing the functioning of the target vocabulary item functions in L1 and L2. This may have encouraged a focus on isolated lexical items and understanding the cross-linguistic information provided, potentially less beneficial for the development of listening strategies and listening comprehension. This was also confirmed by the qualitative data collected. Three out of four learners from the CFoF group who were selected to participate in the interviews reported that they attended mainly to the familiar words in the cross-linguistic information provided by the teacher and use the familiar words to remember the target unknown vocabulary item. In addition, the main differences occurred between the CFoF group and the other three groups on Section Two of the listening comprehension post-test, which as well as containing some challenging vocabulary, also required learners to understand the listening passage globally and to use inferencing skills to overcome vocabulary limitations. It is possible that the other three groups were better equipped to answer those questions correctly than the CFoF group was.

Regarding the "cost-benefits of Lexical Focus-on-Form by the teacher" raised by Tian and Macaro (2012, p. 371), in the present study approximately three minutes were spent teaching one target vocabulary item. This may lead one to expect that the average vocabulary retention rate would be higher on a long-term basis (Schmitt, 2008). In reality, except learners in the CFoF group who retained nearly half of the total number of target lexical items taught (29 out of 60, on average) at the delayed post-test, the long term retention rate for the L2 and the CS group was fairly low (16 out of 60, on average). This is, however, still consistent with what was reported by Laufer (2005), who summarised the results of three of her studies investigating the impact of explicit lexical Focus-on-Form on vocabulary acquisition, indicating that although around 70% of taught vocabulary knowledge was retained at the immediate post-test, the rate of the two-week long-term gains dropped to 21-41%.

Hence, it is still possible to argue that there are merits in adopting explicit lexical Focus-on-Form (Laufer, 2005, 2009) within the classroom context, with the pre- and post-listening comprehension test scores being drawn into the picture. On the one hand, although the CFoF group did not make significant progress in listening comprehension, learners within this group made the largest long-term vocabulary gains among the three treatment conditions. On the other hand, while the CS and the L2 group were less advantaged compared with the CFoF group in vocabulary acquisition, these two groups made significant progress in listening comprehension through explicit lexical Focus-on-Form. In other words, researchers need to weight up the relative cost-benefits of different forms of instruction.

5.4 Does the impact of each type of instruction vary according to learners' English language proficiency?

As the present study did not randomly assign participants to different treatment groups, learners' English language proficiency level was determined by their composite test scores. Data analysis showed that there were group differences in the composite test score before

the intervention. Therefore, in order to address the third research question, learners were assigned to four proficiency levels both across and within each treatment condition.

Regarding the high or the low proficiency level learners' vocabulary learning differences between the three treatment conditions, findings were generally in line with what was found for the first research question. Regardless of the proficiency level, progress of the learners in the CFoF group was significantly better than both the L2 and the CS group for both short-term and long-term learning. While the CS group also significantly outperformed the L2 group for short-term learning, no significant long-term learning differences were confirmed between the CS and the L2 group. However, one exception was found. While the high proficiency level learners in the CFoF group outperformed their counterparts from the CS group for short-term vocabulary learning, there were no significant short-term learning differences for the low proficiency level learners between the CS and the CFoF group.

These findings seem to suggest that unlike the high proficiency level learners who might benefit more from the CFoF explanations than from the CS explanations, both the CFoF and CS explanations are beneficial to the short-term vocabulary acquisition of the low proficiency level learners. It is possible that due to fairly low proficiency level, these learners in the CFoF group had difficulty understanding the cross-linguistic information or were not able to use this information to help them remember the target word. Therefore, their performance was similar to those in the CS group at the immediate post-test.

Findings reveal that vocabulary gains at both post-test and delayed post-test were similar across the four proficiency levels within the CS and CFoF group respectively, meaning that no learners of a particular proficiency level benefited more or less for both short-term and long-term learning under the two types of vocabulary instruction. The situation was the same for the L2 group for two weeks' vocabulary retention. For gain scores at immediate post-test, however, the higher proficiency learners benefited more from the L2 condition than the lower proficiency level did. Furthermore, the low level group was significantly

outperformed by all the other three proficiency level groups (High, Mid-High and Mid-Low).

While the findings for the CS group concurred with those of Tian and Macaro (2012) and Hennebry et al. (2013), the findings for the L2 group contradict with what was found by these two studies. Hennebry et al. (2013) suggest one possible reason explaining why, in their study, results did not confirm that the higher proficiency level learners benefited more from the L2-only explanations than their lower-level counterparts. They argue that the participants' proficiency levels were too similar to make a difference. They then suggested that future studies need to include "much more accentuated levels of proficiency" in order to explore the differences within each type of instruction (p. 15).

In the present study, the mean of the composite scores for the low level learners (109.91) within the L2 group was much lower than the mean for the other three level learners (High: 150.15; Mid-High: 137.31; Mid-Low: 128.44) and as assured by statistical tests, there were significant differences between each two adjacent proficiency levels on their composite scores within the L2 group. A further look into the range of the composite scores for each L2 proficiency level group also confirmed that within the L2 group, there was a much more accentuated level of proficiency (Hennebry et al., 2013). Therefore, the clear differences between each proficiency level allows the present study to provide a clearer picture for the performance of each proficiency level group within the L2-only condition.

In addition, the findings provide further evidence for Kroll and Stewart (1994)'s hierarchical model of bilingual memory. According to the model, although L2 words need to be attached to their L1 translation before access to the underlying mental concepts occurs, direct conceptual links between L2 words and conceptual representations can also be created when further exposure is reached and bilinguals are becoming more fluent in the L2. It is likely that the higher level learners in the L2 group may have bypassed the early stage of language learning, meaning that they may have established direct conceptual links between the L2 words and their concepts, therefore they outperformed the lowest level

learners within the same group. It should, however, be noticed that according to the model, for these more fluent bilinguals, the lexical links from L2 to L1 do not disappear even if direct links are made between L2 and the concepts. This then explains why the vocabulary learning of the L2 group, on the whole, was below the CS group for short-term learning.

5.5 Does the impact of each type of instruction vary according to different word classes and for collocations?

Turning to the fourth research question, investigating whether the impact of each type of instruction varied according to the three word classes (nouns, verbs, adjectives) and for collocations, the present study analysed the data from two perspectives: first, within each treatment group, whether there are any differences in both short-term and long-term learning of collocations and the three word classes. Second, whether the learning outcomes of a specific word class or the collocations were different across the three treatment groups.

Results generated from analysing data from the first perspective initially indicate that within each type of instruction, collocations can be intentionally acquired through oral input, confirming the findings of Laufer and Girsai (2008) and Alali and Schmitt (2012), both of which investigated the impact of different types of classroom instruction on the acquisition of single words and collocations. Findings also reveal that both short-term and long-term learning of collocations significantly outperformed the learning of verbs and adjectives within the L2 and CS conditions, whereas within the CFoF group, the learning of collocations was significantly better than the learning of verbs and nouns on a short-term basis and better than the learning of verbs and adjectives on a long-term basis.

Although the findings in general do not show a clear picture of whether there were any differences between the learning of collocations and nouns within each treatment condition, the retention of collocations was clearly better than that of verbs and adjectives, regardless of the type of instruction. While these findings concur with Laufer and Girsai's

(2008) study in which both immediate and delayed passive and active recall of the target collocations was significantly better than that of the single words under the three types of classroom instruction, they contradict what was found by Alali and Schmitt (2012). In their study, the rates of form recall and meaning recall of the target single words were significantly higher than those of the idioms at both immediate post-test and delayed post-test. It should, however, be noted that the target single words used in their study were extracted from target idioms, so that exposure to single words was double compared to those of idioms. The greater exposure to the single words would possibly lead to the better long-term retention of these single words. In addition, as single words included both nouns and verbs in their study, it was rather unclear whether the learning of idioms was outperformed by the nouns or the verbs.

Another possible reason for why collocations were in general better acquired than single words concerns the characteristics of the collocations selected in the present study. In addition to phrasal verbs, most of the target collocations were formulaic sequences which are "stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray, 2007, p. 9). These formulaic sequences on the one hand tend to be long and abstract, but on other hand impose a higher learning burden and require more effort to learn, and therefore were better acquired.

In addition, within the L2 and the CFoF group in the present study, there were instances where the learning of nouns was significantly better than that of verbs (short-term for the CFoF group while long-term for both the L2 and CFoF group). The finding is supported by van Zeeland and Schimtt (2013a), a study investigating incidental L2 vocabulary learning through listening. In their study, when combining the post-test and delayed post-test results, meaning recall of nouns was significantly better that of verbs. Unlike van Zeeland and Schmitt (2013a), however, who did not find any significant differences between meaning recall of adjectives and verbs, within the CS and the CFoF group in the present study, adjectives showed better short-term gains than verbs. However, this should be considered with caution as this superiority of learning adjectives over learning verbs was

only short-term and disappeared after two weeks' delay. Therefore, it is safe to argue that there is a tendency for nouns to be more easily acquired than verbs, yet the evidence regarding the superiority of learning adjectives over nouns is not conclusive.

Regarding the results from the second perspective, exploring whether the learning of a specific word class and for the collocations is consistent across the three intervention conditions, findings, in general, further confirm what was found by Laufer and Girsai (2008). In other words, there is pedagogical value in comparing and contrasting a target L2 vocabulary item and its L1 translation in the teaching of either L2 single words or collocations. Learners from the CFoF group, who received cross-linguistic information about the target lexical items, not only significantly outperformed the CS group on the retention of both single words and collocations at the delayed post-test, they were also significantly superior to the L2 group on both short-term and long-term learning of the three word classes and for collocations.

5.6 Does the number of repetitions influence vocabulary retention?

In order to explore whether the number of repetitions would influence the retention of single words which received explicit lexical Focus-on-Form through listening, the present study employed a review activity at the end of each intervention session and had two additional review sessions after the intervention procedure was completed. Within the total 43 single words, 28 words were repeated three times, while the remaining 15 words were allocated to three groups (five words in each group): a nine repetition group, a seven repetition group and a five repetition group.

The findings, on the whole, indicate that the higher the number of repetitions, the greater the chance that the target lexical items would be retained. The learning outcomes for five words receiving nine repetitions were significantly higher than the outcomes for their counterparts receiving seven, five or three repetitions. The situation was the same across the three treatment groups as well as within each treatment condition. This is a lower

number of repetitions than what was found in studies investigating incidental vocabulary learning through reading in an L2 learning context (Nation & Wang, 1999; Pigada & Schmitt, 2006; Pellicer-Sánchez & Schmitt, 2010; Webb, 2007) which suggested that at least, ten repetitions were needed in order to improve vocabulary learning to a large extent. However, in the context of incidental vocabulary learning through listening, the nine repetitions used in the present study, although lower than the 11 repetitions suggested by van Zeeland and Schmitt (2013a), is higher than what was found by Vidal (2011), in which L2 listeners' vocabulary gains grew rapidly when five exposures were increased to six.

However, it should be noted that both Vidal (2011) and van Zeeland and Schmitt (2013a) did not suggest a clear picture for the relationship between the rate of meaning recall and the frequency of occurrence of the target words. The former found that, although at the immediate post-test, the meaning recall of the single words which received 11 repetitions was significantly higher than those receiving three and seven repetitions, this advantage disappeared at the two-week delayed post-test. This suggests that vocabulary meaning incidentally acquired through listening does not seem to be influenced by the number of repetitions. Similarly, Vidal (2011), comparing vocabulary gains between incidental learning through reading and through listening, indicates that the frequency of occurrence of the target words is the least important factor, only explaining 24% of the vocabulary gains through listening, while it is the most important factor, explaining 47% of the gains, for vocabulary learning through reading.

Considering the differences between the contexts of incidental learning and intentional learning whereby the former promotes learning without consciously paying attention to target lexical items while the latter aims at committing lexical information to memory, it is not surprising that the target single words in the present study (intentional learning), required fewer repetitions than the items in studies investigating incidental vocabulary learning (e.g., van Zeeland & Schmitt, 2013a). However, the exception of Vidal (2011) still needs to be explained. One possible reason could be that Vidal (2011) used a vocabulary knowledge scale to rate vocabulary gains, from being able to recognise the word form to

being able to recall the word meaning, whereas the present study and van Zeeland and Schmitt (2013a) used a meaning recall test. In the former, vocabulary gains are scored if learners are able to recognise the target word, while in the latter, gains are counted when learners are able to successfully recall the meaning. It is likely that the six repetitions found by Vidal (2011) to be necessary for greater gains is based on the evidence of form recognition rather than meaning recall, therefore is lower than the nine repetitions found as the present study to be more beneficial.

The role of frequency has been studied extensively in relation to incidental learning, but not much in relation to explicit vocabulary instruction. Hence the previous discussion has mainly focused on comparing the findings of the present study with previous studies of incidental learning. There are, however, two studies of explicit vocabulary instruction of note. One by McKeown, Beck, Omanson and Pople (1985) compared vocabulary learning within three types of explicit vocabulary teaching. Target vocabulary items were repeated either four times or twelve times. Results indicated that vocabulary items that occurred twelve times were acquired significantly better than those that were repeated four times. McKeown et al.'s (1985) finding is generally in line with the finding of the present study. As only four and twelve frequency of occurrence were manipulated in their study, the major effects of frequency of occurrence could happen at any lower repetition rate between four and twelve.

Another study by Peters (2014) compared the short-term learning (immediately after the explicit teaching) and long-term learning (one week and two weeks after the explicit teaching) of single words that received one, three or five repetitions. In general, she discovered that the words that received five repetitions were acquired better than those that appeared three times or only once, which is a smaller number of repetitions than what was found in the present study. This is not surprising as Peters (2014) only measured form recall of the target items. However, meaning recall of the target single words was adopted in the present study, which is a relatively more difficult task than form recall.

In reality, for the present study it is possible that the required number of repetitions could be lower, lower than nine repetitions, for significant progress in vocabulary learning through listening to occur. In the present study, more words received three repetitions than received five or seven repetitions. This difference in group size means that it is possible that learning of items in the seven and five repetition groups is not significantly better than the learning of the items in the three repetitions group.

5.7 What learning strategies do learners use in response to the instruction for the target items offered by the teacher in each of the three experimental conditions?

Turning to the final research question, qualitative data were collected through stimulated recall interviews with 12 selected learners from the three treatment groups (four from each group). The interview transcripts were then coded and altogether 25 learner strategies were identified (see Table 4.24 in Section 4.10.3). The range of strategies learners employed in response to the teacher's vocabulary explanation was much larger than that reported by Macaro (2014), a study exploring how English learners of French understand the L2-only or L1-translation vocabulary explanations provided by the teacher. In his study, these L2 French learners mainly used the strategy 'searching for L1 cognates' while ignoring the vocabulary explanation provided by the teacher. This is, however, not surprising. Different from French, Chinese does not have real cognates with English. Therefore, Chinese EFL learners in the present study were not able to rely on finding cognates, which then may have led them to employ a wider range of different strategies.

5.7.1 What are the strategies used?

It was discovered that learners employed different strategies during the classroom intervention for three purposes. First, they initially used different strategies such as *linguistic inferencing*, *translation*, *deduction*, *selective attention on known words*, *monitoring* in order to understand the listening input as well as to discover the meaning of

the unknown lexical items they heard. These are correspondent to those strategies for L2 listening comprehension summarised by Vandergrift and Goh (2012, p. 277-284) and in line with those vocabulary learning strategies for the discovery of a new word's meaning in the vocabulary learning strategies taxonomy developed by Schmitt (1997) (see Table 2.1 in Section 2.4).

Second, in response to the vocabulary explanations provided by the teacher, learners again employed L2 listening comprehension strategies and strategies for the discovery of a new word's meaning. Although the CS and the CFoF group received an L1 translation for the target lexical items, other vocabulary explanations they received (example sentence for the CS group and the cross-linguistic information for the CFoF group) were given in the L2. Hence, they were still trying to understand the L2 listening input and trying to discover the meaning of other unfamiliar words in the vocabulary explanations. Some strategies found to be used at this stage, however, do not fall into the categories of Vandergrift and Goh (2012) and Schmitt (1997), e.g., *waiting*, *making judgement*, *ignoring*, *give up*. This is not surprising as these strategies relate to the teaching procedure used during the intervention. As learners were aware that vocabulary explanations would be provided by the teacher, they adopted the *waiting* strategy to wait for the teacher's explanations. In addition, learners tended to make judgements (*making judgement*) about the way the teacher explained the vocabulary to them, especially for a type of instruction that differed from what they normally experienced in school English classes (e.g., CFoF explanations). For explanations which were too difficult for learners to understand (i.e., L2-only explanations), learners chose to ignore (*ignoring*) or just *give up*.

Finally, learners employed strategies such as *repetition*, *keyword strategy*, *using phonological knowledge*, with the purpose of committing lexical information to memory, which concurs with the strategies for consolidating a word once it has been encountered that are listed in Schmitt's (1997) vocabulary learning strategies taxonomy. It should be noted that learners in the present study were not required by the teacher to remember the lexical information. Although learners may have been aware that they were going to be

tested at the end of the session, it is still surprising that all learners being interviewed, regardless of proficiency levels, mentioned that they were trying to remember the vocabulary explanations provided by the teacher. A possible reason may be that trying to remember a new vocabulary item is a normal learning pattern that learners have been used to since they started learning English. As reported by a more-skilled learner, CSS5HS from the CS group, since primary school, she was familiar with using verbal repetition to remember the meaning of a new word.

5.7.2 What are the differences in overall strategy use across the three treatment conditions?

Findings from the quantitative overview of strategy use (counting the number of instances of different strategies reported) across the three treatment groups reveal that, in general, learners from the L2 group reported more instances of strategy use than learners from the other two groups did. Again, unlike Macaro (2014), in which a very restricted range of strategy use was confirmed, with no group differences between the two treatment conditions (L2-only group and L1-translation group), the present study found a wider range of strategy use pattern and the pattern of strategy use was different across the three treatment conditions.

Findings from the qualitative analysis indicate that two combinations of strategies were shared by learners from the L2 group: one is combining *matching with familiar words* and *translation*, while the other is combining *selective attention on known words*, *translation* and *making judgement*. However, only one type of strategy combination was found in the CS group and the CFoF group respectively. Learners in the CS group used the combination of *waiting*, *force* and *repetition*, while one strategy use pattern was reported by learners from the CFoF group, namely *selective attention on known words*, *waiting* and *translation*.

The two strategy use patterns found in the L2 group suggest that learners in that group used more L2 listening comprehension strategies than vocabulary learning strategies. In

addition, in both of the two patterns, *translation* strategy was used. A possible reason for this may be that learners in L2 group were struggling in understanding the listening input (either the listening comprehension material or the L2-only vocabulary explanation), therefore they were trying to translate the known words in the listening input to help them understand. Compared with these learners from the L2 group, it makes sense that learners in the CS group were just *waiting* and *force*(ing) themselves to remember the meaning through verbal *repetition*, as they were provided with L1 translation of the target lexical items by the teacher. Similarly, learners from the CFoF group also employed the *waiting* strategy to wait for the L1 translation, yet they also combined it with using the *selective attention on known words* and *translation* strategies as learners in the L2 group did. A possible explanation for this may be that additional cross-linguistic information was given to learners in this group. In order to understand this, learners focused on the words they knew in the cross-linguistic information and tried to translate them into Chinese.

These different strategic behaviours discovered across the three treatment groups provide additional information for the quantitative findings regarding the different vocabulary gains and L2 listening progress between groups. Learners from the L2 group, on the one hand, made significant progress in listening as they used more L2 listening comprehension strategies. This, on the other hand, indicates they were struggling with understanding both the listening material and the vocabulary explanations. Therefore, they made the least progress in vocabulary gains. However, the strategy use pattern found in the CS group involved mainly vocabulary learning strategies, they therefore made significant progress in vocabulary gains, and this progress was significantly higher than that of the L2 group. Similarly, learners in the CFoF group also used strategies to learn vocabulary knowledge and specifically attended to the cross-linguistic information. They therefore outperformed the L2 group in vocabulary gains. However, these patterns of strategy use still cannot explain why the L2 and the CS group made significant listening progress while the CFoF group did not. This will be discussed in the below Section 5.7.4.

5.7.3 What are the differences in strategy use between more skilled and less skilled learners?

Findings from the quantitative overview of learners' strategy use indicate that, on average, more skilled learners used more different strategies than less skilled learners, concurring with what was found by Lai (2009) and other empirical evidence (Green & Oxford, 1995; Kim, 2001). Lai (2009), a study exploring the relationship between general language learning strategy use and language proficiency level among 418 EFL learners in Taiwan, found that higher proficiency level learners used more learning strategies than lower level learners and this was true for all the six categorises (memory, cognitive, compensatory, metacognitive, affective, social) of strategies they explored in the study.

In the present study, these differences in strategy use between more and less killed learners were specifically marked for the L2 and the CFoF group. This was then confirmed by a more qualitative analysis. In addition to the differences between the number of different strategies used, more skilled learners tended to use different strategies in an active manner, yet less skilled learners were quite passive when using different strategies in response to the teacher's vocabulary instruction. These different forms of strategy use between more and less skilled learners in the L2 group, somewhat explained why the higher proficiency learners benefited more from the L2 condition than the lower proficiency level did. Although learners in the L2 group all faced difficulty in understanding the teacher's L2 explanations, it is possible that more skilled learners approached the explanations in a more active manner and used more different strategies to understand the explanations. In addition, it was discovered that these more skilled learners in the L2 group, when facing difficulties to understand teacher's vocabulary explanations, tended to hold a questioning attitude towards these. Less skilled learners, however, used a restricted range of strategies in a passive way and tended to ignore and give up trying to understand the teacher's vocabulary explanations, therefore were significantly outperformed by those more skilled learners in vocabulary learning.

This, however, cannot explain why the differences of vocabulary gains were not significant between more and less skilled learners within the CFoF group. A possible explanation can be that compared with those less skilled learners in the L2 group who had difficulties in understanding the L2-only explanations and treated the explanations in a passive way, less skilled learners in the CFoF group, although also approaching the teacher's explanations in a passive manner, received both L1 translation and cross-linguistic information for the target lexical items which require less effort to understand compared with L2-only explanations. Therefore, these less skilled learners made similar vocabulary gains to their more skilled counterparts in the CFoF group.

5.7.4 What are the differences in strategy use across the three treatment conditions?

As discussed in the section 5.7.2, although the differences of patterns of strategy use across the three treatment groups explain some of the quantitative results of vocabulary gains and listening progress, these still cannot explain why the CS group made significant listening progress while the CFoF group did not. A more qualitative investigation of what specific strategies were used provides additional evidence for the above puzzle. Findings reveal that unlike learners from the CFoF group who used *selective attention on known words* focusing more on individual lexical items in the cross-linguistic vocabulary explanations, learners in the L2 and CS group used strategies such as *linguistic inferencing*, *linguistic contextualisation*, *summarisation*, *rely on examples* to gain a more global understanding of the teacher's explanations.

These differences in strategy use for understanding the teacher's vocabulary explanations are correspondent with the teaching procedure for each treatment condition during the intervention. Learners in the L2 group had to gain a more global understanding of the L2-only explanations, and used *summarisation* to summarise the L2-meaning of the word into an L1 translation. Similarly, learners from the CS group also gained a global understanding of the example sentence used for exemplifying the target lexical item and then relied on this example sentence to remember the target item. However, for learners in the CFoF

group, what they received is cross-linguistic information comparing and contrasting the L1 translation and the L2 meaning. It is not surprising that they attended locally to the individual words in that cross-linguistic information in order to make a link between the information and the target lexical item. It is possible that these difference in strategy use for a more global or local understanding of the vocabulary explanations would have made the CFoF group less advantaged in listening than the L2 and CS group.

CHAPTER 6 CONCLUSION

6.1 Significance of the findings

The present study implemented lexical Focus-on-Form through oral input within an EFL communicative language teaching context. The three treatment groups who received lexical Focus-on-Form instruction made significant gains on a short-term and long-term basis and their learning outcomes were significantly higher at post and delayed post-test than those of the Control group who focused on listening comprehension and culture studies. The latter did however make significant pre-post test vocabulary gains, which disappeared by the delayed post-test. It can be concluded that within an EFL CLT classroom in which oral input is a central component, it is worth promoting vocabulary learning through listening. However, in order to facilitate the acquisition of more vocabulary knowledge, this type of learning should be supplemented by explicit Focus-on-Form vocabulary teaching.

Regarding which type of lexical Focus-on-Form instruction can be more valuable, the findings of the immediate vocabulary recall tests for the three treatment groups suggest that learners who were taught through the teacher codeswitching and the contrastive Focus-on-Form approach acquired more vocabulary items than those taught through the L2-only approach. This indicates that there is an advantage of using L1 over L2-only for vocabulary instruction for short-term learning. In addition, although there were no long-term benefits of using teacher CS for vocabulary teaching when compared with L2-only instruction, the vocabulary taught through the CFoF approach was retained significantly better than that taught through the CS and L2-only approach two weeks after the intervention. Therefore, it can be concluded that in order to achieve the greatest vocabulary gains on a long-term basis, the CFoF approach should be promoted.

Regarding learners' listening comprehension performance before and after the experimental sessions, learners in the L2, CS and the Control group all made significant

improvement, yet the CFoF group did not. The progress in listening of the first three groups emphasises the value of vocabulary learning through listening as this approach, in addition to enhancing learners' vocabulary knowledge, also facilitates listening development. By contrast, the CFoF approach, although achieving the most impressive vocabulary gains, was the least beneficial for listening development. Likewise, the Control group made the least progress in vocabulary learning, but showed the most marked improvement in listening. On a pedagogical level, these somewhat contradictory findings suggest that a balanced approach is needed, whereby teachers are clear about what it is they are aiming to achieve from certain pedagogical approaches. Arguably learners need both opportunities to focus on listening in its own right, and to experience oral input with CFoF teacher explanations as a way to enhance vocabulary knowledge.

Looking at whether the impact of the three types of vocabulary instruction varies according to learners' general English language proficiency, results in general reveal that compared with lower-level learners, higher-level learners benefited more from the L2-only and the CFoF vocabulary explanations for short-term vocabulary learning. This indicates that these two types of vocabulary instruction have particular pedagogical value for teaching higher proficiency level learners. Indeed, the L2 meaning and the cross-linguistic information requires more linguistic knowledge and effort to understand. These extra efforts made by the higher proficiency learners then led to a better short-term vocabulary learning.

Analysing the learning by word classes and for collocations, findings first suggest that collocations and nouns tended to be better acquired than verbs and adjectives for both short-term and long-term learning. Therefore, it can be concluded that, at a pedagogical level, compared with teaching nouns and collocations, more time is needed to teach verbs and adjectives. In addition, findings also indicate that both the short-term and long-term learning of collocations and single words by the learners who received CFoF vocabulary explanations was significantly better than those from the L2 and CS group. This further confirms the pedagogical advantage of employing CFoF for vocabulary instruction.

Regarding the impact of different repetitions on vocabulary retention, findings confirm that target lexical items receiving nine repetitions were significantly better retained that those receiving seven, five or three repetitions. Hence, it can be concluded that in order to reach higher retention level, vocabulary recycling should be promoted, whereby specific teaching plans should be made in order to repeat target lexical items for at least nine times within a certain period of time.

Finally, the qualitative analysis regarding learners' strategy use in each of the three treatment conditions firstly suggests that learners from the three groups used L2 listening comprehension strategies to understand the listening input as well as employing vocabulary learning strategies to guess the meaning of the unfamiliar lexical items and to further remember these items. In addition, in general, higher proficiency level learners tended to use more different strategies than low proficiency level learners. Moreover, although certain patterns of strategy use were shared by both higher and lower proficiency level learners within each treatment condition, higher proficiency level learners tended to use these strategy patterns in a more active way, compared with lower proficiency level learners who employed the strategies in a passive manner.

Furthermore, a close look into the specific strategies used suggest that learners from the L2 and CS group employed strategies which can help them to gain a more global understanding of the L2-only and the CS vocabulary explanations. The learners from the CFoF group, however, used strategies focused more on restricted local understandings of specific individual lexical items. Therefore, it can firstly be concluded that there is pedagogical value in paying attention to learners' strategy use within the classroom context. Learners in the present study did not receive any strategy instruction, yet higher proficiency learners were able to employ different strategies spontaneously and effectively in response to the teacher's input. It is possible that teachers might be able to guide lower proficiency level learners to employ strategies more effectively, to use a wider range of strategies as well as to use these strategies in a more active way. Secondly, for teaching vocabulary through CFoF instruction, additional classroom activities should be promoted

to help learners to gain a more global attention to both the listening input and the teacher's vocabulary explanations.

6.2 Methodological contribution and limitations

A mixed methods design was adopted by the present study to avoid potential methodological limitations. Quantitative analyses were employed to investigate the impact of the three types of vocabulary instruction on vocabulary learning, supplemented by stimulated recall interviews afterwards with the purpose of exploring what learning strategies do learners use in response to the instruction for the target items offered by the teacher. While drawing on the research design and research instruments of previous studies, the present study made relevant modifications to them. First, the written vocabulary test used by Tian and Macaro (2012) was replaced with aural vocabulary tests so that the test modality was more consistent with the teaching modality. Second, previous studies did not ensure an equal testing delay for all vocabulary items. The delay ranged from one to three weeks in Hennebry et al. (2013) and from two to seven weeks in Tian and Macaro (2012). The present study was designed so that for all lexical items there was an exact two-week gap between the intervention session and their assessment in a delayed post-test. Third, in order to minimise possible pre-test sensitisation while checking that participants did not know the target items before the intervention, the vocabulary pre-test in the present study was embedded in the general vocabulary levels tests, unlike previous studies which used a simple pre-test of target items (Hennebry et al., 2013; Tian & Macaro, 2012).

Several methodological limitations are to be noted at the same time. First, the present study adopted a quasi-experimental design where participants were not randomly assigned to each treatment group or a control group. Significant differences were found between groups on their general English proficiency level before the intervention. However, as the participants were senior-secondary school EFL learners, it would have been too disruptive to their learning if they were assigned to a different learning environment and therefore

randomisation at the individual level would have affected the ecological validity of the present study. At the same time, actions were taken to minimise this limitation. When analysing the first research question using a two-way mixed ANOVA, the three baseline tests were respectively used as a covariate to see whether they have potential impacts on the test results.

Second, the format of the vocabulary pre-test was not identical to that of the post-test and the delayed post-test as the former assessed meaning recognition while the latter tested meaning recall. Meaning recognition is believed to be easier than meaning recall (Laufer & Goldstein, 2004), therefore it is possible that the participants' vocabulary knowledge was overestimated in the vocabulary pre-test. Although actions were taken to minimise this limitation (see Chapter 3, Section 3.6.4), this could still pose some limitations for the validity of the pre-test. It is worth noting however that this limitation potentially underestimates the vocabulary gains made by all groups. Finally, although modifications were made to the two listening comprehension tests adopted in the present study (for each test, only the first two sections of a standard IELTS listening comprehension test were used), these two tests were still challenging for most of the participants. The mean scores of both tests were relatively low for each group. The reliability for both the tests was also relatively low (.62 for the listening pre-test and .60 for the listening post-test).

Third, due to limited access to teaching facilities and classroom resources, the stimulated recall interviews were only conducted with participants from the three treatment groups. It is possible that if selected learners from the Control group could be interviewed, the present study would have gained a better understanding of the Control group's performance in the vocabulary tests and in the listening comprehension tests.

Regarding these methodological limitations, several recommendations can be made for future research. Initially, future research with participants at secondary school level should aim to recruit a larger sample, using intact classes with similar language proficiency levels for the classroom intervention. Secondly, a more appropriate and reliable test of listening

comprehension should be developed for use in future research with secondary school learners, with several rounds of piloting. In the present study, because of time restrictions, the listening comprehension tests were updated after the initial pilot phase but were not piloted again. Finally, further research could also interview the learners in the Control group, so that more information regarding the differences in vocabulary and listening performance between learners from the Control group and those from the three treatment groups could be gathered.

6.3 Pedagogical implications and concluding remarks

At a broader pedagogical level, results of the present study firstly indicate that there are clear pedagogical advantages of teaching vocabulary through listening. Findings show that not only the three treatment groups made significant vocabulary gains, but learners from the Control group who received culture studies also improved their vocabulary knowledge. In addition, the L2, CS and Control groups also significantly improved in their listening ability. Secondly, findings regarding vocabulary learning and listening comprehension suggest that teachers should be aware of their teaching goals and select an appropriate teaching approach to reach these different goals. The combination of different teaching strategies should be promoted in order to maximise learning.

Thirdly, the results of the present study suggest that more advanced learners tended to benefit more from the English-only vocabulary explanations and from the cross-linguistic information than their less advanced counterparts did. This suggests that more attention should be paid to learners' proficiency level when designing EFL textbooks and classroom teaching procedure. For more advanced learners (e.g., Chinese university-level EFL learners who have received at least 10 years' formal English education in school), there is pedagogical value in providing L2-only vocabulary and of course cross-linguistic vocabulary explanations in both textbook materials and through classroom teaching. Fifthly, vocabulary recycling should be given due attention either when writing new textbooks or drawing up medium and long-term teaching plans so that target lexical items

can reach more repetitions.

Finally, learners should be allowed to follow their own pace in the classroom context. When teaching new vocabulary items, the teacher should make sure that learners have enough time to develop their own strategies to understand the vocabulary explanations. Perhaps more so than for the more advanced learners, less advanced learners may need additional strategy support, either from the teacher or from their peers, which can provide them with more guidance on which strategies to use and on how they can use these strategies more effectively.

In conclusion, it is hoped that the present study will prompt researchers and teachers to devote more attention to vocabulary learning through listening.

REFERENCES

Alali, F. A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words?. *TESOL Journal, 3*(2), 153-180. http://dx.doi.org/10.1002/tesj.13

Anderson, J. R. (2015). *Cognitive psychology and its implications* (8th ed.). NY, US: Worth.

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning, 62*(2), 49-78. http://dx.doi.org/10.1111/j.1467-9922.2012.00706.x

Baddeley, A. (1997). *Human memory: Theory and practice*. UK: Psychology Press.

Bao, Y. (2011). *Codeswitching in beginning Chinese class: An exploratory investigation of teachers' bilingual practices in an Amercian university setting* (Unpublished master's thesis). University of Florida, Gainesville, US.

Best, J. W., & Kahn, J. V. (1998). *Research in education* (8th ed.). London: Allyn and Bacon.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ, 314*, 572. https://doi.org/10.1136/bmj.314.7080.572

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245-261. https://doi.org/10.1191/1362168806lr195oa

Bogaards, P., & Laufer, B. (2004). *Vocabulary in a second language*. Amsterdam: John Benjamins Publishing

Bonate, P. L. (2000). *Analysis of pretest-posttest designs*. US: Chapman & Hall/CRC.

Breeze, R., & Miller, P. (2012). Predictive validity of the IELTS Listening Test as an indicator of student coping ability in Spain. In L. Taylor, & C. J. Weir (Eds.), *IELTS Collected Papers 2* (pp. 487-518). Cambridge, UK: Cambridge University Press.

Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing, 19*(4), 369-394. https://doi.org/10.1191/0265532202lt236oa

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading a Foreign Language, 20*(2), 136-163.

Bryman, A. (1988). *Quantity and quality in social research*. London: Unwin Hyman.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service?. *Bilingualism: Language and Cognition, 13*(3), 359-371. https://doi.org/10.1017/S1366728909990344

Carr, T. H., & Curran, T. (1994). Cognitive factors in learning about structured sequences: Applications to syntax. *Studies in Second Language Acquisition, 16*(2), 205-230. https://doi.org/10.1017/S0272263100012882

Carroll, S. E. (2006). Salience, awareness and SLA. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006)* (pp. 17–24). Somerville, MA: Cascadilla Proceedings Project.

Chen, H. C., & Leung, Y. S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 316-325. http://psycnet.apa.org/doi/10.1037/0278-7393.15.2.316

Cohen, A. D. (1996). *Second language learning and use strategies: Clarifying the issues. CARLA Working Paper Series #3*. Minneapolis, MN: The University of Minnesota.

Cohen, A. D. (1998). *Strategies in learning and using a second language*. London: Longman.

Cook, G. (2010). *Translation in language teaching: An argument for reassessment*. Oxford: Oxford University Press.

Cook, L. K., & Mayer, R. E. (1983). Reading strategies training for meaningful learning from prose. In M. Pressley & J. Levin (Eds.), *Cognitive strategy research* (pp. 87-131). New York: Springer.

Cook, V. (2001). Using the first language in the classroom. *Canadian Modern Language Review, 57*(3), 402-423. https://doi.org/10.3138/cmlr.57.3.402

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268. http://psycnet.apa.org/doi/10.1037/0096-3445.104.3.268

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. London: SAGE Publications.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/10.1007/BF02310555

Cupchik, G. (2001). Constructivist realism: An ontology that encompasses positivist and constructivist approaches to the social sciences. *Forum: Qualitative Social Research/Sozialforschung 2*(1, Art. 7), 1–12.

Daly, W. M. (2001). The development of an alternative method in the assessment of critical thinking as an outcome of nursing education. *Journal of Advanced Nursing, 36*(1), 120–130. https://doi.org/10.1046/j.1365-2648.2001.01949.x

DeKeyser, R. (2003). Implicit and explicit learning. In C.J. Doughty & M.H. Long (Eds.), *Handbook of second language acquisition* (pp. 313-348). London: Blackwell.

Dimitrov, D. M., & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159-165.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning, 24*(1), 37–53. http://dx.doi.org/10.1111/j.1467-1770.1974.tb00234.x

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163-188. https://doi.org/10.1177/0267658309349431

Ellis, N. C. (1994). Vocabulary acquisition: the implicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211–282). London: Academic Press

Ellis, R. (1985). The L1= L2 hypothesis: A reconsideration. *System, 13*(1), 9-24. https://doi.org/10.1016/0346-251X(85)90003-X

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics, 16*(4), 409–441. https://doi.org/10.1093/applin/16.4.409

Faraway, J. J. (2015). *Linear models with R* (2nd ed.). Boca Raton, FL: CRC Press.

Feilzer, Y. M. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research, 4*(1), 6-16. https://doi.org/10.1177/1558689809349691

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE Publications.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*(2), 273−293. http://dx.doi.org/10.2307/40264523

Gass, S. (1997). *Input, interaction, and the second language learner*. Mahway, NJ: Erlbaum.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gathercole, S. E., & Baddeley, A. (1993). *Working memory and language*. NY: Psychology Press.

Godfroid, A. (2010). *Cognitive processes in second language acquisition: The role of noticing, attention and awareness in processing words in written L2 input* (Unpublished doctoral dissertation). Vrie Universiteit Brussel, Brussels, Belgium.

Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the Noticing Hypothesis in the context of vocabulary acquisition. In M. Putz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 169-197). Amsterdam: John Benjamins.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55-75. https://doi.org/10.1016/S0346-251X(99)00060-3

Graham, S., & Santos, D. (2012). Selective listening in L2 learners of French. *Language Awareness, 22*(1), 56-75. https://doi.org/10.1080/09658416.2011.652634

Graham, S., Santos, D., & Vanderplank, R. (2008). Listening comprehension and strategy use: a longitudinal exploration. *System, 36*(1). pp. 52-68. https://doi.org/10.1016/j.system.2007.11.001

Green, J. M., & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly, 29*(2), 261-297. http://dx.doi.org/10.2307/3587625

Gu, Y. (2014). To code or not to code: Dilemmas in analysing think-aloud protocols in learning strategies research. *System, 43*, 74-81. https://doi.org/10.1016/j.system.2013.12.011

Guba, E. G. (1990). The alternative paradigm dialog. In G. E. Guba (Ed.), *The paradigm dialog* (pp. 17-27). CA: SAGE Publications.

Guo, T. (2007). *A case study of teachers' codeswitching behaviours in mainland China's university EFL classrooms and students' reactions to the code-switching* (Unpublished DPhil Thesis). University of Oxford, Oxford, UK.

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple choice formats: Issues with guessing and sampling rates. *ITL-International Journal of Applied Linguistics, 166*(2), 278-306. http://dx.doi.org/10.1075/itl.166.2.04gyl

Harley, H. (2006). *English words: A linguistic introduction*. Oxford: Blackwell.

Hazenberg, S., & Hulstun, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics, 17*(2), 145-163. https://doi.org/10.1093/applin/17.2.145

Hennebry, M., Rogers, V., Macaro, E., & Murphy, V. (2013). Direct teaching of vocabulary after listening: Is it worth the effort and what method is best? *Language Learning Journal, 45*(3), 282-300. https://doi.org/10.1080/09571736.2013.849751

Hill, M. M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching, 41*(2): 87-106. https://doi.org/10.1515/iral.2003.007

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review, 61*(3), 355-382. https://doi.org/10.3138/cmlr.61.3.355

Horst, M., Cobb, T., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.

Hsueh-Chao, M. H., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.

Hu, G. (2002). Recent important developments in secondary English-language teaching in the People's Republic of China. *Language Culture and Curriculum, 15*(1), 30-49. http://dx.doi.org/10.1080/07908310208666631

Hu, G. (2005). English language education in China: Policies, progress, and problems. *Language Policy, 4*(1), 5-24. https://doi.org/10.1007/s10993-004-6561-7

Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley and Sons.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hussein, A. (2009). The use of triangulation in social sciences research: Can qualitative and quantitative methods be combined?. *Journal of Comparative Social Work, 4*(1), Retrieved November 02, 2017, from http://journal.uia.no/index.php/JCSW/article/view/212/147

IELTS. (2015). *Test taker performance 2015*. Retrieved November 02, 2017, from https://www.ielts.org/teaching-and-research/test-taker-performance

Ijaz, I. H. (1986). Linguistic and cognitive determinants of lexical acquisition in a second language. *Language Learning, 36*(4), 401–451. http://dx.doi.org/10.1111/j.1467-1770.1986.tb01034.x

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47-77. https://doi.org/10.1093/applin/21.1.47

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24*(4), 617-637. https://doi.org/10.1017/S0272263102004047

Jiang, N. (2004a). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal, 88*(3), 416-432. http://dx.doi.org/10.1111/j.0026-7902.2004.00238.x

Jiang, N. (2004b). Semantic transfer and development in adult L2 vocabulary acquisition. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 102-126). Amsterdam: John Benjamins Publishing.

Jin, L., & Cortazzi, M. (2002). English language teaching in China: A bridge to the future. *Asia Pacific Journal of Education, 22*(2), 53-64. http://dx.doi.org/10.1080/0218879020220206

Jin, Y., Wu, Z., Alderson, C., & Song, W. (2017). Developing the China Standards of English: challenges at macropolitical and micropolitical levels. *Language Testing in Asia, 7*(1), 1-19. https://doi.org/10.1186/s40468-017-0032-5

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133. https://doi.org/10.1177/1558689806298224

Kim, K. H. (2001). Language learning strategies, learning styles, and beliefs about language learning of Korean university students. *Journal of Pan-Pacific Association of Applied Linguistics, 5*(1), 31–46.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. New York: Prentice Hall.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory*, (Vol. 2, pp. 389-395). London: John Wiley & Sons.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149-174. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373-381. https://doi.org/10.1017/S136672891000009X

Laerd Statistics (2015). Two-way mixed ANOVA using SPSS Statistics. *Statistical tutorials and software guides*. Retrieved 02 November 2017, from https://statistics.laerd.com/

Lai, Y. C. (2009). Language learning strategy use and English proficiency of university freshmen in Taiwan. *TESOL Quarterly, 43*(2), 255-280. http://dx.doi.org/10.1002/j.1545-7249.2009.tb00167.x

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Laufer, B, & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26. https://doi.org/10.1093/applin/22.1.1

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*(4), 567-587. https://doi.org/10.3138/cmlr.59.4.567

Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla Yearbook, 5*(1), 223-250. http://dx.doi.org/10.1075/eurosla.5.11lau

Laufer, B. (2009). Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching, 42*(3), 341-354. https://doi.org/10.1017/S0261444809005771

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694-716. https://doi.org/10.1093/applin/amn018

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language learning, 54*(3), 399-436. http://dx.doi.org/10.1111/j.0023-8333.2004.00260.x

Lawson, M. J., & Hogben, D. (1998). Learning and recall of foreign-language vocabulary: Effects of a keyword strategy for immediate and delayed recall. *Learning and Instruction, 8*(2), 179-194. https://doi.org/10.1016/S0959-4752(97)00016-9

Lee, J. F. (2009). Perceptions of ELT among English language teachers in China. *Education Journal, 37*(1-2), 137-154.

Lee, J. H., & Macaro, E. (2013). Investigating age in the use of L1 or English-only instruction: Vocabulary acquisition by Korean EFL learners. *The Modern Language Journal, 97*(4), 887-901. http://dx.doi.org/10.1111/j.1540-4781.2013.12044.x

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Liu, D., Gong, Y., Zheng, W., Brooks, D., Allingham, N., Sjoquist, R., & Miller, S. (2007). *New senior English for China, Student book 4*. Beijing, China: People's Education Press.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126-141. https://doi.org/10.1093/applin/4.2.126

Louw, B. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. In C. Heffer, H. Sauntson, & G. Fox (Eds.), *Words in context: A tribute to John Sinclair on his retirement* (pp. 48-94). Birmingham, UK: University of Birmingham.

Lyle, J. (1999). Coaches' decision making. In N. Cross, & J. Lyle (Eds.), *The coaching process: Principles and practice for sport* (pp. 210–232). Oxford: Butterworth Heinemann.

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal, 29*(6), 861-878. https://doi.org/10.1080/0141192032000137349

Macaro, E. (2005). Codeswitching in the L2 classroom: A communication and learning strategy. In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges, and contributions to the profession* (pp. 63-84). Boston: Springer.

Macaro, E. (2014). Students' strategies in response to teachers' second language explanations of lexical items, *Language Learning Journal, 45*(3), 352-367. https://doi.org/10.1080/09571736.2014.905971

McKeown, M., Beck, I., Omanson, R., & Pople, M. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*(5), 522-535. https://doi.org/10.2307/747940

McLean, S., Ishii, T., Stoeckel, T., Bennett, P., & Matsumoto, Y. (2016). An edited version of the first eight 1,000-word frequency bands of the Japanese-English version of the vocabulary size test. *The Language Teacher, 40*(4), 3-7.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research, 19*(6), 741-760. https://doi.org/10.1177/1362168814567889

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching, 13*(3-4). 221-246. https://doi.org/10.1017/S0261444800008879

Meara, P. & Milton, J. (2003). *X_Lex, The Swansea Levels Test*. Newbury: Express.

Mecartty, F. H. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning 11*(2), 323–348.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Beverly Hills, CA: SAGE Publications.

Min, H. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning, 58*(1), 73-115. http://dx.doi.org/10.1111/j.1467-9922.2007.00435.x

Ministry of Education, P. R. China. (2000a) *Quanrizhi gaoji zhongxue yingyu jiaoxue dagang (shiyan xiudingban)* [English syllabus for full-time senior secondary schools]. Beijing: People's Education Press.

Ministry of Education, P. R. China. (2000b) *Jiunian yiwu jiaoyu quanrizhi chuji zhongxue yingyu jiaoxue dagang (shiyong xiudingban)* [English syllabus for nine-year compulsory education full-time junior secondary schools]. Beijing: People's Education Press.

Ministry of Education, P. R. China. (2001). *Jiaoyubu guanyu jiji tuijin xiaoxue kaishe yingyu kecheng de zhidao yijian* [Guidelines for promoting primary English language teaching]. Beijing: MOE.

Ministry of Education, P. R. China. (2003). *Putong gaozhong yingyu kecheng biaozhun* [English language curriculum guidelines for senior secondary school]. Beijing: People's Education Press.

Ministry of Education, P. R. China. (2012). *English curriculum of basic education (2011 version)*. Beijing: Beijing Normal University Press.

Mondria, J. A. (2003). The effects of inferring, verifying and memorizing on the retention of L2 word meanings: An experimental comparison of the 'meaning inferred method' and 'meaning given method'. *Studies in Second Language Acquisition, 25*(4), 473-499. https://doi.org/10.1017/S0272263103000202

Myers-Scotton, C. (1993). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon Press.

Myers-Scotton, C. (2001). The matrix language frame model: Developments and responses. In R. Jacobson (Ed.), *Codeswitching worldwide II* (pp. 23-58). Berlin: Mouton de Gruyter.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. NY, US: Newbury House.

Nation, I. S. P. (1996). The four strands in a language course. *TESOL in Context, 6*(2), 7-12.

Nation, I. S. P. (2001). *Learning vocabulary in another language* (1st ed.). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening?. *Canadian Modern Language Review, 63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, I. S. P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language, 12*(2), 355–380.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, US: Heinle.

Nunan, D. (2003). The impact of English as a global language on educational policies and practices in the Asia-Pacific Region. *TESOL Quarterly, 37*(4), 589-613. http://dx.doi.org/10.2307/3588214

Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada journal, 11*(1), 9-29. https://doi.org/10.18806/tesl.v11i1.623

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*(1), 97-130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381-402. https://doi.org/10.1177/1362168815618428

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: do things fall apart?. *Reading in a Foreign Language, 22*(1), 31-55.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research, 18*(1), 75-94. https://doi.org/10.1177/1362168813505384

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1-28.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: toward a typology of code-switching. *Linguistics, 18*(7/8), 581-618. https://doi.org/10.1515/ling.1980.18.7-8.581

Posner, M. I. (1992). Attention as a cognitive and neural system. *Current Directions in Psychological Science, 1*(1), 11-14. https://doi.org/10.1111/1467-8721.ep10767759

Potter, M. C., So, K. F., von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 23-38. https://doi.org/10.1016/S0022-5371(84)90489-4

Punch, K. F. (1998). *Introduction to social research: Quantitative and qualitative approaches*. London: SAGE Publications.

Qian, X., Tian, G., & Wang, Q. (2009). Codeswitching in the primary EFL classroom in China: Two case studies. *System, 37*(4), 719-730. https://doi.org/10.1016/j.system.2009.09.015

Rao, Z. (2013). Teaching English as a foreign language in China: looking back and forward: Reconciling modern methodologies with traditional ways of language teaching. *English Today, 29*(3), 34-39. https://doi.org/10.1017/S0266078413000291

Richards, J. C. & Rodger, T. S. (2001). *Approaches and methods in language teaching*. New York: Cambridge University Press.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. W. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). London: Academic Press.

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmidt, R. W. (2012). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chin, S. Bhatt & I. Walker (Eds.), *Perspectives on individual characteristics and foreign language education* (pp. 27–50). Boston, MA: Mouton de Gruyter.

Schmidt, R. W., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.

Schmitt, N. (1997). Vocabulary Learning Strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. https://doi.org/10.1177/1362168808089921

Schouten-van Parreren, C. (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts?. *AILA Review, 6*(1), 75-85.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology, 45*(1), 21-50. https://doi.org/10.1080/14640749208401314

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin Company.

Silver-Pacuilla, H., Brown, S., Overton, C., & Stewart, A. (2011). *Assistive technology research matters*. Washington, DC: American Institutes for Research.

Silver, R. E., Hu, G., & Iino, M. (2002). *English language education in China, Japan, and Singapore*. Singapore: National Institution of Education, Nanyang Technology University.

Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 3-22. https://doi.org/10.1016/S0022-5371(84)90479-1

Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort? *English Language Teaching Journal, 64*(3), 253-260. https://doi.org/10.1093/elt/ccp059

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577-607. https://doi.org/10.1017/s0272263109990039

Steiner, P. M., Wroblewski, A., & Cook, T. D. (2009). Randomized experiments and quasi-experimental designs in educational research. In K. E. Ryan & J. B. Cousins (Eds.), *The SAGE international handbook on educational evaluation* (pp. 75-95). London: SAGE Publications.

Stewart, J., & White, D. A. (2011). Estimating guessing effects on the vocabulary levels test for differing degrees of word knowledge. *TESOL Quarterly, 45*(2), 370-380. http://dx.doi.org/10.5054/tq.2011.254523

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research, 69*(3), 261-285. https://doi.org/10.3102/00346543069003261

Tang, E., & Nesi, H. (2003). Teaching vocabulary in two Chinese classrooms: Schoolchildren's exposure to English words in Hong Kong and Guangzhou. *Language Teaching Research, 7*(1), 65-97. https://doi.org/10.1191/1362168803lr113oa

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*(1), 90-105. https://doi.org/10.1093/applin/11.1.90

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. CA: SAGE Publications.

Tian, L. (2011). *Teacher codeswitching in a communicative EFL context: Measuring the effects on vocabulary learning*. Oxford: VDM.

Tian, L., & Macaro, E. (2012). Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university students: A Lexical Focus-on-Form study. *Language Teaching Research, 16*(3), 367-391. https://doi.org/10.1177%2F1362168812436909

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*(2), 183–203. https://doi.org/10.1017/S0272263100012870

Tseng, W. T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning, 58*(2), 357-400. http://dx.doi.org/10.1111/j.1467-9922.2008.00444.x

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language learning, 53*(3), 463-496. http://dx.doi.org/10.1111/1467-9922.00232

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Oxford, UK: Routledge.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition?. *Applied Linguistics, 24*(1), 56-89. https://doi.org/10.1093/applin/24.1.56

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219-258. http://dx.doi.org/10.1111/j.1467-9922.2010.00593.x

Wang, W., & Lam, A. S. (2009). The English language curriculum for senior secondary school in China: Its evolution from 1949. *RELC Journal, 40*(1), 65-82. https://doi.org/10.1177%2F0033688208101447

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics, 28*(1), 46-65. https://doi.org/10.1093/applin/aml048

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53*(1), 13-40.

Widdowson, H. G. (2003). *Defining issues in English language teaching*. Oxford: Oxford University Press.

Wilcox, R. R. (1998). The goals and strategies of robust methods. *British Journal of Mathematical and Statistical Psychology, 51*(1), 1-39. http://dx.doi.org/10.1111/j.2044-8317.1998.tb00659.x

Willis, J. W. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. Thousand Oaks, CA: SAGE Publications. http://dx.doi.org/10.4135/9781452230108

Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly, 52*(2), 203-226. http://dx.doi.org/10.1002/rrq.163

Wu, Y. (2001). English language teaching in China: Trends and challenges. *TESOL Quarterly, 35*(1), 191-194. https://doi.org/10.2307/3587867

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics, 27*(1), 103-129. https://doi.org/10.1093/applin/ami045

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review, 57*(4), 541-572. https://doi.org/10.3138/cmlr.57.4.541

Van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609-624. https://doi.org/10.1016/j.system.2013.07.012

Van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension. *Applied Linguistics, 34*(4), 457-479. https://doi.org/10.1093/applin/ams074

Zhang, X. (2013). The I don't know option in the Vocabulary Size Test. *TESOL Quarterly, 47*(4), 790-811. http://dx.doi.org/10.1002/tesq.98

Zhao, T., & Macaro, E. (2014). What works better for the learning of concrete and abstract words: teachers' L1 use or L2-only explanations?. *International Journal of Applied Linguistics, 26*(1), 75-98. http://dx.doi.org/10.1111/ijal.12080

APPENDICES

Appendix A Listening passages (adapted from Liu et al., 2007), listening comprehension tasks and target lexical items

A student of African wildlife

For forty years, Jane Goodall, a researcher of African wildlife, has been **outspoken** about making the rest of the world understand the life of animals. One **prominent** thing she discovered was that chimps hunt and eat meat, which **altered** the way people think about chimps. Until then everyone had thought chimps ate only fruit and nuts. She actually observed chimps as a group hunting a monkey and then eating it. She also discovered how chimps communicate with each other, and her study of their body language helped her **figure out** their social system.

Jane has **argued** that wild animals, like chimps, should be left in the wild and not used for **entertainment** or advertisements. She has helped to set up special **shelters** where they can live safely. She is **leading a busy life** but she says: "Once I stop, it all comes **crowding in** and I remember the chimps in laboratories. It's terrible. It affects me when I watch the wild chimps. I say to myself, 'Aren't they lucky?" And then I think about small chimps in cages though they have done nothing wrong. Once you have seen that you can never forget ...". Jane has achieved everything she wanted to do: working with animals in their own environment, gaining a doctor's degree and showing that women can live in the forest as men can. She **inspires** those who want to cheer the achievements of women.

Listening comprehension

1. According to the speaker, what do people usually think that chimps eat?

A. Fruits and vegetables
B. Meat
C. Nuts and fruits

2. Which of the followings was NOT studied by Jane Goodall?

A. What is chimps' social system?
B. How do chimps communicate with each other?
C. How to use chimps for advertisements?

3. How did Jane feel about those chimps living in the wild?

A. They are terrible.
B. They are lucky.
C. They make her happy.

A Master of Nonverbal Humour

Charlie Chaplin grew more and more popular when he began making films, as his charming character, the little tramp, became known throughout the world. The tramp, a poor, homeless man with a **moustache**, wore large trousers, **worn-out** shoes and a small round black hat. He walked around **stiffly** carrying a walking stick. This character was a social failure but was loved for his optimism and determination to **overcome** all difficulties. He was the underdog who was kind even when others were unkind to him.

How did the little tramp make a sad situation entertaining? Here is an example from one of his most famous films, The Gold Rush. It is toward the end of the nineteenth century and gold has just been discovered in Alaska. Like so many others, the little tramp has rushed there in search of gold, but without success. Once he and another man are hiding in a small **hut** during a snowstorm with nothing to eat. They are so hungry that the little tramp tries boiling one of his **leather** shoes for dinner. The little tramp **cuts off** the leather top of the shoe and shares the shoe with the other fellow. He tries cutting and **chewing** the bottom of the shoe as if it were the finest steak. Then he **picks out** the lace of the shoe and eats it as if it were spaghetti. He eats each mouthful with great enjoyment. The acting is so **convincing** that it makes you believe that it is one of the best meals he has ever tasted!

Listening comprehension

1. Why was the character, the little tramp, loved by so many people?

A. Because the character is played by Charlie Chaplin.
B. Because of his optimism and determination.
C. Because of his funny dressing style.

2. Why did the little tramp rush to Alaska at the end of the 19th century?

A. To search for gold
B. To change the sad situation he was suffering
C. To search for success

3. What did the little tramp and his friend eat for dinner during a snow storm?

A. Spaghetti
B. Steak
C. Leather shoes

Communication: No problem?

Yesterday, another student and I, **representing** our university's student **association**, went to the Capital International Airport to meet this year's international students. They were coming to study at Beijing University. We would take them first to their **dormitories** and then to the student canteen. After waiting for half an hour**,** I saw several young people enter the waiting area looking around **curiously**. I stood for a while and then went to greet them.

Tony Garcia from Colombia was the first person to arrive, closely followed by Julia Smith from Britain. After I met them and then introduced them to each other, I was very surprised. Tony **approached** Julia, touched her shoulder and kissed her on the cheek! She stepped back appearing surprised and put up her hands, as if **in defence**. I guessed that there was probably a major misunderstanding. **On the contrary**, when Darlene Coulon from France came **dashing** through the door, she recognised Tony Garcia's smiling face. They shook hands and then kissed each other twice on each cheek, since that is the French custom when adults meet people they know.

As I get to know more international friends, I learn more about this cultural "body language". These actions are not good or bad, but are simply ways in which cultures have developed. I have seen, however, that cultural customs for body language are very general - not all members of a culture behave in the same way. **In general**, though, studying international customs can certainly help avoid difficulties in today's world of cultural **crossroads**!

Listening comprehension

1. Why did the speaker go to the Capital International Airport?

A. To meet a group of international friends
B. To pick up some visitors from a university abroad
C. To meet a group of international students

2. How long did the speaker wait for?

A. Half an hour
B. An hour
C. Fifteen minutes

3. How did the speaker feel about the body language used in different cultures?

A. They are difficulty to understand.
B. They are easy to learn.
C. They are very general, not good or bad.

Showing our feelings

Body language is one of the most powerful means of communication. People around the world show all kinds of **inner** feelings, wishes and attitudes that they might never speak aloud. It is possible to "read" others around us, even if they do not intend for us to catch their unspoken communication. Of course, body language can be misread, but many gestures and actions are universal.

The most universal facial expression is, of course, the smile – its **function** is to show happiness and put people **at ease**. It does not always mean that we are truly happy, however. There are unhappy smiles, such as when someone "**loses face**" and smiles to hide it. From the time we are babies, we show unhappiness or anger by **frowning**. In most places around the world, frowning and **turning one's back to** someone shows anger. Making a fist and shaking it almost always means that someone is angry and **threaten**ing another person.

Being respectful to people is **subjective**, but in general it is probably not a good idea to give a hug to a boss or teacher. In almost every culture, it is not usually good to stand too close to someone of a higher **rank**. Standing at a little distance with open hands will show that I am willing to listen.

With so many cultural differences between people, it is great to have some **similarities** in body language. We can often be wrong about each other, so it is an amazing thing that we understand each other as well as we do!

Listening comprehension

1. Which of the following is NOT TRUE about body language?

A. It is a way that people communicate with each other.
B. It can sometimes cause misunderstanding.
C. It is always used together with spoken language.

2. According to the speaker, which of the followings is NOT appropriate?

A. Give a hug to your teacher
B. Stand at a little distance to your boss
C. Give a smile to someone you are familiar with

3. What did the speaker think about body language?

A. People may easily make mistakes about it.
B. It is easy to learn.
C. It is different because of different cultures.

Theme parks

Which **theme** park would you like to visit? There are **various** kinds of theme parks, with a different park for almost everything: food, culture, science, cartoons, movies or history. Some parks are famous for having the biggest or longest roller coasters, others for showing the famous sights and sounds of a culture. Whichever and whatever you like, there is a theme park for you!

The theme park you are probably most familiar with is Disneyland. It can be found in several parts of the world. As you **wander** around the **fantasy** amusement park, you may see Snow White or Mickey Mouse in a parade or on the street. Of course Disneyland also has many exciting rides, from giant **swing**ing ships to terrifying free-fall drops. With all these attractions, **no wonder tourism** is increasing. If you want to have fun and more than fun, come to Disneyland!

If you want to experience the **ancient** days and **great deeds of** English princes and queens, then England's Camelot Park is the place for you. Every area of the park **is modelled after** life in the days of King Arthur and the Knights of the Round Table. In one place, you can watch magic shows with Merlin the Wizard. If you want to see fighting with swords, then the jousting area is a good place to visit. Do you like animals? Then visit the farm area, and learn how people in ancient England ran their farms. To enter a world of fantasy about ancient England, come to Camelot Park!

Listening comprehension

1. According to the speaker, which of the following is a theme park NOT likely to have?

A. A restaurant
B. A cartoon factory
C. A shopping mall

2. Which of following is true about Disneyland?

A. It is famous for having the longest roller coasters.
B. It has terrifying free-fall drops.
C. It is no longer a great place for tourism.

3. What people can do in Camelot Park?

A. Experience the life of English princes and queens.
B. Fight with swords in the jousting area
C. Have the chance to run their own farms

The noodle harvest

April Fool's Day, or April 1st, is known in many countries as a day for **playing jokes on** others. One of the most famous jokes in England took place on a British TV show, Panorama, in 1957. This show **explored** problems and progress all over the world, so nobody was surprised when it began with a report on the excellent noodle harvest in south Switzerland. The programme mentioned two reasons for a high **output**: an unusually warm winter and the disappearance of the **insect** that attacked the noodle **crop** every year. The reporter showed many noodle trees with the farmers **pulling** noodles **off** them and putting them into baskets. The show also explained that it was the results of many years' patient research with the trees to produce noodles of exactly the same shape. But even so they explained, the life of noodle farmers was not easy. "The last two weeks of March are an **anxious** time for noodle farmers. There is always a chance of very cold weather **spoiling** their crop."

Many people in England believed the story and rang BBC to find out how to grow their own noodle trees. They were told to "place a piece of noodle in a tin of tomato **sauce** and hope for the best." This may seem **weird**, but in the 1950s not many British people travelled abroad and very few of them ate noodles. The show itself was very convincing for its careful research and convincing information, so they were shocked to find the next day that they had all believed an April Fool's joke.

Listening comprehension

1. What does the passage mainly tell us?

A. A story of a noodle harvest.
B. The reasons why the farmers got a noodle harvest
C. One of the best April Fool's jokes ever

2. Which of the following is a reason for the noodle harvest?

A. An unusually warm winter
B. Patient research with the noodle trees
C. Abundant rainfall

3. How did people react the next day of the show?

A. They rang BBC to find out more information.
B. They started to grow their own tree.
C. They were very surprised.

Appendix B Listening comprehension pre-test

Listening comprehension test **Name_____ Class_____**

Section 1 Questions 1-5

Example question	Answer
Destination?	Harbour City

Complete the notes below.

Write no more than **two words and/or a number** for each answer.

Questions: transport from Bayswater
1 Express train leaves at (1) …….
2 Nearest station is (2) …….
3 Number 706 bus goes to (3) …….
4 Number (4) ……. bus goes to station
5 Earlier bus leaves at (5) …….

Questions 6–10

Complete the table below.

Write no more than **one word and/or a number** for each answer.

Transport	Cash fare	Card fare
Bus	(6) $ ……	$1.50
Train (peak)	$10	$10
Train (off-peak) – before 5pm or after (7) …… pm	$10	(8) $ ……
(9) …… ferry	$4.50	$3.55
Tourist ferry ((10) ……)	$35	–
Tourist ferry (whole day)	$65	–

1_____ 2_____ 3_____ 4_____ 5_____

6_____ 7_____ 8_____ 9_____ 10_____

Section 2 Questions 11–14

Which counsellor should you see?

Write the correct letter, A, B or C, next to questions 11–14.

A	Louise Bagshaw
B	Tony Denby
C	Naomi Flynn

Questions
11) if it is your first time seeing a counsellor
12) if you are unable to see a counsellor during normal office hours
13) if you do not have an appointment
14) if your concerns are related to anxiety

Questions 15-20

Complete the table below.

Write no more than **two words** for each answer.

Workshop	**Contact**	**Target group**
Adjusting	what you need to succeed academically	(15) …… students
Getting Organised	use time effectively, find (16) …… between study and leisure	all students
Communicating	talking with staff, communicating across cultures	all students, especially (17) ……
Anxiety	(18) ……, breathing techniques, meditation, etc.	students about to sit exams
(19) ……	staying on track for long periods	(20) …… students only

11_____ 12_____ 13_____ 14_____ 15_____ 16_____

17_____ 18_____ 19_____ 20_____

Appendix C Listening comprehension post-test

Section 1:
You will hear a man telephoning a sports club to ask about membership and facilities.
First you have some time to look at **Questions 1 to 4**.

Questions 1-4
Complete the notes below.

Write **NO MORE THAN TWO WORDS** for each answer.

Notes on sports club

Example	*Answer*
Name of club:	**Flagstone**
Facilities available:	Golf
	1 .....
	2 .....
Classes available:	• kick-boxing
	• **3** .....
Additional facility:	**4** ..... (restaurant opening soon)

Questions 5-8

Complete the table below.

Write **NO MORE THAN TWO NUMBERS** for each answer.

MEMBERSHIP SCHEMES					
Type	**Use of facilities**	**Cost of classes**	**Times**	**Joining fee**	**Annual subscription fee**
GOLD	**All**	**Free**	**Any time**	**£250**	**5** £.....
SILVER	**All**	**6** £.....	from **7** ..... to	**£225**	**£300**
BRONZE	**Restricted**	**£3**	**from 10.30 to 3.30 weekdays only**	**£50**	**8** £.....

Questions 9 and 10

Complete the sentences below.

Write **ONE WORD ONLY** for each answer.

9 To join the centre, you need to book an instructor's ..... .

10 To book a trial session, speak to David ..... (0458 95311).

Section 2:

You will hear a local radio broadcast about the Rivenden City Theatre.

First you have some time to look at **Questions 11 to 16**.

Now Listen carefully and answer **Questions 11 to 16:**

What change has been made to each part of the theatre?

Choose **SIX** answers from the box and write the correct letter, **A-G**, next to questions **11-16**.

RIVENDEN CITY THEATRE

A	doubled in number
B	given separate entrance
C	reduced in number
D	increased in size
E	replaced
F	strengthened
G	temporarily closed

Part of the theatre

11 box office ..... .

12 shop ..... .

13 ordinary seats ..... .

14 seats for wheelchair users ..... .

15 lifts ..... .

16 dressing rooms ..... .

Questions 17-20

Complete the form below.

Write **NO MORE THAN TWO WORDS AND/OR A NUMBER** for each answer.

Royal Hunt of the Sun	October 13th to **17**	**18** ..... pm	for **19** ..... and	**20** £.....

Appendix D General vocabulary size test and vocabulary pre-test

词汇测试 第一部分　　姓名_____　　班级_____

1.	9.	17.	25.	33.
A. 钱	A. 穿过	A. 银行	A. 直言不讳的	A. 涌入
B. 食物	B. 推	B. 海洋动物	B. 开朗的	B. 渗透
C. 时间	C. 吃得太快	C. 学校	C. 有信心的	C. 流动
D. 朋友	D. 等待	D. 房子	D. 聪明的	D. 拥挤
2.	10.	18.	26.	34.
A. 石头	A. 真实的	A. 画画	A. 悲惨的	A. 接受
B. 板凳	B. 老的	B. 说话	B. 奢侈的	B. 借出
C. 地毯	C. 圆形的	C. 成长	C. 显著的	C. 与…相处的好
D. 嫩芽	D. 其他的	D. 痛哭	D. 和谐的	D. 鼓舞
3.	11.	19.	27.	35.
A. 贫穷	A. 任何的	A. 睡衣	A. 改变	A. 钱包
B. 感到开心	B. 没有	B. 钟	B. 传递	B. 胡须
C. 非常有兴趣	C. 好的	C. 花	C. 投票	C. 力量
D. 高的	D. 老的	D. 面包	D. 询问	D. 眼睛
4.	12.	20.	28.	36.
A. 游泳	A. 很长时间	A. 打开	A. 预料到	A. 直白的
B. 学习	B. 非常快	B. 记得	B. 解释	B. 暴露的
C. 扔球	C. 远	C. 处理	C. 弄清楚	C. 完美的
D. 开车	D. 去你家	D. 相信	D. 忽视	D. 破旧的
5.	13.	21.	29.	37.
A. 浮动	A. 食物	A. 海洋	A. 打扫	A. 僵直的
B. 跳跃	B. 故事	B. 营地	B. 粘贴	B. 喜欢的
C. 公园	C. 一群人	C. 医院	C. 与…相处得好	C. 特别的
D. 跑	D. 游戏	D. 旅馆	D. 争论	D. 正常的
6.	14.	22.	30.	38.
A. 父母	A. 导致	A. 湖泊	A. 书本	A. 克服
B. 钱包	B. 修理	B. 小孩	B. 娱乐活动	B. 离开
C. 笔	C. 解释	C. 领导	C. 电影	C. 靠近
D. 鞋子	D. 理解	D. 安静的地方	D. 歌剧	D. 穿过
7.	15.	23.	31.	39.
A. 会议	A. 什么都没有	A. 过去	A. 电影院	A. 土丘
B. 旅行	B. 足够	B. 震惊的事件	B. 商场	B. 简陋小屋
C. 测试	C. 一些	C. 晚上	C. 庇护所	C. 山
D. 计划	D. 许多	D. 夏天	D. 饭店	D. 喷泉
8.	16.	24.	32.	40.
A. 非常坏的东西	A. 何时	A. 友好的	A. 快乐地成长	A. 丝绸
B. 什么也没有	B. 因为什么	B. 非常大的	B. 过着忙碌的生活	B. 布
C. 非常好的东西	C. 何处	C. 非常快的	C. 异想天开	C. 铁
D. 某物	D. 何种方式	D. 圆的	D. 引领着潮流	D. 皮革

词汇测试 第二部分

1.
A. 维持
B. 放大
C. 改善
D. 获取

2.
A. 问题
B. 一段时间
C. 做的事情
D. 书本

3.
A. 脚后跟
B. 等级
C. 价钱
D. 标准

4.
A. 答案
B. 休息的地方
C. 下一步骤
D. 基础

5.
A. 强大的
B. 有名的
C. 有钱的
D. 生气的

6.
A. 抽屉
B. 车库
C. 冰箱
D. 笼子

7.
A. 酒吧
B. 银行
C. 购物中心
D. 游泳池

8.
A. 草稿
B. 空白处
C. 圆圈
D. 大洞穴

9.
A. 间谍
B. 傻瓜
C. 作家
D. 专家

10.
A. 商人
B. 学生
C. 木匠
D. 士兵

11.
A. 对的时间
B. 问题
C. 钱
D. 结果

12.
A. 修理
B. 看了两次
C. 想难了
D. 抵制

13.
A. 借出
B. 潦草地写
C. 打扫
D. 写名字

14.
A. 归还
B. 考虑
C. 拒绝
D. 留下

15.
A. 演讲
B. 短跑比赛
C. 音乐
D. 食物

16.
A. 钱
B. 时间
C. 压力
D. 脏话

17.
A. 支持
B. 让...先走
C. 提及
D. 回答

18.
A. 黑白的动物
B. 书架
C. 邻居
D. 军队

19.
A. 小孩
B. 膝盖
C. 钱
D. 拥有

20.
A. 绳子
B. 钻洞器
C. 钱包
D. 梯子

21.
A. 舞会
B. 尝试
C. 休息室
D. 品牌

22.
A. 修理
B. 密封
C. 仔细看
D. 打开

23.
A. 推开
B. 邀请
C. 警告
D. 卷入战争

24.
A. 储备
B. 烤箱
C. 债务
D. 雇员

25.
A. 切断
B. 弯折
C. 撕裂
D. 连接

26.
A. 沉重地打击
B. 咀嚼
C. 学习
D. 模仿

27.
A. 拿走
B. 捡起
C. 挑选出
D. 安装上

28.
A. 令人着迷的
B. 多愁善感的
C. 简单的
D. 令人信服的

29.
A. 代表
B. 购买
C. 使用
D. 计算出

30.
A. 小说
B. 健身房
C. 协会
D. 艺术学校

31.
A. 餐厅
B. 艺术馆
C. 宿舍
D. 舞台

32.
A. 好奇的
B. 古怪的
C. 神奇的
D. 使人开心的

33.
A. 改变
B. 弥补
C. 完全理解
D. 靠近

34.
A. 暴露
B. 防卫
C. 逃避
D. 面对

35.
A. 与之相反
B. 简单来说
C. 除此以外
D. 进一步讲

36.
A. 可怜的
B. 令人尊敬的
C. 迷人的
D. 诡异的

37.
A. 因此
B. 大体上
C. 比如
D. 事实上

38.
A. 十字路口
B. 聚会
C. 走廊
D. 门厅

39.
A. 内在的
B. 突出的
C. 疯狂的
D. 简单的

40.
A. 玩具
B. 功能
C. 想法
D. 储藏室

词汇测试 第三部分 姓名_____ 班级_____

1.
A. 重复
B. 转移
C. 减价
D. 修复

2.
A. 同意
B. 混合物
C. 公司
D. 假设

3.
A. 牧师
B. 理由
C. 后者
D. 答案

4.
A. 阻塞
B. 分享
C. 赋予黄金边界
D. 铺平

5.
A. 治疗药物
B. 餐厅
C. 烹饪方法
D. 方程

6.
A. 细菌
B. 红花植物
C. 骆驼
D. 被盗物品

7.
A. 观众
B. 行为
C. 大量的钱
D. 岛屿

8.
A. 燃料
B. 止痛片
C. 衣物
D. 隔热材料

9.
A. 丝绸
B. 硬木料
C. 毛皮
D. 亮金属

10.
A. 告诉别人
B. 解释
C. 想出
D. 批判

11.
A. 博物馆
B. 习惯
C. 传奇
D. 周期性事件

12.
A. 完全改变
B. 在其他东西中
C. 仿造
D. 强制实行

13.
A. 事件
B. 支持
C. 问题
D. 解决方案

14.
A. 发现
B. 检查
C. 努力工作
D. 庆祝

15.
A. 独立
B. 独处
C. 能量
D. 骄傲

16.
A. 隧道
B. 柱子
C. 连字符
D. 窗帘

17.
A. 称赞
B. 帮助
C. 奖赏
D. 观众

18.
A. 审核
B. 接受
C. 复制
D. 删除

19.
A. 模式
B. 速度
C. 态度
D. 数量

20.
A. 椅子
B. 空气质量
C. 员工
D. 雇主

21.
A. 有效率的
B. 生气的
C. 有能力的
D. 容易受伤的

22.
A. 装饰
B. 独立
C. 毁灭
D. 污染

23.
A. 建筑物
B. 意见一致
C. 观点
D. 成分

24.
A. 编织
B. 组装铁制品
C. 说服人
D. 捉弄人

25.
A. 焦虑
B. 放松
C. 全神贯注
D. 迷惑

26.
A. 绝望
B. 沮丧
C. 丢脸
D. 冷静下来

27.
A. 跳跃
B. 站起来
C. 走过去
D. 皱眉头

28.
A. 质问
B. 转身
C. 迎合
D. 解释

29.
A. 跳跃
B. 弯腰
C. 坚持
D. 握拳

30.
A. 相当多的
B. 巨大的
C. 主观的
D. 正确的

31.
A. 等级
B. 阶段
C. 总结
D. 城市

32.
A. 相似
B. 决心
C. 同情心
D. 信念

33.
A. 名字
B. 方位
C. 距离
D. 主题

34.
A. 稀缺的
B. 各种各样的
C. 平静的
D. 礼貌的

35.
A. 露营
B. 停留
C. 游荡
D. 离开

36.
A. 条件
B. 凳子
C. 笔记本电脑
D. 幻想

37.
A. 静止不动
B. 爆炸
C. 裂开
D. 摆动

38.
A. 难怪
B. 假设
C. 总之
D. 没预料到

39.
A. 相机
B. 旅游业
C. 隧道
D. 高速公路

40.
A. 古老的
B. 同样的
C. 有名的
D. 伤心的

词汇测试 第四部分

1.
A. 合同
B. 理念
C. 方法
D. 法案

2.
A. 详细的
B. 极好的
C. 简单的
D. 相似的

3.
A. 项目
B. 调查
C. 会议
D. 方面

4.
A. 架构
B. 层
C. 要素
D. 合伙人

5.
A. 补偿
B. 排除
C. 定位
D. 召集

6.
A. 国内的
B. 专业的
C. 注册的
D. 成功的

7.
A. 未知的
B. 外部的
C. 客观的
D. 随后的

8.
A. 逻辑
B. 目标
C. 图案
D. 条款

9.
A. 合作
B. 迁徙
C. 集合
D. 进化

10.
A. 妥协
B. 优先权
C. 出版物
D. 继任者

11.
A. 相反情况
B. 依次
C. 曲调
D. 及时

12.
A. 武断的
B. 至关重要的
C. 坚硬的
D. 合适的

13.
A. 明显的
B. 成熟的
C. 相互的
D. 克制的

14.
A. 可供替代的
B. 任务
C. 评论
D. 交流

15.
A. 主题
B. 计划
C. 同事
D. 论文

16.
A. 合法的
B. 可获得的
C. 明显的
D. 主要的

17.
A. 因素
B. 估计
C. 地点
D. 出口

18.
A. 获得
B. 控制
C. 开始实行
D. 寻求

19.
A. 使融入
B. 宣传
C. 协调
D. 保留

20.
A. 系列
B. 阶段
C. 方面
D. 参数

21.
A. 追寻
B. 改变
C. 观察
D. 促进

22.
A. 披露
B. 利用
C. 承认
D. 恢复

23.
A. 精确的
B. 最小的
C. 灵活的
D. 多种多样的

24.
A. 意识形态
B. 等级制度
C. 化学制品
D. 时尚

25.
A. 模棱两可的
B. 相同的
C. 歪曲的
D. 当代的

26.
A. 感激
B. 积累
C. 补充
D. 遵从

27.
A. 开发
B. 操纵
C. 放弃
D. 降低

28.
A. 性质的
B. 同时发生的
C. 临时的
D. 死板的

29.
A. 回忆往事
B. 近况
C. 英雄事迹
D. 生平介绍

30.
A. 理解
B. 模仿
C. 批评
D. 忽视

31.
A. 开玩笑
B. 看不起
C. 离开
D. 一起玩乐

32.
A. 偷窃
B. 探索
C. 买
D. 学习

33.
A. 钱
B. 速度
C. 兴趣爱好
D. 产量

34.
A. 恐龙
B. 昆虫
C. 爬行动物
D. 鸽子

35.
A. 瀑布
B. 溪流
C. 草
D. 庄稼

36.
A. 拼接
B. 植入
C. 放下
D. 拉出

37.
A. 有信心的
B. 确定的
C. 焦虑的
D. 怀疑的

38.
A. 推
B. 挤压
C. 毁掉
D. 操作

39.
A. 酱汁
B. 盐
C. 啤酒
D. 鸡肉

40.
A. 正确的
B. 怪异的
C. 自愿的
D. 令人担忧的

Appendix E Vocabulary post-test

姓名: ______________________ **班级:** _______________________

请听老师读出某个单词或词组，同时你还会听到一个包含该单词的句子。**如果你不知道该单词或词组的意思，请选 0**。如果你认识该单词或词组，**请在横线上写出该单词或词组的含义**（中英文都可以），并且从 1-5 中选择你对所给出单词或词组意思的确定程度，1 为非常不确定，5 为非常确定，从 1-5 程度依次递增，并请告知在上这堂课之前你是否认识该单词。

0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否
0	________________________	1	2	3	4	5	是	否

Appendix F Vocabulary delayed post-test and final vocabulary delayed post-test

姓名: ______________________ **班级:** _______________________

请听老师读出某个单词或词组，同时你还会听到一个包含该单词的句子。**如果你不知道该单词或词组的意思，请选 0**。如果你认识该单词或词组，**请在横线上写出该单词或词组的含义**（中英文都可以），并且从 1-5 中选择你对所给出单词或词组意思的确定程度，1 为非常不确定，5 为非常确定，从 1-5 程度依次递增。

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

0 ________________________ 1 2 3 4 5

Appendix G Stimulated recall interview questions

L2 group

Watch a video of teaching the target word 'convincing'

1. 当你第一次听到 convincing 这个单词的时候，你感觉怎样？
(How did you feel when you first heard the sentence containing the word 'convincing'?)
2. 听到该单词后，你接下来做了什么？
(What did you do then?)
3. 当老师在解释 convincing 含义的时候，你脑海里在想什么？
(What was in your mind when the tutor was explaining the meaning of 'convincing' to you?)
4. 对于老师所提供的 convincing 的英文解释，你感觉怎样？
(How did you feel about the English-only explanations of 'convincing' provided by the tutor?)
5. 因为当时你并不能写下这个单词以及它的意思，你是如何应对老师所给出的该单词的英文解释的？
(Since you were not allowed to write down the word and its meaning, how did you cope with the English explanation provided by the tutor?)
6. 在接受了老师这种解释单词的方法之后，你对于未来是否可以正确使用 convincing 这个词有什么看法吗？
(How did you feel about the possibility to successfully use 'convincing' in the future after receiving this type of vocabulary instruction?)
7. 就帮助你去理解听力材料这一方面来说，你对于老师所提供的对 convincing 这个单词的解释有什么看法？
(What was your opinion about the vocabulary instruction for the word 'convincing' that you received in relation to understanding relevant listening materials?)
8. 总的来说，你对于老师对该单词解释的方式感觉怎么样？
(In general, how did you feel about the way that the tutor explained the word to you?)

CS group

Watch a video of teaching the target word 'convincing'

1. 当你第一次听到 convincing 这个单词的时候，你感觉怎样？
(How did you feel when you first heard the sentence containing the word 'convincing'?)
2. 听到该单词后，你接下来做了什么？
(What did you do then?)
3. 当老师在解释 convincing 含义的时候，你脑海里在想什么？
(What was in your mind when the tutor was explaining the meaning of 'convincing' to you?)
4. 对于老师所提供的 convincing 的中文解释，你感觉怎样？
(How did you feel about the CS explanations of 'convincing' provided by the tutor?)
5. 因为当时你并不能写下这个单词以及它的意思，你是如何应对老师所给出的该单词的英文解释的？
(Since you were not allowed to write down the word and its meaning, how did you cope with the English explanation provided by the tutor?)
6. 在接受了老师这种解释单词的方法之后，你对于未来是否可以正确使用 convincing 这个词有什么看法吗？
(How did you feel about the possibility to successfully use 'convincing' in the future after receiving this type of vocabulary instruction?)
7. 就帮助你去理解听力材料这一方面来说，你对于老师所提供的对 convincing 这个单词的解释有什么看法？
(What was your opinion about the vocabulary instruction for the word 'convincing' that you received in relation to understanding relevant listening materials?)
8. 总的来说，你对于老师对该单词解释的方式感觉怎么样？
(In general, how did you feel about the way that the tutor explained the word to you?)

CFoF group

Watch a video of teaching the target word 'convincing'

1. 当你第一次听到 convincing 这个单词的时候，你感觉怎样？
(How did you feel when you first heard the sentence containing the word 'convincing'?)
2. 听到该单词后，你接下来做了什么？
(What did you do then?)
3. 当老师在解释 convincing 含义的时候，你脑海里在想什么？
(What was in your mind when the tutor was explaining the meaning of 'convincing' to you?)
4. 对于老师所提供的 convincing 的对比该单词中文和英文用法的解释，你感觉怎样？
(How did you feel about the CFoF explanations of 'convincing' provided by the tutor?)
5. 因为当时你并不能写下这个单词以及它的意思，你是如何应对老师所给出的该单词的英文解释的？
(Since you were not allowed to write down the word and its meaning, how did you cope with the English explanation provided by the tutor?)
6. 在接受了老师这种解释单词的方法之后，你对于未来是否可以正确使用 convincing 这个词有什么看法吗？
(How did you feel about the possibility to successfully use 'convincing' in the future after receiving this type of vocabulary instruction?)
7. 就帮助你去理解听力材料这一方面来说，你对于老师所提供的对 convincing 这个单词的解释有什么看法？
(What was your opinion about the vocabulary instruction for the word 'convincing' that you received in relation to understanding relevant listening materials?)
8. 总的来说，你对于老师对该单词解释的方式感觉怎么样？
(In general, how did you feel about the way that the tutor explained the word to you?)

Appendix H The initial code book

Strategies used while listening to the sentence including unfamiliar vocabulary items (Base on to Vandergrift & Goh, 2012, p. 277-284)

No.	Strategy	Definition	Example
1	Linguistic inferencing	Using known words in an utterance to guess the meaning of unknown words.	S11: I recognised the word was unfamiliar to me, so I tried to use the other information in this sentence which was familiar to me to guess the meaning of this unknown word.
2	Selective attention on known words	Attending to known or familiar words in the language input that assist in understanding and/or task completion	S9: (heard the collocation 'figure out') Actually I thought it was 'finger when I heard it for the first time.
3	Monitoring	Checking, verifying, or correcting one's comprehension or performance in the course of a task	S9: After you showed the sentence in PowerPoint, I suddenly realised that I was wrong. (The student realised that the actual collocation is 'figure out', not 'finger out'.)
4	Translation	Rendering ideas from one language to L1 in a relatively verbatim manner.	S5: I know the Chinese meaning of every single word in this sentence. I can immediately form the meaning of the whole sentence.
5	Waiting	Waiting for explanation, verification, rephrasing, or examples about the language and/or task.	S2: Nothing at all, just wait for your explanation.
6	Selective attention on nearest information	Attending to language input or situational details which is in a fairly short input distance to assist in understanding and/or task completion	S12: (heard 'put people at ease') Because it is at the end of the sentence. I could often recognise the words at the end of each sentence. (The student is in the low-proficiency level. During the interview, she said she could only recognise an unfamiliar lexica item under two circumstances: 1) the sentence itself is very short; 2) the unfamiliar lexical item appears at the end the sentence)
7	Linguistic contextualisation	Relating a word or a phrase heard to an environment where the	S6: (heard 'figure out') I was quite sure that it was a collocation as I can tell from the grammar structure of

		word has appeared before.	the sentence, ‘help sb. do sth.’. Firstly, I heard ‘help her’ and ‘their social system’. In addition, ‘out’ may not appear separately, so I supposed that ‘figure out’ was a collocation.
8	Deduction	Consciously applying learned or self-developed rules to understand the target language.	S5: I would use another word which I know to replace this word, even if the meaning the word is different from the original word. In this way, I could form a general meaning of this sentence. That’s what I usually do.
9	Attend to spelling	Try to spell the unfamiliar words	S2: I was wondering how to spell the word ‘figure’.
10	Give up	Think of nothing, blank in mind	S12: no. I didn’t. I felt that I couldn’t recognise anything.

Strategies for understanding teacher's vocabulary explanation

No.	Strategy	Definition	Example
1	Translation	Translate English explanation into L1	S2: I translated your explanation, but it didn't make sense to me.
2	Selective attention on known words	Attending to known or familiar words in the language input that assist in understanding and/or task completion	S2: In your explanation, there is a word 'sure'. That is '确信' (sure). So the meaning of 'convincing' is '使某人感到确信' (that makes sb. feel sure).
3	Ignoring	Ignore the vocabulary explanation provided	S1: I waited for your explanation. If I can't understand what you explained, I choose to ignore that part.
4	Summarisation	Use a simple L1 word to summarise the English explanation	S4: (hesitation) because the English explanation that you provided has a few English words, I tried to translate them word by word and to find a Chinese word which can cover the translation of the English meaning.
5	Searching	Search for other familiar words which have similar meanings	S7: I just felt that the Chinese meaning, '弄明白' (figure out) that you provided was very familiar to me. I kind of remember that I must know other words which may have the similar meaning.
6	Confirming	Refer back to the original sentence to confirm the meaning.	S4: yes. Since you said before that 'association' is a noun and '俱乐部' is a noun in Chinese, I assumed that association means '俱乐部'. Then, I referred back to the original sentence, 'University student association'.
7	Making judgement	Making judgement on the vocabulary explanation provided	S11: I felt that the first part of your explanation was useful. However, the second part was not useful.
8	Waiting	Waiting for explanation, verification, rephrasing, or examples about the language and/or task	S10: (the target word is 'convincing') it's hard to say. I followed your explanation and thought about 'convince'. And then use 'convince' to remember.

Strategies for consolidating a vocabulary item once it has been encountered

No.	Strategy	Definition	Example
1	Repetition	Repeat the vocabulary item and its meaning verbally	S5: nothing special. I just read your PowerPoint and repeat the word and its meaning in my mind for several times?
2	Force	Force oneself to remember the meaning provided	S8: (hesitation) I forced myself to remember.
3	Segment word	Segment a word according to its pronunciation to assist remembering	S5: I would divide the word into different sections according to its pronunciation, that is 'con' 'vin' 'cing'.
4	Remember as a collocation	remember the meaning of a collocation which includes the word.	S6: I decided to remember this word as a collocation 'student association'.
5	Deduction	Consciously applying learned or self-developed rules to understand the target language.	S6: In the example that you gave us, there is '解释一个问题' (figure out a question). I would link 'question' with 'why'. If there is a 'why', there should be a 'because'. So I finally link 'figure out' with 'because' and then remembered 'figure out' as '弄明白' (figure out).
6	Translation	Refer back to the original sentence and translate the whole sentence	S5: I simply followed your instruction and referred back to the original sentence and translated it into Chinese word by word in my mind.
7	Making connections	connect to familiar words	S7: So I linked this word 'tourism' with the other two words that I knew, 'tour' '旅游', 'tourist' '游客'. Then I remembered the three words together 'tour' '旅游', 'tourist' '游客' and 'tourism' '旅游业'.
8	synonyms	Search for synonyms	S6: after you provided me with the meaning '放松' (at ease), I thought about another word which also means '放松' (at ease). That is 'relax'.
9	Rely on examples	Use an example sentence or phrase including the new vocabulary item	S5: I felt that the example sentence that you provided was very easy to remember for me. So I actually used the example sentence to help me to remember.

10	Keyword strategy	Generate an L1 word "that is similar in sound or appearance to the new foreign language word; and development of an interactive image involving that keyword and the definition of the new word". (Lawson & Hogben, 1998, p. 179)	S6: I have my own solution. As you can see, there is a stress at end the word, 'ten', which I can link to '咬牙切齿' (very angry). In this way, I can link to '恐吓' (threaten).
11	Remember part of a collocation	Remember part of a vocabulary item to assist remembering	S11: I actually use the word 'out' to help me remember. I assumed that 'out' means '清楚，明白' (clear) to me. Then, 'figure out' means '弄清楚，弄明白' (figure out).

Appendix I A coded sample transcript of a stimulated recall interview

S2 is a high proficiency level learner from the L2 group.

Researcher - R; Student2 - S2

First section - Charlie Chaplin - convincing

Play the video of explaining the word “convincing”

R: OK. let’s begin. This is part of the video when we were having the “Charlie Chaplin” class. So when that part of the listening recording was played at our class, what was on your mind?

S2: Silence.

R: Did you recognise that the word “convincing” or maybe other words are unfamiliar to you?

S2: Yes.

R: What did you do then?

S2: Nothing at all, just wait for your explanation.

R: Ok. Since you have watched the video showing that I am explaining the word convincing. Did you understand my explanation?

S2: Yes.

R: Can you tell me more about it? For example, What were you thinking when I was explaining this word to you?

S2: In your explanation, there is a word “sure”. That is “确信” (sure). So the meaning of “convincing” is “使某人感到确信” (that makes sb. feel sure).

R: Oh, you actually translate the English explanation that I provided into Chinese and then understand it?

S2: Yes. If you didn’t provide “sure” in your English explanation, I am afraid that I would not be able to understand the word.

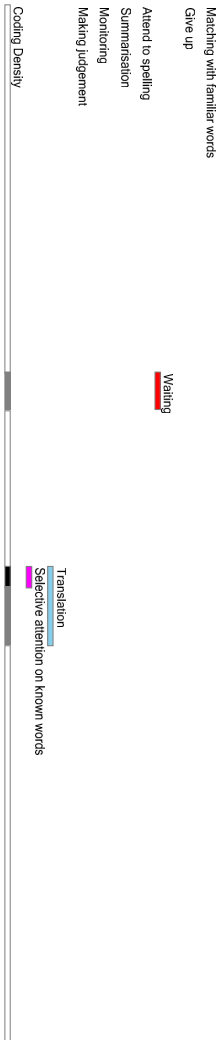
R: So the way you understand the meaning of convincing is to first search for the familiar keywords in the English explanation that I provided…

S2: Yes.

R: and then translated those keywords into Chinese to help you to understand. Is that correct?

S2: Yes. Yes.

R: Since you were not allowed to write down anything during the class, did you do anything after I explain this word to you?



S2: No. I did nothing and I certainly forgot this word after your explanation and didn’ t provide the correct answer in the test at the end of the class.

R: How did you feel about my explanation for this word?

S2: I can manage to understand it.

R: What exactly do you mean that you can manage to understand? According to what you explained previously, is there anything to do with Chinese? Maybe you have gathered a correct Chinese meaning from that English explanation?

S2: Actually I feel that I have just gathered a general Chinese meaning from your English explanation. So I can manage to understand it.

R: Ok. Are you confident to use this word in the future?

S2: I prefer to use the word that I am familiar with, for example, “sure”.

R: Maybe you prefer to use much simpler ones?

S2: Yes. I am not confident to spell convincing correctly in the future.

R: Did my explanation help you to understand the meaning of the whole sentence?

S2: I can manage to understand it, but it might take me a while. I have to read the sentence for several times before I understand it.

R: Does the explanation that I provided make sense to you when using that to rephrase “convincing” in the sentence?

S2: Yes. It makes sense to me.

R: How did you feel about the way I explain this word to you?

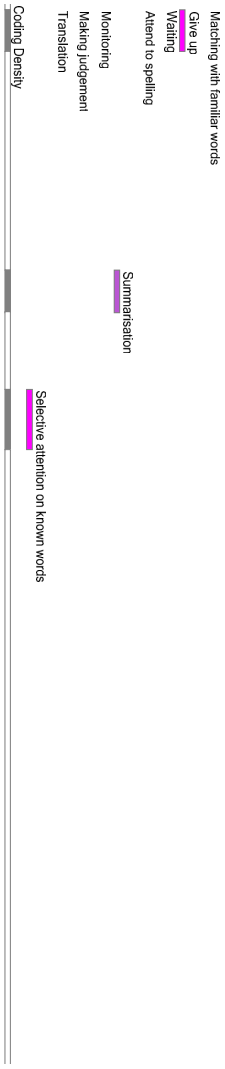
S2: (Silence)

R: You can say anything. The explanation is too long, too difficult, too boring…

S2: I have no idea what to say…

R: Ok. You mentioned that you catch the word “sure” in my explanation first. Would it be better if I only give you “sure” to explain this word?

S2: Quite the same to your original explanation.



R: Maybe give you the Chinese meaning of the word directly?

S2: That will certainly be better.

R: OK. Thank you!

Second section - A student of African wildlife - figure out

Play the video of explaining the collocation “figure out”

R: This is the way how I explained the collocation “figure out” to you. Can you now understand the meaning of this collocation?

S2: (Hesitation) I can understand.

R: Ok. When listened to the sentence including “figure out”, can you recognise maybe “figure out” is an unknown part for you?

S2: Yes. I can recognise.

R: Did you actually recognise “figure” as an unfamiliar single word or “figure out” as an unknown collocation.

S2: As an unknown collocation.

R: What did you do then?

S2: I was wondering how to spell the word “figure”.

R: Did you think about other things?

S2: (Silence)

R: Maybe any other familiar words to you?

S2: Yes. Yes. I felt whether this word could be “finger”.

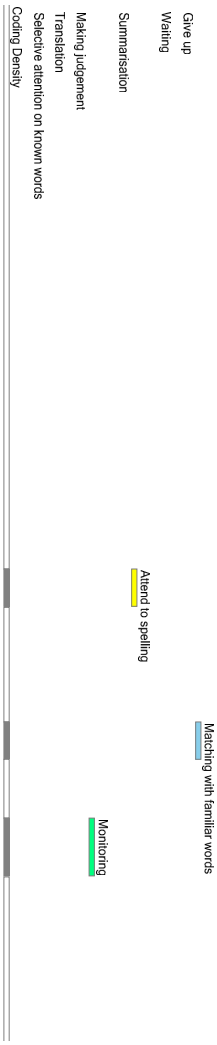
R: Oh, maybe you were actually confused by the pronunciation of the two different words since they have different pronunciations?

S2: Yes. At that time, I was even thinking about whether there were any connections between “finger” and “out”.

R: How did you feel about the explanation that I provided?

S2: (Hesitation) I only feel…

R: Could you please speak a bit louder? I hope I can clearly record your voice.



S2: OK.

R: How did you feel? Because you use a certain way to understand the explanation of “convincing”, did you actually do anything to help you understand the meaning of “figure out”?

S2: I only catch the word “understand” in your explanation.

R: I hope this was not what I guided you to say so…

S2: No. No. That’s what I actually did.

R: Ok. The next question: Did my explanation actually help you to understand the whole sentence?

S2: Yes, your explanation was helpful and helped me to understand the sentence.

R: Can you actually tell me the meaning of the sentence please?

S2: “她对它们肢体语言的研究帮助她理解了猩猩的社会体系” (Her study of their body language helped her figure out their social system).

R: OK. So your understanding of “figure out” is actually “理解” (understand) in Chinese?

S2: Yes.

R: Do you feel confident to use the collocation in the future?

S2: The same as convincing. I would prefer to use “understand”.

R: OK. Does this mean that you prefer to use words and phrases that you are familiar with?

S2: Yes. I would be more confident to use those words that I am familiar with.

R: In general, how did you feel about the way I explain this word to you?

S2: (Silence)

R: You can rate it. From 1-5, 1 means very poor and 5 means very good.

S2: (Laugh). I would rate it 1.5.

R: OK. Can you explain why?

S2: (Silence)

R: 1.5 means that you may not satisfy with the way I explained. What explanation do you actually prefer to have?

S2: Chinese meaning. Very simple Chinese meaning.

R: Did you feel that the way I explain the words to you may be a bit complicated?

S2: Yes. A bit. Chinese is our first language.

R: Ok. For the previous two English explanations, can I say that you can partly understand the meaning I provided because you actually prefer to catch the keywords within the explanation and you actually can understand the keywords?

S2: Yes.

Third Section - Communication, no problem? - association

Play the video of explaining the word “association”

R: My explanation for association is very simple “society or club”. Can you understand it?

S2: I understand what is “club”. However, I didn’t have a clear meaning about “society”.

R: ok. Did you try to rephrase “association” as “club” in the original sentence?

S2: “学生俱乐部” (student association) or other meaning?

R: Yes. So my explanation actually helped you to understand the meaning of the sentence?

S2: (Hesitation) A bit weird.

R: why?

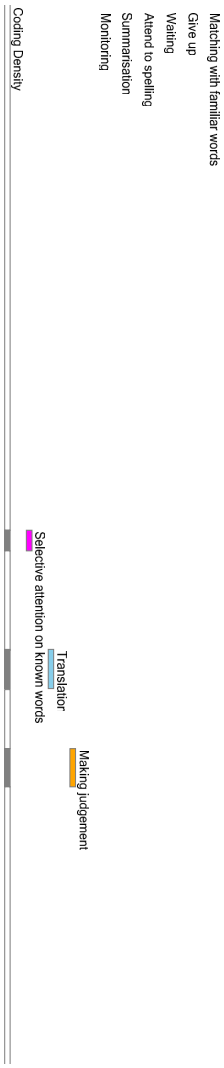
S2: (Silence)

R: Is it because “学生俱乐部” (student club) doesn’t make sense in Chinese?

S2: Yes. Yes.

R: what if I say “学生会” (student association)?

S2: That makes sense.



R: Did you think about rephrase it into “学生会”?

S2: No, I didn’t. I just use the English meaning you provided to translated it into “学生俱乐部” (student club).

R: since you were not allowed to write down the meaning, what did you do to remember it?

S2: (Silence)

R: This is actually in our fourth lesson. For the collocation “figure out”, did you try to remember it?

S2: Of course.

R: What did you do then?

S2: That is very easy to remember because in your class, you actually didn’t explain many collocations. I felt collocations would be much easier to remember than single words.

R: So what about “association”? Did you successfully remember it?

S2: I remembered it when you were explaining it, yet I might forget later.

R: What did you do to remember it?

S2: I would remember it as “club”.

R: did you really remember it as “club” or as “俱乐部” (club).

S2: “俱乐部” (club). But I kind of felt weird when putting “俱乐部” into the sentence.

R: Ok. I understand.

R: do you feel confident to use the word later?

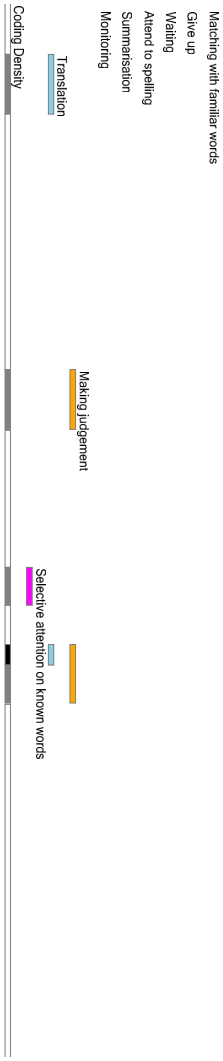
S2: No.

Fourth section - Theme parks - tourism

Play the video of explaining the word “tourism”

R: This is how I explained the word “tourism” to you. When you first heard this sentence, did you recognise that “tourism” maybe an unfamiliar word to you?

S2: Yes.



R: what did you think in you mind?

S2: Nothing.

R: In general, when you heard an unfamiliar English word, what would you do then?

S2: I didn't do anything. Just wait for you to explain.

R: so when I gave you the English explanation for this word to you. What did you do then?

S2: I translated your explanation, but it didn't make sense to me.

R: how you translate it?

S2: I translated it as "产业的观光和假日" (the industry's sightseeing and holiday).

R: ok. This is how you translate and understand "the industry of sightseeing and holiday"?

S2: Yes. I was wondering what's this… it totally didn't make any sense to me.

R: Can you now understand the English meaning?

S2: No. I can't.

R: Would the example that I gave you help you to understand the word?

S2: It still didn't make sense to me when I translated the example sentence into Chinese. I had no idea what this is.

R: After I explained this word to you, did you try to do anything to remember it?

S2: No. I didn't. Since the meaning didn't make sense to me, how exactly I could remember it? I felt even if I remember it, the meaning that I gathered would still be wrong.

R: how confident do you feel to use this in the future?

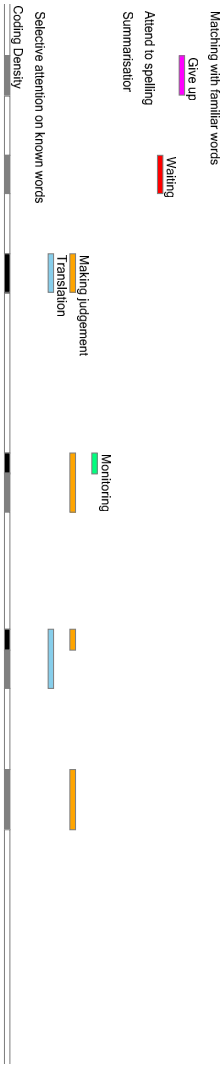
S2: No. I wouldn't be able to use.

R: what do you think of the way that I explained this word to you?

S2: I feel difficult to understand it.

R: Would it be too complicated?

S2: (Hesitation) Yes.



R: So the way actually how you understand the meaning I provided was to directly translate the English meaning into Chinese.

S2: Yes, but I failed to do so!!

Fifth Section - Showing our feelings - at ease

Play the video of explaining the verb “at ease”. When playing the video, the student told me that she remembered the meaning of the collocation.

R: when you first heard the sentence, can you recognise “at ease”?

S2: Yes, I can.

R: you can recognise the word “ease” or the collocation “at ease”?

S2: at ease. I actually know the meaning of “ease”, yet had no idea about the meaning of “at ease”.

R: what is the meaning of “ease”?

S2: oh, no. I know the word “easy”, not “ease”.

R: you actually thought “ease” was “easy” when you first heard the sentence, didn’t you?

S2: yes.

R: when I was explaining the collocation to you, what did you do then? What were you thinking?

S2: (hesitation). “Relax” means “放松”, so I thought “at ease” could be “放松”.

R: so you actually caught the keyword “relax” in my explanation and translated into Chinese to help you understand the collocation.

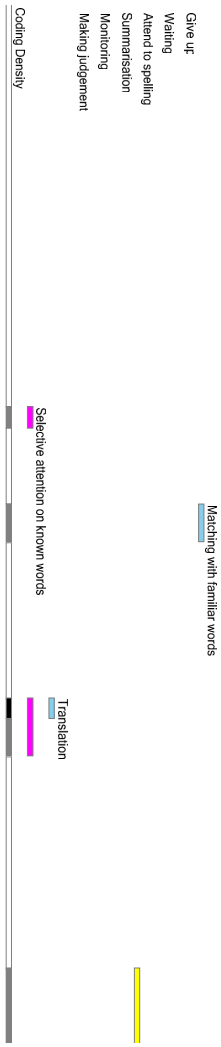
S2: yes.

R: Did my explanation help you to understand the whole sentence?

S2: yes.

R: do you feel confident to use the collocation in the future?

S2: yes. Of course. I can use it. I can successfully spell the collocation. However, for the previous a few vocabulary items, the



main reason that I am not confident is they are difficult to spell correctly.

R: So compared to relax, you actually prefer to use at ease?

S2: yes.

R: How did you feel about the way I explained the collocation.

S2: I feel it is fine to me.

Sixth Section - Showing our feelings - threaten

Play the video of explaining the verb “threaten”.

R: did you recognise that “threaten” is an unfamiliar word to you?

S2: yes.

R: what did you do then?

S2: I just wait.

R: After I explained the word to you, what did you do then?

S2: (Silence)

R: don’t be nervous. Just try to recall what you did then.

S2: I actually searched for the keywords which I am familiar with in the English explanation that you provided.

R: what are the keywords?

S2: “kill” and “hurt”

R: what did you do then?

S2: I tried to translate the English meaning into Chinese and successfully did it.

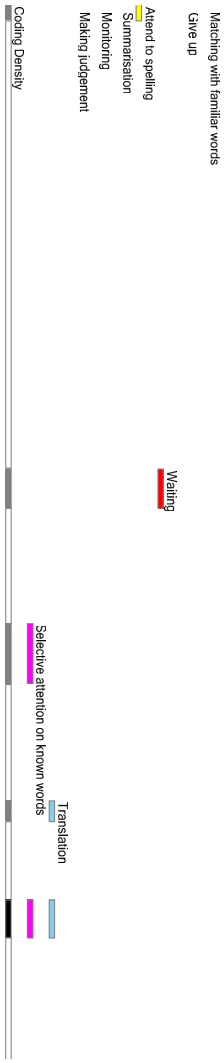
R: Did you remember what meaning did you give at the test?

S2: I remembered that I provided “伤害” (hurt).

R: So you actually use “伤害” (hurt) which is only part of my explanation to help you to understand the word?

S2: Yes, that’s what I usually do.

R: How did you feel about the explanation that I gave?



S2: it is too long…

R: did my explanation help you to understand the whole sentence?

S2: (Silence and shake head).

R: no help at all?

S2: yes.

R: do you feel confident to use it?

S2: no, the word itself is too long.

R: in general, how did you feel about the way I explain the new words or collocations to you?

S2: too tired.

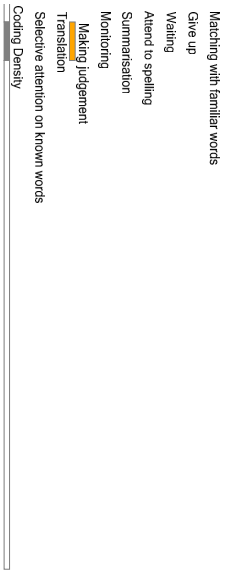
R: a bit difficult?

S2: yes.

R: which way would you prefer for explaining new words to you?

S2: give Chinese meaning directly.

R: ok. Thank you very much!



Appendix J Ethical Approval documents

University of Reading
Institute of Education
Ethical Approval Form A

University of Reading

Tick one:

Staff project: ___
Postgraduate project: PhD √ EdD_____

Name of applicant (s): **Pengchong Zhang**

Title of project: **Comparing different types of intentional EFL vocabulary acquisition for Chinese high-school learners of English**

Name of supervisor (for student projects): **Professor Suzanne Graham and Professor Jeanine Treffers-Daller**

Please complete the form below including relevant sections overleaf.

	YES	NO
Have you prepared an Information Sheet for participants and/or their parents/carers that:		
a) explains the purpose(s) of the project	√	
b) explains how they have been selected as potential participants	√	
c) gives a full, fair and clear account of what will be asked of them and how the information that they provide will be used	√	
d) makes clear that participation in the project is voluntary	√	
e) explains the arrangements to allow participants to withdraw at any stage if they wish	√	
f) explains the arrangements to ensure the confidentiality of any material collected during the project, including secure arrangements for its storage, retention and disposal	√	
g) explains the arrangements for publishing the research results and, if confidentiality might be affected, for obtaining written consent for this	√	
h) explains the arrangements for providing participants with the research results if they wish to have them	√	
i) gives the name and designation of the member of staff with responsibility for the project together with contact details, including email . If any of the project investigators are students at the IoE, then this information must be included and their name provided	√	
k) explains, where applicable, the arrangements for expenses and other payments to be made to the participants	√	
j) includes a standard statement indicating the process of ethical review at the University undergone by the project, as follows: 'This project has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct'.	√	
k)includes a standard statement regarding insurance: "The University has the appropriate insurances in place. Full details are available on request".	√	
Please answer the following questions		
1) Will you provide participants involved in your research with all the information necessary to ensure that they are fully informed and not in any way deceived or misled as to the purpose(s) and nature of the research? (Please use the subheadings used in the example information sheets on blackboard to ensure this).	√	
2) Will you seek written or other formal consent from all participants, if they are able to provide it, in addition to (1)?	√	
3) Is there any risk that participants may experience physical or psychological distress in taking part in your research?		√
4) Have you taken the online training modules in data protection and information security which can be found here: http://www.reading.ac.uk/internal/imps/InformationComplianceTraining/imps-information-compliance-training.aspx	√	
5) Have you read the Health and Safety booklet (available on Blackboard) and completed a	√	

Risk Assessment Form to be included with this ethics application?			
6) Does your research comply with the University's Code of Good Practice in Research?	√		
	YES	NO	N.A.
7) If your research is taking place in a school, have you prepared an information sheet and consent form to gain the permission in writing of the head teacher or other relevant supervisory professional?	√		
8) Has the data collector obtained satisfactory DBS clearance?	√		
9) If your research involves working with children under the age of 16 (or those whose special educational needs mean they are unable to give informed consent), have you prepared an information sheet and consent form for parents/carers to seek permission in writing, or to give parents/carers the opportunity to decline consent?	√		
10) If your research involves processing sensitive personal data[1], or if it involves audio/video recordings, will you obtain the explicit consent of participants/parents?	√		
11) If you are using a data processor to subcontract any part of your research, have you got a written contract with that contractor which (a) specifies that the contractor is required to act only on your instructions, and (b) provides for appropriate technical and organisational security measures to protect the data?			√
12a) Does your research involve data collection outside the UK?	√		
12b) If the answer to question 11a is "yes", does your research comply with the legal and ethical requirements for doing research in that country?			√
13a. Does the proposed research involve children under the age of 5?		√	
13b. If the answer to question 13a is "yes": My Head of School (or authorised Head of Department) has given details of the proposed research to the University's insurance officer, and the research will not proceed until I have confirmation that insurance cover is in place.	√		
If you have answered YES to Question 3, please complete Section B below			

PLEASE COMPLETE **EITHER SECTION A OR B** AND PROVIDE THE DETAILS REQUIRED IN SUPPORT OF YOUR APPLICATION, THEN SIGN THE FORM (SECTION C)

A: My research goes beyond the 'accepted custom and practice of teaching' but I consider that this project has **no** significant ethical implications.	√

Give a brief description of the aims and the methods (participants, instruments and procedures) of the project in up to 200 words. Attach any consent form, information sheet and research instruments to be used in the project (e.g. tests, questionnaires, interview schedules).

Please state how many participants will be involved in the project: 150 (18 for the pilot), aged 15.
This form and any attachments should now be submitted to the Institute's Ethics Committee for consideration. Any missing information will result in the form being returned to you.

The primary aim of the study is to explore the impact of different kinds of vocabulary instruction (lexical Focus-on-Form) within English as a Foreign Language teacher-learner interactions using oral texts (i.e. listening material). It is unclear from the literature whether such lexical Focus-on-Form is best delivered entirely in the target language (i.e. in the language being taught), or by contrast, through teacher codeswitching, whereby part of the explanation of the words encountered in the spoken texts is given in the learners' mother tongue. A third area to consider is contrastive Focus-on-Form, in which cross-linguistic information about the target vocabulary is provided by the teacher through a communicative task, so that learners gain understanding of "the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system" (Laufer & Girsai, 2008, p. 696).

With consent from the head teacher, class teacher, parents and students, participation would involve 150 learners, aged 15, from three selected classes (50 learners for each class) in high school Grade 1 for the main study and 18 learners for the pilot study. These three classes (18 learners for the pilot study) would be randomly assigned to three treatment groups: target language group, teacher codeswitching group and contrastive focus-on-form group.

The intervention would be divided into two identical teaching phases. In each teaching phase, learners would initially receive a vocabulary knowledge pre-test as well as a listening comprehension test, both in written format. In the vocabulary test, learners would hear the teacher reading out relevant lexical items and write down their meanings either in English or in Chinese. The format of the listening comprehension test would be exactly the same as the listening tests which learners are used to receiving in their school, in which multiply-choice questions focusing on the main idea and details of the listening materials would be asked. An

[1] Sensitive personal data consists of information relating to the racial or ethnic origin of a data subject, their political opinions, religious beliefs, trade union membership, sexual life, physical or mental health or condition, or criminal offences or record.

intervention for each treatment group would be administered afterwards. Each treatment group would receive three listening comprehension classes in each phase, taught by the researcher himself. During the class, each group would receive a listening comprehension task and additional vocabulary instruction through target language only, teacher codeswitching or contrastive focus-on-form. At the end of each class, a vocabulary post-test would be employed. All classes will be video recorded with consent from the school, parents as well as the participants themselves, but the camera will focus primarily on the teacher. At the end of each phase, all participants will complete a delayed post-test on vocabulary knowledge as well as a listening comprehension test.

After the second teaching phase, four participants within each group will be selected and invited to undertake a stimulated recall interview. Participants will be selected according to their test results to give a range of proficiency levels. They will firstly watch a video clip of the researcher teaching a specific item of vocabulary and will be asked a series of recall questions in order to gain insights into how they were thinking about the vocabulary explanation during the lesson. One month after the second teaching phase, an additional vocabulary delayed post-test as well as a listening comprehension test will be administered.

Please note that the intention is to use the same information/consent forms for both the pilot and main study, modified as appropriate. The information/consent forms, instruction for all tests as well as the interview questions will be further translated into Chinese.

B: I consider that this project **may** have ethical implications that should be brought before the Institute's Ethics Committee.	

Please provide all the further information listed below in a separate attachment.

1. title of project
2. purpose of project and its academic rationale
3. brief description of methods and measurements
4. participants: recruitment methods, number, age, gender, exclusion/inclusion criteria
5. consent and participant information arrangements, debriefing (attach forms where necessary)
6. a clear and concise statement of the ethical considerations raised by the project and how you intend to deal with then.
7. estimated start date and duration of project

This form and any attachments should now be submitted to the Institute's Ethics Committee for consideration. Any missing information will result in the form being returned to you.

C: SIGNATURE OF APPLICANT:

I have declared all relevant information regarding my proposed project and confirm that ethical good practice will be followed within the project.

Signed Print Name Pengchong Zhang Date 15/03/2015

STATEMENT OF ETHICAL APPROVAL FOR PROPOSALS SUBMITTED TO THE INSTITUTE ETHICS COMMITTEE

This project has been considered using agreed Institute procedures and is now approved.

Signed: Print Name Andy Kempe Date 9th April 2015
(IoE Research Ethics Committee representative)*

* A decision to allow a project to proceed is not an expert assessment of its content or of the possible risks involved in the investigation, nor does it detract in any way from the ultimate responsibility which students/investigators must themselves have for these matters. Approval is granted on the basis of the information declared by the applicant.

Last updated: 18 November 2014

University of Reading
Institute of Education
Risk Assessment Form for Research Activities February 2014

University of Reading

Select one:

Staff project: ☐ PGR project: ☑ MA/UG project: ☐

Name of applicant (s): **Pengchong Zhang**

Title of project: **Comparing different types of intentional EFL vocabulary acquisition for Chinese high-school learners of English**

Name of supervisor (for student projects): **Professor Suzanne Graham and Professor Jeanine Treffers-Dalle**

A: Please complete the form below

Brief outline of Work/activity:	**The primary aim of the study is to explore the impact of different kinds of vocabulary instruction (lexical Focus-on-Form) within English as a Foreign Language teacher-learner interactions using oral texts (i.e. listening material). It is unclear from the literature whether such lexical Focus-on-Form is best delivered entirely in the target language (i.e. in the language being taught), or by contrast, through teacher codeswitching, whereby part of the explanation of the words encountered in the spoken texts is given in the learners' mother tongue. A third area to consider is contrastive Focus-on-Form, in which cross-linguistic information about the target vocabulary is provided by the teacher through a communicative task, so that learners gain understanding of "the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system" (Laufer & Girsai, 2008, p. 696).** **With consents from head teacher, class teacher, parents and students, participation would involve 150 learners from three selected classes (50 learners for each class) in high school Grade 1 for the main study and 18 learners for the pilot study. These three classes (18 learners for the pilot study) would be randomly assigned to three treatment groups: target language group, teacher codeswitching group and contrastive focus-on-form group.** **The intervention would be divided into two identical teaching phases. In each teaching phase, learners would initially receive a vocabulary knowledge pre-test as well as a listening comprehension test, both in written format. In the vocabulary test, learners would hear the teacher reading out relevant lexical items and write down their meanings either in English or in Chinese. The format of the listening comprehension test would be exactly the same as the listening tests which learners are used to receiving in their school, in which multiply-choice questions focusing on the main idea and details of the listening materials would be asked. An intervention for each treatment group would be administered afterwards. Each treatment group would receive three listening comprehension classes in each phase, taught by the researcher himself. During the class, each group would receive a listening comprehension task and additional vocabulary instruction through target language only, teacher codeswitching or contrastive focus-on-form. At the end of each class, a vocabulary post-test would be employed. All classes will be video recorded with consent from the school, parents as well as the participants themselves, but the camera will focus primarily on the teacher. At the end of each phase, all participants will complete a delayed post-test on vocabulary knowledge as well as a listening comprehension test.** **After the second teaching phase, four participants within each group will be selected and invited to undertake a stimulated recall interview. Participants will be selected according to their test results to give a range of proficiency levels. They will firstly watch a video clip of the researcher teaching a specific item of vocabulary and will be asked a series of recall questions in order to gain insights into how they were thinking about the vocabulary explanation during the lesson. One month after the second teaching phase, an additional vocabulary delayed post-test as well as a listening comprehension test will be administered.**

Where will data be collected?	**One high school in China**

Significant hazards:	**None identified. The schools themselves have a duty to maintain a safe area of work within the school.** **During the video recording I will attempt to ensure that there are no trailing leads from the**

	video equipment. If this is unavoidable then I will follow the school's guidelines on dealing with trailing leads to prevent a trip hazard. The video equipment will be loaned from the University and will therefore have the necessary safety checks.

Who might be exposed to hazards?	**N/A**

Existing control measures:	**The rooms fall within the school's Health & Safety responsibilities.**

Are risks adequately controlled:	Yes ☑ No ☐

If NO, list additional controls and actions required:	Additional controls	Action by:

B: SIGNATURE OF APPLICANT:

I have read the Heath and Safety booklet posted on Blackboard, and the guidelines overleaf.
I have declared all relevant information regarding my proposed project and confirm risks have been adequately assessed and will be minimized as far as possible during the course of the project.

Signed: Print Name Pengchong Zhang Date…….

STATEMENT OF APPROVAL TO BE COMPLETED BY SUPERVISOR (FOR UG AND MA STUDENTS) **OR** BY IOE ETHICS COMMITTEE REPRESENTATIVE (FOR PGR AND STAFF RESEARCH).

This project has been considered using agreed Institute procedures and is now approved.

Signed: Print Name Andy Kempe Date 9th April '15

* A decision to allow a project to proceed is not an expert assessment of its content or of the possible risks involved in the investigation, nor does it detract in any way from the ultimate responsibility which students/investigators must themselves have for these matters. Approval is granted on the basis of the information declared by the applicant.

Appendix K Information sheets and consent forms for all parties (Chinese and English versions)

Researcher: Pengchong Zhang
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

University of Reading

Parent/carer information sheet

Research Project: Comparing different types of EFL instruction for Chinese high-school learners of English

Researcher: Pengchong Zhang

Supervisor: Professor Suzanne Graham

Dear parent/carer

I am a PhD candidate at the University of Reading. As part of the data collection stage of my dissertation, I am writing to invite your child to take part in a research study about teaching and learning English.

What is the study?

The study aims to compare the different ways in which English is being taught in a high school EFL context and the impact these might have on learning outcomes. It hopes to make recommendations regarding how teachers can best help learners to make progress in English learning.

Why has my child been chosen to take part?

Your child has been invited to take part because his/her English teacher has expressed an interest in being involved in this research study, and because s/he has experienced approximately three years' English-only EFL teaching mode. Additionally, all other learners who are taught English by Mr/Mrs in high school Grade 1 are being invited to take part.

Does my child have to take part?

It is entirely up to you whether you give permission for your child to participate. You may also withdraw your consent to participation at any time during the project, without any repercussions to you, by contacting the researcher; e-mail: pengchong.zhang@pgr.reading.ac.uk

What will happen if my child takes part?

With your agreement, your child would experience six 45-min English listening comprehension classes as part of their normal English instruction. In each class they would complete one or two language tests (up to 30 minutes) and a listening comprehension test (105minutes each) in class time. The classes would involve the researcher administering a listening comprehension task and providing additional vocabulary explanations which would help your child to understand the relevant listening material. With the consent of all parties, the class would be video-recorded, with the camera focusing primarily on the teacher rather than your child. In addition, a small number of children, representing a range of English proficiency levels, would be interviewed.

The interview session would last approximately 15 minutes and with your consent would be audio-recorded and transcribed. Finally, in order to set the students' English learning in context, it would be helpful to have details of your child's performance levels in his/her most recent English examinations.

All the tasks would take place during school time. All efforts would be made to seek an appropriate time for your child to take part and to ensure that his/her normal study is not adversely affected.

Researcher: Pengchong Zhang
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

What are the risks and benefits of taking part?

The information given by your child in the study will remain confidential and will only be seen by the researcher himself and his supervisor. Neither you, you child nor the school will be identifiable in any published report resulting from the study. Information about individuals will not be shared with the school.

Participants in similar studies have found it interesting to take part. We anticipate that the findings of the study will be useful for teachers in planning how they teach English at high school level.

What will happen to the data?

Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications. The records of this study will be kept private. No identifiers linking you, your child or the school to the study will be included in any sort of report that might be published. Children will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet and on a password-protected computer and only the researcher will have access to the records. The data will be destroyed securely once the findings of the study are written up.

Who has reviewed the study?

This research study has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct. The University has the appropriate insurances in place. Full details are available on request.

What happens if I/my child change my mind?

You/your child can change your mind at any time without any repercussions. During the research, your child can stop completing the activities at any time. If you change your mind after data collection has ended, we will discard your child's data.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you can contact the researcher's supervisor, Professor Suzanne Graham; e-mail: s.j.graham@reading.ac.uk

We do hope that you will agree to your participation in the study. If you do, please complete the attached consent form.

Thank you for your time.

Yours sincerely

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2009 & 2011

Researcher: Pengchong Zhang
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

University of Reading

Research Project: Comparing different types of EFL instruction for Chinese high-school learners of English

Parent/Carer Consent Form

I have read the Information Sheet about the study and received a copy of it.

I understand what the purpose of the study is and what is required of my child and me. All my questions have been answered.

Name of child: _____

Name of parent:_____

Please tick as appropriate:

I consent to my child completing language tests and listening comprehension tests for use within this study ☐

I consent to the school giving the researcher details of my child's recent English examination results. ☐

I consent to my child participating in a 15 minutes individual interview. ☐

I consent to the audio-recording of my child's interview. ☐

I consent to the video-recording of my child's English listening comprehension classes. ☐

Signed: _____

Date: _____

The Queen's Anniversary Prizes
For Higher and Further Education
2009 & 2011

研究人员: 张鹏翀
邮件: pengchong.zhang@pgr.reading.ac.uk
导师: Professor Suzanne Graham
邮件: s.j.graham@reading.ac.uk

University of Reading

家长信息告知书

研究项目: 比较几种英语教学方法在高中英语交流性课堂中的作用

研究人员: 张鹏翀

导师: Professor Suzanne Graham（格雷厄姆教授）

尊敬的家长：

您好！本人目前在英国雷丁大学攻读教育学博士学位。因博士论文课题的需要，特此邀请您的孩子参与一项关于高中英语教学的实验。该实验的主要目的是对比几种不同的英语教学方法对高中生英语学习的影响。

选择您的孩子参与此次实验主要有两个原因。首先，您孩子的老师对此次实验较有兴趣，觉得孩子能够通过参与，从中获益；其次，您的孩子之前已经接受了 4 至 5 年的英语基础教育，体验过以英语为教学语言的英语课堂模式。

在您给予答复前，我先将此次实验的具体内容告知您。实验开始前，您的孩子将接受一次英语基础知识测试（30 分钟）和英语听力测试（15 分钟）。随后的六周，您的孩子将分别接受每周一次，为时 45 分钟的课堂干预，即由研究者本人给他们上一堂 45 分钟的高中英语听力课。在课上，孩子会首先接受听力背景知识补充并完成一个听力练习，随后由研究者给其分析听力材料。课堂最后会进行一个随堂测试。课堂所用的英语教材，均选取自教育部审核通过的高中英语课本。在课堂干预完成的一周以及四周后，您的孩子将会分别被邀请参与另一次的英语基础知识测试和英语听力测试。整个课堂会在获得您以及孩子本人允许的前提下录像。录像的目的是记录研究者本人上课的情况，并不会针对您的孩子。此外，根据研究需要，在您以及孩子本人允许的前提下，您的孩子可能会被邀请参与一次为时 15 分钟的一对一访谈。访谈将被录音，以供未来数据分析使用。为了更好地在实验前了解您孩子的英语能力，望您能够允许学校提供您孩子近期的英语考试成绩，以供参考。所有的实验时间将全力配合您孩子以及学校的学习教学计划，力求在不影响正常学习的情况下来实施。

在此，我承诺实验涉及的所有材料以及工具都已通过雷丁大学学术委员会的实验风险以及道德评估。您以及您孩子的任何信息将不会出现在未来可能发表的刊物上。收集到的所有实验数据包括各类测试的结果，仅研究者本人和其导师能够接触，并且将会在论文提交审核后妥善销毁。此外，雷丁大学有健全的保险系统，覆盖其人员的学术实验工作，如您需要，我很乐意提供更多的信息。

如果在实验进行的任何阶段，您或您的孩子改变主意，都可以随时和我联系，无条件地撤回您的授权同意以及中止实验。我的联系方式是：

电子邮件：pengchong.zhang@pgr.reading.ac.uk

此外，您在实验实施过程中，有任何的投诉或者抱怨，您可以直接与我的导师格雷厄姆教授（Professor Graham）联系，她的电子邮件地址是：s.j.graham@reading.ac.uk

此致

敬礼！

The Queen's Anniversary Prizes
For Higher and Further Education
2009 & 2011

研究人员: 张鹏翀
邮件: pengchong.zhang@pgr.reading.ac.uk
导师: Professor Suzanne Graham
邮件: s.j.graham@reading.ac.uk

家长同意书

我已收到一份信息告知书并已经阅读了其中关于此次实验的内容。
我已被明确告知此次实验的各项内容以及我需要给予的帮助。

孩子姓名：_____
家长姓名：_____

请勾选合适内容：

我同意我的孩子完成相关的听力测试以及英语基础知识测试。 ☐

我同意学校提供我孩子近期的英语考试成绩。 ☐

我同意我的孩子参与一次为时 15 分钟的一对一访谈。 ☐

我同意对我孩子参与的课堂干预进行录像。 ☐

我同意对我孩子参与的访谈进行录音。 ☐

签名：_____

日期：_____

Researcher: Pengchong Zhang
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

University of Reading

Student information sheet

Research Project: Comparing different types of EFL instruction for Chinese high-school learners of English

Researcher: Pengchong Zhang

Supervisor: Professor Suzanne Graham

Dear Student

I am a PhD candidate at the University of Reading. As part of the data collection stage of my dissertation, I am writing to invite you to take part in a research study about teaching and learning English.

What is the study?

The study aims to compare the different ways in which English is being taught in a high school EFL context and the impact these might have on learning outcomes. It hopes to make recommendations regarding how teachers can best help learners to make progress in English learning.

Why has I been chosen to take part?

You have been invited to take part because your English teacher has expressed an interest in being involved in this research study, and because you have experienced approximately three years' English-only EFL teaching mode. Additionally, all other learners who are taught English by Mr/Mrs in high school Grade 1 are being invited to take part.

Do I have to take part?

It is entirely up to you whether you give permission to participate. You may also withdraw your consent to participation at any time during the project, without any repercussions to you, by contacting the researcher; e-mail: pengchong.zhang@pgr.reading.ac.uk

What will happen if I take part?

With your consent, you would experience six 45-min English listening comprehension classes as part of your normal English instruction. In each class you would complete one or two language tests (30 minutes each) and a listening comprehension test (15 minutes each) in class time. The classes would involve the researcher administering a listening comprehension task and providing additional vocabulary explanations which would help you understand the relevant listening material. With your consent, the class will be video-recorded, with the camera focusing primarily on the teacher rather than you. In addition, a small number of learners, representing a range of English proficiency levels, will be interviewed.

The interview session will last approximately 15 minutes and with your consent will be audio-recorded and transcribed. Finally, in order to set the students' English learning in context, it would be helpful to have details of your performance levels in your most recent English examinations.

All the tasks will take place during school time. All efforts will be made to seek an appropriate time for you to take part and to ensure that your normal study is not adversely affected.

Researcher: Pengchong Zhang
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

University of Reading

What are the risks and benefits of taking part?

The information given by you in the study will remain confidential and will only be seen by the researcher himself and his supervisor. Neither you nor the school will be identifiable in any published report resulting from the study. Information about individuals will not be shared with the school.

Participants in similar studies have found it interesting to take part. We anticipate that the findings of the study will be useful for teachers in planning how they teach English at high school level.

What will happen to the data?

Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications. The records of this study will be kept private. No identifiers linking you or the school to the study will be included in any sort of report that might be published. You will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet and on a password-protected computer and only the researcher will have access to the records. The data will be destroyed securely once the findings of the study are written up.

Who has reviewed the study?

This research study has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct. The University has the appropriate insurances in place. Full details are available on request.

What happens if I change my mind?

You can change your mind at any time without any repercussions. During the research, you can stop completing the activities at any time. If you change your mind after data collection has ended, we will discard your data.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you can contact the researcher's supervisor, Professor Suzanne Graham; e-mail: s.j.graham@reading.ac.uk

We do hope that you will agree to your participation in the study. If you do, please complete the attached consent form.

Thank you for your time.

Yours sincerely

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2009 & 2011

Researcher: Pengchong Zhang
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

University of Reading

Research Project: Comparing different types of EFL instruction for Chinese high-school learners of English

Student Form

I have read the Information Sheet about the study and received a copy of it.

I understand what the purpose of the study is and what is required of me. All my questions have been answered.

Name: _____

Please tick as appropriate:

I consent to completing language tests and listening comprehension tests for use within this study ☐

I consent to the school giving the researcher details of my recent English examination results. ☐

I consent to participate in a 15 minutes individual interview. ☐

I consent to the audio-recording of my interview. ☐

I consent to the video-recording of the English listening comprehension classes. ☐

Signed: _____

Date: _____

The Queen's Anniversary Prizes
For Higher and Further Education
2009 & 2011

研究人员: 张鹏翀
邮件: pengchong.zhang@pgr.reading.ac.uk
导师: Professor Suzanne Graham
邮件: s.j.graham@reading.ac.uk

University of Reading

学生信息告知书

研究项目: 比较几种英语教学方法在高中英语交流性课堂中的作用

研究人员: 张鹏翀

导师: Professor Suzanne Graham（格雷厄姆教授）

尊敬的同学：

你好！本人目前在英国雷丁大学攻读教育学博士学位。因博士论文课题的需要，特此邀请你参与一项关于高中英语教学的实验。该实验的主要目的是对比几种不同的英语教学方法对高中英语学习的影响。

你被选择参与此次实验主要有两个原因。首先，你的老师对此次实验较有兴趣，觉得你能够通过参与，从中获益；其次，你在上高中以前，已经接受了至少五年的英语基础教育，体验过以英语为教学语言的英语课堂模式。

此次实验内容主要包括以下几个部分。首先，实验开始前，你将完成一次英语基础知识测试（30 分钟）和英语听力测试（15 分钟）。随后的六周，你将会被邀请参与，每周一次，每次 45 分钟的由研究者本人提供的高中英语听力课。在课上，你会首先接受听力背景知识补充并完成一个听力练习，随后由研究者给你们分析听力材料。课堂最后会进行一个随堂测试（5 分钟）。在听力课结束的一周以及四周后，你将会分别被邀请参与另一次的英语基础知识测试和英语听力测试。整个课堂会在获得你本人允许的前提下录像。录像的目的是记录研究者本人上课的情况，不会针对任何你的课堂表现。此外，根据研究需要，你可能会被邀请参与一次为时 15 分钟的一对一访谈。访谈将在你允许的前提下录音，以供未来数据分析使用。为了更好地在实验前了解你的英语能力，希望你能够允许学校提供你近期英语考试的成绩，以供参考。所有的实验时间将全力配合你以及学校的学习教学计划，力求在不影响你正常学习的情况下来实施。

在此，我承诺实验涉及的所有材料以及工具都已通过雷丁大学学术委员会的实验风险以及道德评估。你的任何信息将不会出现在未来可能发表的刊物上。收集到的所有实验数据包括你参与并提供的各类测试结果，仅研究者本人和其导师能够接触，并且将会在论文提交审核后妥善销毁。此外，雷丁大学有健全的保险系统，覆盖其人员的学术实验工作，如你需要，我很乐意提供更多的信息。

如果在实验进行的任何阶段，你改变主意，都可以随时和我联系，无条件地撤回你的授权同意以及中止实验。我的联系方式是：

电子邮件：pengchong.zhang@pgr.reading.ac.uk

此外，你在实验实施过程中，有任何的投诉或者抱怨，可以直接与我的导师格雷厄姆教授（Professor Graham）联系，她的电子邮件地址是：s.j.graham@reading.ac.uk

此致

敬礼！

The Queen's Anniversary Prizes
For Higher and Further Education
2009 & 2011

研究人员: 张鹏翀
邮件: pengchong.zhang@pgr.reading.ac.uk
导师: Professor Suzanne Graham
邮件: s.j.graham@reading.ac.uk

学生同意书

我已收到一份信息告知书并已经阅读了其中关于此次实验的内容。
我已被明确告知此次实验的各项内容以及我需要给予的帮助。

学生姓名：_____

请勾选合适内容：

我同意完成相关的听力测试以及英语基础知识测试。 ☐

我同意学校提供我近期的英语考试成绩。 ☐

我同意参与一次为时 15 分钟的一对一访谈。 ☐

我同意对我参与的课堂干预进行录像。 ☐

我同意对我参与的访谈进行录音。 ☐

签名：_____

日期：_____

Researcher: Pengchong Zhang
Phone: 18068970575
Email: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
Email: s.j.graham@reading.ac.uk

University of Reading

Teacher information sheet

Research Project: Comparing different types of EFL instruction for Chinese high-school learners of English

Researcher: Pengchong Zhang

Supervisor: Professor Suzanne Graham

Dear Teacher

I am a PhD candidate at the University of Reading. As part of the data collection stage of my dissertation, I am writing to invite your students to take part in a research study about teaching and learning English.

What is the study?

The study aims to compare the different ways in which English is being taught in a high school EFL context and the impact these might have on learning outcomes. It hopes to make recommendations regarding how teachers can best help learners to make progress in English learning.

Why have my students been chosen to take part?

Your students have been invited to take part because they have experienced approximately three years' English-only EFL teaching mode, which will facilitate the collection of appropriate data.

Do my students have to take part?

It is entirely up to you whether you give permission for your students to participate. You may also withdraw your consent to participation at any time during the project, without any repercussions to you, by contacting the researcher; Tel: 18068970575, e-mail: pengchong.zhang@pgr.reading.ac.uk

What will happen if my students take part?

With your agreement, your students would experience six 45-min English listening comprehension classes as part of their normal English instruction. In each class they would complete one or two language tests (30 minutes each) and a listening comprehension test (15 minutes each) in class time. The classes would involve the researcher administering a listening comprehension task and providing additional vocabulary explanations which would help your students to understand the relevant listening material. With the consent of all parties, the class would be video-recorded, with the camera focusing primarily on the teacher rather than your students. In addition, a small number of students, representing a range of English proficiency levels, would be interviewed.

The interview session would last approximately 15 minutes and with your consent would be audio-recorded and transcribed. Finally, in order to set the students' English learning in context, it would be helpful to have details of your students' performance levels in their most recent English examinations.

All the tasks would take place during school time. All efforts would be made to seek an appropriate time for your students to take part and to ensure that their normal study is not adversely affected.

Researcher: Pengchong Zhang
Phone: 18068970575
Email: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
Email: s.j.graham@reading.ac.uk

What are the risks and benefits of taking part?

The information given by your students in the study will remain confidential and will only be seen by the researcher himself and his supervisor. Neither you, the students nor the school will be identifiable in any published report resulting from the study. Information about individuals will not be shared with the school.

Participants in similar studies have found it interesting to take part. We anticipate that the findings of the study will be useful for teachers in planning how they teach English at high school level.

What will happen to the data?

Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications. The records of this study will be kept private. No identifiers linking you, your student or the school to the study will be included in any sort of report that might be published. Students will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet and on a password-protected computer and only the researcher will have access to the records. The data will be destroyed securely once the findings of the study are written up.

Who has reviewed the study?

This research study has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct. The University has the appropriate insurances in place. Full details are available on request.

What happens if I/my student change my mind?

You/your students can change your mind at any time without any repercussions. During the research, your students can stop completing the activities at any time. If you change your mind after data collection has ended, we will discard your students' data.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you can contact the researcher's supervisor, Professor Suzanne Graham; e-mail: s.j.graham@reading.ac.uk

We do hope that you will agree to your participation in the study. If you do, please complete the attached consent form.

Thank you for your time.

Yours sincerely

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2009 & 2011

Researcher: Pengchong Zhang
Phone: 18068970575
Email: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
Email: s.j.graham@reading.ac.uk

University of Reading

Research Project: Comparing different types of EFL instruction for Chinese high-school learners of English

Teacher Consent Form

I have read the Information Sheet about the study and received a copy of it.

I understand what the purpose of the study is and what is required of my student and me. All my questions have been answered.

Name of teacher: _____

Please tick as appropriate:

I consent to the involvement of my students in the project as outlined in the Information Sheet. ☐

I consent to the video-recording of the intervention lessons. ☐

I consent to the audio-recording of interviews with students. ☐

Signed: _____

Date: _____

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2009 & 2011

研究人员: 张鹏翀
电话: 18068970575
邮件: pengchong.zhang@pgr.reading.ac.uk
导师: Professor Suzanne Graham
邮件: s.j.graham@reading.ac.uk

老师信息告知书

研究项目: 比较几种英语教学方法在高中英语交流性课堂中的作用

研究人员: 张鹏翀

导师: Professor Suzanne Graham（格雷厄姆教授）

尊敬的老师：

您好！本人目前在英国雷丁大学攻读教育学博士学位。因博士论文课题的需要，特此邀请您的学生参与一项关于高中英语教学的实验。该实验的主要目的是对比几种不同的英语教学方法对高中生英语学习的影响。

选择您的学生参与此次实验主要原因是，在上高中以前，您的学生已经接受了 4 至 5 年的英语基础教育，体验过以英语为教学语言的英语课堂模式，符合实验的设计初衷。

在您给予答复前，我先将此次实验的具体内容告知您。实验开始前，您的学生将接受一次英语基础知识测试（30 分钟）和英语听力测试（15 分钟），测试内容均选自普通高中英语教材。随后的六周，您的学生将分别接受每周一次，每次 45 分钟的课堂干预，即由研究者本人给他们上一堂 45 分钟的高中英语听力课。在课上，他们会首先接受听力背景知识补充并完成一个听力练习，随后由研究者给其分析听力材料。课堂最后会进行一个随堂测试（5 分钟）。课堂所用的英语教材，均选取自教育部审核通过的高中英语课本。在课堂干预完成的一周以及四周后，您的学生将会分别被邀请参与另一次的英语基础知识测试和英语听力测试。整个课堂会在获得学生家长以及学生本人允许的前提下录像。录像的目的是记录研究者本人上课的情况，并不会针对个别学生。此外，根据研究需要，在学生本人以及家长允许的前提下，您的学生可能会被邀请参与一次为时 15 分钟的一对一访谈。访谈将被录音，以供未来数据分析使用。为了更好地在实验前了解您学生的英语能力，望您能够提供学生近期英语考试的成绩，以供参考。所有的实验时间将全力配合学生以及学校的学习教学计划，力求在不影响学生正常学习的情况下来实施。

在此，我承诺实验涉及的所有材料以及工具都已通过雷丁大学学术委员会的实验风险以及道德评估。您以及您学生的任何信息将不会出现在未来可能发表的刊物上。收集到的所有实验数据包括各类测试的结果，仅研究者本人和其导师能够接触，并且将会在论文提交审核后妥善销毁。此外，雷丁大学有健全的保险系统，覆盖其人员的学术实验工作，如您需要，我很乐意提供更多的信息。

如果在实验进行的任何阶段，您或您的学生改变主意，都可以随时和我联系，无条件地撤回您的授权同意以及中止实验。我的联系方式是：

电话：18068970575；电子邮件：pengchong.zhang@pgr.reading.ac.uk

此外，您在实验实施过程中，有任何的投诉或者抱怨，您可以直接与我的导师格雷厄姆教授（Professor Graham）联系，她的电子邮件地址是：s.j.graham@reading.ac.uk

此致

敬礼！

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2009 & 2011

研究人员: 张鹏翀
电话: 18068970575
邮件: pengchong.zhang@pgr.reading.ac.uk
导师: Professor Suzanne Graham
邮件: s.j.graham@reading.ac.uk

老师同意书

我已收到一份信息告知书并已经阅读了其中关于此次实验的内容。
我已被明确告知此次实验的各项内容以及我需要给予的帮助。

老师姓名：_____

请勾选合适内容：

我同意我的学生参与此次实验。 ☐

我同意对我学生参与的课堂干预进行录像。 ☐

我同意对我学生参与的访谈进行录音。 ☐

签名：_____

日期：_____

Researcher: Pengchong Zhang
Phone: 18068970575
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

University of Reading

Head Teacher information sheet

Research Project: Comparing different types of EFL instruction for Chinese high-school learners of English

Researcher: Pengchong Zhang

Supervisor: Professor Suzanne Graham

Dear Head Teacher,

I am a PhD candidate at the University of Reading. As part of the data collection stage of my dissertation, I am writing to invite your school to take part in a research study about teaching and learning English.

What is the study?

The study aims to compare the different ways in which English is being taught in a high school EFL context and the impact these might have on learning outcomes. It hopes to make recommendations regarding how teachers can best help learners to make progress in English learning.

Why has this school been chosen to take part?

Your school has been selected because the English-only instruction mode adopted by your school for English language teaching will facilitate the collection of appropriate data.

Does the school have to take part?

It is entirely up to you whether you give permission for the school to participate. You may also withdraw your consent to participation at any time during the project, without any repercussions to you, by contacting the researcher; Tel: 18068970575, e-mail: pengchong.zhang@pgr.reading.ac.uk

What will happen if the school takes part?

With your agreement, participation would involve 150 learners from three selected classes (50 learners for each class) in high school Grade 1. These three classes would be randomly assigned to three treatment groups. Each group would receive six sessions of English instruction over six weeks as an intervention, divided into two phases. The sessions would be taught by the researcher himself, each with a slightly different form of listening comprehension and vocabulary teaching. For all groups, the intervention sessions would include a listening comprehension task, vocabulary instruction, and a brief language test at the end. With your permission, and also that of parents and students, the teaching sessions would be video-recorded, with the camera focusing primarily on the teacher rather than the learners. Extracts from these recordings would be used in individual interviews with 12 students from across the three different groups, with each interview taking place at a time convenient to the students and lasting approximately 15 minutes. With the consent of all parties, these interviews would be audio-recorded and transcribed.

The learners would also complete an additional language test and a listening comprehension test before and after each phase of English instruction, with a final language test and a listening comprehension test a couple of weeks after the last intervention session. Each of these tests would be completed in class time, and would take up to 25 minutes each. Finally, in order to set the students' English learning in context, it would be helpful for us to have details of their performance levels in their most recent English examinations.

Researcher: Pengchong Zhang
Phone: 18068970575
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

If you agree to the school's participation, we will seek further consent from parents/carers and the students themselves, as well as from their class teacher.

What are the risks and benefits of taking part?

The information given by participants in the study will remain confidential and will only be seen by the researcher himself and his supervisor. Neither you, the students nor the school will be identifiable in any published report resulting from the study. Information about individuals will not be shared with the school.

Participants in similar studies have found it interesting to take part. We anticipate that the findings of the study will be useful for teachers in planning how they teach English at high school level.

What will happen to the data?

Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications. The records of this study will be kept private. No identifiers linking you, the students or the school to the study will be included in any sort of report that might be published. Participants will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet and on a password-protected computer and only the research team will have access to the records. The data will be destroyed securely once the findings of the study are written up.

Who has reviewed the study?

This research study has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct. The University has the appropriate insurances in place. Full details are available on request.

What happens if I change my mind?

You can change your mind at any time without any repercussions. If you change your mind after data collection has ended, we will discard the school's data.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you can contact the researcher's supervisor, Professor Suzanne Graham; e-mail: s.j.graham@reading.ac.uk

We do hope that you will agree to your participation in the study. If you do, please complete the attached consent form.

Thank you for your time.

Yours sincerely

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2009 & 2011

Researcher: Pengchong Zhang
Phone: 18068970575
E-mail: pengchong.zhang@pgr.reading.ac.uk
Supervisor: Professor Suzanne Graham
E-mail: s.j.graham@reading.ac.uk

University of Reading

Research Project: Comparing different types of EFL instruction for Chinese high-school learners of English

Head Teacher Consent Form

I have read the Information Sheet about the project and received a copy of it.

I understand what the purpose of the project is and what is required of me. All my questions have been answered.

Name of Head Teacher: _____

Please tick as appropriate:

I consent to the involvement of my school in the project as outlined in the Information Sheet. ☐

I consent to the video-recording of the intervention lessons. ☐

I consent to the audio-recording of interviews with students. ☐

Signed: _____

Date: _____

The Queen's Anniversary Prizes
For Higher and Further Education
2009 & 2011

研究人员: 张鹏翀
电话: 18068970575
邮件: pengchong.zhang@pgr.reading.ac.uk
导师: Professor Suzanne Graham
邮件: s.j.graham@reading.ac.uk

校长信息告知书

研究项目: 比较几种英语教学方法在高中英语交流性课堂中的作用

研究人员: 张鹏翀

导师: Professor Suzanne Graham

尊敬的校长：

您好！本人，张鹏翀，目前在英国雷丁大学攻读教育学博士学位。因博士论文课题的需要，特此邀请贵校参与一项关于高中英语教学的实验。该实验的主要目的是对比几种不同的英语教学方法对高中生英语学习的影响。

在您给予答复之前，我先将此次前期实验的具体内容告知您。本实验需要在贵校高一年级选取三个班级的学生接受一次英语基础知识测试（30 分钟）和英语听力测试（15 分钟）。此后的六周，每个班级的学生将分别接受每周一次，为时 45 分钟的课堂干预，即由研究者本人给他们上一堂 45 分钟的高中英语听力课。在课上，学生会首先接受听力背景知识补充并完成一个听力练习，随后由研究者给其分析听力材料。课堂最后会进行一个随堂测试。课堂所用的英语教材，均选取自教育部审核通过的高中英语课本。在课堂干预完成的一周以及四周后，所有参与实验的学生将会被邀请参与另一次的英语知识测试和听力测试。整个课堂会在获得各方允许的前提下全程录像。录像的目的是记录研究者本人上课的情况，并不会针对个别学生。此外，三组学生中会分别抽取四位学生，在家长以及学生本人的允许下，进行一次为时 15 分钟的一对一访谈。访谈将被录音，以供未来数据分析使用。为了更好地了解参与此次实验的学生的英语能力，在家长和学生同意的情况下，望贵校能够提供学生的中考英语成绩以供参考。所有的实验时间将全力配合学生以及学校的教学计划，力求在不影响学生正常学习的情况下来实施。

在此，我承诺实验涉及的所有材料以及工具都已通过雷丁大学学术委员会的实验风险以及道德评估。您学校以及学生的任何信息将不会出现在未来可能发表的刊物上。收集到的所有实验数据，仅研究者和其导师能够接触，并且将会在论文提交审核后妥善销毁。此外，雷丁大学有健全的保险系统，覆盖其人员的学术实验工作，如您需要，我很乐意提供更多的信息。

如果您能同意贵校学生参与此次实验，我将进一步征得个别学生，学生家长以及班级相关老师的同意。同时，如果您在实验进行的任何阶段，改变主意，都可以随时和我联系，撤回您的授权同意。我的联系方式是：

电话：18068970575；电子邮件：pengchong.zhang@pgr.reading.ac.uk

此外，您在实验实施过程当中，有任何的投诉或者抱怨，您可以直接与我的导师（Professor Graham）联系，她的电子邮件是：s.j.graham@reading.ac.uk

此致

敬礼！

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2009 & 2011

研究人员: 张鹏翀
电话: 18068970575
邮件: pengchong.zhang@pgr.reading.ac.uk
导师: Professor Suzanne Graham
邮件: s.j.graham@reading.ac.uk

校长同意书

我已收到一份信息告知书并已经阅读了其中关于此次实验的内容。
我已被明确告知此次实验的各项内容以及我所在的学校需要给予的帮助。

校长姓名：_____

请勾选：

我同意学校学生参与此次实验。 ☐

我同意对学校学生参与的课堂干预进行录像。 ☐

我同意对学校学生参与的访谈进行录音。 ☐

签名：_____

日期：_____